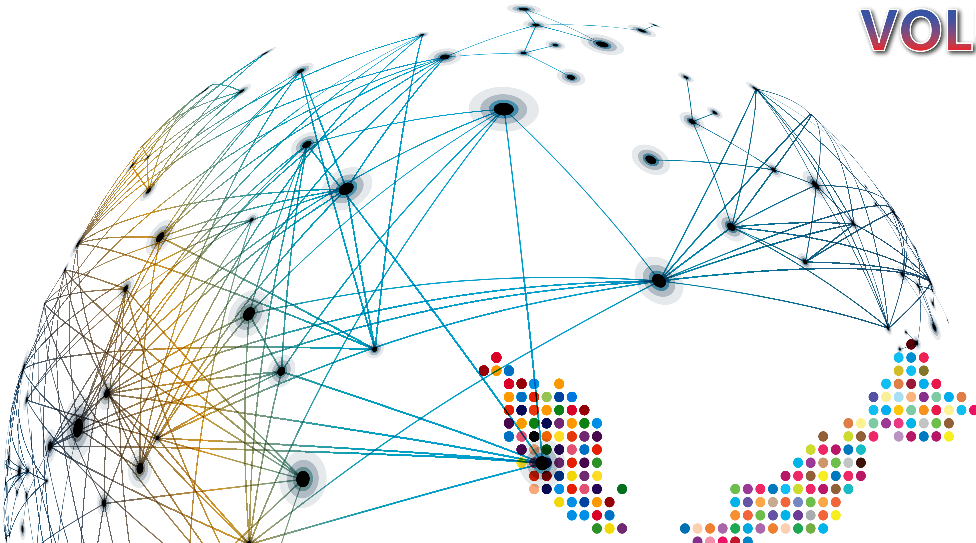
62nd ISI World Statistics Congress
ISI 2019
Kuala Lumpur
18 - 23 August 2019

PROCEEDING

INVITED PAPER SESSION

VOLUME 1



62nd ISI WORLD STATISTICS CONGRESS 2019

18 - 23 August 2019, Kuala Lumpur

Come | Connect | Create

www.isi2019.org

PROCEEDING

ISI WORLD STATISTICS CONGRESS 2019

INVITED PAPER SESSION (VOLUME 1)

Published by:
Department of Statistics Malaysia
Block C6, Complex C
Federal Government Administrative Centre
62514 Putrajaya
MALAYSIA

Central Bank of Malaysia
Jalan Dato' Onn
P.O. Box 10922
50929 Kuala Lumpur
MALAYSIA

Malaysia Institute of Statistics
Department of Mathematical Sciences
Faculty of Sciences and Technology
43600 UKM Bangi, Selangor
MALAYSIA

Portal : https://www.isi2019.org
Email : lpc@isi2019.org

Published in February 2020

Suggested citation:
Department of Statistics Malaysia (DOSM). 2019. Proceeding of the 62nd ISI World Statistics Congress 2019: Invited Paper Session: Volume 1, 2019. 482 pages

Preface

The 62nd International Statistical Institute World Statistics Congress (ISI WSC 2019) has a long tradition since 1887, held for the first time in Kuala Lumpur, Malaysia on 18 to 23 August 2019. ISI WSC 2019 is a global gathering of statistical practitioners, professionals and experts from industries, academia and official authorities to share insights in the development of statistical sciences.

The congress attracted an overwhelming number of participants across the regions. The scientific sessions were delivered over five days with parallel sessions and e-poster sessions running all day long. The scientific program reaches across the breadth of our discipline that comprised of Invited Paper Sessions (IPS), Special Topic Sessions (STS) and Contributed Paper Sessions (CPS). Papers presented exhibit the vitality of statistics and data science in all its manifestations.

I am very honoured to present the proceedings of ISI WSC 2019 to the authors and delegates of the congress. The proceedings contain papers presented in IPS, STS and CPS which were published in fourteen (14) volumes. Scientific papers were received from August 2018 and were carefully reviewed over few months by an external reviewer headed by Scientific Programme Committee (SPC) and Local Programme Committee (LPC). I am pleased that the papers received cover variety of topics and disciplines from across the world, representing both developed and developing nations.

My utmost gratitude and appreciation with the expertise and dedication of all the reviewers, SPC and LPC members for their contributions that helped to make the scientific programme as outstanding as it has been.

Finally, I wish to acknowledge and extend my sincere thanks to the member of National Organising Committee of ISI WSC 2019 from Department of Statistics Malaysia, Bank Negara Malaysia, Malaysia Institute of Statistics and International Statistical Institute for endless support, commitment and passion in making the ISI WSC 2019 a great success and congress to be remembered.

I hope the proceedings will furnish the statistical science community around the world with an outstanding reference guide.

Thank you.

Dr. Mohd Uzir Mahidin
Chairman
National Organising Committee
62nd ISI WSC 2019

Scientific Programme Committee of the 62nd ISI WSC 2019

Chair
Yves-Laurent Grize, Switzerland

Vice-Chairs
Kwok Tsui, Hong Kong, China
Jessica Utts, USA

Local Programme Committee Chair and Co-Chair
Rozita Talha, Malaysia
Ibrahim Mohamed, Malaysia

Representatives of Associations
Rolando Ocampo, Mexico (IAOS)
Cynthia Clark , USA (IASS)
Bruno de Sousa, Portugal (IASE)
Sugnet Lubbe, South Africa (IASC)
Mark Podolskij, Denmark (BS)
Beatriz Etchegaray Garcia, USA (ISBIS)
Giovanna Jona Lasinio, Italy (TIES)

At-Large Members
Alexandra Schmidt, Canada/Brazil
Tamanna Howlader, Bangladesh

Institutional/Ex Officio
Helen MacGillivray, ISI President, Australia
Fabrizio Ruggeri, Past SPC Chair, Italy

Liaison at ISI President's Office
Ada van Krimpen, ISI Director
Shabani Mehta, ISI Associate Director

Local Programme Committee of the 62nd ISI WSC 2019

Chairperson
Rozita Talha

Co-Chairperson
Prof. Dr. Ibrahim Mohamed

Vice Chairperson
Siti Haslinda Mohd Din
Daniel Chin Shen Li

Members
Mohd Ridauddin Masud
Dr. Azrie Tamjis
Dr. Chuah Kue Peng
Dr. Tng Boon Hwa
Prof. Dr. Ghapor Hussin
Prof. Madya Dr. Yong Zulina Zubairi
Prof. Madya Dr. Wan Zawiah Wan Zin @ Wan Ibrahim
Dr. Nurulkamal Masseran
Siti Salwani Ismail
Rosnah Muhamad Ali
Dr. Ng Kok Haur
Dr. Rossita Mohamad Yunus
Muhammad Rizal Mahmod

TABLE OF CONTENTS

The use of registers in the German censuses 2011, 2021 and beyond

Stefan Dittrich, Thomas Koerner, Markus Zwick
Federal Statistical Office, Germany

Abstract
The history of population and housing censuses in Germany is special for a number of reasons. The most obvious feature is that after the last traditional census in 1987 the next population census did only take place in 2011. The census in 2011 was prepared under the restrictions of a judgment of the German constitutional court that preceded the census in 1987 and imposed strict data protection rules. Still under the impression of the public debate in 1987, in 2011 a combined census model was launched integrating elements of register use, a conventional census (on housing), and a sample survey. The results from the sample survey are used in order to correct the data for errors due to over- and undercoverage as well as to collect such variables that were not available from registers. The paper, in a first part, gives insight into the backgrounds to the creation of this specific model and provides an overview of the experiences made in the 2011 census. In its second part, the paper presents the lessons learned for the implementation of a similar combined model in 2021 and introduces the current plans to move to a purely register-based census model in the post-2021 era.

Keywords
Combined census; register-assisted census; register-based census; correction of over- and undercoverage in population registers; census post-2021 developments

1. Introduction

Introducing a combined census model in Germany was not straightforward for several reasons. Firstly, fewer suitable registers are available than in countries that are adopting register-based approaches. Secondly, strict data protection regulations that were established in the context of the last traditional German census in 1987 make it challenging to find solutions to link registers from different areas. Against this background neither a person ID nor a dwelling ID have been introduced so far, which makes any linkage between registers a burdensome undertaking. At the same time, a traditional census based on interviewer-administered data collection is not popular among stakeholders due to the sheer size of the cost.

Since 1983, the traditional census data collection has also enjoyed only limited popularity among the respondents. The last traditional census in

Germany finally took place in 1987, but only after some protracted delay. As a consequence of this special context in Germany, a combined census model was developed, tested in a large-scale test in 2001 and finally implemented in 2011.

This paper briefly recapulates the origins that led to the creation of this specific census model (section 2), outlines the basic features of the combined model as well as the lessons learned during its implementation in 2011 and the current plans for the 2021 round (section 3). Finally it provides an outlook as regards the perspectives for the post-2021 age in Germany (section 4).

2. Legal and institutional background

The creation of the combined model for the German census can only be understood against the background of the difficult implementation of the last traditional census, which was stopped by the German constitutional court only few weeks prior to its implementation in 1983 (for a short history of censuses in Germany before the 1980s see Scholz, R. D., &Kreyenfeld, M. (2016)). It was implemented in modified form in 1987. After the controversial discussions of the census during the period 1983-1987, the Federal Government was reluctant to engage in a traditional census again. So, instead of carrying out a full census in the 2001 census round, a large-scale census test was conducted to assess the viability of a register-assisted approach, that combined data obtained from registers with a number of primary data collections.

The new model had to comply with the judgement of the German constitutional court that was delivered on the occasion of the planned census 1983. This judgement has, since then, had a major impact on data protection regulation in Germany. It stated that the right of informational self-determination directly follows from the fundamental right of personal freedom, guaranteed by article 2 of the constitution. Any data collection required from the public therefore is only considered constitutional if justified by a legal basis, which needs to be specific and clear as well as commensurate compared to the public interest at stake. While data for administrative purposes may only be collected for specific, well justified and commensurate purposes, collection for official statistics, given its specific role, is allowed for a certain stock of information that can be used for multiple purposes. Consequently, data collected for statistical purposes must be used for statistical purposes only and under no circumstances can be transferred to other public bodies (“Rückspielverbot”) (Bundesverfassungsgericht (1983)).

3. The census models 2011 and 2021

3.1 The combined model implemented in 2011

The basic idea of the combined census model in Germany was to use the data in the fields of demography and employment from the available administrative registers (such as the population registers maintained by the municipalities, and the employment statistics register of the Federal Employment Agency). Together with a complete enumeration of buildings and dwellings (as no sufficient register information was available on such units) and a supplementary sample survey (for variables on persons not available from registers), a 'census-typical' data set was to be constructed. A census test carried out in 2001 revealed that it was also necessary to use the supplementary sample survey to correct for the errors detected in the registers (see Statistische Ämter des Bundes und der Länder (2004)).

The demographic information available from the decentralised population registers were completed – and where necessary, corrected by merging it with information from other registers and mandatory primary surveys. By combining different data sources and methods of automatic data generation, a distinct data set containing all required census information could be created for each person, each household and each building with dwellings.[1]

In order to merge the data of the different parts of the census data collection, first a basic register was established, containing a list of all addresses where buildings with residential space existed at the census reference day. This address and building register was the key link for all data collections during the census. It was also used as the statistical population for the sampling procedure of private households and for the housing census.

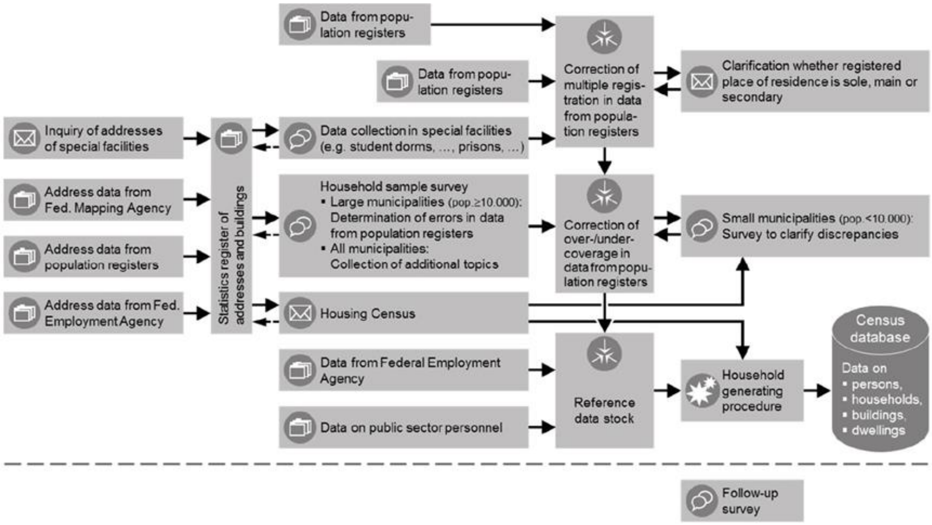
The main data sources used in the combined model were the following (see figure 2):

- The *population registers* provided the main demographic data as well as information on family relationships for all individuals that belong to the target population (about 86 million data records). The data from the municipal population registers were collected at the census reference day (9 May 2011) and were updated three months later in order to take into account delayed register entries and delayed deregistrations. The register data were merged in a nationwide data set and it was subsequently tested whether people were registered at a secondary place of residence only.

[1] The following presentation is based on the English summaries provided by Bechtold (2013, 2016). A more detailed (German) description of the methodology applied in 2011 can be found in Statistische Ämter des Bundes und der Länder (2015).

Figure 2: The German Census Model in 2011 from Bechtold (2016)

- The *supplementary household sample survey*, covering almost 10 percent of the population



Source: own design

was used to adjust the register data in municipalities with 10,000 inhabitants or more, after the registers have been corrected for multiple residences. For the calculation of the population of large municipalities, the level of error of the population registers (over- and under-coverage) detected by the household survey was taken into account. The sample was designed to ensure that the population figures of large municipalities meet an error margin of 1 percent target at a 95 percent confidence level. The method applied to optimise the sampling process was dedicated individually to each municipality and the sample size ranged between 2.1 percent and 45.6 percent and differed significantly even for municipalities of a similar size. For municipalities with less than 10,000 inhabitants, a survey was carried out among those households that had been identified as needing clarification by comparing results of the survey of buildings and dwellings and the population registers.

In addition to the objective to establish the population figures, the supplementary household survey was also used to cover further compulsory census variables of the EU that are not available from registers (in particular the labour market participation and the educational attainment). The additional census topics were collected in all municipalities (not just those with 10,000 inhabitants or more). The sample size was designed to allow publications at NUTS-3 level.

- For persons living in *special facilities*, e.g. a communal accommodations, care institutions, dormitories or similar types of living quarters, census information was collected using a complete enumeration. Addresses carrying stigmatising information, e.g. in the case of psychiatric hospitals or prisons ("confidential special facilities"), were distinguished from non-confidential special facilities, e.g. student dormitories. In confidential special facilities, the privacy of data collection was secured by a special procedure and only a reduced set of variables was collected.

- As there are no registers of buildings and dwellings covering the whole of Germany, the variables of the housing census needed to be obtained through *a postal survey of buildings and dwellings* that was conducted among all property owners (for the total of just under 20 million buildings with residential space, data were collected at approximately 19 million owners). In addition, the census of buildings and housing covers auxiliary variables (number of persons living in a dwelling and names of two persons) which were used in the household generation procedure (see below).
- Information on the employment of the population has been taken from registers of the *Federal Employment Agency* (for about 36 million employees subject to social insurance contributions) and from the *administrative files of the public service agencies with personnel* (for about 3 million public officials, judges and soldiers). These registers were equally used to supplement the demographic information obtained from the population registers, the household sample survey and the survey of addresses with non-confidential special facilities. Together with the register of addresses and buildings, this information constituted the reference data stock.
- To obtain information about household and family structures and their housing conditions (that information is not included in registers) data from the various census components had to be combined in a so-called *household generating procedure*. In this multistage procedure, information about persons from the population registers, the household sample survey and the survey conducted at special facilities was used to form households and to link them to dwellings collected in the housing census.[2]

Merging data sets from different sources for individual persons was one of the great challenges of the 2011 census, because it had to be accomplished without an existing personal identification number available in the different registers. An already existing set of ID numbers for the purpose of the tax authorities was available in some of the registers, but could not be used due to legal restrictions. Therefore individual and address-based information such as name, sex, date of birth, municipal code, post code, street name, and house number were used to link respective records of different data sets.

[2] For further details, see https://www.zensus2011.de/EN/2011Census/Methodology/Methodology_household_generating_procedure_node.html

3.2 Lessons learned of the 2011 census

Retrospectively, the modular concept of the German census 2011 combining register use and primary statistical data collections was successful. A qualitative evaluation came to the result that the model worked well, and (while identifying proposals for an improved implementation) should be the basis for the census 2021. At the same time, it was concluded that the quality of the population registers was still not considered sufficient for the purposes of a census so that the supplementary household sample survey was maintained as an element to correct for over- and under-coverage in the registers (and to collect the data for variables not covered by the registers). Against this background, the household sample survey remains an integral elemental part with an opitimization of the model considering census 2011 results. The model was scientifically accepted and achieved a high precision. At the same time the model reduced the burden and the cost on data collection for the statistical offices and the respondents compared to the former traditional census considerably. The household generating procedure was another new element which turned out for being a precise way to determine family and household data down to the local level.

3.3 current objectives for the 2021 census

Changes for the census 2021 consider census 2011 experiences in quality aspects and possibilities to reduce complexity and thus to be more timely with results. Additionally, modifications aim to make the results easier to understand and raise general acceptance of the model. The main amendments for the census 2021 are:

- The supplementary *household sample survey* used to adjust the register data in municipalities with 10,000 inhabitants or more will be carried out in all municipalities in 2021. An analysis of the 2011 data showed that the special survey on addresses with implausibilites in smaller municipalities did not fulfil the adjustment needed. It turned out that the overall assumption of a better register quality in smaller municipalities was correct, but the correcting rate was still higher than assumed before. To avoid a large scale of small municipalites with high sample rates by stretching the survey concept and the relative error margin of big municipalities 1:1, the relative error margin will be switched to an absolute error margin using a target function for the municipality size in between. Secondly, for small cities using the same administration for their population registers the sample survey will be drawn on the joined municipality level. The results on this level will be splitted for each municipality afterwards using a two-step estimation model with survey results and register data.
- The *interaction of the different census components* need to be designed early on to allow a comprehensive technical approach that liaise the individual parts

of the model together. The results of the different surveys and census components will therefore be linked in a central data stock instead of storing them separately as in 2011. In doing so, data can be cross-checked and validated at an early stage of data processing. Inconsistencies and implausibilities can be removed by rules or even by manual checks. This helps to improve data quality and helps reducing efforts to link the data with each other consecutively.

- The use of paper questionnaires has to be reduced as much as possible by an "*online-first*" *strategy*. This is an important part in being on time in the next census round, which can not only contribute to an important cost reduction, but also decrease response burden because respondents are specifically guided through the questionnaire.
- Building up the *address register* has to start earlier and one of the data sources will not be used any more. In 2011, three main sources were leveraged to collect addresses: the Population Registers, data of the Federal Mapping Agency and data of the Federal Employment Agency. The latter will not be used in 2021 anymore as there were no further addresses added by this source, but many cross-checks were necessary due to different spelling of cities, streets and house numbers.
- In 2011, data of the *Federal Employment Agency* were furthermore used to generate data on employment: The data were of high quality but users complained the complexity of analyses, since different employment figures were released depending on whether they were based on the combined model or the household survey only. Looking at employment in a broader sense, this source had to be analyzed in combination with the household survey to cover self-employed or unemployed as well.
- The *weighting scheme* of the supplementary household survey was targeted primarily at a highly precise number of inhabitants. The production of results for census variables that were not available from registers was only considered as a second priority in the development of the estimators. The weighting procedure needs to be optimized in order to minimize any risk of bias in case of the census variables not available from the registers.

4. Future perspectives towards a register-based census

Although the combined census model was implemented successfully, already before conducting 2021 census work, preparations for a change-over to a fully register-based census following the 2021 census round have already started. These developments are driven by several considerations, which are presented in the following paragraphs.

As demonstrated by the German Statistical Council and recent discussions at the European level, users require census data more frequently, more timely, and in more differentiated regional breakdown (Eurostat (2017, 2018)). The size of the operations used for data collection in traditional, but also many combined censuses goes along with rather long production times so that results can only be produced with some delay. As the experiences of countries with register-based census illustrate, production time can be shortened by using a register-based approach. At the same time, a frequency of ten years is considered too long by many users. Providing annual results for a limited subset of census variables could help to remedy this concern. Finally, the potential of the census to inform local and regional decisions regarding population and housing can only be tapped if data are provided in grid-based form allowing for variable tabulations also for non-administrative areas.

Despite the fact that the combined census model in Germany already led to important cost reductions, its components related to traditional data collection (in particular the complete enumeration of buildings and dwellings and the relatively large supplementary household sample survey) still go along with relatively high cost. Experiences from countries with purely register-based censuses indicate that important cost reductions can be achieved (UNECE 2014).

With further efforts to make progress in the digitisation of public administration, the register infrastructure in Germany is under constant development. This development may lead to further harmonisation and linking of registers, which could open new opportunities for the use in official statistics and in particular the census. With new technologies applied for data protection and encryption, some of the legal challenges in connection with the constitutional courts judgement of 1983 could possibly be solved.

The introduction of a fully register-based census presupposes that a number of preconditions are being met regarding register access, data quality and the possibility to link the records from several registers:

- **Data access**: To move to a register-based census, the necessary data in the fiields of demography, buildings and dwellings, household and family types, as well as labour market and education must exist and be accessible to official statistics. While a significant part of the information of interest is already available in registers, some still need to be created. For instance, no nationwide registers on buildings and dwellings as well as on the educational attainment of the population are currently existing.
- **Data quality**: To be used for statistical purposes, registers need be of sufficient quality and approaches need to be developed in order to correct the data for errors. This is of particular importance in the case of administrative registers, which are usually not kept for statistical purposes and therefore often need data processing. As shown above, the census 2011 has shown that the population registers are subject to significant over- and undercoverage. We are currently exploring suitable ways to correct for these errors by linking the

population register data with other registers containing signs of administrative life and those identifying potential errors.

- **Approaches to link registers**: The development of suitable approaches to link data from different registers is a basic precondition to a register-based census. This is not only necessary as not all the variables required for a population and housing census are available in one single register. Also quality assurance requires reliable and easily implementable linkage procedures to check for administrative signs of life, but also to analyse the consistency of information included in more than one register. For a register-based census, two kinds of record linkage are of particular importance: First, at the level of persons, it must be possible to link the data of persons included in several registers, ideally supported by a constant identifier. Second, for production of data on housing, household and family types, approaches need to be developed to link the persons to the dwellings they are living in by the use of a building and dwelling identifier. Further units requiring record linkage procedures include addresses as well as institutions.

Assuming that the legal, organisational, and technological possibilities to link registers for statistical purposes will increase, this will also lead to new opportunities not only for data production, but also for consistency checks between the values included for the same units in different registers.

References

1. Bechtold. S. (2013). Lessons learned from a mixed-mode census for the future of social statistics. Paper presented at the 97th DGINS Conference "New conceptual design for household and social statistics", Wiesbaden, 27 September 2011. Available at https://www.destatis.de/EN/AboutUs/Events/DGINS/Document_Destatis_MixedModeCensus.pdf?__blob=publicationFile (accessed 14 July 2017)
2. Bechtold. S. (2016). The 2011 Census Model in Germany. In: Comparative Population Studies D1-D9 (date of release: 04.08.2016). Available at DOI: 10.12765/CPoS-2016-07en (accessed 14 July 2017)
3. Bergmann. N. (2009). Volkszählung und Datenschutz. Proteste zur Volkszählung 1983 und 1987 in der Bundesrepublik Deutschland. Hamburg. Deutschland. Diplomica.
4. Bundesbeauftragter für den Datenschutz und die Informationsfreiheit. BfDI (2009). 25 Jahre Volkszählungsurteil Datenschutz – Durchstarten in die Zukunft!. Bonn.
5. Bundesverfassungsgericht (1983). Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983. (BVerfG 65, 1) Available at https://www.zensus2011.de/SharedDocs/Downloads/DE/Gesetze/Volksz

aehlungsurteil_1983.pdf?__blob=publicationFile&v=9 (accessed 4 July 2017)

6. Eurostat (2017). Budapest Memorandum „Population Movements and Integration Issues – Migration Statistics“. 103rd Conference of the Directors General of the National Statistical Institutes. Budapest.
7. Eurostat (2018). Strategy for the Post 2021 Census. Doc. DSS/2018/Mar/3.1, Meetting of the European Directors of Social Statistics, Luxembourg, 2-3 March 2018.
8. Scholz. R. D. &Kreyenfeld. M. (2016). The register-based census in Germany: historical context and relevance for population research. In: Comparative Population Studies - Zeitschrift für Bevölkerungswissenschaft 41 (2016), 2, pp. 175-204. Available at: DOI: http://dx.doi.org/10.12765/CPoS-2016-08en (accessed 14 July 2017)
9. Statistische Ämter des Bundes und der Länder (2004). Results of the census test. Statistisches Bundesamt. Wiesbaden. Deutschland.
10. Statistische Ämter des Bundes und der Länder (2015). Zensus 2011. Methoden und Verfahren. Statistisches Bundesamt. Wiesbaden. Deutschland. Available at https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaetze_Archiv/2015_06_MethodenUndVerfahren.pdf?__blob=publicationFile&v=6 (accessed 14 July 2017)
11. Würzberger. P., Störtzbach. B. &Stürmer. B. (1986). Volkszählung 1987. Rechtliche Grundlagen und Konzept nach dem Urteil des Bundesverfassungsgerichts vom 15. Dezember 1983. In: Wirtschaft und Statistik 12/1986, pp. 927-957.

Developing Methodology for the Register-based Census in Estonia

Diana Beltadze
Statistics Estonia

Abstract
The next census in Estonia is planned as a register-based one, all necessary information will be derived from 25 registers. During the preparation period, the algorithms for creating census variables from register variables are prepared and tested, as well as criteria for quality checks are developed. While working out the methodology for the register-based census it became obvious that the list of data sources is subject to change. It may require patience to obtain data from new sources. Even if the source of relevant data is available and the data holder has agreed to share the data, the process can be time-consuming.

Keywords
registers system; cross validation of registers; indexes

Introduction

Change in the census method in countries has been quite slow. Since the 1970 round of censuses, statisticians have considered it important to search for alternative methods to traditional census, in order to reduce the cost of censuses and the burden on respondents. However, the introduction of new methods has proved difficult.

Prior to 2000 censuses, the potential of using registers for census was extensively studied by Statistics Estonia. Although there were many registers in the country that could be accessed, these were often not adopted because of the insufficient quality. The data inclusion rate of the registers was too low, the concepts did not correspond to the needs of the population census, and many of the necessary characteristics were unavailable in the registers. In addition, there were difficulties with matching the records, the legislative base was often inadequate, and there was opposition by the public. The improvement of the registers on the basis of the latest census data became an important task.

Before selecting the method for the 2011 census in Estonia, a decision had to be made regarding the long-term development direction of censuses in Estonia, i.e. whether to:

1) move as fast as possible towards a fully register-based census, or

2) maintain (initially) an independent position of the census in the society, developing the traditional census through the use of new technologies and other innovations.

Statistics Estonia has decided to increase the use of administrative registers for statistical purposes. The target is still to prepare for a register-based census using an information technology solution to maximise automated data collection.

For a register-based census in Estonia, a number of conditions must be met at national level, the most important of which are listed below:

1) availability of a population register of sufficient quality at least at the local level;
2) availability of building and dwelling registers of sufficient quality at least at the local level;
3) availability of sufficiently high-quality registers or databases that allow determining economic activity of the population (tax, pension, social protection and employment service registers as well as business register);
4) a link between the various registers and databases, which requires the adoption of different identifiers (identifiers for persons, families, households, dwellings and buildings);
5) a legislative basis supporting the creation and continuous updating of the registers, providing access to registers for the production of statistics and the right to combine data from different data sources;
6) adequate harmonisation of register and census concepts, i.e. with the aim that the definitions of register and census characteristics are not significantly different;
7) good cooperation between the producers of statistics and administrative bodies.

It is clear that registers can be of high quality when people accept the need for their data to be held in registers and that this is useful and necessary for them. However, the trusting relationship between citizens and the state is not as strong in transition countries (Central and Eastern European countries) compared to e.g. the Nordic countries (Puur, 2013). The management of high-quality registers requires sufficient financial resources for the continuous development and updating of the registers.

Chapter 1
Preparation for the register-based census in Estonia in 1996-2008

The possibility of a register-based census in Estonia has been considered since 1996. In general terms, it can be said that the problem is as old as the national registers of Estonia. The issues that have been discussed more are the use of standards (address data, classifications) and that registers lack characteristics necessary for census. It has been stressed that the choice of the register-based census methodology requires virtually all mandatory census characteristics to be covered by registers, that there is a system whereby all objects observed by census have been identified and that address data are used.

The issue of registers emerged in the Statistics Estonia correspondence of authorities again in 2005 – a year when systematic exercise of assessing the quality of registers can be observed. The reason why the topic of registers was not addressed could be associated, among other things, with strict legal environment for producing statistics, by which e.g. it was prohibited by law to use the 2000 census data for analysis (this applied until the entry into force of the new act in 2010). This meant that data validation work using register data was not possible either.

A new initiative was launched in 2007. This involved preparations for a new census round. Statistics Estonia analysed the availability of the population and housing census indicators in state databases on the basis of a self-assessment questionnaire of the registers (11). The study concluded that the use of register data, considering the compiled census programme, was not feasible due to the fact that databases and registers did not have sufficient information for the indicators required for the census (Paut, 2007). The main shortcomings concerned the execution of the census programme, which consisted of the mandatory output of the EU, as well as the needs of Estonian users. There was a complete lack of data on household composition and actual family status, working time, religion, foreign language skills and on the agricultural activities of households. For the whole population, there was no data on occupation, mother tongue, links between households and household members, the number of children born to woman, educational attainment, migration and the place of birth of parents. The data indicating the living conditions was lacking. It was unknown whether the actual place of residence and the registered place of residence of persons coincided. There was no overview of people living in institutions (monasteries, children's homes, etc.). It was not possible to link the population register to the register of construction works and buildings; in addition, a situation was possible that a person was registered in a dwelling which did not exist; also, metadata in the registers were incomplete.

It was hard to explain to the public and the registrars that data needed for the census were not available in the registers. It took some explaining before it became clear that it was too early to have a register-based census.

In conclusion for the period 1996–2008, it should be noted that there were two approaches to the use of registers for census: a pessimistic one, emphasising a lack of data quality and big data gaps in the registers, and an optimistic one, which highlighted the possibility to save on census costs by using modern information technology tools and possibilities to integrate data of various registers, reorganising the dispersed way in which information was handled by different authorities (Beltadze, 2016). The view of the representatives of the pessimistic approach sprung from the responsibility for the quality of census results, gaps in legislation, the lacking opportunities to update hardware and software platforms, including databases and systems, for the production of statistics. At the same time, there was a desire and willingness to cooperate more closely with various authorities and agencies to study the data of registers and find ways for improving the situation. In order to improve the data quality of registers, the producer of statistics recommended specifying data quality requirements and introducing auditing of the registers. It was also reminded that there was a need to ensure regular updating of the data in the registers and it is crucial to harmonise the definitions and classifications used in different registers.

Chapter 2
Preparation for the register-based census in 2009–2015

New impetus was given to the register-based census in the context of entry into force of the Official Statistics Act in 2010. Statistics Estonia was given the task to start preparations for a register-based census. The task of comparing 2011 census data with data in registers was stipulated by law (Official Statistics Act, chapter Census). The data inclusion of the databases and registers was insufficient with regard to required indicators before the previous census (Beltadze, 2016). It was settled that the assessment and improvement of the quality of registers using census data was a prerequisite for the transition to a fully register-based census. A comparative analysis of datasets was done during 2010-2013, after all census results were published. Based on the analysis it was concluded that combining registers and census, 2011 data produces an effect which results in improved data quality indicators of register data for a number of characteristics during the census data processing stage. This concludes that registers can be used for a census. At the same time, similarly to the results of the methodological report in 2013, the analysis results pointed to insufficient data quality.

Table 1. Person, household and dwelling characteristics in registers in 2013

Assessment of coverage in registers	Census characteristics
A – complete	Sex, age, country of birth, nationality
B – partial	Labour status, occupational status Useful area/number of rooms, building type of dwelling
C – limited	Permanent residence, residence abroad and year of arrival, previous residence, relationships between household members. Housing arrangements, type of dwelling, basis for use of dwelling, ownership status, number of persons living in the dwelling, technical characteristics of the dwelling (water supply system, toilet, washing facilities, heating type, time of construction).
D – none	Occupation, place of work

Source: REGREL methodological report in 2013

There are references in the 2013 methodological report to the need to start work to improve data quality in the registers; however, the question remained how the collected data should be improved in the registers. Registers are document-based in Estonia, which sets a limit on their clean-up, as there is a need for a legal basis. Legislation for this is complicated, as in fact, the situation cannot be solved by neither the register keeper nor Statistics Estonia – there are no good solutions at this point in time.

By 2014 of the second period, a new level had been reached in terms of the interoperability of state information systems:

1) almost all census characteristics, excl. occupation and place of work, were covered by registers;
2) there was a system of personal identification codes, on the basis of which the majority of personal data registers had identifiers;
3) there were codes for address objects and commercial register codes in use.
4) Negotiations were started for amendments to the legislation to commence collecting data into the employment register (place of work and occupation) and the population register (residential address).

Considering the new situation, Statistics Estonia decided to start preparing a register-based census and test the situation in the two trial censuses in 2016 and 2019.

New goals were set to:

- resolve questions of substantive / legal / technical possibility of a register-based census;
- settle the relationship between the definitions of register and census characteristics;
- create a concept and plan for capture of census data;
- develop a methodology for determining the population of the census, i.e. who should be enumerated;
- develop a framework for quality improvement that would ensure coverage, accuracy, regular updating of data in registers.

In conclusion, during the period of discussions about the register-based census in 2009–2015, progress was made towards networking and interoperability of information systems. An address standard has been adopted by the population register, data quality is measured by both database keepers and Statistics Estonia, there is a secure data exchange environment, registers have owners and the owners have tasks, census data can be used for data methodology work and quality standards have been established for data.

The development of methodology for a register-based census in Estonia began in 2010 with an inter-institutional methodology project, which resulted in a relatively negative assessment of Estonia's system of registers and raised questions about Estonia's ability to conduct a register-based census.

Chapter 3
Preparations in 2015–2019

Analysing the methodology report and results of the last census, the decision was made to continue development of the methodology within the EU grant works.

From 2015 to 2019, the methodology team for register-based census at Statistics Estonia has worked in the following main directions:

1. Cooperation with state registers to verify the quality of data therein, identify shortcomings and support efforts to improve the quality;
2. Development and testing of algorithms for calculation of census characteristics based on the information found in registers;
3. Development of methodology (indexes) for correcting inaccuracies in register data through models based on information obtained from multiple registers and other sources of information.

The main concern in a register-based census pertains to the poor quality of registers, which results from incorrect data submitted by the population.

Estonia's greatest problem in this respect is the inaccuracy of residence data in the Population Register.

This has forced Statistics Estonia to develop an 'index methodology' to verify and specify the register data on the basis of a large number of other registers and data sources. In the REGREL pilot census in 2019, this methodology will be tested in three particular cases – residency index, partnership index, and placement index (Tiit et al., 2017a, b).

All these indexes use Estonia's administrative databases as sources of information, which can be combined to form an interoperative data system with common identifiers. Assuming that, in the present day, a person living in Estonia inevitably leaves certain traces of activity in the form of records in different databases, it is possible to verify the person's residence in the country, as well as connections between persons and their locations, on an annual basis. Such verification is based on signs of life, signs of partnership and signs of placement that are recorded in registers every year (Tiit et al., 2017 b). The annual indexes are established as linear combinations of the respective signs, which makes it possible to trace the change in a person's status in different years.

The indexes are calculated for all persons who have received an Estonian personal identification code. This makes it possible to monitor transnational persons who have left Estonia, incl. to detect whether they have returned or how trans-boundary commuters move between their homeland and other countries (Tiit, et al., 2018).

Even though the general indexing principles have been established and model parameters have undergone empirical assessment, the methodology itself is still developing and new signs can be added depending on new information (incl. big data) becoming available(Beltadze, 2018). The accuracy of the index-based estimates is assessed through use and additional surveys, and the results are provided with potential estimation error values. Addition of new information (further signs) will result in consistent improvement of the accuracy of index-based estimates.

The index-based methodology has been presented in research articles and at international conferences. Population statisticians of several countries who face similar problems have expressed interest in the practical applicability of the indexes.

The baseline situation for a register-based census can be quite diverse in different countries but there are international requirements and standards for the outputs of register-based population and housing censuses. These requirements are the same, irrespective of the particular census methodology.

Considering this background, it is very important to plan and execute the necessary number of successful pilot censuses before the first register-based population and housing census.

The rehearsal of 2019 is piloting of a full-scale register-based census where nearly all output characteristics are calculated on the basis of register information.

This option facilitates testing of:

- availability of information in registers and transportability of the data;
- quality and coverage of the register information in relation to the total population;
- performance and accuracy of the algorithms developed for the calculation of census characteristics;
- capacity of model-based indexes (residency index, partnership and location index) to generate estimates that reflect the actual situation.

The quality of the results of the pilot census will be assessed according to developed rules and norms both with regard to individual characteristics as well as sets of characteristics (cubes and marginal cubes).

If the pilot census produces adequate results and outputs that meet the international requirements, it means that a register-based census is feasible in Estonia in 2021.

If the results indicate that some census characteristics

- cannot be calculated on the basis of registers, or
- coverage or quality do not meet international requirements,

then another option for census will be implemented.

There is the strategic risk that the results of the register-based census will not meet the needs of Estonian users, because it is difficult to find a workable solution to the problem of data quality of the place of residence in the register.

Indexes were tested in a separate survey in 2018, which demonstrated that despite the fact that indexes improve the deficiencies in the data, they are not 100% reliable. Data should be improved by another ca 5% in order to achieve a better and acceptable consistency for the accuracy. New sources of data should also be found for that.

For this reason, Statistics Estonia needs to explore additional data sources to identify the potential 'markers of life' and 'markers of partnership' in order to improve index-based methodology results.

All relevant census characteristics are available in the registers. We are searching for solutions for the users; if these are not acceptable, we need to find a method that will ensure the expected census output.

Conclusion

Developing methodology for the register-based census is a time-consuming activity. Before using data from registers for censuses, the quality of the data is verified in reference to basic statistical criteria. When data are included in different registers, they can be used for verifying the quality of data, on the one hand, and for selecting the most reliable values in accordance with the developed methodological rules, on the other hand.

Generally, census characteristics cannot be acquired directly from registers, because registers have been designed for other, non-statistical purposes and most of the definitions used differ from statistical definitions. It means that data from multiple registers have to be used in order to form certain census characteristics (e.g., the characteristic of 'activity status' requires data from more than 10 registers), while some characteristics are covered by duplicate information in several registers. The methodologists have solved the following main problems connected with the forming of census characteristics.

This work has demonstrated that a register-based population and housing census is feasible and the preparations for the census have been purposeful. But we could not say that it is possible to conduct a register-based census in 2021. Because the biggest problem for Estonia is the difference between registered and actual places of residence. This affects the breakdown of the lowest level of the place of usual residence (municipality) and all household and family characteristics.

Along with data also the methodology of census statistics is being developed, i.e. new possibilities will emerge for processing data. New data categories and data formats require improvements in the methodology and new methodological approaches.

References

1. Puur, A.(2013).Register-based census methodology report.In Estoniann.Tallinn
2. Beltadze, D.(2016).Securing interoperability for the first registerbased census.Tallinn University.
3. Paut, H.(2007).Summary of the analysis on 11 registers.In Estonian.Tallinn.
4. Beltadze, D.(2016). Information technology and its impact on productivity.Paper for the workshop on the census, Geneva,Switzerland.
5. Maasing, E. (2015). Permanent residency status determination in register-based census. Master's thesis. University of Tartu, Faculty of Mathematics and Computer Sciences, Institute of mathematical statistics. [www] http://dspace.utlib.ee/dspace;/handle/10062/47557
6. Tiit, E.-M., Vähi, M. (2017a)Indexes in demographic statistics: a methodology using nonstandard information for solving critical problems. Papers on Anthropology XXVI/1, pp. 72–87
7. Tiit, E.-M., Visk, H., Levenko, V. (2018). Partnership index. Bulletin of Statistics Estonia No 2.

8. Tiit, E.-M., Maasing, E. (2016 b). Residency index and its applications in censuses and population statistics. Quarterly Bulletin of Statistics Estonia No 3, pp 53–60
9. Tiit, E.-M., Vähi, M. *2(2017b).*Indexes in demographic statistics: a methodology using nonstandard information for solving critical problems. Papers on Anthropology XXVI/1, pp. 72–87
10. Beltadze, D.(2018). Census Methodology in Estonia. Paper for the workshop on the census, Geneva,Switzerland.
11. Statistics Estonia.(2013).REGREL methodology report.Tallinn

Predictive Analytics of Big Neuroscience Data

Ivo D. Dinov, Nina Zhou, Syed Husain, Alexandr Kalinin, Yi Zhao, and Simeone Marino

Statistics Online Computational Resource (SOCR), Departments of Health Behaviour and Biological Sciences (HBBS) and Computational Medicine and Bioinformatics (DCMB), Michigan Institute for Data Science (MIDAS), University of Michigan, Ann Arbor, MI 8109, USA

Abstract

This work present some of the Big neuroscience data research and education challenges and opportunities. Specifically, I identify the core characteristics of complex neuroscience data, discuss strategies for data harmonization and aggregation, and show case-studies using large data of normal and pathological cohorts. Examples of the demonstrated techniques include DataSifter, which enables secure sharing of sensitive data, compressive big data analytics, which facilitates inference on multi-source heterogeneous datasets, and model-free prediction providing forecasting of clinical features or derived computed phenotypes. Simulated data as well as clinical data (e.g., UK Biobank (UKBB), Alzheimer's Disease Neuroimaging Initiative (ADNI), and amyotrophic lateral sclerosis (ALS) case-studies) are used for testing and validation of the techniques. In support of open-science, result reproducibility, and methodological improvements, all datasets, statistical methods, computational algorithms, and software tools are freely available online.

Keywords

Big Data; Model-based analytics; Model-free inference; Neurodegenerative disorders; Data science; Open-science

1. Introduction

This paper aims to present some of the contemporary Big neuroscience data challenges, provide examples of solutions for specific problems, and identify research, computational, and educational opportunities. We will begin by defining data science and predictive analytics and examining the common characteristics of Big datasets. Focusing on several driving biomedical and health challenges, we will pinpoint some concrete barriers to data sharing. We will briefly review two complementary strategies to enable data computing on sensitive information, -differential privacy (Dwork 2009) and homomorphic encryption (Gentry 2009). Then, we will describe a recently introduced technique for statistical obfuscation of sensitive data (DataSifter) and demonstrate its approach to balancing data security and data-utility (Marino, Zhou et al. 2018). We will conclude by examining three biomedical and health applications using neurodegenerative aging disorders, paediatric pathological brain development, and exploratory census-like population neuroscience.

Figure 1 shows a schematic that illustrates the relation between census-like population-based view of natural processes (left), their Big Data proxy representation (middle), and classical (small) sampling based process description. By examining many dozens of complex biomedical and health case-studies we identified the common characteristics of Big Data (Dinov 2018), **Table 1**.

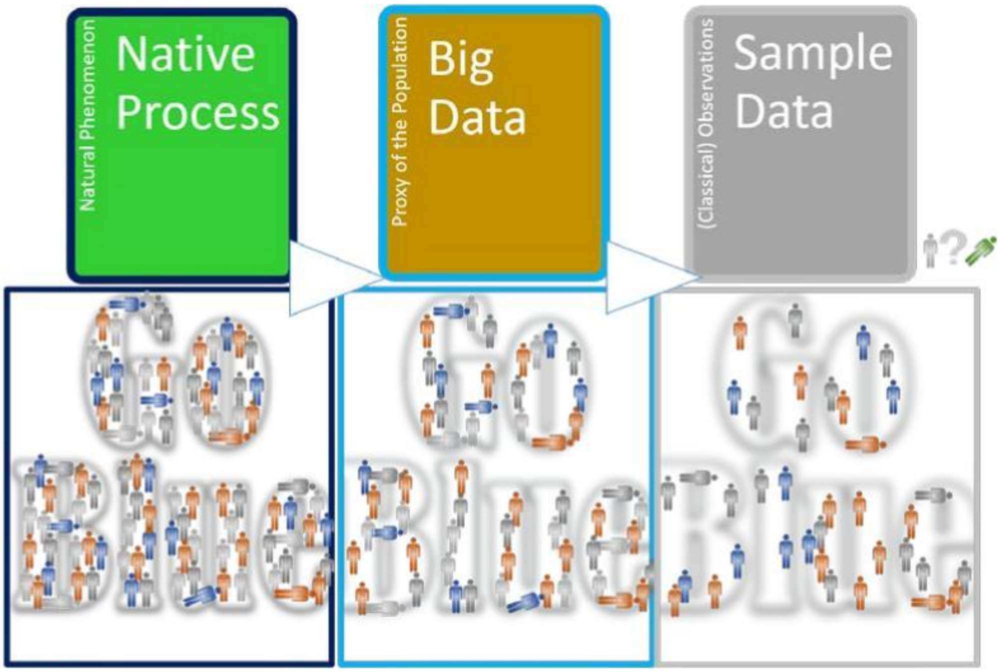


Figure 1: Schematic of the relation between native processes (left), their Big Data representations (middle), and traditional sampling based process characterization. Note that (1) the ideal population view of the process is often unobservable and intractable, the Big Data proxy of the process often requires substantial data management, harmonization, aggregation, preprocessing and wrangling before it can be analysed, and (3) the sample data may facilitate rapid and effective data analytics, but may also represent a limited view of the entire process.

Table 1: Common characteristics of Big biomedical and healthcare datasets.

Dimensions of Big Data	Properties and Tool specifications
Size	Harvesting and management of vast amounts of data
Complexity	Wranglers for dealing with heterogeneous data
Incongruency	Tools for data harmonization and aggregation
Multi-source	Transfer and joint modeling of disparate elements
Multi-scale	Macro to meso to micro scale observations
Time	Techniques accounting for longitudinal patterns in the data
Incomplete	Reliable management of missing data

2. Methodology

There are a few complementary strategies that enable scientific computing on sensitive datasets. Examples of these include ε-differential privacy (Dwork 2009), homomorphic encryption (Gentry 2009), and statistical obfuscation via *DataSifter* (Marino, Zhou et al. 2018). Below we review each of these techniques.

2.1 ε-differential privacy (ε-DP)

ε-DP provides a mechanism to mine information in databases without compromising privacy. By estimating the theoretical limits on the balance between information utility and risk of sharing data, this technique enables data governors to quantify the potential risks of information re-identification. However, it is difficult to apply on high-dimensional, unstructured, skewed, or categorical data (Dwork 2009).

Assume we have a dataset including measurements of the following features: $\{ C_1, C_2, \ldots , C_k \}$, which can be categorical or numerical. Relational databases (DBs) store lists of cases $\{ x_1, x_2, \ldots , x_n \}$, $x_i \in C_1 \times C_2 \times \ldots , \times C_k$, $1 \leq i \leq n$. ε-Differential privacy relies on adding noise to the data in the database, which adds protection against reidentification of individual records. An algorithm f is called ε-*differentially private* if for all possible inputs (datasets or DBs) D_1, D_2 that differ on a single record and all possible f outputs, y, the probability of correctly guessing D_1 knowing y is not significantly different from the corresponding probability of D_2 given y. In other words,

$$\frac{P(f(D_1) = y)}{P(f(D_2) = y)} \leq e^{\varepsilon}, \quad \forall y \in Range(f).$$

Clearly the small positive number, $\varepsilon > 0$ and $e^{\varepsilon} \sim 1$, controls the level of uncertainty about reidentification of the source data (D_1 or D_2) from the known observation, y. The global sensitivity of f is the smallest number

S(f), such that $\forall D_1$, D_2 that differ on at most one element $\| f(D_1) - f(D_2) \|_1 \leq$ S(f). There are many differentially private algorithms, e.g., random forests, decision trees, k-means clustering, etc. For instance, if $f : D = DB \to R^m$, the algorithm outputting $f(D) + (\eta_1, \eta_2, \ldots, \eta_m)$, with $\eta_i \in$ *Laplace* $\left(\mu = 0, \sigma = \sqrt{2}\frac{S(f)}{\varepsilon}\right)$, $\forall i$ is ε-differently private.

2.2 Fully-Homomorphic Encryption (FHE)

FHE security is based on preprocessing the data by encryption to allow subsequent program execution and data-driven inference using the encrypted information (Gentry 2009). As a result, the process outputs are encrypted and their interpretation requires ability to *decrypt* the information following the data analytics. It represents an elegant and powerful mathematical framework for bijective (encoding/decoding) processing and analytics. Albeit, it is very fast, FHE has some limitations, e.g., deriving the ***f '*** – commutative analytic evaluators – is never a trivial task and requires close cooperation between data governor and data user. **Figure 2** shows schematically the process of data analytics using fully-homomorphic encryption.

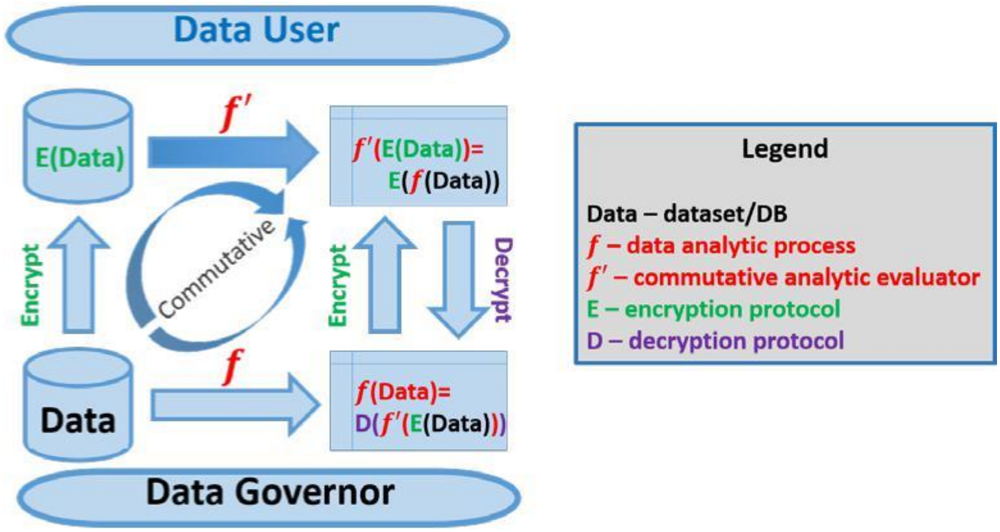


Figure 2: Data analytics via fully-homomorphic encryption.

2.3 *DataSifter* Statistical Obfuscation

The process of data-masking using statistical obfuscation is the core of the *DataSifter* technique. It combines artificial random missingness with partial information alterations using data swapping within subjects' neighbourhoods. These operations have minimal impact on the joint distribution of the obfuscated (sifted) output data as the controlled rate of missingness is introduced completely at random and nearest neighbourhoods tend to have consistent distributions. The *DataSifter* algorithm preserves the exact data structure as well as the bulk of the total data energy of the original data in terms of conserving the overall distribution of the original data features. Simultaneously, the method obfuscates the individual cases sufficiently to protect against the risks of subject re-identification. The *DataSifter* technique includes several user-controlled parameters that allow the data governor the flexibility to control the level of obfuscation, trading privacy protection and preservation of signal energy (Marino, Zhou et al. 2018). **Figure 3** shows a schematic of the *DataSifting* protocol.

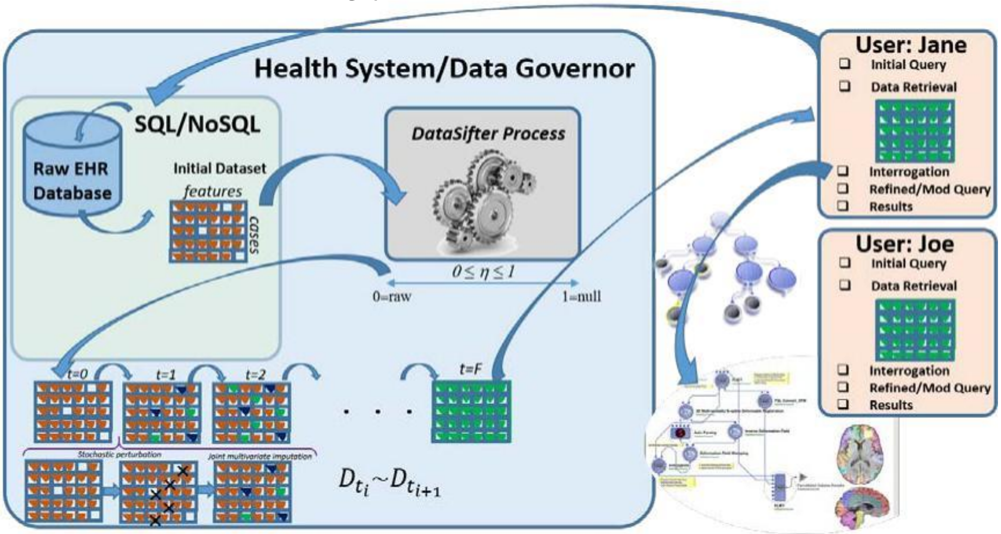


Figure 3: Summary of the *DataSifter* protocol.

Figure 4 illustrate the validation results of applying the DataSifter to a specific clinical case-study. In this case we obfuscated a large Autism Brain Imaging Data Exchange (ABIDE) dataset including 1,098 volunteers and 2,400 features (http://fcon 1000.projects.nitrc.org/indi/abide) (Di Martino, Yan et al. 2014, Torgerson, Quinn et al. 2015). The results include the Percent of Identical Feature Values (PIFV), vertical axis, for different DataSifter obfuscation levels. Each box represents all subjects in the ABIDE sub-cohort and random forest prediction of a specific binary clinical outcome - autism spectrum disorder – (ASD) status (ASD vs. control).

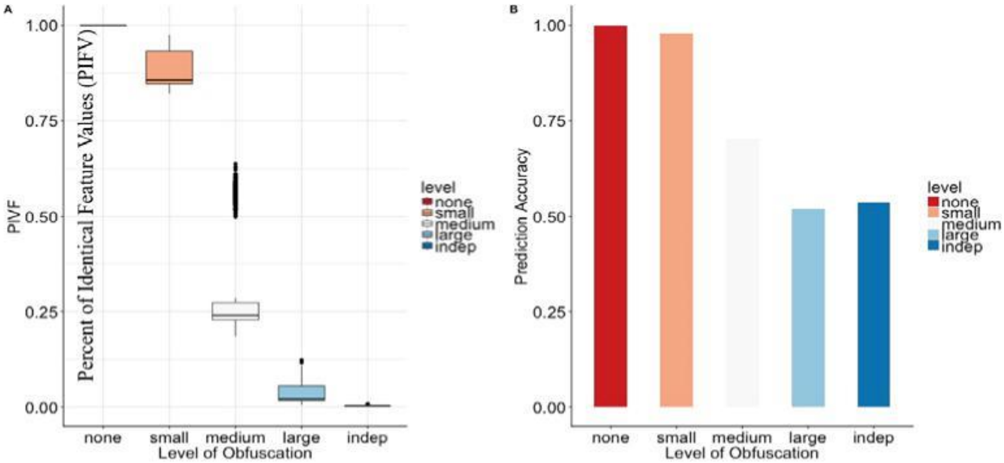


Figure 4: DataSifter obfuscation – trade-offs between privacy protection and preservation of data utility.

In addition, we use established model-based and model-free techniques to interrogate the data (Dinov 2016, Dinov 2016, Dinov 2018, Gao, Sun et al. 2018, Kalinin, Allyn-Feuer et al. 2018, Marino, Xu et al. 2018, Tang, Gao et al. 2018, Zhao, Matloff et al. 2018). These include both confirmatory (hypothesis driven) and exploratory (visual analytics) inferential techniques to extract knowledge, identify patterns, forecast trends, and forecast univariate outcomes of interest and derived computed phenotypes.

3. Results

Open-science relies heavily on data sharing, findability, accessibility, interoperability, and reusability (FAIR) (Wilkinson, Dumontier et al. 2016), open-source development (Feller and Fitzgerald 2002), and transdisciplinary cooperation (Kreps and Maibach 2008, Dinov 2018). **Figure 5** presents some examples of recent results illustrating the power of advanced mathematical modelling techniques, statistical inferential methods, and machine learning strategies to analyse complex, multisource, heterogeneous, and incomplete datasets.

Cluster	Consistency	Variance	Cluster-Size	Silhouette
1	1	0	565	0.58
2	0.99	0.02	427	0.63
3	0.96	0.05	699	0.5
4	0.99	0.02	733	0.5

The SOCR Data Dashboard enables aggregation of multisource data and visual data query and analytics (Husain 2015).

Amyotrophic Lateral Sclerosis (ALS) study aiming to predict disease progression using clinical and lab test information of 2,424 participants and over 2,400 features (Huang, Zhang et al. 2017, Tang, Gao et al. 2018).

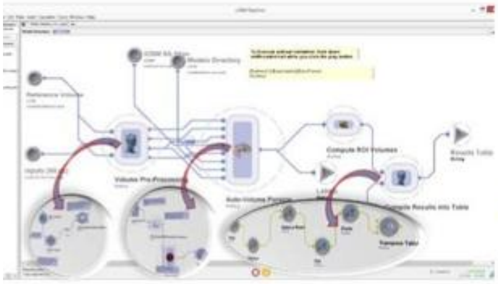


Figure 5: Examples of recent Big health data analytic studies.

A study examining over 10,000 participants in the UKBB cohort identified deep phenotypic traits in the population related to mental health using unsupervised machine learning methods (Zhao, Zhao et al. 2019). The left panel above shows the automated end-to-end computational pipeline workflow deriving thousands of brain morphometric features. The panel on the right shows a decision tree illustrating a simple clinical decision support system providing machine guidance for identifying depression feelings based on categorical variables and neuroimaging biomarkers. Each terminal node, includes the percentage of subjects being labelled as “no” and “yes”, in this case, answering the question “Ever depressed for a whole week.” The p-values listed at branching nodes indicate the significance of the corresponding splitting criterion.

4. Discussion and Conclusion

There are many remaining data science “open problems” including establishing the fundamentals of data representation, modelling, and analytics, quality control and data value metrics, and effectively strategies for data wrangling, harmonization, aggregation, and joint understanding. There also are terrific opportunities for scientific discoveries, basic science developments, ubiquitous range of applications, development of effective educational resources, and designing learning modules to engage a wider cross-section your investigators. All these activities demand substantial community, institutional, state, federal, international, and philanthropic support to advance data analytic methods, enhance the computing infrastructure, train and support students and fellows, and tackle the *Kryder Law >> Moore Law* trend (Dinov 2014).

Acknowledgements
This research is supported in part by NSF grants 1734853, 1636840, 1416953, 0716055 and 1023115; NIH grants P20 NR015331, U54 EB020406, P50 NS091856, P30 DK089503, UL1TR002240; and the Elsie Andresen Fiske Research Fund.

References

1. Di Martino, A., C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer and M. Dapretto (2014). "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism." Molecular psychiatry **19**(6): 659.
2. Dinov, I. (2016). "Methodological Challenges and Analytic Opportunities for Modeling and Interpreting Big Healthcare Data." GigaScience **5**(12): 1-15.
3. Dinov, I. (2018). Data Science and Predictive Analytics: Biomedical and Health Applications using R, Springer International Publishing.
4. Dinov, I., Petrosyan, P, Liu, Z, Eggert, P, Hobel, S, Vespa, P, Woo, Moon S, Van Horn, JD, Franco, J, and Toga, AW. (2014). "High-Throughput Neuroimaging-Genetics Computational Infrastructure." Frontiers in Neuroinformatics **8**(41): 1-11.
5. Dinov, I. D. (2016). "Volume and value of big healthcare data." Journal of Medical Statistics and Informatics **4**(1): 1-7.
6. Dwork, C. (2009). The differential privacy frontier. Theory of Cryptography Conference, Springer.
7. Feller, J. and B. Fitzgerald (2002). Understanding open source software development, Addison-Wesley London.
8. Gao, C., H. Sun, T. Wang, M. Tang, N. I. Bohnen, M. L. T. M. Müller, T. Herman, N. Giladi, A. Kalinin, C. Spino, W. Dauer, J. M. Hausdorff and I. D. Dinov (2018). "Model-based and Model-free Machine Learning Techniques for Diagnostic Prediction and Classification of Clinical Outcomes in Parkinson's Disease." Scientific Reports **8**(1): 7129.
9. Gentry, C. (2009). A fully homomorphic encryption scheme, Stanford University.

10. Huang, Z., H. Zhang, J. Boss, S. A. Goutman, B. Mukherjee, I. D. Dinov, Y. Guan and A. L. S. C. T. C. for the Pooled Resource Open-Access (2017). "Complete hazard ranking to analyze right-censored data: An ALS survival study." PLOS Computational Biology **13**(12): e1005887.
11. Husain, S., Kalinin, A, Truong, A, Dinov, ID (2015). "SOCR Data dashboard: an integrated big data archive mashing medicare, labor, census and econometric information." Journal of Big Data 2(13): 1-18.
12. Kalinin, A. A., A. Allyn-Feuer, A. Ade, G.-V. Fon, W. Meixner, D. Dilworth, S. S. Husain, J. R. de Wett, G. A. Higgins, G. Zheng, A. Creekmore, J. W. Wiley, J. E. Verdone, R. W. Veltri, K. J. Pienta, D. S. Coffey, B. D. Athey and I. D. Dinov (2018). "3D Shape Modeling for Cell Nuclear Morphological Analysis and Classification." Scientific Reports **8**(1): 13658.
13. Kreps, G. L. and E. W. Maibach (2008). "Transdisciplinary science: The nexus between communication and public health." Journal of Communication **58**(4): 732-748.
14. Marino, S., J. Xu, Y. Zhao, N. Zhou, Y. Zhou and I. Dinov (2018). "Controlled Feature Selection and Compressive Big Data Analytics: Applications to Biomedical and Health Studies." PLoS Bioinformatics **13**(8): e0202674.
15. Marino, S., N. Zhou, Y. Zhao, L. Wang, Q. Wu and I. D. Dinov (2018). "HDDA: DataSifter: statistical obfuscation of electronic health records and other sensitive datasets." Journal of Statistical Computation and Simulation **HDDA**: 1-23.
16. Tang, M., C. Gao, S. A. Goutman, A. Kalinin, B. Mukherjee, Y. Guan and I. D. Dinov (2018). "Model- Based and Model-Free Techniques for Amyotrophic Lateral Sclerosis Diagnostic Prediction and Patient Clustering." Neuroinformatics.
17. Torgerson, C., C. Quinn, I. Dinov, Z. Liu, P. Petrosyan, K. Pelphrey, C. Haselgrove, D. Kennedy, A. Toga and J. Van Horn (2015). "Interacting with the National Database for Autism Research (NDAR) via the LONI Pipeline workflow environment." Brain Imaging and Behavior: 1-15.
18. Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos and P. E. Bourne (2016). "The FAIR Guiding Principles for scientific data management and stewardship." Scientific data **3**.
19. Zhao, L., W. Matloff, K. Ning, H. Kim, I. D. Dinov and A. W. Toga (2018). "Age-Related Differences in Brain Morphology and the Modifiers in Middle-Aged and Older Adults." Cerebral Cortex: bhy300-bhy300.
20. Zhao, Y., L. Zhao, N. Zhou, Y. Zhao, S. Marino, T. Wang, H. Sun, A. Toga and I. Dinov (2019). "Predictive Big Data Analytics using the UK Biobank Data." Scientific Reports **in press**.

US Statistical System, the Congress, Fundamental Principles- How Does It Work

Hermann Habermann[1], Katherine Wallman[2], Misha Belkindas[3]
[1]Statistical Consultant, Arlington, VA, USA
[2]Consultant, McLean, VA, USA
[3]ODW Consulting, Rockville, MD, USA

Abstract

The foundation of government in the United States rests on the principle of three co-equal branches of government - legislative, executive and judicial. It is the tensions between these branches that determine how well and how smoothly the government operates. The legislature enacts laws and it is the responsibility of the executive branch to carry out these laws. Tension often arises on the interpretation of the laws by the executive branch. The statistical system in the US is on the far decentralized end of an organization spectrum from centralized to decentralized. In that sense, the statistical system in the USA is quite different from systems in the rest of the world. There are over 100 agencies collecting statistical information with 13 so called principal statistical agencies. Each agency is located in an individual cabinet department with an oversight and coordinating agency in the Executive Office of the President. Fundamental Principles of Official Statistics exist in a sphere outside of the government itself in that they are a common ethos developed by official statisticians and adopted by the United Nations General Assembly. This paper will examine the operational difficulties in implementing the Fundamental Principles. The decennial Census will be used as an example to demonstrate these issues and to show how the Fundamental Principles are and are not followed.

Keywords

National statistical system; Congress; fundamental principles; independence

1. Introduction

We will first briefly review the development and importance of the fundamental principles and introduce the concept of independence. Next we will describe the United States federal statistical system and how it operates in relation to the three branches of United States government: executive, legislative and judicial. We will describe how the tensions between these branches and the authorities granted to leaders of the federal statistical system give rise to problems in implementing the Fundamental Principles. Finally we will use the recent controversy in the United States over the proposed addition of a question on citizenship to the decennial census as an illustration of the problems in implementation of the Fundamental Principles in the United States.

2. Fundamental Principles, Credibility and Public Trust

As described on the United Nations web site[1] the need for a set of principles governing official statistics became apparent at the end of the 1980s when countries in Central Europe began to changenfrom centrally planned economies to market-oriented democracies. It was essential to ensure that national statistical systems in such countries would be able to produce appropriate and reliable data that adhered to certain professional and scientific standards. Towards this end, the Conference of European Statisticians developed and adopted the Fundamental Principles of Official Statistics in 1991 (CES/702), which were subsequently adopted in 1992 at the ministerial level by ECE as decision C (47). Statisticians in other parts of the world soon realized that the principles were of much wider, global significance. Following an international consultation process, a milestone in the history of international statistics was reached when the United Nations Statistical Commission at its Special Session of 11-15 April 1994 adopted the very same set of principles – with a revised preamble – as the United Nations Fundamental Principles of Official Statistics.

At its forty-second session in 2011, the Statistical Commission discussed the Fundamental Principles of Official Statistics and acknowledged that the Principles were still as relevant today as they had been in the past and that no revision of the 10 Principles themselves was necessary. The Commission recommended, however, that a Friends of the Chair group revise and update the preamble of the Fundamental Principles in order to take into account new developments since the time when the Principles were first formulated. At its forty-fourth session in 2013, the Statistical Commission adopted the revised preamble.

At the same session the Commission recommended to the Economic and Social Council the adoption of a draft resolution on the Fundamental Principles of Official Statistics. In accordance with that recommendation, the Council endorsed the Fundamental Principles in its resolution 2013/21 of 24 July 2013. In the same resolution, the Council recommended the Fundamental Principles to the General Assembly for endorsement. Pursuant to the recommendation of the Economic and Social Council, the representative of Hungary, together with 48 co-sponsors, introduced a draft resolution on the matter at the sixty-eighth session of the General Assembly. After a short informal consultation process, the Assembly, in its resolution 68/261 of 29 January 2014 endorsed the Fundamental Principles of Official Statistics.

[1] https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx

The recognition of the importance of agency credibility and public trust in official statistics is of course found in the Fundamental Principles. This document states, inter alia:

> "To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data."

The Office of Management and Budget (OMB), the oversight agency of the federal government, while recognizing the importance of trust, also emphasized the role of professional independence and professional judgment in generating that trust. This was codified in OMB Statistical Policy Directive No. 1[2]. The Directive states that the four "Fundamental Responsibilities" of a federal statistical agency are: (1) "produce and disseminate relevant and timely information," (2) "conduct credible and accurate statistical activities," (3) "conduct objective statistical activities," and (4) "protect the trust of information providers by ensuring the confidentiality and exclusive statistical use of their responses."[3]

Since 1992 the Committee on National Statistics of the National Academies of Sciences, Engineering and Medicine has issued a periodic report on Principles and Practices for a Federal Statistical Agency[4]. These reports echo the importance of the Fundamental Principles and the OMB principles to the US statistical system and they note:

> "To be credible and unhindered in its mission, a statistical agency must maintain a widely acknowledged position of independence from undue external influences. It must avoid even the appearance that its collection, analysis, or reporting processes might be manipulated for political purposes or that individually identifiable data collected under a pledge of confidentiality might be turned over for administrative, regulatory, or law enforcement uses."

It is important to note that both the Policy Directives issued by OMB and the Principles and Practices developed by the Committee on National Statistics (CNSTAT) introduce a concept not used in the Fundamental Principles. That concept is independence. It can be argued that Fundamental Principle 2 relating to decisions by statistical agencies based solely on scientific and professional considerations implies independence. However, both OMB and

---

[2] Federal Register, Vol. 79, No. 231, p.71612 (December 2, 2014).

[3] Ibid. at pp.71614-71615.

[4] National Academies of Sciences, Engineering, and Medicine. (2017). Principles and Practices for a Federal Statistical Agency, Fifth Edition, Washington, DC: The National Academies Press. https://doi.org/10.17226/24810.

CNSTAT found it important to specifically delineate the importance of independence.

The use and value of federal statistics – including the decennial census – depend on their being seen as accurate and unbiased. If statistics are seen to be biased and inaccurate, they will not be used and therefore they will be of little or no value. Statistics are trusted when the agencies that produce the data are seen as making decisions based on professional not political considerations. For their data to be credible, statistical agencies must gain and hold the trust of the nation.

Professional independence is a foundation for building this trust and ensures that decisions about statistical matters are free of any real or perceived political interference.[5] [6] Professional independence is important not only for the credibility of the statistics agency but also for the credibility of the decisions made by political appointees.

The reason for professional independence is to ensure that decisions based on statistical outputs are not tainted by real or perceived political interference.

3. Characteristics of the US federal statistical system

In the United States there is not a single federal statistics agency that collects and disseminates statistical information.[7] Rather, the United States has a decentralized statistical system with over 100 agencies that conduct statistical activities, of which 13 are designated "principal statistical agencies" by the Office of Management and Budget. These agencies are located in their respective Departments (e.g. Bureau of Labor Statistics in the Labor Department, Census Bureau in the Commerce Department) with an oversight and coordinating agency (the Statistical Policy Office) in the Office of Management and Budget (OMB). The head of each statistical agency may be a career official (e.g. National Center for Health Statistics) or a Presidential Appointee with Senate Confirmation (e.g. Census Bureau) who reports to

[5] The Office of Management and Budget, which coordinates the federal statistical system, has identified several fundamental responsibilities of federal statistical agencies, including that they maintain both impartiality and the perception of impartiality. Office of Management & Budget, Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units, Federal Register Vol. 79, No. 231, pp. 71610-71616 (December 2, 2014).

[6] The Committee on National Statistics of the National Academies of Sciences has also identified independence from political and other undue external influence as a core principle for federal statistical agencies. National Academies of Sciences, Engineering, and Medicine. (2017). Principles and Practices for a Federal Statistical Agency, Sixth Edition, Washington, DC: The National Academies Press. https://doi.org/10.17226/24810

[7] Statistical information is that which can only be used for aggregate or summary purpose and which protects the confidentiality of individuals information.

senior officials in the Department in which the statistical agency is located. While OMB does not have line management authority over individual statistical agencies, OMB approval is needed for any data collection that is promulgated to ten or more respondents. Moreover, as codified most recently elin the Paperwork Reduction Act of 1995, OMB develops system-wide standards to ensure federal statistics are of high quality and that the burden on the public is minimized.

In addition to being more decentralized at the federal level than most other national statistical systems, the US system also differs from many in both the degree of oversight by the legislative branch and the extent of the authority of the heads of the statistical agencies. The United States Constitution is predicated on three branches of government: legislative, executive and judicial. This separation of powers was created because of the suspicion of the founders of an all too powerful executive which could become tyrannical and on the need to balance power in the government. The legislative branch (the Congress of the United States) enacts laws and both authorizes domestic activities and appropriates funds. It is the executive branch which is responsible for implementing the laws and regulations enacted by the Congress. Finally the judicial branch interprets the laws and adjudicates disputes between the other branches. Congress also has oversight responsibilities over the operations of the executive branch. Tensions exist over the extent and appropriateness of these oversight functions and these are exacerbated when different political parties control the Congress and the executive branch.

Each statistical agency must obtain yearly approval for its budget and proposed activities, and any significant changes in the agency's program during the year often must be approved by designated committees in the Congress. As a result, decisions which might be thought of as "professional decisions" best left to the heads of statistical agencies and the career staff in the Statistical Policy Office of the OMB often involve political appointees in the executive branch, elected members of Congress and career statisticians.

For example:

- The questions to be asked on the census and the operational activities of the census have to be approved by Congressional committees,
- Approval to suspend a survey or make significant changes in the sample size must be approved by political appointees in the Executive Branch and often by Congressional committees,
- Race and ethnicity classification standards are developed in consultation with Congress.

It is this environment that precipitates the concern with independence of statistical offices. Because of the importance of official statistics it is to be expected that elected politicians and their political appointees in the Executive Branch as well as the members of Congress who are also elected would be concerned about the activities of the statistical agencies and the numbers that they produce. Moreover, a process which involves these disparate parties including professional statisticians as well as elected politicians lends credibility and unity to the resulting decisions. What must be guarded against is a situation in which decisions are made as a result of partisan beliefs. There is no easy way to avoid this and the directives by the Office of Management and Budget and the Committee on National Statistics as well as global support from the Fundamental Principles may be seen as necessary but may not be sufficient. Transparency of the decision process and the voices of the professional statisticians, particularly if they have dissenting opinions, must be heard by the public. We will conclude with an example demonstrating how the Congress, political forces in the Executive Branch, the judicial system and the views of professional statisticians can come together.

4. Adding a Citizenship Question to the Decennial Census

On March 26, 2018 the Secretary of Commerce[8] concluded that a citizenship question should be added to the 2020 Decennial Census. It was asserted that this proposal was intended to produce information on citizenship at the census block level[9]. The information was said to be necessary for implementation of the Voting Rights Act. Currently, citizenship information is available from the American Community Survey at the census block group level[10]. The Census Bureau can provide estimates of block data from block group data by using statistical modelling techniques.

As a result of the Secretary's decision several lawsuits[11] were filed to overturn this decision. These lawsuits were filed by a variety of interested parties including multiple states, organizations such as the American Civil Liberties Union, and private parties. They alleged, inter alia, that adding the

[8] The Constitution provides that the conduct of the decennial census shall be determined by the Congress and the, at that time, Republican controlled congress delegated that authority to the Secretary of Commerce.

[9] A census block is the smallest geographic unit used by the Census Bureau for tabulation of 100-percent data (data collected from all houses, rather than a sample of houses).

[10] A census block group is a geographical unit used by the Census Bureau which is the next largest geographic area than a block. It is the smallest geographic unit for which the bureau publishes sample data, i.e. data which is only collected from a fraction of all households. A block-group generally contains between 600 and 3000 people

[11] Two of the authors of this paper – Katherine Wallman and Hermann Habermann – served as experts in one or more of the lawsuits.

question would cause a reduction in participation in the census by minorities. This would result in a less equitable political representation for some groups and an unjust disparity in the allocation of federal funds. As a result of disclosure forced by the lawsuits it was discovered that prior to the Secretary making his decision the Census Bureau had counselled the Secretary against adding the citizenship question. The Census Bureau had determined that:

- There was insufficient justification of the need for citizenship data at the block level,
- Even if one accepted the need for block level data there was a less-costly and better-quality alternative based on administrative records, and
- There was evidence that minority response on the decennial census would be adversely affected.

Two of the lawsuits have concluded, and in each case a Federal judge has ruled for the plaintiffs and determined that partisan political factors influenced the Secretary's decision to add a question on citizenship and have ordered that the question be removed from the decennial census. The United States government has appealed to the Supreme Court of the United States and their decision is expected before the ISI meeting in August. The authors believe that the Secretary's decision to add a citizenship question against the advice of Census Bureau professionals, and his decision to further create a current, comprehensive statistical reference list on citizenship, risks undermining the credibility of the Census Bureau and the 2020 Decennial census as well as the professional staff of the Census Bureau.

5. Conclusion

As this example illustrates, the question remains open as to whether the statistical agencies in the United States have the necessary degree of independence from a partisan political process. Although, as mentioned previously, the concept of independence is not explicitly considered the Fundamental Principles the international community may want to give some attention to examining the concept. Certainly, independence by statistical agencies must have limits. These agencies are funded in whole or mostly by tax payer funds and so one can expect not only oversight and accountability but also interest in the operations and procedures of the statistical agencies. How far should this go? Should the questions on the decennial census be approved by the legislature? If a statistical agency wants to use sampling to identify vacant houses which should be deleted should the legislature become involved? Beyond the decennial census, should changes in the System of National Accounts that involve accounting for climate changes be subject to veto by elected officials? How much authority in hiring should heads of agencies have? Should these be different from other civil service executives? Should the head of an agency have the ultimate authority to determine what data are released? As information becomes more important and official statistics struggles to remain relevant these are only some of the questions that will continue to need attention.

Parliament and official statistics in a multinational context: An EU perspective

Maria João SANTOS, Cristina PEREIRA DE SÁ
Eurostat, European Commission

Abstract
The European Parliament is a key player in the democratic process. The paper describes the multiple roles of the European Parliament in EU official statistics: providing a forum for political debate and decision-making; acting as a co-legislator adopting and amending legislative proposals; deciding on the EU budget together with the Council; supervising the work of the European Commission and other EU bodies; and cooperating with national parliaments of EU countries.

Keywords
Parliament; Official statistics; Governance; Accountability

1. Introduction

Across the world, countries, regions and multinational / supranational entities have found different ways of establishing the governance for the development and production of official statistics. One of the most important actors, the statistical office, can be located within the administrative structure but also as an autonomous agency outside the main branch of the executive. The delimitation of the statistical system (be it national or multinational) is crucial to ensure credibility and thus enable users, providers and respondents to trust the producers, the processes and the products of official statistics.

The relationships between all actors in the statistical system (producers, users, supervisors, etc.) are established via a clear governance structure or organisation imposing obligations and reporting procedures that are binding for the whole system. In addition, non-legally binding instruments also set standards in that respect; examples are the UN Fundamental Principles for Official Statistics which the global statistical system is encouraged to follow, and the European Statistics Code of Practice to which the members of the European Statistical System commit themselves to adhere.

The Cambridge Dictionary defines governance as 'the way that organisations or countries are managed at the highest level, and the systems for doing this'. As a consequence, good governance may be achieved by both the creation and use of arrangements that ensure consistency of processes, transparent decision-making and adequate (external) oversight. This must cascade from the highest level of the organisation. Good governance has positive benefits for the reputation of the statistical office, its general culture (a culture of excellence), the transparency of its decision-making procedures, risk management, etc.

2. The European Parliament's role in the EU

The European Parliament (EP) is an important actor within the governance established by the European Union (EU) for the development and production of official statistics, as a key forum for political debate and decision-making at the EU level. The Members of the European Parliament are directly elected by voters in all Member States to represent people's interests and to make sure that EU institutions work democratically.

Apart from its legislative power as co-legislator, the EP also has budgetary powers, both shared with the representatives of EU governments in the Council. In this regard, Parliament and Council decisions about annual expenditure and revenue must fall within the annual spending limits laid down in the EU's long-term financial plan, the Multiannual Financial Framework (MFF), negotiated once every seven years.

Once the EU budget is adopted, the European Commission is responsible for its implementation (other institutions are in charge of their own administrative budgets). The European Parliament exercises democratic oversight to make sure that the Commission and the other institutions deal properly with European funds: on a recommendation from the Council of the European Union, it decides whether to grant the 'discharge', i.e. final approval of how the budget for a specific year has been implemented. The Parliament may also make recommendations to the Commission on the execution of the budget. In fact, it deals in a similar manner with the approval of the accounts of other institutions, including its own administrative budget.

Moreover, the EP has a range of supervisory and control powers, which allow it to exercise oversight over other institutions, to monitor the proper use of the EU budget and to ensure the correct implementation of EU law.

Regarding the European Commission (the guardian of the treaties and the EU's executive arm), the EP has the right to approve and dismiss it. Since 1994, commissioners-designate have been required to appear before the EP hearing and upon appointment.

The EP ensures democratic control over the Commission, which regularly submits reports to Parliament including an annual report on EU activities and on the implementation of the budget. Once a year, the Commission President gives a State of the Union address at a plenary parliamentary session. The EP regularly invites the Commission to initiate new policies and the Commission is required to reply to oral and written questions from Members of the EP.

3. Annual and multiannual programming

The Commission (Eurostat) prepares its multiannual work programme in collaboration with Member States. The Director-General of Eurostat has the responsibility to define the process by which the various stakeholders and producers provide the information necessary for the planning process. Eurostat and the European Statistical System obtain information about user needs, different data sources and the outputs.

The multiannual European Statistical Programme focuses on long-term challenges and strategic developments of the statistical system as a whole, including elements such as vision and priorities for the future, activities and projects implementing those priorities, aspects needing improvement, needs expressed by users and stakeholders, and resources needed to achieve those objectives.

4. Accountability of European statistics

Accountability to the supervisory bodies, to users and to the public at large for what the national statistical system has produced with the given resources is essential, and the adopted programmes are the benchmark against which such a performance is assessed. Therefore, in the middle and at the end of the period covered by a programme the Commission prepares and presents to the European Parliament and the Council reports evaluating the extent of implementation.

In addition, the European Parliament holds the annual hearings of the Director-General of Eurostat with the relevant EP committee (Committee on Economic and Monetary Affairs, ECON), as foreseen in Regulation 223/2009 on European statistics. This statistical dialogue is a useful opportunity for the Members of the EP to gain a better insight into − and an understanding of − key statistical issues.

Accountability is also ensured via the annual report of the European Statistical Governance Advisory Board. ESGAB was established by Decision 235/2008/EC of the European Parliament and of the Council of 11 March 2008 to provide an independent overview of the European Statistical System (ESS) as regards the implementation of the European Statistics Code of Practice. Its tasks include the preparation of an annual report for the European Parliament and the Council on the implementation of the Code by Eurostat and the ESS as a whole. ESGAB also advises the Commission on appropriate measures to facilitate the implementation of the Code.

The 2018 ESGAB Annual Report was transmitted to the European Parliament and the Council on 22 January 2019, followed by an exchange of views between the ECON Committee and the Chair of ESGAB on 28 January. The report mainly focuses on the achievements of the ESS over the past years and on the new challenges and opportunities such as an increased engagement with stakeholders and particularly with users, the development of partnerships with the private sector, the fight against disinformation, etc. Importantly, it also reports on progress by Eurostat and the ESS as a whole as regards implementation of previous ESGAB recommendations, and by Member States on the implementation of the improvement actions emanating from the 2013-2015 round of peer reviews on assessment of compliance with the European Statistics Code of Practice.

5. The European Parliament in the statistical legislative process

The EU uses a variety of legislative procedures to adopt laws. The procedure followed for a legislative proposal depends on the type and subject of the proposal. The EP and Council jointly adopt the vast majority of EU laws (ordinary legislative procedure), while in specific cases a single EU institution can adopt legislation alone. The national parliaments of EU countries are consulted on all Commission proposals, and any changes to the EU treaties require the agreement of every EU country.

Most EU laws in the statistical domain are adopted using the ordinary legislative procedure, in which the European Parliament (directly elected) and the Council of the EU (representatives of the 28 EU countries) have equal say. The Commission submits a legislative proposal to the Parliament and Council, who must agree on the text for it to become EU law.

Via a series of readings of a proposed law, Parliament and Council review and amend the text. If the two institutions agree on the amendments, the proposed law is adopted. If the Parliament and Council cannot agree on amendments, a second reading takes place. If no agreement is reached at the second reading, the proposal is put before a 'conciliation committee' made up of equal numbers of Parliament and Council representatives. Commission representatives also attend the meetings and contribute. Once the committee reaches an agreement, the text is sent to the Parliament and Council for a third reading, so it can finally be adopted as law.

6. The European Parliament as user of official statistics

Official statistics underpin the development and implementation of EU policies. Making the appropriate policy choices is key to delivering higher and fairer growth, better jobs and a stronger capacity to smoothen the impacts of global economic cycles. A consistent set of priorities is essential to guide national reform plans and complement efforts made at EU level. Without high quality official statistics, such objectives are not possible.

The EP uses official statistics in the drafting of its numerous reports and opinions. It has a dedicated service to draft many of those reports, studies and opinions - the European Parliamentary Research Service (EPRS)[1].

The EPRS mission is to provide Members of the European Parliament and parliamentary committees with independent, objective and authoritative analysis of, and research on, policy issues relating to the European Union, in order to assist them in their parliamentary work. The EPRS provides a comprehensive range of products and services, backed by specialist internal expertise and knowledge sources in all policy fields, so empowering Members and committees through knowledge and contributing to the Parliament's effectiveness and influence as an institution.

Eurostat receives around 120 requests from the EP services per year regarding methodological or data support. Eurostat works closely with the EPRS by providing data for many of its reports. One recent example of this collaboration was the interactive infographic on lifelong learning in the EU published recently by the EP[2]. For this graphic, the EPRS used data from a variety of sources, including data published by Eurostat. Another recent example is "Gender equality in the EU's digital and media sectors[3]".

The EP is also a member of the European Statistical Advisory Committee (ESAC), established by Decision No 234/2008/EC of 11 March 2008 to assist the EP, the Council and the Commission in ensuring that user requirements are taken into account in determining the strategic objectives of statistics in the European Union. According to Art 3 of that Decision, "At the request of the European Parliament, the Council and the Commission the Committee shall deliver an opinion on matters relating to user requirements and costs incurred by data suppliers in the development of the Community's statistical information policy, in the priorities of the Community statistical programme, in the evaluation of existing statistics, in data quality and in dissemination policy."

The importance of identifying priorities and facilitating exchanges among user groups has grown in recent years in the light of the challenges facing statistical offices: reduced budgets and the pressure for efficiency gains whilst coping with demands for new statistics. ESAC's main purpose is to bring together the views of users and suppliers of statistics with those of producers in the EU. How is this done in practice? During the preparation of the EU statistical programmes, ESAC is consulted and puts forward its view on the

[1] http://www.europarl.europa.eu/thinktank/en/researchbypublicationtype.html

[2] http://www.europarl.europa.eu/thinktank/infographics/lifelonglearning/index.html

[3] http://www.europarl.europa.eu/RegData/etudes/BRIE/2018/614695/EPRS BRI(2018)614695 EN.pdf

proposed statistical programmes, taking account of users' requirements and producers' concerns in the production and dissemination of official statistics.

Via its membership of ESAC, the EP, an important user of official statistics, has a channel for influencing in advance the development and production of official statistics.

7. Conclusions

We have described the multiple aspects of the role of the European Parliament in the statistical domain in the European Union. After this description, the following question could be raised: what would be the optimal model for involvement of the parliament in the governance of the statistical systems?

As key institutions in a democratic system and by their representative role, parliaments can promote good governance by encouraging citizens' participation and involvement in the decision-making processes and, in turn, by making political institutions accountable to the people. Parliaments are no longer bodies with the single role of enacting legislation: they are the voice of the citizens who have elected them and are increasingly becoming closer to citizens.

Ensuring balanced priority-setting, providing a high degree of legitimacy and being vigilant about response burden are crucial means to achieving good governance. Parliaments can further contribute to that by strongly pushing for accountability, monitoring and reporting and by empowering citizens to be more engaged in such a process.

To achieve this goal, parliaments need to be given the necessary means (capacity and resources) but also to fully integrate their public participation role with their other 'more traditional' structures and functions.

Sound Methodology for Sound Official Statistics

Eric Rancourt
Statistic Canada

Abstract

The information society has evolved to a point where data and pieces of information of all kinds are readily available at our finger tips. All these digital fragments can be informative or misleading. It is simply not possible to understand and/or defend a data point when it stands on its own; it requires a context and a purpose to become a meaningful piece of information. In this data day and age, contexts, purposes as well as tools, technology and methods are overwhelmingly abundant. However, how does one know which are the most appropriate for their purpose? While National Statistical Offices (NSO) might have been lighthouses whose beacon people readily followed, they now have to explain and justify themselves for constituents to see the value of the information they emit. It is the responsibility of the NSO to be able to not only produce quality data, but also to enable potential users to understand what it produces. This article will use the example and values of the Methodology Branch to provide an example of a frame for meeting some of the rigour and transparency challenges that NSOs are facing.

Keywords

Valid statistical inference; Rigour; Quality; Scientific approach.

1. Introduction

National Statistical Offices produce information that constitute Official Statistics. This information stems from data needs that are usually expressed in statistics acts (e.g. Canada, 2018) or from data needs expressed by decision makers for which data was gathered and then statistics produced. The description of processes and methods used to produce such information is usually provided and explanations are readily available to users who are interested in increasing their understanding related to these processes (see for example the Statistics Canada's Policy on informing users of data quality and methodology, Statistics Canada, 2000). Sometimes the quality of statistics or the actual production of statistics might be questioned, requiring that the processes and methods are not only explained/detailed but also justified and defended. In other situations, nothing is asked and nothing needs to be defended; it is simply that attention is drawn elsewhere - to other data sources that may not have the quality that is produced by the NSO. In all three cases, it is important to have solid frameworks on which to stand, whether it is for subject matter, methodology, standards, governance, data stewardship, communication or any other area of expertise. This paper considers the case of methodology at Statistics Canada.

2. Modernization at Statistics Canada

Over the last few years, society has evolved to become a digital world where data is a center piece to decision making and everyday life. Nowadays, people are providing and receiving information through their cellular phones and other devices almost on a continuous basis. This data revolution has dramatically altered the statistical landscape and left NSOs with little choice but to change their methods of operation. As the traditional approaches are not sustainable to meet information needs in a rapidly changing environment, Statistics Canada has embarked on a modernization agenda (Arora, 2018) built around five pillars:

- A more user-centric approach to data needs;
- Increased and much more dynamic collaboration with partners;
- Using leading-edge methods;
- Providing capacity-building support to other actors of the national statistical system;
- A modern workforce and workplace that are suited to the times and more flexible to changing contexts.

This agenda is being partly implemented through pathfinder projects aimed at adopting new approaches to producing information, all the while moving forward in order to learn by doing. Four program areas have been identified as pathfinders:

- Measuring aspects of cannabis consumption (in the pre- and post-legalisation context);
- Producing housing statistics (based on administrative data from cities integrated with information already available at Statistics Canada);
- Enhancing the measurement of tourism-related activities (using alternative data for levels of expenses and movement of people);
- Increasing the output of statistics to better assess low-carbon economy (by better integrating more sources of data).

These ground-breaking programs have all been undertaken and the results each has produced are paving the way for Statistics Canada's modernization in various other areas.

3. Sound Methodology

3.1 Vision

The core values of effective statistical systems are defined by Fellegi (1996) as legitimacy and credibility. One of the elements of credibility is the use of sound methodology. As part of the modernization initiative, the Methodology Branch within Statistics Canada has adopted a new vision in 2019 that is based on a number of guiding principles that are consistent with the corporate modernization approach. The new Methodology vision is centered on three statistical values:

- Promoting and supporting the production of *valid statistical inference* in a more explicit fashion. That is, ensuring that estimates are supported by sound statistical methods that allow users to draw the right conclusions.
- Developing *quality* approaches as well as quality measures both for the statistics produced and the various steps of data processing in the new environment of increased uses of administrative data by dwelling on current approaches;
- Injecting *rigour* into all aspects of statistical production in a more deliberate and documented way. This means using very thorough and careful approaches in all steps of statistical programs.

These values underpin and guide our decisions for priorities and ensure sound and coherent approaches across the various programs. The values are complemented by two principles:

- Prospective work: There is a concerted effort to increase significantly the percentage of work devoted to research and development, enabling us to more proactively position the statistical programs in order to meet future needs. It also means that more resources are to be directed towards activities that have impacts on future cycles of surveys. This process takes place as part of both annual budget discussions as well as regular progress review meetings.
- Open Scientific Management: Transparency and the scientific approach are to be applied not only to statistical issues, but also to the management process of both projects and human resources.

3.2 Scientific Approach

Scientific principles are part of the United Nations Fundamental Principles of Official Statistics (Principle 2.). In order to anchor the vision in a solid basis, the scientific approach, which is already the underlying system of the work in Methodology at Statistics Canada, has been made more explicit in programs and is being enhanced. This section presents the approach and its enhancements.

3.2.1. The basic scientific approach

Application of the scientific approach means that a series of steps are followed with discipline to ensure that rigour and defendable approaches are used in statistical programs and in management practice. The steps of the approach are summarized below:

1. Needs: A problem or issue is identified and requires data to be examined/collected or a decision to be made;
2. Observation: Existing or passed data are analyzed, the literature and previous approaches are studied, and available methods are considered;
3. Hypotheses: Hypotheses are formulated and/or a design is elaborated with the intent of testing a certain view;
4. Test / Do: A test is conducted on the first iteration, and possibly in more than one, and then the approach is implemented;
5. Analysis: Results are studied and conclusions are drawn;
6. Communication: Findings are communicated to peers and open for constructive criticism;
7. Iteration: The process is iterated after the test for subsequent tests or for implementation.

Needs ⇒ Observation ⇒ Hypotheses ⇒ Test/Do ⇒ Communication

To enrich the scientific approach, the three values of the Branch are made present in these steps as:

- Valid statistical inference is promoted and imbedded into step 3 when formulating hypotheses and developing the design;
- Quality is sought at all steps but plays a more prominent role in step 4 where it is measured (or estimated);
- While rigour is implicit in the scientific approach, it is actively pursued at each step.

3.2.3 An Enriched Scientific Approach

The scientific approach has always served the Methodology Branch well, but to provide a broader perspective on needs, issues, methods and approaches, a check point is added to consult in order to ensure that the rest of the sequence is viable and appropriate. The sequence of steps is:

1. Needs: A problem or issue is identified and requires data or a decision;
2. Observation: Existing or past data are analyzed, the literature and passed approaches are studied, and available methods are considered;

3. Hypotheses: Hypotheses are formulated and/or a design is elaborated with a view to test;
4. **Check point: Once hypotheses and a design are developed, a pause is made to ensure risks, sensitivity, ethics, legal or other considerations are adequately taken into account;**
5. Test / Do: A test is conducted on the first and possibly in more than one iteration, and then the approach is implemented;
6. Analysis: Results are studied and conclusions are drawn;
7. Communication: Findings are communicated to peers and open for constructive criticism;
8. Iteration: The process is iterated after the test for subsequent tests or for implementation.

Needs ⇒ Observation ⇒ Hypotheses ⇒ Check point ⇒ Test/Do ⇒ Communication

Finally, progress is/should not only be communicated at the end of the cycle, but throughout the cycle.

3.3 Materialization of the approach in the Methodology Branch

At Statistics Canada, methodology services are centralized in the Methodology Branch where expert resources are assigned to support projects. This enables the development of consistent methods in surveys and facilitates knowledge exchange between methodologists working on different programs. Still in this structure, there are numerous interactions between methodologists and subject matter and other specialists. The Methodology Branch is centralized, but does not work in isolation.

Whether the work is oriented towards research, development or production, in all cases the scientific approach is followed. This approach has traditionally been implicitly imbedded into the methodological practices, but substantial efforts are deployed to make the approach even more explicit. The aim is not only to increase transparency but also to actively promote rigour through a consistent principled-based approach. Concretely, the scientific approach materializes itself as follows:

1. Needs. A proactive approach, centered on users and focusing on prospective work to identify needs sooner than before, has been increasingly adopted in the wake of Statistics Canada's modernization and Methodology's new vision. This allows methodologists to gain a head start in the application of the scientific approach.
2. Observation. In Methodology, this often takes the form of a literature review, accompanied by research on what was done in the past for the

same or similar programs and how international colleagues have approached the issues. It may also involve conducting descriptive analysis using prior information from surveys or administrative files.

3. Hypotheses. Based on the evidence obtained from prior information, from the literature, methods are developed and planned. This may require development or extension of methods; it may also involve the adoption of already-existing methods or an adaptation of these.
4. Check point. Once hypotheses are formulated and plans are established, the approach / method is presented to a technical committee and, when appropriate, to Statistics Canada's Advisory Committee on Statistical Methods (ACSM) for further guidance. Depending on the impact and the reach of the method, it may also be presented to a subject matter steering committee, a subject matter advisory committee and/or a senior management committee. These committees provide the breadth and depth needed for scientific, practical, management, ethical and social appraisal.
5. Test / Do. It is part of our best practices to test the approach on a small scale in order to gather further information that will be used to refine the hypotheses. This could mean conducting a pilot, producing a beta version of a computer program, or conducting simulation studies on real or artificial data. If a first iteration of the scientific approach has taken place, then this step is the actual implementation of the method.
6. Analysis. Study of the results takes place and decisions are made. At this step a method is selected over other options and implementation can begin.
7. Communication. The methods implemented are documented and shared with colleagues as well as subject matter specialists. In some cases, methods and findings will lead to an internal working paper, or be submitted to a conference or a refereed journal. Similarly, research studies will lead to an internal report, but may often become a more official paper presented at a conference or submitted to a refereed journal.

   Communication is not limited to the "communication" step as it need to pervade to the whole approach. For example, discussion taking place with subject matter specialists may lead to new projects being launched. Or, once a set of hypotheses have been formulated, discussion among peers and external experts ought to take place. The check point is, by definition, a communication process and the test will involve planning requiring careful/effective communication.
8. Iteration. Normally the first loop is a test and the second one is the implementation of new methods. However, there could be more than one loop.

4. Sound Official Statistics

Application of the scientific approach takes place regularly throughout Statistics Canada. One could consider numerous activities such as price index development, production of national accounts and field and processing operations, IT development and even HR, Finance and Communications as a non-exhaustive list where the scientific approach is applied or could be more explicitly applied.

We present below a brief outline of the scientific approach to the modernization agenda for Statistics Canada to illustrate how it is implemented:

1. Needs. As society is rapidly changing and becoming increasingly complex through the data revolution, there are new needs for real-time and more granular information to be released.
2. Observation. Statistics Canada has conducted an environmental scan to assess the current state of data needs, the pertinent stakeholders, the activities of other NSOs and academics as well as the types of data that are already available.
3. Hypotheses. A modernization strategy was developed to focus on user-centric service delivery based on such pillars as leading-edge methods and data integration; statistical capacity building and leadership; sharing and collaboration; a modern workforce and a flexible workplace. The hypothesis is that modernizing statistical programs in this way will improve several aspects of Statistics Canada such as relevance and timelines.
4. Check point. The strategy was discussed in committees within Statistics Canada as well as with other departments, the Minister of Innovation, Science and Economic Development and with other national statistical offices. A strong governance was put in place with a modernization head and a modernization steering committee.
5. Test / Do. Four pathfinder projects were identified to test the modernization agenda. They were the measurement of cannabis, transition to a low-carbon economy, international tourism to Canada and foreign-owned housing. Tests also involve new leading-edge methods such as modelling in the form of projects involving small area estimation or machine learning.
6. Analysis. Lessons learned were studied and a decision was made to continue and expand modernization based on intelligence gathered by the pathfinder projects.
7. Communication. Employees have been fully informed during the process as were also stakeholders, decision makers and Canadians. For example, an internal modernization bulletin is regularly issued to keep all employees informed.

8. Iteration. Now that the pathfinder projects were successfully accomplished, the modernization approach has been expanded to many more programs. As well, lessons learned related to new methods (e.g. machine learning) and new approaches (e.g. housing program based on no survey data); only administrative data) are promoted to other statistical programs.

5. Conclusion

This version of the scientific approach is increasingly used more formally within Statistics Canada's Methodology Branch as a frame for its management activities and development of statistical methods. Such a frame provides a solid backbone for all activities and enables statistical and managerial activities to be rigorous, transparent and defendable. As a result, it contributes to maintaining or increasing trust. Adaptations of this framework could be used by other NSOs to enhance their own communication practices, which have a direct influence on the level of trust they inspire.

Acknowledgements

The author would like to thank Stéphane Dufour, Claude Girard, Andrea Leigh MacMillan, Claudia Sanmartin and Nora Taki for their comments on the paper.

References

1. Arora, A. (2018). *Modernizing the National Statistical System – Stakeholder Consultations*. Catalogue 89200003. Statistics Canada. ISBN 978-0-660-31580-5.
2. Canada (2018). *Statistics Act*. Minister of Justice, http://laws-lois.justice.gc.ca, Ottawa.
3. Fellegi, I.P. (1996). Characteristics of an Effective Statistical System. *International Statistical Review*. Vol 64, No.2, pp.165-187.
4. Statistics Canada (2000). *Policy on informing users of data quality and methodology*. Statistics Canada, Ottawa.
5. United Nations (2014). *Fundamental Principles of Official Statistics*. Resolution adopted by the General Assembly. A/RES/68/261.

Building and maintaining trust in today's data landscape

Joanne Hillermann
Statistical Data Integration Division
Australian Bureau of Statistics

Abstract
The availability of big data to National Statistical Offices (NSOs) and their innovative use has created opportunities to enhance the quality of existing statistical series, often at lower cost; as well as enable new insights into long standing public policy questions, primarily by combining big data from various sources, including with Population Census data. The opportunities presented by big data to NSOs also come with many challenges, particularly at a time when the public's trust in the ability of public and private institutions to keep their data safe is relatively low. This paper briefly discusses the opportunities of big data in the Australian context and then primarily focuses on the need for NSOs to build and maintain trust in the current data landscape. Trust in data use, which is sometimes referred to as "social licence", is defined and examined through multiple lenses. A framework for building and maintaining trust is also presented for use by NSOs. The framework considers a number of elements, including transparency and communication, engaging with providers, the role of expert advisors, privacy, and risk management. This paper concludes with an outline of the Australian Bureau of Statistics' (ABS) experiences developing this framework and plans to operationalise.

Keywords
Social license; Big data; Data integration

1. Introduction

The Australian Government Pubic Data Policy Statement declares that "data held by the Australian Government is a strategic national resource that holds considerable value for growing the economy, improving service delivery and transforming policy outcomes for the Nation".[1]

As data volumes are growing exponentially, so too is the potential value of this data. Effectively harnessing the value of data is vital to remain competitive in the global, digital economy.

[1] Australian Government Public Data Policy Statement. https://www.pmc.gov.au/sites/default/files/publications/austgovtpublicdatapolicystatement1.pdf (Link as of 24/04/2019.)

In this environment of increased demand for data and statistics, National Statistical Offices (NSOs) are being challenged to deliver the best possible statistical program, in more efficient and innovative ways. This demand presents both opportunities and challenges to make greater use of existing data sources.

2. Opportunities and Challenges

The availability of big data and large administrative data sources provides innovative opportunities for the ABS. Existing administrative or market data can be used to reduce costs, reduce provider load and improve the sustainability of statistical outputs. It can also improve the accuracy, relevance, consistency, interpretability and timeliness of the statistical work program. New statistical products can enable new insights into long standing public policy questions, particularly from combining big data from various sources, and combining with collected data such as the Population Census or national surveys.

The ABS has used big data sources to replace direct collection of data. For example, since 2014, the ABS has used scanner and web-scraped data in the production of the Australian Consumer Price Index (CPI). This has reduced the cost of the collection, increased the size of the sample data, and ultimately increased the accuracy of the CPI.[2]

Big data sources are being investigated for use in improving statistical processes. For the 2021 Australian Census of Population and Housing, administrative data sources are being investigated for a range of purposes, including use in assessing whether dwellings are occupied. Such data can make enumeration processes more efficient by focussing follow-up activity.

The ABS also makes extensive use of large administrative datasets for data integration projects. The Data Integration Partnership for Australia is an investment to maximise the use and value of the Australian Government's data assets. Data integration and analysis enabled through this partnership has created new insights into important and complex policy questions.[3]

However, innovative uses of data can bring challenges in terms of public acceptance of the use of their data in new ways. Increased data use is happening at a time when levels of trust in Government in Australia (as evidenced by results from the Edelman Trust Barometer) are lower than the

---

[2] Leigh Merrington, ABS, Continuing to unlock the potential of new and existing data sources, IAOS Conference, September 2018. http://www.oecd.org/iaos2018/programme/IAOS-OECD2018 Merrington.pdf (Link as of 24/04/2019.)

[3] Another paper for this conference – A Case Study of Data Integration in Australia – highlights some key data integration projects and the potential for policy insights from analysis of the integrated data assets.

global average and have declined over the past decade. The barometer reported the percentage trust in Government in Australia has fallen from 53% in 2009 to 42% in 2019, with a low of 35% in 2018. In comparison, the global level of trust in Governments was 47% in 2019.[4]

A study on Data Governance in Australia[5] has shown that citizens are concerned about Government use of their data, particularly public trust in keeping their data safe. Citizens also desire transparency in data use, and for new data uses to be fair. The study reported that 29% of Australians think the Government can be trusted to use data responsibly and 27% believe the Government is open and honest about how data are collected, used and shared. The study also reported that, of the organisations asked about, the ABS is the most trusted with regard to data.

3. What is Public Trust in Data Use?

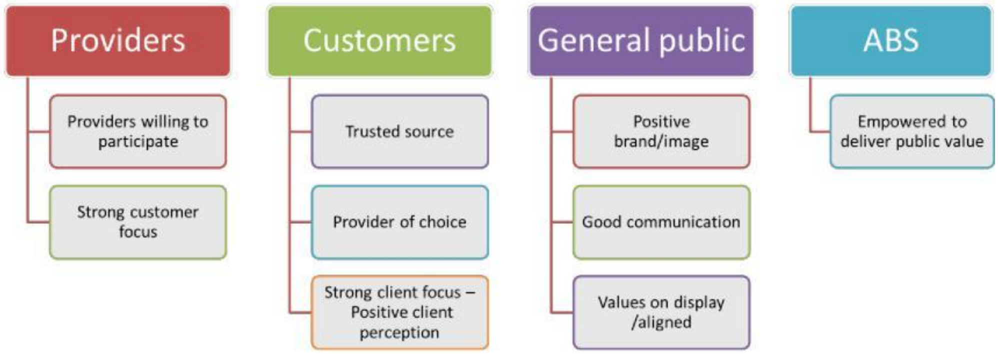
Public comfort or community acceptance with how their data is used is often referred to as "social licence". ABS prefers to use the term "trust" or "community acceptance" in data use rather than social licence as we believe trust is a more general term that is readily understood by the public. It also avoids any possible connotations that social licence can be bought. Trust must be built and maintained through behaviours and actions.

Diagram 1 outlines what trust looks like from various perspectives for the ABS. From the provider perspective (i.e. the people and business that provide their data to the ABS), high levels of trust mean they would experience a strong customer focus from ABS and be willing to provide accurate data. High levels of trust in ABS from the general public perception would be shown by a positive brand/image, a good level of understanding about how data is used, and a strong alignment of our values with public perceptions.

[4]2019 Edelman Trust Barometer. https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019EdelmanTrustBarometerGlobalReport.pdf?utmsource=website&utmmedium=globalreport& utm campaign=downloads (Link as of 24/04/2019.)

[5] Biddle, Edwards, Gray and McEachern, Public Attitudes Towards Data Governance in Australia, 2019. http://csrm.cass.anu.edu.au/research/publications/public-attitudes-towards-data-governance-australia-1 (Link as of 24/04/2019.)

Diagram 1. Perspectives of trust in ABS



In 2017, the Data Futures Partnership in New Zealand published useful guidelines for trusted data use which present eight key questions that organisations can answer to explain how they collect and use data, to better build trust with clients and the wider community.[6] The questions focus on value, protection and choice, and indicate areas of possible public concern about data use initiatives.

4. Need to Build Public Trust in Data Use

The ABS recognises the increasing need to build the trust of citizens in the ABS and use of their data. This is particularly important to an NSO like the ABS, as we rely on the willing cooperation of citizens and businesses to provide accurate information in order to produce high quality statistical products that inform Australia's important decisions. Data is our core business, not a by-product of an administrative process. We need to keep pace with, or positively influence, community attitudes to data use.

Recent 'data' events in Australia and around the world have provided a number of important insights for the ABS. These insights include:

- Conduct large data collection activities (e.g. Population and Housing Census) as a major change program;
- Deliver a good user experience that meets service delivery expectations;
- Be alert to a disruptive, changing environment, which impacts the public response; and
- Address public privacy and use of data concerns.

The lessons highlight the importance of good communications, citizen-centred design and engagement approaches for major data activities. In particular, NSOs need to explain how data will be used, the key benefits of data use, and how it will be kept safe.

---

[6] Path to Social Licence: Guidelines for Trusted Data Use. http://datafutures.co.nz/our-work-2/talking-to-new-zealanders/ (Link as of 24/04/2019.)

It is also important to recognise that the ABS operates as an entity within the Australian data system. Data initiatives or issues in one sector or jurisdiction can quickly change public perceptions about data use in general. For example, the possible introduction of new data sharing and release legislation in Australia later this year is likely to promote greater public awareness and discussion about data sharing and use and influence public opinion.

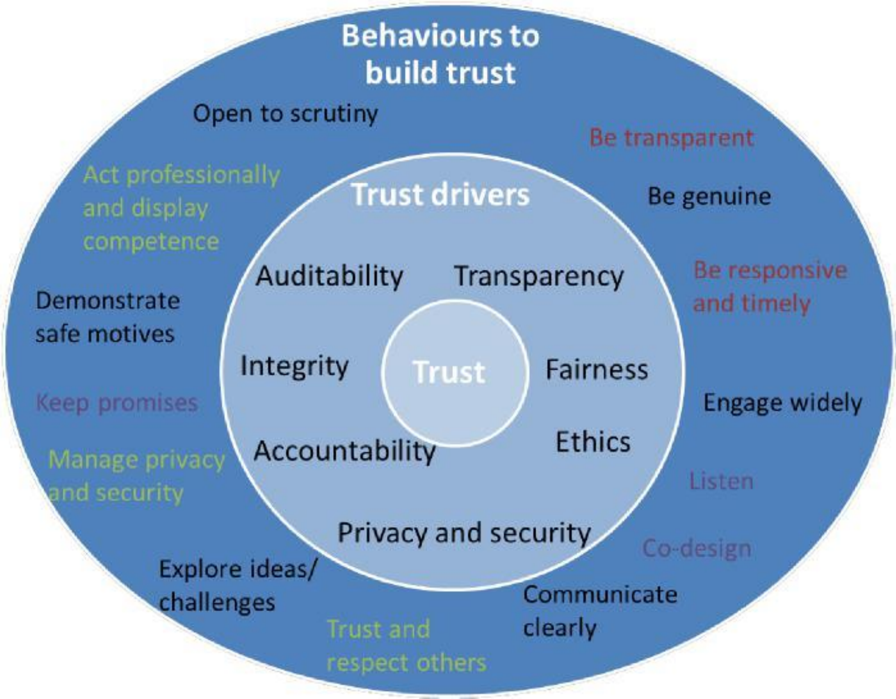
5. Abs Building Trust Model

The ABS Building Trust model comprises a conceptual model (that shows trust drivers and behaviours to build trust) and a strategic model (that outlines the domains of action for activities that will help build public trust in ABS data use).

Diagram 2 illustrates the conceptual model of trust. At a fundamental level, trust in a data organisation can be understood as the organisation having the right motives, keeping promises and demonstrating professional competence. These three characteristics show the level of trustworthiness of an organisation. However, trustworthiness is a necessary, but not sufficient condition for gaining community acceptance or trust in use of data. Trustworthiness needs to be accompanied by actions to demonstrate trustworthiness, which together will influence levels of community acceptance or trust. The model shows actions as trust drivers and behaviours to build trust. The trust drivers in the model are based on six essential principles for ensuring a trustworthy data system, recently published by the World Economic Forum.[7]

[7] World Economic Forum, Data Policy in the Fourth Industrial Revolution: Insights on personal data https://www.weforum.org/whitepapers/data-policy-in-the-fourth-industrial-revolution-insights-on-personal-data (Link as of 24/04/2019.)

Diagram 2. Conceptual model of trust in data use



While trust is sometimes seen as a binary condition, that is, you either have trust or you don't, trust is context specific and an organisation might be trusted in one area, but not in another. Also, from an aggregate level, there is a distribution of trust across the population. It would be very unlikely to gain the trust of 100% of the population for a particular data initiative, so we need to consider what level of community acceptance of an initiative is acceptable for the initiative to go ahead unchanged.

Trust can be viewed like capital, or a bank of credit, that can be drawn upon when needed. High levels of public trust can protect an organisation from reputational damage when risks or issues are realised.

The conceptual model provides a common language about trust and its drivers and is likely to remain fairly unchanged over time. In comparison, the strategic model outlined in Diagram 3 may change over time as the actions designed to build trust in data use evolve and change with experience.

The strategic model shows the six broad domains of action that are likely to drive ABS efforts to improve public trust in data use. The ABS already delivers in each of these domains but is actively considering options for improvement. Better communication and engagement with the public are expected to have the greatest impact on building community acceptance and trust in ABS data use.

Diagram 3. Strategic Model of trust in data use

Some of the ideas under active consideration are:

- Be transparent and communicate clearly
    - Improve communication with stakeholders and providers in early stages of consideration and design of projects and further explain how decisions are made
    - Provide more accessible information that demonstrates the value of data use, how data is protected and accountability mechanisms
- Meaningfully engage with citizens and act on their advice
    - Introduce a citizenpanel to take citizen views into account when determining acceptable (and ethical) uses of data and be more transparent
    - Greater use of co-design or deliberative engagement for selected data projects
- Manage risks and issues
    - Provide centralised support for issues management
- Provide good customer service
    - Improve provider and customer experiences in delivery of our services
- Be accountable and open to external scrutiny

  - Invite independent scrutiny and use of expert advisors to review our work program and practices
- Keep data safe
  - Improve cyber security protections
  - Improve communication about how data is kept safe

6. CONCLUSION

The ABS is acutely aware of the need to maintain and build the community's trust – both in the institution as a whole; and how the ABS collects, uses and disseminates data.

This paper provides both a conceptual model and a set of practical steps the ABS and other National Statistical Organisations (NSOs) can utilise to build trust with communities. Building trust is a multi-dimensional activity, with action needed across the spectrum of communication, engagement, privacy, data protection (to name some elements). Building trust in the NSO will enable the value of public data to be maximised for the benefit of communities.

The ABS will, in coming months, operationalise the model and progress the concrete actions that are expected to have the greatest impact. These will be reviewed and evaluated and further actions developed over time.

The mood of the twitterers in Mexico

Gerardo Leyva, Abel Coronado
National Institute of Statistics and Geography (INEGI)

Abstract

The mood of the twitterers in Mexico generates statistical information from georeferenced tweets that are posted daily from any point of the Mexican territory and reports the "positivity ratio", which relates the tweets with a positive emotional load to those that have a negative emotional load. The process uses sentiment analysis techniques that involve the labeling of a subset of tweets by human beings and the subsequent machine learning process that allows automating the identification of the underlying emotional charge, which happens to match the human criterion more than 80% of the time. This allows generating the highest frequency time series published by INEGI, since every day the computer updates the classification of about 100,000 tweets with information up to the previous day. The system, freely available on the Internet, allows consulting at national and regional level in annual, quarterly, monthly, weekly, daily and even hourly series. Also, it is possible to see the hashtags that dominated the scene and directly access news sites that help to associate the variations in the series with specific events. With a series of time that begins in 2016, the system shows dynamics consistent with what should be expected for a number of clearly identified events, such as the results of the US Presidential Elections on November 11, 2016 (negative), the increase in national petrol prices at the beginning of 2017 (negative) and the two large earthquakes of September 2017 (both negative), the triumph in Soccer of Mexico against Germany on June 2017 (positive) and the award of "The Form of Water" with the Oscar for the best film of 2018 (positive). This is a practical exercise in the generation of statistical information on inferred aggregate subjective well-being using entirely non-conventional statistical information, which can be refined to address particular topics in relation to how they are felt by the twitterers. This product is the result of combined research efforts of INEGI and national and international academic institutions and has been the basis for the development of other data science projects at INEGI.

Keywords

Georeferenced tweets; sentiment analysis techniques; machine learning process; positivity quotient; big data

1. Introduction

Official Statistics can be seen as a building supported by 3 pillars: censuses, surveys, and administrative registers. The so-called "data deluge", a by-product of the digital revolution, has generated the possibility of counting with a fourth pillar: the so-called "big data". We still don't know how important big data will become for the production of official statistics, relative to the other pillars, but now we can see that it is, at least, very promising. Even its more sceptical critics now accept that big data should play a role in the production of official statistics. Tough, problems of practical implementation remain; one of them being the access to the sources of information, which frequently is intermediated by enterprises whose main purpose is not necessarily the supply or generation of information for the common good.

Big data is frequently associated with the three "V": Volume, Variety, and Velocity. Two additional "V"s have been added more recently: Veracity and Value. All these "V"s are of course valid elements for the characterizations of a slippery concept, but we think they tend to put most of the weight on data itself rather than on the way it is used. Even if big data emerged from a series of technological innovations, its essence is more on the side of the way data is approached. So we can even think about a big data paradigm, according to which big data can be seen as a flexible approach to use and re-use the totality of a data set, structured or not, in a diversity of possible purposes, normally different to those that originated the information set in the first place.

It´s clear that extensive use of the term "big data" may turn it into a buzz concept, which is precisely what Dan Ariely captures in his famous cynical anti-definition: "Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it…". But beyond definitions, the fact is that more and more people around the world are getting familiar with what big data is, and many more are participating in its generation or simply using it. National Statistical Offices are obliged to look for ways to incorporate big data sources in their supply of information, for more than one reason: it is cheap and normally comes in high frequency and granularity. Among the many ways to engage with the possibilities involved in big data, this paper shows the way followed by the National Institute of Statistics and Geography (INEGI) to exploit Twitter in order to generate a web service reporting the mood of tweeterers in México on a daily basis.

So, INEGI can also claim to be doing it. It has decided to venture into the world of "big data" to explore the usefulness of non-traditional sources of information in order to link them with the generation of statistical and geographic information. As the first step in this direction, INEGI has undertaken in an experimental way the "sentiment analysis" by means of the implementation of machine learning techniques to the interpretation of the emotion that underlies the messages that are published on Twitter.

The machine learning techniques allow to train a computer to replicate the human criterion in identifying the emotional load of each tweet, either negative or positive; This, in turn, allows classifying the tweets and build an indicator that relates the number of tweets associated with a positive emotional charge (tweets positive) for each tweet associated with a negative emotional charge (negative tweets). We call this indicator the "positivity ratio" and define it as the number of positive tweets over the number of negative tweets for a given geographic area over a given period of time.

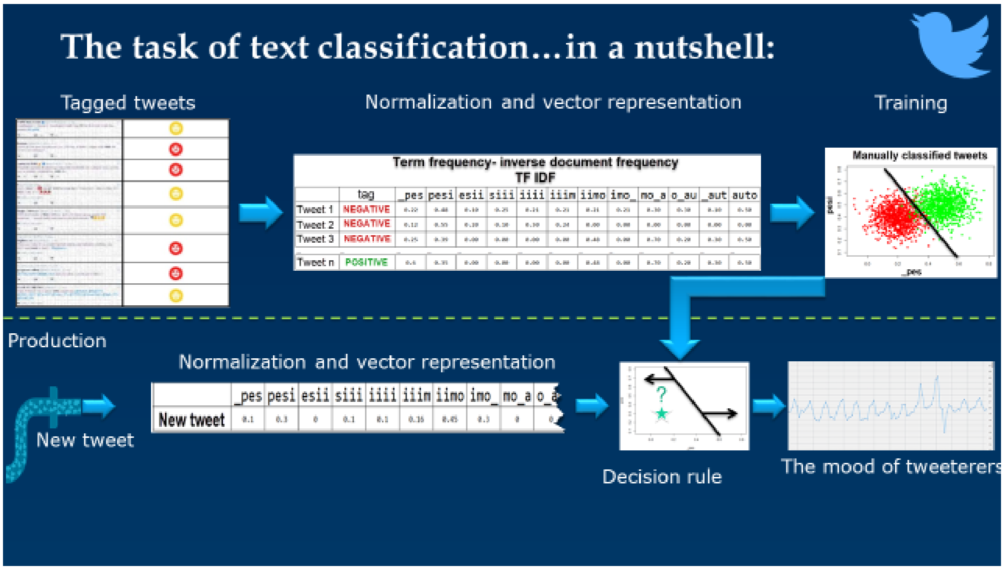
2. Methodology

To build the Mood of the Tweeters it was necessary to download hundreds of millions of tweets georeferenced to the Mexican territory, available through the free Twitter API. We started by manually labelling a relatively small set of training tweets. Human labellers (or taggers), were asked to gauge the underlying emotional charge (positive, negative or neutral) of each tweet. The resulting set of tagged tweets was then used to train a computer to replicate the human criterion needed to classify tweets in two categories: positive and negative. The training set was composed of about 54.000 standardized tweets, tagged by 5000 students of University TecMilenio and 5000 of our INEGI co-workers. Both sets of taggers were scattered all over the 32 federate entities (states) of the country. Human taggers from each state received tweets generated in the same state in such a way as to facilitate the proper interpretation of regionalisms. Each tweet was assessed by different taggers and each tagger evaluated the same tweet several times, so that information from inconsistent qualifiers could be identified and disposed of. Additionally, the underlying emotional charge of each tweet could be identified as each tweet was systematically qualified by different people as positive, negative or neutral, which helped to achieve a more robust qualification.

The tweets classified by humans were then used to train the computer, for which it was necessary to incorporate an assembly of classifiers developed with the support of experts in data science from INFOTEC and Centro Geo, two Mexican centres of research. After comparing among many alternatives, it was found that an assembly of 33 different Support Vector Machine (SVM) estimation procedures, each one run with a different 80% of the tweets tagged by humans and available for training and validation.

The classification was not run over words but instead, we used q-grams of different lengths. In fact, we found optimal to use q-grams of order 3, 4, 5, and 7. We also normalized text, gave polarity values to emoticons, implemented number, URL and user substitution, and transformed words to basic linguistic roots. TF-IDF scheme was used to achieve vectorization necessary for the classification. This allowed us to identify clusters of tweets based on chains of q-grams of different orders derived from each tweet. Since we trained the computer with a set of tweets tagged by humans, we were able to use SVM to estimate the multidimensional hyperplanes that best separate tweets that we knew were positive from those that we knew were negative.

Using a set of tweets classified by humans that were not used in the training of the machine, it was possible to establish that such assembly of classifiers allows the computer to properly classify 80 of each 100 tweets, which is a percentage of particularly high success among experiences in the field of sentiment analysis. Once with the computer properly trained to classify tweets, INEGI was given to the task of exploiting the results and presenting them. It should be noted that although the information published on Twitter is public, INEGI only reports aggregate data and at no time reports nominative or individualized tweets. Even in the tweets classification system (Pioanálisis) the tweets were anonymized before they were presented to the human taggers. So, every day INEGI reports the tweets from the day before, automatically classified by the computer. To do this, the computer simply uses the 31 SVM hyperplanes previously estimated and decides the positivity or negativity of each tweet according to a majority report rule. The following image summarizes the procedure:



Besides generating the automatic classification of tweets, one additional challenge was to present the results in an interactive, agile, friendly and attractive setting. So, through the web page of INEGI we presented a tool with the following characteristics:

- Ensuring Continuous collection of tweets in parallel, having an automated contingency plan for when failures occur.
- Adapting the automatic classification to the institutional platform and make every night the automatic interpretation of the emotional charge (positive or negative) of the tweets collected in the day.
- Storing the tweets labelled daily according to their emotional load in a NoSql database, process and add the tweets leaving the results available for consultation.
- Developing a new visualization that would allow the results to be represented yearly, quarterly, monthly, weekly or daily, and even per hour, both at national and state level.
- Distinguishing in each federative entity among the individuals that habitually tweet from it (local) and those that habitually tweet from another federal entity (visitors).
- Achieving a WEB application that is available on any platform, including mobile phones.
- Incorporating a filter component for dates, allowing the user to select a specific day range.
- Incorporating functionality to share visualizations on various social networks, as well as to download data.
- Incorporating an automatic tour that guides the user in the different functionalities of the system.
- Reporting the status of tweets collection over time.
- Presenting a menu to select series by federal entity, as well as their maps, which are updated and ordered automatically on a daily basis.
- Incorporating an updated methodological document.
- Offering a cloud of words with the #Hashtags most used on that day, redirecting directly to the Twitter site to see the detail of a particular #Hashtag, as well as links to external news sites, which are not the responsibility of INEGI, but could help to contextualize the positivity or negativity of the tweeters.
- Including a link to "Pioanálisis" which is the application in which new tweets are manually classified. The idea of this is allowing users to volunteer for helping to classify tweets.

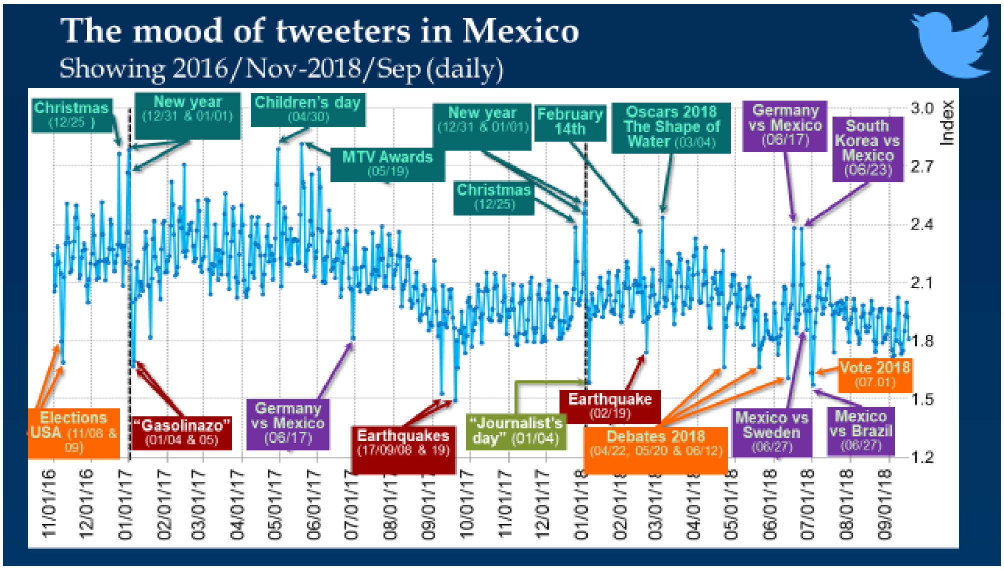
3. Result

The Mood of the Tweeterers in Mexico reports a positivity ratio from January 2016 with daily automatic updating, which can be viewed for the country as a whole and for each of its federal entities with annual frequency, quarterly, monthly, weekly, daily and even by the hour. In this sense, it is the exercise of generating statistical information with the highest frequency that

INEGI has generated so far. Also, due to its characteristics, the *Mood of the Tweeterers in Mexico* is a unique tool worldwide.

The *Mood of the Tweeterers* is part of an effort to measure welfare "beyond GDP" and in particular is linked to the agenda for measuring the subjective wellbeing, which seeks to extract from individuals information on how they are doing in life, although in this case it is not about self-reported wellbeing, as happens with traditional survey generated statistics. Instead, in this case we are talking about a form of wellbeing that is inferred from the underlying emotional charge present in the messages published on Twitter. We are taking advantage from the fact that people expressing themselves when using Twitter are letting us to know much more than the mere message: they are allowing us to glimpse into their hearts and minds. In this way, the evolution of the quotient of positivity allows to visualize how the positive and negative moods diffuse or become generalized to a greater or lesser extent among the whole twitting population and how this greater or lesser diffusion changes from one period to another and from a federative entity (state) to another. This allows us to take a look at aspects of national reality that acquire greater relevance to the extent that the use of digital social networks is extended over the Internet.

With a time series beginning in 2016, the system shows a behavior consistent with what should be expected for a number of clearly identified events, such as the results of the US Presidential Elections on November 11, 2016 (negative), the increase in national petrol prices at the beginning of 2017 (negative) and the two large earthquakes of September 2017 (both negative), the triumph in Soccer of Mexico over Germany in June 2017 (positive) and the award of "The Form of Water" with the Oscar for the best film of 2018. More relevant, the Mood of the Tweeterers in Mexico showed its capability to automatically capture the emotional impact of relevant changes in the environment with the earthquakes of 8 and 19 September 2017. The first quake had its epicentre in the southern part of the country, while the epicentre of the second one was very close to Mexico City. Both affected Mexico City, but the send one hit harder over the capital of the country. In both cases the positivity ratio showed sharp declines, expressing the negative emotional impact of the disasters. The following image shows some of the main dates that can be identified at the national level with the help of the system of “The Mood of Tweeterers in Mexico”:



In Addition, the tool allows to visualize how differentiated reactions were given by federative entity, so that while the City of Mexico (affected by both quakes) presents notable lows in the positivity both the 8th and the 19th of September, in Chiapas (affected only by the first earthquake) we only observe a fall the in the early morning of September 8th and in Chihuahua, a northern state completely unaffected by the earthquakes, there is no noticeable change in any of the two dates.

It is important to say that not all movements in the positivity ratio series have an obvious interpretation. In this sense users can complement the information they have with the one offered by the system to enrich their explanation of the mood variations of the tweeters. To the extent that users become familiar with the system and to the extent that they have longer time series, the same users will be able to refine the interpretation of the variations in time and space of the positivity quotient, as well as adapting this interpretation to their own analytical needs, which will enrich the knowledge about the aspects of reality that govern the dynamics of the mood of the tweeterers.

4. Discussion and Conclusion

It´s important for users to be clear that the Mood of the Tweeterers in Mexico is an experimental statistical exercise in which INEGI makes an effort to reuse information from social networks to transform it into Statistical information of public utility. Unlike traditional statistical sources such as surveys or censuses, social media information does not respond to a conceptual, statistical and operational design aimed at knowing specific aspects of our reality, but we can harness it to do just that. Thus far we have found the automatic classification of tweets gives out meaningful results.

Exploring big data as a fourth pillar for official statistics involves several advantages such as: reducing the costs of raising information by traditional methods such as surveys, alleviating the burden of respondents, having information with greater opportunity, contribute to modernization and innovation in the production of statistics, as well as devising new ways of measuring and monitoring Sustainable Development Goals (ODS or SDG).

However, the incorporation of Big Data to the generation of official statistics implies many challenges derived precisely from using for statistical purposes information that was not originally created with this in mind. The technical challenges implicit in the domestication for statistical purposes of non-traditional sources of information are not negligible. Finally, non-traditional sources of information tend to be in the hands of companies whose main mission is not the generation of information of public utility, which must be persuaded to collaborate.

The Mood of Tweeterers in Mexico is available on INEGI´s web page (www.inegi.org.mx) at: http://www.beta.inegi.org.mx/app/animotuitero/.

Granger causality in yield curves of different markets

Rituparna Sen
Applied Statistics Division. Indian Statistical Institute, India

Abstract
We develop time series analysis of functional data observed discretely, treating the whole curve as a random realization from a distribution on functions that evolve over time. The method consists of principal components analysis of functional data and subsequently modeling the principal component scores as vector ARMA process. We justify the method by showing that an underlying ARMAH structure of the curves leads to a VARMA structure on the principal component scores. We derive asymptotic properties of the estimators, fits and forecast. For term structures of interest rates, this provides a unified framework for studying the time and maturity components of interest rates under one set-up with few parametric assumptions. We apply the method to the yield curves of USA and India. We compare our forecasts to the parametric model of Diebold and Li (2006). We then use the method of Granger causality on the VARMA of Principal components scores for the two markets to test if one market drives the other.

Keywords
Functional Principal Component; Vector ARMA; Prediction; Asymptotics; Granger causality

1. Introduction

Functional data analysis (see Ramsay and Silverman (2005) for an overview) is an extension of multivariate data analysis to functional data where each observation is a curve, rather than a vector in R^n. An important feature of FDA is its ability to handle dependencies within each observation, especially smoothness, ordering and neighborhood. Actual observations can be at discrete and irregular points within the curve. The first step of FDA is to replace these actual observations by a simple functional representation. Spline-based approximation is the most commonly used method. Kernel or wavelet-based approximations are also used.

An important tool of functional data analysis (FDA) is functional principal component analysis (FPCA, see Castro et al. (1986); Rice and Silverman (1991)). Functional processes can be characterized by their mean function and the eigenfunctions of the autocovariance operator. This is a consequence of the Karhunen-Loève representation of the functional process. The components of this representation can be estimated. Individual trajectories are then represented by their functional principal component scores, which are available for subsequent statistical analysis. This often leads to substantial dimension reduction.

Most of the development in FDA has been with independent and identical replications of function valued data. This permits the use of information from multiple curves to identify patterns. However, in certain situations, it is unrealistic to assume that the functions are independent. We do need some process structure. One idea to follow up here is to work with the replication principle implicit in stationary time series, where the values of the process are functions. This problem of dependent functional observations is gaining popularity only recently. Besse et al. (2000) develop an AR(1) model for forecasting climatic variations. Kargin and Onatski (2008) use an AR(1) model for forecasting Eurodollar futures. H¨ormann and Kokoszka (2010) study weakly dependent functional processes, but they ignore the issue of smoothing. This is common to a lot of work following Bosq (2012), eg. Bosq (2014) and Aue et al. (2015), where the theory is developed assuming that the functions are observed continuously. In practice, however, we only observe the functions at a dense but discrete subset of the support, often with measurement error. Then we need to interpolate smoothly to infer about the whole function. This raises new questions about the behavior of the estimators. We develop the theory, where the functions follow a stationary ARMA(p; q) model and are observed discretely with noise. We start with kernel smoothing, followed by dimension reduction using FPCA. Based on the time series of the first few significant principal components, we fit a VAR or VARMA model. We provide techniques for estimation of the model parameters and selection of the optimal model.

The methods developed are applied to the modeling and forecast of yield curves. Hays et al. (2012) have previously used FDA for yield curve modeling. The method in Hays et al. (2012) is factor analysis with penalized likelihood for estimation, assuming an AR(p) model for the factors. In this paper the method is PCA with state space modeling for time series, assuming VARMA(p,q) evolution of the components.

2. Methodology

Consider a sample of n smooth random trajectories $\left(f_t\,(u)\right)_{u\in\mathcal{T}}$

for $t = 1, \ldots, n$ generated from a process f. Throughout we assume that f is an element of the Hilbert space $\mathfrak{H} := L^2\,(\mathcal{T})$ endowed with the inner product $\langle f, g\rangle_{\mathfrak{H}} = \int_{\mathcal{T}} f(u)g(u)du$ and the norm $\|f\| = \sqrt{\langle f, f\rangle_{\mathcal{H}}} < \infty$ a.s.. The observe measurement are support point u_{ti} on the domain $\mathcal{T} = [a_1, a_2]$ with additive white noise error W_{ti} which is independent of the underlying process. The measurements are for $t = 1, \ldots, n$ and $i = 1, \ldots, m$:

$$\widetilde{f}_t(u_{ti}) = f_t(u_{ti}) + W_{ti} \text{ with } E\,(W_{ti}) = 0,\ VAR(W_{ti}) = \sigma^2 \qquad (1)$$

2.1 Principal Components Analysis of Functional Data

We represent the smooth functional f in terms of its decomposition into functional principal components, a common approach in FDA. For the domain $\mathcal{T}$, setting

$$\mathbf{G}_f(u,v) = Cov\,(f(u), f(v)), \qquad E(f(u)) = \mu_f(u), \quad u, v \in \mathcal{T}, \tag{2}$$

the functional principal components are the eigenfunctions of the auto-covariance operator $\mathbf{G}_f : \mathcal{H} \mapsto \mathbb{R}$, a linear operator on the space $\mathcal{H}$, that is given by

$$G_f(g)(u) = \int_{\mathcal{T}} \mathbf{G}\mathrm{T}_f(u,v) g(v) dv.$$

We denote the orthonormal eigenfuncitons by ϕ_k, with associated eigenvalues λ_k for $k = 1, 2, \ldots$, such that $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\Sigma_k \lambda_k < \infty$. The Karhunen-Loève theorem (see Rice and Silverman (1991)) provides a representation of individual random trajectories of the functional f, given by

$$f(u) = \mu_f\,(u) + \textstyle\sum_{k=1}^{\infty} \xi k \phi k(u), \qquad u \in \mathcal{T}, \tag{3}$$

Where the ξ_k are uncorrelated random variables that satisfy

$$\xi_k = \int (f(u) - \mu_f(u))\, \phi_k(u) du, \;\; E\xi_k = 0, \quad VAR(\xi_k) = \lambda_k. \tag{4}$$

Under the data generating mechanism in (1), one has with indicator function *I(.)*

$$\mathrm{E}(\widetilde{f}_t(u)) = \mu_f(u), \quad \mathrm{Cov}(\widetilde{f}_t(u), \widetilde{f}_t\,(v)) = \mathbf{G}_f(u,v) + \sigma^2 I\,(u = v). \tag{5}$$

This implies that the smooth mean function u_f and the smooth covariance surface $\mathbf{G}_f$ can be consistenly estimated from available data by pooling the sample of n trajectories and smoothing the resulting scatterplot. The exception for targeting points on $\mathbf{G}_f$ with $u = v$ in (5) is necessitated by the presence of W. This does not pose a problem, since it follows from the smoothness of the surface $\mathbf{G}_f$ that the areas of $\mathbf{G}_f(u,v)$, for which $u = v$, can still be consistently estimated. Well-known procedures exist to infer eigenfunctions and eigenvalues (Rice and Silverman (1991); Müller et al. (2006)).

Processes f are then approximated by substituting estimates and using a judiciously chosen finite number *K* of terms in sum (3). This choice can be made using one-curve-leave out cross-validation (Rice and Silverman (1991)), pseudo-AIC criteria Yao et al. (2005) or a scree plot, a tool from multivariate analysis, where one uses estimated eigenvalues to obtain a prespecified fraction of variance explained as a function of *K* or looks for a change-point.

The above procedure is also known in numerical analysis under the acronym *proper orthogonal decomposion* and as such it is used to price and hedge financial derivatives on forward curves; see Hepperger (2010) for examples from the energy market and further references.

2.2 Time Series of Funcitonal Data

In this section we show that an ARMA (*p,q*) structure on the principal components scores. Starting with the setup as described in (1), we assume that the series of functions follows a stationary ARMAH (*p,q*) model with mean $\mu \in \mathcal{H}$:

$$f_t(\cdot) - \mu_f = \psi_1(f_{t-1}(\cdot) - \mu_f) + \cdots + \psi_p(f_{t-p}(\cdot) - \mu_f) + \epsilon_t(\cdot). \quad (6)$$

Where $\epsilon_t(\cdot) = \eta_t + \theta_1 \eta_{t-1}(\cdot) + \cdots + \theta_q \eta_{t-q}(\cdot), and\ \eta_t(\cdot)$ is $\mathfrak{H}$ white noise. $\psi_1, \cdots, \psi_p\ and\ \theta_1, \cdots, \theta_q$ are linear continuous functions. Combining (6) and (3) we have,

$$\sum_{k=1}^{\infty} \xi_{kt}\phi k(\cdot) = \psi_1 \sum_{k=1}^{\infty} \xi kt - 1\phi k(\cdot)) + \cdots + \psi_p(\sum_{k=1}^{\infty} \xi_{kt-p}\phi_k(\cdot)) + \epsilon_t(\cdot). \quad (7)$$

Using linearity and continuity of $\psi_1, \cdots, \psi_p$, this implies,

$$\sum_{k=1}^{\infty} \xi kt\phi k(\cdot) = \sum_{k=1}^{\infty} \xi kt - 1\,\psi 1\,(\phi k(\cdot)) + \cdots + \sum_{k=1}^{\infty} \xi_{kt-p}\psi_p(\phi_k(\cdot))\,\epsilon_t(\cdot). \quad (8)$$

Using vector notation, we have:

$$\Phi(\cdot)\Xi_t = \psi_1(\Phi(\cdot))\Xi_{t-1} + \cdots + \psi_p(\Phi(\cdot))\Xi_{t-p} + \epsilon_t(\cdot). \quad (9)$$

where $\Phi = (\phi_1, \phi_2, \cdots)$ and $\Xi = (\xi_{1t,}\xi_{2t,} \cdots)^T$. Since the eigenfuncitons ϕk are orthonormal, we can premultiply equation (10) by Φ^T to get:

$$\Xi_t = \Phi^T\psi_1(\Phi(\cdot))\Xi_{t-1} + \cdots + \Phi^T\psi_p(\Phi(\cdot))\Xi_{t-p} + Z_t. \quad (10)$$

It remains to show that $Z_t = \Phi^T \epsilon(\cdot)$ is an MAH(q) process. This can be proved by verifying that the autocovariances of Z vanish for lags of order greater than *q*. This is immediate as

$$\mathrm{Cov}(Z_t, Z_s) = \Phi^T \mathrm{Cov}(\epsilon_t, \epsilon_s) = \Phi$$

and ϵ is itself an MAH(q) process.

This implies a VARMA(*p,q*) structure on the vector of principal component score Ξ_t.

Moreover, since $\Xi_t = \Phi^T f_t$, stationarity of f implies stationarity of Ξ.

2.3 Granger Causality

The concept of the Granger causality was originally introduced by Wiener(1956) and formulated by Granger(1969). Given two stationary time series, if the variance of the prediction error for the second time series at the present time is reduced by including past measurements from the first time series in the linear regression model, then the first time series can be said to Granger-cause the second time series.

The theory of Granger causality in multivariate time series was developed by Geweke(1982). Consider vector time series Z that has autoregressive representation

$$Z_t = \sum_{s=1}^{\infty} \boldsymbol{B}_s Z_{t-s} + \epsilon_t$$

Suppose not that $Z_t : m \times 1$ has been partitioned into $k \times 1$ and $l \times 1$ subvectors $\mathbf{x}_t$, and $y_t, z'_t = (x'_t, y'_t)$, reflecting an interest in relationship between X and Y. We are interested in the Granger causality of Y on X.

X has autoregressive representation as follows:

$$\mathbf{x}_t = \sum_{s=1}^{\infty} \mathbf{E}_{1s} \mathbf{x}_{t-s} + \mathbf{u}_t , \qquad var(\mathbf{u}_t) = \Sigma_1$$

We partition the linear projection of $\mathbf{x}_t$ on $\mathbf{X}_{t-1}$ *and* $\mathbf{Y}_{t-1}$ as

$$\mathbf{x}_t = \sum_{s=1}^{\infty} \mathbf{E}_{2s} \mathbf{x}_{t-s} + \sum_{s=1}^{\infty} \mathbf{F}_{2s} \mathbf{y}_{t-s} + \mathbf{v}_t \qquad var(\mathbf{v}_t) = \Sigma_2$$

The measure of linear feedback from Y to X is $\mathbf{F}_{Y\to x} = In(|\Sigma_1|/|\Sigma_2|)$ *where* $|A|$ *denotes the determinant of* A. The estimation is done by truncating the infinite AR representation at finite p and then using OLS. If the disturbances are independent and identically distributed, the conventional large-sample distribution theory may be used to test the null hypothesis that a given measure of feedback is zero. If $\mathbf{F}_{Y\to x} = 0$, then

$$In(|\hat{\Sigma}_1|/|\hat{\Sigma}_2) \Longrightarrow \chi^2(klp).$$

3. Empirical Examples

The yield curve of two different economies, USA and India, are studied for comparative purpose. The US department of treasury webpage lists the daily yield curve from 1990 till date for certain maturities from 1 month to 30 years. The Indian government bond historical data can be obtained from in.investing.com for each maturity separately from 3 months to 15 years. The specific maturities are listed in table 1. We separate the data into years because for long time horizons the stationarity assumption of the time series may not be valid. We present the results for the year 2015 for USA and India. They are representative of the other years. In Figure 1, we present the raw data for the countries. For each weekday of the year we have data of dimension 11(for US data) and 17(for Indian data). We think of it as a time series of functions. It is observed that the US curves are pretty smooth whereas the Indian data has more fluctuations, both with respect to maturity and in time.

The initial fitting of functional data to obtain mean, covariance and principal components is done by employing the PACE package for functional data analysis written in Matlab. We use the Gaussian kernel. The package is available at

http://www.stat.ucdavis.edu/PACE/

VAR model fitting and diagnostics is done using the econometrics toolbox in Matlab.

VARMA and related state space model computations are done using the Dynamic Systems Estimation (dse) package in R available at

http://cran.r-project.org/web/packages/dse/index.html.

It should be noted that in all the actual data applications, the models chosen by AIC criterion had the MA degree zero.

Using multivariate Granger causality toolbox in matlab, we test for Granger causality between yield curves of USA and India. The package also does the using trace instead of determinant. We find that there is no causality in either direction.

References

1. Aue, A., D. D. Norinho, and S. Hörmann (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association 110*(509), 378–392.
2. Besse, P. C., H. Cardot, and D. B. Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics 27*(4), 673–687.
3. Bosq, D. (2012). *Linear processes in function spaces: theory and applications,* Volume 149. Springer Science & Business Media.
4. Bosq, D. (2014). Computing the best linear predictor in a hilbert space. applications to general armah processes. *Journal of Multivariate Analysis 124,* 436–450.
5. Castro, P. E., W. H. Lawton, and E. A. Sylvestre (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics 28*, 329–337.
6. Hays, S., H. Shen, and J. Z. Huang (2012). Functional dynamic factor models with application to yield curve forecasting. *Annals of Applied Statistics 6*(3), 870–894.
7. Hepperger, P. (2010). Option pricing in hilbert space valued jump-diffusion models using partial integro-differential equation. *SIAM Journal on Financial Mathematics 1(1),* 454– 489.
8. Hörmann, S. and P. Kokoszka (2010). Weakly dependent functional data. *The Annals of Statistics 38*, 18451884.

9. Kargin, V. and A. Onatski (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis 99,* 2508–2526.
10. Muller, H. G., U. Stadtmuller, and F. Yao (2006). Functional variance processes. *Journal of the American Statistical Association 101,* 1007–1018.
11. Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis.* New York, NY: Springer Series in Statistics.
12. Rice, J. A. and B. W. Silverman (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of Royal Statistical Society Series B 53*, 233–243.
13. Yao, F., H. G. Muller, and J. L. Wang (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100,* 577–590.

Data integration of unbalanced social media and survey information via calibration

Luciana Dalla Valle[1], Ron Kenett[2]
University of Plymouth, Plymouth, UK
KPA Ltd, Neaman Institute, Technion and Hebrew University, Raanana, Israel

Abstract

We live in an age of unprecedented amounts of information, generated in every sector, including business, government and health care, delivered at high speed and available in a wide variety of forms and formats. This data may come from many different sources, such as social media posts, digital pictures and videos, cell phone GPS, purchase transaction records and signals sensors used to gather climate information. This information - high volume, diverse and fast - is what is called Big Data. Social media are amongst the most prolific generators of big data and allow billions of people all around the world to daily interact, post and share contents and give spontaneous feedback on specific topics. As opposed to traditional media such as newspapers, books and television, social media is freely accessible, which means everyone can publish content and control how the information is generated and shared. Through social media, people express their opinions and sentiments towards specific topics, products and services. The ability to harness big data and social media data is an opportunity to obtain more accurate analyses and to improve decision-making in industry, government and many other organizations. However, handling big data may be challenging and proper data integration is a key dimension in achieving high information quality. We propose a novel approach to data integration that calibrates online generated big data with interview-based customer survey data. A common issue of customer surveys is that responses are often overly positive, making it difficult to identify areas of weaknesses in organizations. On the other hand, online reviews are often overly negative, hampering an accurate evaluation of areas of excellence. The proposed methodology calibrates the levels of unbalanced responses in different data sources via resampling and performs data integration using Bayesian Networks to propagate the new re-balanced information. We show, with a case study example, how the novel data integration approach allows businesses and organizations to get a bias-corrected appraisal of the level of satisfaction of their customers. The application is based on the integration of online data of review blogs and customer satisfaction surveys from the San Francisco airport. We illustrate how this integration enhances the information quality of the data

analytic work in four of InfoQ dimensions, namely, Data Structure, Data Integration, Temporal Relevance and Chronology of Data and Goal.

Keywords
Bayesian networks; Big Data; Information quality (InfoQ); Resampling techniques

1. Introduction

The growing availability of abundant masses of data in every sector, including business, government and health care, is posing new analytic and statistical challenges. In recent years, advances in the literature of big data analysis have been significant. Amongst recent contributions to sentiment analysis, Stander et al. (2016a) and Stander et al. (2016b) extracted Facebook data to analyze sentiment scores and voting patterns about the June 2016 EU referendum in the UK. Zhang et al. (2011) used sentiment analysis techniques to predict stock market indicators using Twitter data. Asur and Huberman (2010) predicted box-office movie revenues, performing an analysis of sentiments from comments posted on social media. However, big data analysis and social media mining may be challenging. The main issues are related to the quality of data collected and reported and to the integration of multiple datasets. In particular, social media big data often contain biased information, especially online blogs describing opinions and sentiments about specific products and services. Indeed, online reviews generally include overly negative comments and feedback, since users tend to feel more free to express their dissatisfaction online, rather than in other contexts. On the other hand, traditional reviews generally include overly positive comments, since people tend not to feel comfortable to voice their opinions in surveys and may not be completely honest about their discontent. In both cases, the levels of the variables expressing customers' views are (sometimes strongly) unbalanced, preventing a correct evaluation of customer satisfaction. In handling these challenges, data integration is key, especially where data come in both structured and unstructured formats and need to be integrated from disparate sources stored in systems managed by different departments. In most cases, the efficient aggregation and correlation of multiple datasets of considerable dimensions may be very complex (Daniel, 2015). Foresti et al. (2012) agree that data aggregation from multiple information sources is key to decision-makers and describe a regression-based data integration methodology applied to public and private financial databases. Dalla Valle (2016) illustrates a different approach for blending information from official statistics and organizational data, based on the generalization of Heckman's method where inference is performed according to the Bayesian framework. Dong and Srivastava (2015) describe the big data integration techniques of schema mapping, record linkage and data fusion and identify a range of open problems in this research area. Chakraborty et al. (2015) define a novel approach to integrate diverse data types, such as historic data, survey data, management planning data, expert knowledge and incomplete data, by converting data into Bayesian probability forms. Dalla Valle (2014 and 2017a) and Dalla Valle and Kenett (2015) introduced an innovative approach to integrate survey data with official statistics data based on calibration using copulas and nonparametric Bayesian Networks (BNs). For an overview about copulas and their applications to finance, see Dalla Valle (2017b and 2017c) and references therein. For an introduction to BNs see, for example, Pearl (2009), Jensen (2001), Ben Gal (2007), Koski and Noble (2009) and Pourret et al. (2008). In this paper, we propose a novel methodology that calibrates social media information with online review data via resampling and performs integration using BNs. This approach allows businesses and organizations to correctly analyze the sentiments of online users on social media, facilitating an accurate evaluation of the satisfaction of their customers. Such an integration, combining different overlapping data sources, enhances the information quality of the data analytic work in four dimensions: Data Structure, Data Integration, Temporal Relevance and Chronology of Data and Goal (Kenett and Shmueli, 2016).

2. Methodology

The methodology proposed in this paper aims at achieving data integration of traditional customer satisfaction survey data with social media data via resampling using BNs, expanding the approach presented in Dalla Valle and Kenett (2015). We perform data integration emphasizing blog-type data, which is a big data environment source. However, our approach is scalable to other social media and big data sources. As mentioned above, properly handling data integration is a key dimension in achieving high information quality (Kenett and Shmueli, 2016). The proposed data integration methodology aggregates customer survey data with information extracted from social media, performing calibration of different datasets. The idea is in the same spirit of external benchmarking used in small area estimation (Pfeffermann, 2013). In small area estimation benchmarking robustifies the inference by forcing the model-based predictors to agree with a design-based estimator. Similarly, our methodology is based on qualitative data calibration performed via resampling, where the variables levels are balanced and customer survey estimates are updated to agree with more timely social media data estimates. Calibration is implemented by altering the class distribution of customers' reviews in one of the datasets to obtain a re-balanced sample, which reflects the distribution of the second dataset.

This approach involves the selection of a calibration link variable and the creation of a new artificial data set by suitably resampling the observations belonging to the classes of the calibration link. In particular, the calibration link variable is resampled by oversampling with replacement the minority class and by undersampling without replacement the majority class. More formally, the resampling approach can be described as follows. Let us consider the variables denoted by the pairs ($\boldsymbol{x}$, y), where $\boldsymbol{x}$ represents a set of measured characteristics and y is a target (or key) variable. Here, we consider the specific case where $\boldsymbol{x}$ is defined in a d-dimensional space X being the product set between discrete domains, and the target variable y, which is affected by class imbalance, takes values in the categorical domain Y= {Y_{min}, Y_{maj}}, where Y_{min} is the minority class and Y_{maj} is the majority class. Suppose that a sample D_n = ($\boldsymbol{x}_1$, y_1), ..., ($\boldsymbol{x}_n$, y_n), of the pairs ($\boldsymbol{x}$, y), whose generic row is ($\boldsymbol{x}_i$, y_i), $i = 1, ..., n$, is observed on n individuals or objects. The class labels y_i belong to the set {Y_{min}, Y_{maj}} and $\boldsymbol{x}_i$ are some related attributes supposed to be realizations of a random vector $\boldsymbol{x}$. Let the number of units in class Y_j, $j = min, maj$, be denoted by $n_j < n$ and the corresponding class proportions be denoted by $p_j = n_j/n$. The resampling procedure for generating a new artificially re-balanced dataset, consists of the following steps:

1) Select $y^* = Y_j$ with probability 1/2.
2) Select ($\boldsymbol{x}_i$, y_i) $\in D_n$, such that $y_i = y^*$, with probability $1/n_j$.
   a. If $y^* = Y_{min}$, oversample with replacement by adding ($\boldsymbol{x}_i$, y^*) to D_n;
   b. If $y^* = Y_{maj}$, undersample without replacement by removing ($\boldsymbol{x}_i$, y^*) from D_n.

Repeat steps 1 and 2 until the desired class proportions are achieved or until the minority class reaches the desired size. This procedure produces a new rebalanced dataset $D_m{}^*$, of size m, where the desired proportions of observations belong to the two classes. For more details about the class imbalance problem and resampling techniques see, for example, Chawla (2005) and Menardi and Torelli (2014).

In the present work, the resampling approach described above is applied to interview- and online-based imbalanced datasets to achieve data integration. Following this bias correction, BNs are built to identify the main determinants of customer satisfaction.

The proposed data integration methodology is structured in three phases:

1) *Data structure modelling*. Let D^{SU} denote the interview-based survey dataset and D^{SM} denote the social media dataset. This phase consists in implementing BNs to construct the causal relationships between the variables of both the customer survey, D^{SU}, and social media, D^{SM}, datasets, separately. BNs are chosen amongst other data modelling techniques for their flexibility and ability to encode probabilistic relationships among variables of interest, allowing an easy identification of the determinants of

customer satisfaction. However, the presence of unbalanced samples can affect the correct assessment and evaluation of customer satisfaction and may lead to misleading conclusion. Data integration, implemented by rebalancing the unbalanced levels of D^{SU} with the levels of D^{SM} (or viceversa), allows us to accurately analyze customer satisfaction.

2) *Identification of the calibration link*. In the second phase a calibration link, in the form of one or more unbalanced key variables, is identified between customer survey and social media data. Denoting with ($\boldsymbol{x}^{SU}$, y^{SU}) the variables of D^{SU} and ($\boldsymbol{x}^{SM}$, y^{SM}) the variables of D^{SM}, then let y^{SU} be the calibration link of D^{SU} and y^{SM} be the calibration link of D^{SM}. We suppose that calibration links are unbalanced variables, with y^{SU} taking values in the categorical domain $Y^{SU}= \{Y_{min}^{SU}, Y_{maj}^{SU}\}$, with proportions $p^{SU}= \{p_{min}^{SU}, p_{maj}^{SU}\}$, and y^{SM} in $Y^{SM}= \{Y_{min}^{SM}, Y_{maj}^{SM}\}$, with proportions $p^{SM}= \{p_{min}^{SM}, p_{maj}^{SM}\}$, where Y_{min}^{SU} and Y_{min}^{SM} are the minority classes and Y_{maj}^{SU} and Y_{maj}^{SM} the majority classes of the interview- and blog-based surveys. Calibration links can be target variables expressing overall satisfaction or can be other variables influencing the overall satisfaction.

3) *Performing calibration*. In the last phase calibration is performed by suitably resampling the datasets, based on the distribution of the calibration link variables. In this phase, one of the dataset, for example D^{SU}, is rebalanced following the resampling approach described above, until $p^{SU} \approx p^{SM}$. Therefore, a new rebalanced dataset D^{SU*} with the desired proportions of the calibration link variable will be generated. Similarly, calibration can be performed on D^{SM}, obtaining the new rebalanced dataset D^{SM*}. BNs are then updated for the re-balanced datasets D^{SU*} or D^{SM*}, allowing the calibrated information to be propagated to achieve data integration. This approach will allow us to properly analyze customer satisfaction surveys and to achieve the goal of accurately identifying pockets of dissatisfaction and areas of excellence within an organization.

3. Result

We illustrate the application of the methodology by integrating airport passengers' data collected via interview-based survey with data extracted from an online review website. The context of this example is an analysis focused on improving the Temporal Relevance of a customer satisfaction survey by linking its results to online reviews that are continuously updated. The data integration methodology described here provides information to decision makers that is both up to date and comprehensive. In this sense, the Data Integration supports proper Chronology of Data and Goal. The example therefore enhances the information quality in four of the InfoQ dimensions: Data Structure, Data Integration, Temporal Relevance and Chronology of Data and Goal.

The first dataset we analyze is a subset of the 2016 customer survey administered to the passengers of San Francisco International Airport (SFO)[1]. The passenger dataset contains information pertaining to customer demographics and satisfaction with airport facilities, services, and initiatives. The data was collected in May 2016 through interviews with 3,087 customers in each of SFO's terminals and boarding areas. Customers were asked to rate the airport in several categories, including cleanliness ratings. Additional data collected include customers' income, mode of arrival to the airport, travel style, and various other categories. The SFO dataset comprises demographic and satisfaction variables, including a variable expressing customers' overall satisfaction. The satisfaction variables included in the SFO dataset express the passengers' judgements on a five-point scale. For comparison purposes, we transformed the original customers' ratings into dichotomous variables. The variables were dichotomized following two different schemes. The first of these schemes is called *BOT1+2* and it is constructed by aggregating customers who responded '1' or '2' (corresponding to extreme dissatisfaction and dissatisfaction, respectively). The second scheme is called *TOP5* and identifies customers who responded '5' (corresponding to extremely satisfied) on the five-point scale. BOT1+2 is very effective in identifying pockets of dissatisfaction and areas of improvements, while TOP5 emphasizes areas of excellence. For more on statistical analyses using the two dichotomizing schemes see Kenett and Salini (2011).

The second dataset, that we named Skytrax dataset, contains information extracted from the reviews published online by passengers of the SFO airport[2]. For comparative purposes, only recent reviews of SFO passengers were analyzed. The dataset includes demographic and satisfaction variables, with judgements on individual characteristics and on the airport as a whole. For the sake of comparison, we applied the BOT1+2 and TOP5 dichotomization schemes to the Skytrax satisfaction variables.

After transforming the original data, we applied the three phases of the data integration methodology described in Section 2 to the SFO customer survey and to the Skytrax social media datasets using the BOT1+2 as well as the TOP5 dichotomization. Initially, from SFO as well as Skytrax, two new datasets were generated according to the BOT1+2 and TOP5 dichotomization schemes. Then, the data integration methodology was applied twice: once to the BOT1+2 datasets and once to the TOP5 datasets, to illustrate the use of different calibration functions.

---

[1] The data are publicly available on the website http://www.flysfo.com/media/customer-survey-data

[2] The data are publicly available on the website http://www.airlinequality.com

3.1 Data Integration of BOT1+2 Datasets:

In the first phase of data integration, we analyzed the SFO customer satisfaction survey data with BNs, implemented using the GeNIe software V 2.1 (University of Pittsburgh, Pittsburgh, USA).

The data modelling phase consists in the construction of BNs for both the SFO and the Skytrax datasets.

Then, we select the OVERALL variable as calibration link for the BOT1+2 dichotomized datasets. The percentage of dissatisfied passengers in the SFO survey dataset is only 2%, while the same percentage in the Skytrax online dataset is almost 50%. Therefore, the levels of OVERALL in the SFO survey dataset need to be re-balanced by resampling, to make the distribution similar to that of the Skytrax online dataset. The SFO customer survey dataset was resampled, as explained in Section 2, using the R package ROSE (Lunardon et al., 2014). The BN was updated via parameter learning and hence calibrated to reflect the information contained in the online reviews. Figure 1 (left panel) illustrates the BN of the BOT1+2 SFO customer satisfaction survey dataset, after calibration of the OVERALL node via resampling. The distribution of the overall satisfaction is now balanced, with a higher proportion of dissatisfied customers, as appears in online reviews. This calibrated BN shows that the percentages of passengers who are dissatisfied with cleanliness, walkways, shopping areas and the free Wi-Fi are 19%, 23%, 33% and 14%, respectively. These results highlight, much more clearly than those based on the original unbalanced dataset, the weaknesses and corresponding areas of improvement of the airport.

Skytrax dataset. However, there is an imbalance in its classes, since the percentage of 'excellent' answers is only 24%. The same variable appears to be well-balanced in the SFO survey dataset, where the percentage of 'excellent' is close to 50%. Therefore, the Skytrax dataset needs to be resampled, in order to re-balance the distribution of QUEUING according to the distribution of the SFO survey dataset. In order to re-balance the QUEUING variable, the Skytrax online reviews dataset was resampled, to reflect the distribution of a similar variable (PASSTHRU) in the SFO customer survey dataset. Calibration between the two datasets was performed and the BN of the TOP5 Skytrax dataset was updated via parameter learning. Figure 1 (right panel) illustrates the BN of the TOP5 Skytrax reviews social media dataset, after calibration of the QUEUING node via resampling. The distribution of passengers' satisfaction with queuing is now balanced, with a higher proportion of extremely satisfied passengers, as appears in the SFO customer survey dataset. This calibrated BN shows that the percentages of passengers who are extremely satisfied with cleanliness, restaurants, shopping and seating areas have increased and are equal to 44%, 37%, 34% and 14%, respectively. In addition, the percentage of very satisfied passengers overall is 34%. These results calibrate the overly negative online reviews and underline the areas of excellence of the airport.

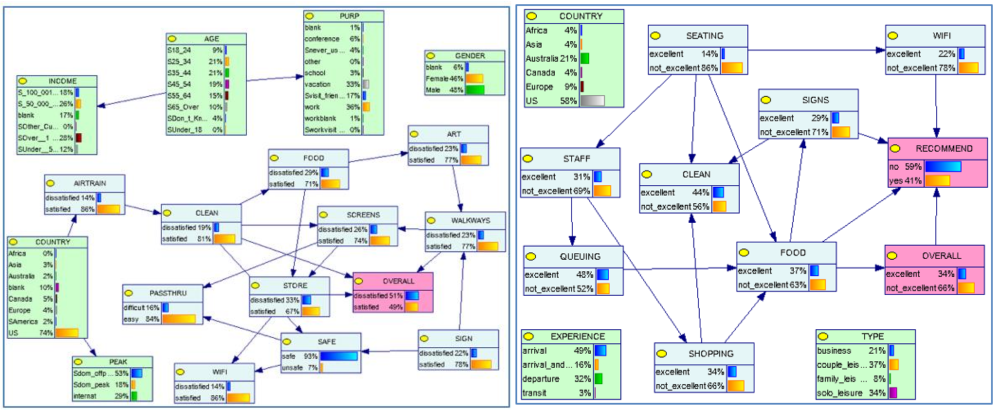


Figure 1: (Left panel) BN of the BOT1+2 SFO customer satisfaction survey dataset, after calibration of the OVERALL node via resampling. (Right panel) BN of the TOP5 Skytrax reviews social media dataset, after calibration of the QUEUING node via resampling.

4. Discussion and Conclusion

With the growing exploitation of big data, integration of data sources becomes a key capability. Traditional integration methods rely on extract transform and load (ETL) and record linkage techniques (Kenett and Raanan, 2010). In this paper, we propose a novel approach to data integration that combines online big data with a comprehensive survey. The methodology is derived from resampling and modeling the data using BNs, and identifying overlapping links that are used for calibration. We show,

with an example, how data integration between online blogs and a customer satisfaction survey supports proper chronology of data and goal. The example demonstrates of such data integration enhances the information quality of a study in four of the InfoQ dimensions: Data Structure, Data Integration, Temporal Relevance and Chronology of Data and Goal.

References

1. Asur, S. and Huberman, B.A. (2010). Predicting the future with social media, Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference, Vol. 1, pp. 492-499.
2. Ben Gal, I. (2007). Bayesian Networks, in Encyclopedia of Statistics in Quality and Reliability, Ruggeri, F., Kenett, R. S. and Faltin, F. (editors in chief), Wiley, UK.
3. Chakraborty, S., Mengersen, K., Fidge, C., Ma, L., and Lassen, D. (2015). Multifaceted modelling of complex business enterprises. PloS one, Vol. 10, No.8, e0134052.

4. Chawla, N.V. (2005) Data mining for imbalanced datasets: An overview. In: Data mining and knowledge discovery handbook. Springer US, p. 853-867.
5. Dalla Valle, L. (2017a) Data Integration. in: Wiley StatsRef: Statistics Reference Online, ed. M. Davidian, B. Everitt, R. Kenett, G. Molenberghs, W. Piegorsch and F. Ruggeri, John Wiley & Sons, pp. 1-6.
6. Dalla Valle, L. (2017b) Copula and Vine Modeling for Finance, in: Wiley StatsRef: Statistics Reference Online, ed. M. Davidian, B. Everitt, R. Kenett, G. Molenberghs, W. Piegorsch and F. Ruggeri, John Wiley & Sons, pp. 1-5.
7. Dalla Valle, L. (2017c) Copulas and Vines, in: Wiley StatsRef: Statistics Reference Online, ed. M. Davidian, B. Everitt, R. Kenett, G. Molenberghs, W. Piegorsch and F. Ruggeri, John Wiley & Sons, pp. 1-5.
8. Dalla Valle, L. (2016) The Use of Official Statistics in Self-Selection Bias Modeling. Journal of Official Statistics, Vol. 32, No. 4, pp. 887–905.
9. Dalla Valle, L. (2014). Official statistics data integration using copulas. Quality Technology and Quantitative Management; 11(1), 111-131.
10. Dalla Valle, L. and Kenett, R.S. (2015). Official Statistics Data Integration for Enhanced Information Quality, Quality and Reliability Engineering International, Vol. 31, No. 7, pp. 1281-1300.
11. Daniel, B. (2015). Big Data and analytics in higher education: opportunities and challenges. British Journal of Educational Technology, 46(5), pp. 904-920.
12. Dong, X.L. and Srivastava, D. (2015). Big Data Integration. Synthesis Lectures on Data Management. Morgan Claypool Publishers.
13. Foresti, G., Guelpa, F. and Trenti, S. (2012). Enterprises in a globalized context and public and private statistical setups. SIS Scientific Meeting.
14. GeNIe. Decision Systems Laboratory. University of Pittsburgh: USA, 2006. http://genie.sis.pitt.edu
15. Jensen, F. V. (2001). Bayesian Networks and Decision Graphs, Springer.
16. Kenett, R.S. and Raanan, Y. (2010). Operational Risk Management: a practical approach to intelligent data analysis, John Wiley and Sons, Chichester, UK.
17. Kenett, R.S. and Salini, S. (2011). Modern Analysis of Customer Satisfaction Surveys: with applications using R, John Wiley and Sons, Chichester: UK.
18. Kenett, R.S. and Shmueli, G. (2016). Information Quality: The Potential of Data and Analytics to Generate Knowledge, John Wiley and Sons. www.wiley.com/go/information_quality.
19. Koski, T. and Noble, J. (2009). Bayesian Networks – An Introduction, John Wiley and Sons, Chichester, UK.

20. Lunardon, N., Menardi, G., and Torelli, N. (2014). ROSE: A Package for Binary Imbalanced Learning. R Journal, Vol. 6, No. 1, pp. 82-92.
21. Menardi, G., and Torelli, N. (2014). Training and assessing classification rules with imbalanced data. Data Mining and Knowledge Discovery, Vol. 28, No, 1, pp 92–122.
22. Pearl, J. (2009). Causality: Models, Reasoning, and Inference, 2nd ed., Cambridge University Press, UK.
23. Pfeffermann, D. (2013). New Important Developments in Small Area Estimation, Statistical Science, 28, pp. 40-68.
24. Pourret, O, Naïm P. and Marcot, B. (2008). Bayesian Networks: A Practical Guide to Applications, John Wiley and Sons, Chichester: UK.
25. Stander, J., Dalla Valle, L. and Cortina Borja, M. (2016a). Sentiments, surnames and so long EU. Communicator, Special Supplement: Science Communication, Autumn 2016, pp. 19–23.
26. Stander, J., Dalla Valle, L., Eales, J., Baldino, A. and Cortina Borja, M. (2016b). The EU referendum: extracting insights from Facebook using R, Significance Magazine, available on-line at https://www.statslife.org.uk/politics/2889-what-information-can-we-extract-from-social-media-about-the-uk-s-eu-referendum
27. Zhang, X., Fuehres, H. and Gloor, P.A. (2011). Predicting stock market indicators through twitter "I hope it is not as bad as I fear". Procedia – Social and Behavioral Sciences, Vol. 26, pp. 55-62.

A complete measure of wealth and wealth inequality

Arjan Bruil
Statistics Netherlands

Abstract
In the past years, linking micro data to macro aggregates of the system of national accounts (SNA) received a lot of attention. Most efforts focused on adding distributions of income and consumption to macro aggregates of the household sector. In this article, we will include the wealth dimension as well. In measures of wealth inequality, often pension entitlements are not included. In the Dutch System of National Accounts however, these are the largest asset of households. These entitlements relate to work-related and voluntary pensions, but exclude public pensions. Recently, estimates of household pension wealth are extended with implicit pension wealth. This is the result of ongoing efforts to make pension systems more comparable over countries. In this article, we will show the leveling effects of the pension schemes, and compare the new wealth measures with income inequality in the Netherlands.

Keywords
distributions; national accounts; net worth; pension entitlements

1. Introduction

Studies on inequality have long focused on income. There is ample data available to construct long time series, with a great level of detail. For wealth, the data availability is less abundant. When wealth inequality is considered (for example: OECD, 2015), these often cover the micro statistics. In recent years, micro data are linked more and more to National Accounts totals. The biggest difference between micro and macro concepts of wealth is whether pension entitlements are included. In the Netherlands, these are the largest financial asset of the household sector in the System of National Accounts (SNA), amounting to 205% of GDP in 2016. The public pension schemes are not included in the net worth of the SNA[1]. The micro statistics exclude both work-related and public pension schemes from the assets. There might be reasons to neglect pension entitlements, as they differ from other wealth components such as bank savings or equity. Pension entitlements are not freely accessible and households cannot bequeath it. However, in lifecycle terms, it is by design a means of future consumption. Apart from this conceptual discussion,

[1] Since the most recent SNA update (2008), the public pension plans are measured in a supplementary pension table, as if they were funded.

researchers are also limited by the data availability of pension entitlements. Saez and Zucman (2016) solve the absence of observed wealth by capitalizing income flows for the United States. This allows them to analyse a time series going back to 1913. They acknowledge the importance of pension entitlements by imputing a distribution of the SNA pension entitlements, based upon distributions of wages and pension benefits. Other studies use survey data on wealth (Vermeulen, 2018), often limited by the absence of the very wealthy, or data on pension wealth.

Van Bavel and Frankema (2017) argue that wealth inequality in the Netherlands, just as in other welfare states, is relatively high, while income inequality is low. They call this the inequality paradox of Northern European welfare states, and suggest that one of the reasons might be that households' asset portfolios miss the large collective arrangements. Wilterdink (2015) mentions that international comparisons of wealth inequality is difficult, because of data availability, quality issues in both micro and macro data, but also due to the different approaches in the unit of observation. The author does mention that developments over time are less hampered by these limitations. However, where pension entitlements decrease inequality, the influence of these entitlements on developments remains unclear (Wilterdink, 2015, p.358).

2. Methodology

The starting point of our methodology is the household database as described by Bruil (2018). This database, in which all Dutch residents are included, covers the entire SNA household sector. For each individual it is known to which household he or she belongs. For all the individuals and households the sector accounts are constructed, using a large number of micro data sources. These data sources are linked, preferably using a record linking technique, using a personal identifier that is unique over all data sets. Data sources that do not consist this personal identifier are imputed in the dataset using a common characteristic of the household, individual, or group of individuals. We add the balance sheets, using the wealth components from the Integral Income and Wealth Studies (IIWS), the Pension Claims Statistics (PCS), and the Household Finance and Consumption Survey (HFCS). For the extended net worth concept we add public pension entitlements as well, following the methodology laid out in the Technical Compilation Guide (Eurostat, European Central Bank, 2011) and the further work carried out for the Netherlands (CBS, 2018). Where disposable income is constructed from micro data, this is not possible for net worth because of timeliness of the microdata. We use the microdata to breakdown the balance sheets.

The IIWS is an integral register, largely based on tax records. It covers the wealth on the first of January, which corresponds with the opening balance sheets in the SNA. The IIWS is used for Currency and Deposits, and Loans. Data gaps between micro and macro totals are allocated as much as possible to the right individuals, households, or groups. For instance, the coverage of loans in the IIWS is incomplete. Assets related to mortgages with a savings or investment component are not observed, while they are included in the National Accounts. We allocate part of the data gap to mortgage-payers. Another gap occurs because the loans of the self-employed are netted in their business value, while these are made explicit in the SNA. This part of the data gap is estimated by the interest receipts and payments of self-employed and allocated accordingly. We use the macro data from custodians to estimate debt securities and equity and investment fund shares, but use the IIWS to break this down over households. The other accounts receivable/payable are allocated over individuals in relation to their current accounts.

The IIWS does not included data on pension entitlements. We use the PCS, which aims to give an overview of the mandatory (employment related) pension entitlements of the population. We use the accrued to date entitlements to calculate the net present value (NPV) of the future benefits receivable. For each individual we combine the PCS with their life expectancy[2] and legal retirement age. Using the Interest Rate Term Structure (IRTS) for the discount rate, as prescribed by the supervisor[3], we calculate the NPV of these entitlements. This calculation does not lead to a perfect match with the SNA totals for some reasons. First, we used the PCS of 2014 for which we assumed developments to 2015 and 2016. Second, the PCS is a sample survey, taken from pension funds, and we need to impute the missing individuals. Third, these entitlements are old age only and partner pensions and disability pensions are excluded. Finally, pension funds could have used other parameters for life expectancy. Using for example socio-economic class could possibly lead to different outcomes. The remaining difference between our estimate and the NA total is allocated proportionally over the individuals. This approach differs from Saez and Zucman (2016), who impute the pension entitlements for 60% to retirees and the remaining 40% to workers. With our method we find that these shares are approximately the other way around.

We further include the }IFCS of 2014 for the voluntary pension entitlements. The }IFCS' main aim is to gather micro-level structural information on euro area households' assets and liabilities. The survey also collects other information in order to analyse the economic decisions taken by households. Only private households and their current members residing in the national territory at the time of data collection are included in the survey, which is held once every three years. The HFCS gives us an average amount of

[2] Life expectancy is imputed in the data set for each combination of age and gender

[3] The Dutch Central Bank prescribes the IRTS monthly, see: https://statistiek.dnb.nl/

voluntary pensions by age and gender, which we impute in our household database and balance it proportionally to arrive at the macro total.

With regard to the non-financial assets, we distinguish dwellings, land underneath dwellings, transfer costs, and company assets of self-employed. For the first three items we use the Valuation of Immovable Property Act (In Dutch: WOZ), for the distributional information. This data source is available as a register, and can be linked to individuals using the unique key. Company assets are distributed over individuals using the consumption of fixed capital in accordance with the current accounts.

For the macro estimate of public pension entitlements, the population is divided in two groups: those who accrue entitlements, and those who receive a pension benefit. For the latter we have actual benefits from the IIWS, for the former we assume that entitlements increase up to the average pension entitlement in that year. In the Netherlands, the first pillar does not depend on labour history, simply on the years one has been a resident in the country. In fifty years a full benefit can be accrued, hence we assume a two percent yearly increase[4]. Similar to the approach for the work-related pension schemes, this is combined with the life expectancy and the foreseen retirement age, to estimate a NPV of all entitlements. This requires discounting, which is the most influential parameter in the calculation. Eurostat proposes a 3% real discount rate. We deviate from this proposal with the main argument that this would make the work-related and the public pension entitlements incomparable. Pension funds in the Netherlands are obliged to use the IRTS, which compares poorly with the Eurostat proposal. First, because the Eurostat approach treats each entitlement the same, while the IRTS differentiates by loan term. Second, the 3% real rate seems too positive in recent years. Loan terms of four years and shorter even had a negative discount rate in the IRTS. Entitlements with a loan term far in the future have the highest discount rate, but with a nominal 2.407% in 2016, this was far from the 3% real interest rate. Apart from the discount rate, we also deviate from the proposed approach for estimating the public pension entitlements. Two approaches are possible, the projected benefit obligations approach (PBO), and the accrued benefit obligations approach (ABO). The PBO approach estimates the entitlements including future wage increases or indexation of pensions. The ABO approach does not include these. Eurostat proposes the former approach, but we favour the latter. Again we argue that ABO best reflects SNA practice, and that it makes our resulting public pension entitlements comparable with the work-related pensions.

[4] We assume that any benefit-gap, because one may have resided outside the Netherlands for some time, is included in the average benefit in a year and thus in the average accrued entitlement in that year.

3. Result

The benefit of the Eurostat approach is that the results have a smooth development over time (Table 1). The slight dip in 2012 compared to 2011 is the result of the retirement age increase that was enacted in 2012. Our preferred method is more volatile, because of the changes in the discount rate. In 2013, the difference between both calculations was largest, amounting to 148 billion euro. Between 2013 and 2014, the discount rate dropped for all loan terms, causing the entitlements to increase with 171 million euros. Towards the end of 2016, the IRTS dropped further, which resulted in a difference of more than 100 billion euros between both estimates, but now the preferred method is higher. From the comparison with the Eurostat proposed method, the effect of the discount rate is not completely clear, because also the approach differs. A sensitivity analysis on the IRTS shows that a 1%-point change of the discount rate results in entitlements a 13%-19% change in entitlements.

Table 1: Estimates of public pension entitlements (closing balance sheets; million euros)

approach	*discount rate*	*2011*	*2012*	*2013*	*2014*	*2015*	*2016*
PBO	3%	1,104,594	1,049,045	1,072,632	1,108,912	1,153,286	1,172,967
ABO	IRTS	971,594	949,894	924,875	1,095,792	1,184,437	1,283,321
ABO	IRTS-1%	1,146,166	1,113,092	1,079,030	1,291,763	1,406,027	1,530,720
ABO	IRTS+1%	836,111	822,025	803,636	942,968	1,013,291	1,093,237

Our distributional information covers the opening balance sheet of 2015 and 2016. In 2015, net worth according to SNA was equal to 2.7 trillion euros, or 347 thousand euros on average per household. However, of the 7.9 million households, 411 thousand had a negative net worth. The share of the 10%-poorest (measured by their SNA net worth) households in total net worth was -1.6%. The top 10% owned 47.2%, and the top 1% even 15.0% of total net worth. The top 10% owned almost 17 times as much as the bottom 40%. The 20/20 ratio, that divides the net worth of the top 20% by that of the bottom 20%, was negative. This was because the bottom 20% had negative net worth.

In 2016, net worth according to SNA increased with 151 billion euros for the households sector as a whole. This benefitted households in the bottom end of the distribution, as fewer households had a negative net worth. The shares and ratios show that the distribution was more even; the bottom decile was less negative, while the share of the top 1% and top 10% decreased. This was also reflected in the ratios.

Table 2: Wealth inequality by concept (opening balance sheets)

		2015					2016				
		Net worth according to SNA	Net worth excluding pension entitlements	Extended net worth (IRTS)	Extended net worth (IRTS +1%)	Extended net worth (IRTS -1%)	Net worth according to SNA	Net worth excluding pension entitlements	Extended net worth (IRTS)	Extended net worth (IRTS +1%)	Extended net worth (IRTS -1%)
Sum	mn euros	2,744,070	1,358,779	3,839,863	3,687,038	4,035,833	2,894,916	1,419,289	4,079,353	3,908,207	4,300,943
Mean	euros	346,849	171,749	485,356	466,039	510,127	363,024	177,980	511,553	490,091	539,340
No. of households	*thousands*	7,911	7,911	7,911	7,911	7,911	7,974	7,974	7,974	7,974	7,974
... with positive net worth		7,501	6,981	7,683	7,640	7,733	7,599	7,138	7,790	7,746	7,839
... with neg or 0 net worth		411	931	229	271	178	375	837	185	229	135
Net worth shares											
D1 - lowest net wor	*share in total*	-1.6%	-5.6%	-0.6%	-0.8%	-0.4%	-1.3%	-5.1%	-0.4%	-0.6%	-0.2%
D2		0.6%	0.1%	1.1%	1.0%	1.3%	0.7%	0.1%	1.4%	1.2%	1.6%
D3		1.4%	0.2%	2.4%	2.2%	2.6%	1.6%	0.3%	2.7%	2.6%	3.0%
D4		2.4%	0.6%	3.7%	3.5%	3.9%	2.8%	0.8%	4.1%	3.9%	4.3%
D5		3.8%	1.5%	5.2%	5.1%	5.5%	4.3%	2.1%	5.6%	5.5%	5.9%
D6		5.9%	3.4%	7.3%	7.1%	7.5%	6.4%	4.3%	7.7%	7.5%	7.9%
D7		8.7%	6.8%	9.9%	9.8%	10.0%	9.3%	7.9%	10.2%	10.1%	10.3%
D8		12.7%	11.6%	13.3%	13.3%	13.3%	13.1%	12.6%	13.5%	13.5%	13.5%
D9		19.0%	18.6%	18.5%	18.7%	18.3%	19.1%	19.6%	18.4%	18.6%	18.2%
D10 - highest net worth		47.2%	62.9%	39.2%	40.2%	38.0%	44.0%	57.5%	36.7%	37.6%	35.5%
Top 1%	*share in total*	15.0%	26.2%	11.2%	11.6%	10.7%	13.1%	21.8%	9.8%	10.2%	9.4%
Indicators											
Gini-coefficient		0.665	0.860	0.571	0.584	0.553	0.635	0.816	0.542	0.556	0.524
Hoover index		0.488	0.631	0.413	0.424	0.400	0.464	0.597	0.392	0.402	0.378
Ratios											
10/40		16.6	-13.3	5.9	6.7	5.1	11.7	-14.5	4.7	5.3	4.1
20/20		-70.9	-14.6	110.6	249.9	62.4	-104.0	-15.2	56.8	86.5	38.4

The results in Table 2 show that including (public) pension entitlements in the net worth concept matters for the level of inequality. When the extended concept is considered, SNA net worth increases with the public pension entitlements from Table 1[5]. As a result, the number of households with negative net worth drops to 185 thousand. In addition, all inequality measures that are presented, show a levelling effect. The share in total net worth of the top 1% is 57.5% when pensions are disregarded, but 36.7% when public, work-related, and voluntary pensions are included (in 2016). Van Bavel and Frankema (2017) mentioned wealth inequality in the Netherlands was high, compared to income inequality. In our database for 2016, the Gini-coefficient of gross disposable income was 0.399, while that of primary income was 0.560. The difference between these two are for a large part in the redistributive schemes of the government, and pension funds. The Gini of 0.816, for net worth excluding pension entitlements, is indeed much higher than income inequality. When the extended net worth concept is considered the Gini drops to 0.542, a difference of 0.274. This is still higher than income inequality, but these welfare schemes have a levelling effect that is not reflected in the wealth measure used in micro data. The developments from 2015 to 2016 are consistent over all measures and concepts. In 2016, the Gini-coefficient is closer to zero than a year before, meaning that inequality decreased. The Hoover index shows that less net worth should be reallocated to achieve

[5] We use the closing balance sheet estimate for 2014 in Table 1 for the opening balance sheet of 2015 in Table 2 to align the reference periods.

equality, and the income shares of the top 1 and top 10% decrease, where those of the bottom 10% improve.

While in income inequality studies it is common to account for economies of scale, because households differ in size and composition, this is less often done so in studies on wealth inequality. An important reason for this is that there is no information about when this wealth will be used, and by who. Income is used in the reference period it concerns, but wealth is postponed consumption, and can even be bequeathed to future generations (Sierminska & Smeeding, 2005).

Table 3 shows different perspectives to household wealth. When net worth is considered per capita, inequality measured by the Gini-coefficient and Hoover index increases, compared to Table 2. Standardising household net worth by the oxford scale, the oxford modified scale, or the square root scale[6] mainly decreases inequality. The exception is the net worth concept excluding pension entitlements. The sensitivity analyses on the extended net worth yield different levels of inequality, but the change compared to Table 2 is similar. When the individual is considered as the unit of analysis, there is a need to allocate the net worth over the individuals within the household. We included only the core members of the household (i.e. the head of the household and the partner), because they are the ones most likely to possess the wealth. We allocated an even amount to partners, even though wealth can be possessed by only one of them. This perspective is similar to the per capita measure of the household, only that some of the household members are excluded. This leads to the lowest measures of inequality in both years.

Table 3: Different perspectives of household net worth

			2015					2016				
			Net worth according to SNA	Net worth excluding pension entitlements	Extended net worth (IRTS)	Extended net worth (IRTS +1%)	Extended net worth (IRTS -1%)	Net worth according to SNA	Net worth excluding pension entitlements	Extended net worth (IRTS)	Extended net worth (IRTS +1%)	Extended net worth (IRTS -1%)
Gini-coefficient	Household	Per capita	0.673	0.869	0.577	0.591	0.559	0.646	0.830	0.550	0.565	0.531
		Oxford scale	0.664	0.862	0.566	0.581	0.548	0.636	0.821	0.538	0.553	0.518
		Oxford modified sca	0.661	0.859	0.563	0.577	0.544	0.631	0.817	0.533	0.549	0.514
		Square Root	0.661	0.858	0.564	0.578	0.545	0.632	0.815	0.535	0.550	0.516
	Individual	Per capita	0.633	0.838	0.530	0.545	0.511	0.599	0.793	0.497	0.513	0.477
Hoover index	Household	Per capita	0.499	0.643	0.419	0.430	0.405	0.476	0.610	0.398	0.410	0.383
		Oxford scale	0.490	0.635	0.410	0.421	0.395	0.466	0.601	0.388	0.400	0.373
		Oxford modified sca	0.486	0.632	0.407	0.418	0.392	0.462	0.597	0.384	0.396	0.369
		Square Root	0.487	0.631	0.408	0.419	0.394	0.462	0.597	0.386	0.397	0.371
	Individual	Per capita	0.458	0.605	0.376	0.388	0.361	0.431	0.569	0.352	0.364	0.337

Negative net worth is not necessarily troublesome, for instance when it concerns households in the early phase of their lifecycle, who have hardly accumulated assets. Also, households can be wealth-poor, but income-rich. Stiglitz *et al.* (2009) explicitly recommend to consider income jointly with wealth,

[6] The Oxford scale assigns the value of 1 to the household head, 0.7 to each additional adult in the household, and 0.5 to each child. For the Oxford Modified Scale the weights of the additional adult and child are 0.5 and 0.3. The square root scale takes the square root of the number if household members.

to get a complete picture of household welfare. The used household dataset allows us to do so on the macro and micro level. It includes all income transactions that add up to macro disposable income, on the individual and household level. This allows for a wealth distribution over gross disposable income deciles. 315 thousand households (4% of the population) are both in the bottom decile of the income distribution, and in the bottom decile of the wealth distribution. Furthermore, we find that none of the income deciles has negative net worth (Table 4). The share in net worth of the top 10% and top 1% of the income distribution is much lower than it was when the simple wealth distribution was considered. The ratios show this as well. The 10/40 ratio was 4.7 (Table 2), this becomes 0.7 in the joint analysis, meaning that the bottom 40% has more assets than the top 10%. The 20/20 ratio drop from 56.8 to 2.8.

4. Discussion and Conclusion

We have shown a large number of wealth inequality measures. Building on the work of Bruil (2018) who created a micro dataset for the Dutch SNA household sector, we add net worth to the analysis. We showed that the concept of net worth greatly affects the levels of inequality.

Table 4: Joint income and wealth inequality

		2015					2016				
		Net worth according to SNA	Net worth excluding pension entitlements	Extended net worth (IRTS)	Extended net worth (IRTS +1%)	Extended net worth (IRTS -1%)	Net worth according to SNA	Net worth excluding pension entitlements	Extended net worth (IRTS)	Extended net worth (IRTS +1%)	Extended net worth (IRTS -1%)
Gross Disposable Income shares											
D1 - lowest income	*share in total*	4.2%	4.0%	5.5%	5.4%	5.6%	4.5%	4.1%	5.7%	5.6%	5.9%
D2		4.0%	3.1%	6.2%	6.1%	6.3%	4.0%	3.1%	6.2%	6.1%	6.3%
D3		5.8%	5.3%	7.5%	7.4%	7.6%	5.8%	5.0%	7.4%	7.3%	7.5%
D4		7.2%	5.9%	8.3%	8.2%	8.3%	7.4%	6.1%	8.4%	8.4%	8.5%
D5		8.3%	7.0%	8.9%	8.8%	8.9%	8.7%	7.6%	9.2%	9.2%	9.2%
D6		9.0%	7.7%	9.2%	9.1%	9.2%	9.2%	7.9%	9.3%	9.3%	9.3%
D7		9.9%	8.5%	9.7%	9.6%	9.7%	10.0%	8.7%	9.7%	9.7%	9.8%
D8		11.2%	10.0%	10.6%	10.6%	10.6%	11.5%	10.5%	10.8%	10.8%	10.8%
D9		13.8%	12.9%	12.5%	12.5%	12.4%	14.3%	14.1%	12.8%	12.9%	12.7%
D10 - highest income		26.6%	35.7%	21.7%	22.2%	21.3%	24.8%	32.8%	20.4%	20.8%	20.0%
Top 1%	*share in total*	7.1%	12.6%	5.4%	5.6%	5.2%	5.7%	9.6%	4.3%	4.5%	4.2%
Ratios											
10/40		1.3	1.9	0.8	0.8	0.8	1.1	1.8	0.7	0.8	0.7
20/20		4.9	6.8	2.9	3.0	2.8	4.6	6.6	2.8	2.9	2.7

Wealth inequality is largest when pension entitlements are excluded, which is often the used concept of wealth studies. Including all pension entitlements has a levelling effect, but wealth inequality remains higher than income inequality. The choices made in the construction of these pension entitlements influence the macro level, but also the inequality measure. The sensitivity of the discount rate, by far the most influential parameter, shows that the Gini would drop by an additional 0.018 if the discount rate used for the public pension entitlements would be 1%-point lower, and that the Gini would increase by 0.014 if the discount rate would be 1%-point higher. This effect is shown explicitly for the public pension entitlements, but the same would hold for the work-related pension entitlements. The requirements for the work-related entitlements in the

SNA are adopted and used for the public entitlements as well, making both schemes comparable.

It does matter whether wealth is considered for the total household, per capita, equivalized, or whether the individual is the unit of analysis. The results differ in levels of inequality, but developments from 2015 to 2016 are all in the same direction. A great benefit of our approach is that we have a fully consistent dataset on micro and macro level, allowing us to make a joint analysis of income and wealth inequality. This offers a completely different insight compared to a separate income or wealth analysis. In the end wealth inequality can be very high or very low, depending on concepts, perspective, or the dimensions included

References

1. Bruil, A. (2018). Adding Inequalities to the SNA Framework: How Macro Disposable Income Benefits and Differs from Micro Disposable Income. Paper prepared for the 35th IARIW General Conference. Copenhagen. Retrieved from http://www.iariw.org/copenhagen/bruil.pdf
2. CBS. (2018). The supplementary pension table: 301407 -ESTAT grant 2015 04.2.21-2. Grant report.
3. Eurostat, European Central Bank. (2011). Technical Compilation Guide for Pension Data in National Accounts. Eurostat Methodologies & Working papers.
4. OECD. (2015). In It Together: Why Less Inequality Benefits All. Paris: OECD Publishing. doi:http://dx.doi.org/10.1787/9789264235120-en
5. Saez, E., & Zucman, G. (May, 2016). Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data. Quarterly Journal of Economics(Vol.131 Issue 2), 519-578.
6. Sierminska, E., & Smeeding, T. (2005). Measurement issues: equivalence scales, accounting framework and reference unit. Luxembourg Income Study. Luxembourg.
7. Stiglitz, J., Sen, A., & Fitoussi, J. (2009). Report by the Commission on the Measurement of Economic and Social Progress. Retrieved from http://www.stiglitz-sen-fitoussi.fr/en/index.htm
8. van Bavel, B., & Frankema, E. H. (2017). Wealth inequality in the Netherlands, c. 1950-2015.: The paradox of a northern European welfare state. Tijdschrift voor Sociale en Economische Geschiedenis, 14(2), 29-62.
9. Vermeulen, P. (June, 2018). How fat is the top tail of the wealth distribution? Review of Income and Wealth, Series 64(Number 2), 357-387. doi:10.1111/roiw.12279
10. Wilterdink, N. (2015). Vermogensongelijkheid in Nederland: ontwikkelingen sinds 1850. Amsterdam: Van Gennep.

Income, consumption and wealth data integration - a household perspective

Friderike Oehler, Sigita Grundiza, Francesca Tartamella
European Commission, Eurostat, Luxembourg

Abstract

Disparities in income, consumption and wealth (ICW) are increasingly analysed, not only by the research community and international organisations but also by the public. The joint distribution of income, consumption and wealth data provides links between the three economic dimensions. These data help to describe more thoroughly material well-being and households' economic vulnerability. Income and consumption aggregates drawn from national accounts (macro-level data) describe the situation of households as an institutional unit in the macroeconomic context. Income distribution (from the European Union Statistics on Income and Living Conditions, EU-SILC) and consumption data (from the Household Budget Survey), on the other hand, are based on micro-level data and used to measure inequalities in the context of social policies. Eurostat has been working on the two work streams of the ICW project: The joint distribution of ICW (based on household surveys) and micro (survey statistics) –macro (national account) data links for households' ICW. Differing concepts and data collection practices mean that the analysis of these different sources do not necessarily lead to the same conclusions as regards people's prosperity. The work has been done in close cooperation with Organisation for Economic Co-operation and Development and European Central Bank.

The paper describes the methods and results of the ICW project. The random hot-deck method was used for the statistical matching of ICW data from the surveys. By comparing micro- and macro-level statistics on households, we can understand their complementarities and differences and build robust links between the data sources.

Keywords

Households perspective; households' surveys; distribution of national accounts; statistical matching

1. Introduction

The European Commission has stressed the need to bring social indicators on a par with the EU reinforced macroeconomic governance. A key part of the strategy is the availability of harmonised statistics at EU level covering the distributional aspects of households' Income, Consumption and Wealth (ICW). Eurostat has been working on the two work streams of the ICW project:

- The joint distribution of ICW[1] (based on household surveys[2]).
- Micro (survey statistics) –macro (national accounts) data links for households' income and consumption[3].

This paper will cover the methodology and experimental results of both work strands.

2. Methodology

2.1. Micro-Macro links for income and consumption

Differing concepts and data collection practices between national accounts and social surveys mean that these different sources do not necessarily lead to the same conclusions as regards people's prosperity. In 2018 Eurostat published a new experimental statistics webpage "Income and consumption: social surveys and national accounts". It includes a conceptual and numerical comparison of income and consumption indicators between social surveys and national accounts as regards household income and consumption. This entails:

- analysing methodological concepts in data sources and quality of underlying data, in order to create categories of income and consumption that are similar in both data sources,
- estimate the total of income and consumption for these categories (using the categorisation established for each data source),
- calculate the data gaps between these categories,
- distribute the data gap (results are not yet published by Eurostat) to derive distributional indicators derived from surveys and benchmarked to national accounts totals.

With regard to the latter, the choice of the method for distributing the gap can significantly impact the results. According to Zwijnenburg J. (2016), the main reasons for data gaps are the quality of micro data used to derive distributions (including measurement and estimation errors), the quality of macro data, the methodological consistency between micro and macro data sources, the quality of the adjustments to the national accounts totals and the correction for the underground economy and illegal activities. The soundness of the distributional results needs to be complemented by metadata. This should reflect the quality of input, consistency of concepts and actual data

[1] https://ec.europa.eu/eurostat/web/experimental-statistics/income-consumption-and-wealth

[2] Income- EU-SILC (European Union Statistics on Income and Living Conditions); consumption – HBS (Household Budget Survey); wealth – HFCS (Household Finance and Consumption survey, ECB)

[3] https://ec.europa.eu/eurostat/web/experimental-statistics/ic-social-surveys-and-national-accounts

(national accounts and households surveys). The method for allocating the gap depends on the previously mentioned factors. If the initial micro or macro data quality is low, the quality of the distributional result will be low. Next, the consistency of concepts and actual micro and macro data needs to be compared. If the consistency of concepts and data is high, the distributional results could be estimated to be good, however if the consistency of concepts or data is not satisfactory the quality of distributional results would be medium or low.

2.2. Joint distribution of ICW

Few countries run integrated surveys for collecting data on income, consumption and/or wealth at once. This is because such a survey would be excessively long and households reluctant to answer. Thus, in most countries individual income, consumption and wealth surveys collect data from different households. As a consequence, there is no way to directly link the records of these surveys and we need statistical matching methods to join the data from the different sources together into a single data set using the categorical variables they have in common.

In previous experiments, results produced by different matching methods (random hot-deck, rank hot-deck, distance hot-deck, conditional mean, mixed approach), which have been described by D'Orazio et al. (2006), had been compared. The random hot-deck method turned out to be well suited to match EU-SILC and HBS data. For joining HFCS to the matched EU-SILC-HBS data set, we make use of the gross income variable available in both EU-SILC and HFCS to apply the rank hot deck method. It should be kept in mind though that both these methods rely on the Conditional Independence Assumption (CIA), assuming that the variables of interest (total disposable income, total consumption expenditure and total assets) are fully explained by the matching variables and independent from each other. Since the CIA might be challenged, indicators based on the joint ICW micro data set are purely experimental at this stage.

A prerequisite of statistical matching methods is the comparability of potential matching variables. Therefore, we first define the reference person of households (following the definition adopted by the Canberra group on household income statistics, UNECE 2011) and harmonise common categorical variables. These potential matching variables are then compared using the Hellinger Distance. We consider variables of the different data sets "equally distributed" if the Hellinger Distance is below 0.05. Subsequently, we run a backward regression to select those matching variables with the highest explanatory power predicting consumption variations. Both EU-SILC and HBS data are stratified according to these matching variables. Within each stratum, HBS donor observations are randomly selected to match EU-SILC recipient observations. If there are less than 3 donor observations for each recipient, the backward selection of matching variables is reiterated reducing the number of variables. The whole matching process is replicated 100 times. The last step then consists of re-calibrating the EU-SILC weights in the matched data set adjusting them to a number of consumption margins. This step is also repeated 100 times.

To join wealth data from HFCS to the matched income-consumption data set, data are stratified according to three matching variables: Household type, food consumption quintile and tenure status. Within each of these strata, the two data sets are ranked according to gross income data. Then, for each receiving observation the closest donor observation of the cumulative distribution function is selected. Again, the process is repeated 100 times.

3. Results

3.1. Micro-Macro links for income and consumption

Figure 1 shows the data gaps for household disposable income (HGDI) for the EU Member States and EFTA countries. In 2015, the HGDI average data gap for the EU-28 between EU-SILC and National Accounts was 27%. In 2015, for Austria the data gap for HGDI was 20%, corresponding to 80% coverage rate for EU-SILC and national accounts data. Figure 1 shows the contribution of each income component to the HGDI data gap. The largest contributors to the data gaps on average were: operating surplus (9 percentage points), property income (8 pp) and self-employed income (8 pp). Operating surplus is not part of HGDI definition in EU-SILC. Data suggest that the wealthiest part of population is not very well covered in EU-SILC. In addition, both property and self-employed income have medium/low conceptual consistency in the both data sources. It should be noted that even if there was no aggregated HGDI data gap, there could be data gaps for income components that offset each other. Denmark and Norway provide such examples.

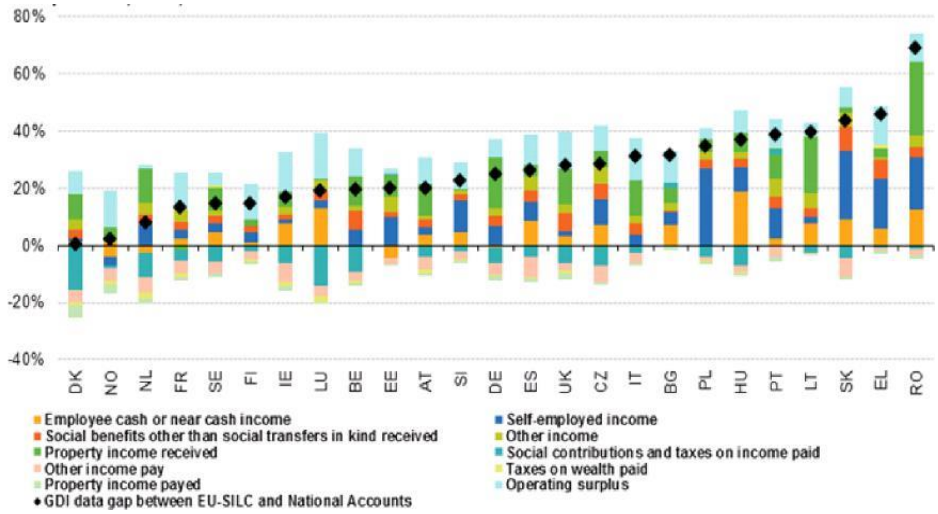


Figure 1: HGDI data gap between EU-SILC and National Accounts-contribution of income components 2015, %

In analogy to income Figure 1, Figure 2 shows the data gap between the HBS and National Accounts together with the relative contribution of each consumption component to the gap for 2010. The data gap for total household consumption is 27 % on average. Although the categories contributing to the data gap vary from country to country, there are several common patterns: for the countries with the largest data gaps between sources, the biggest contributors (in absolute terms) to the gaps are housing, transport, restaurants and hotels, and miscellaneous goods and services consumption. The housing item includes imputed rents that some countries did not provide in the HBS, which could explain the data gap for this consumption item.

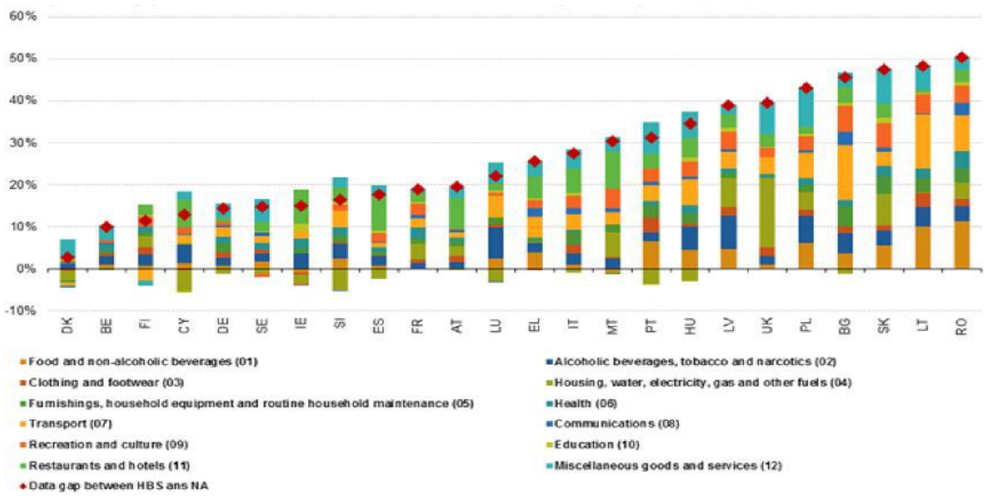


Figure 2: consumption data gap between HBS and National Accounts-contribution of consumption components 2010, %

In order to allocate the data gaps for total income several data allocation methods are explored:

1) Proportional allocation of the data gaps (this test is done for each income component and for total income);
2) Minimising and maximising inequality – allocation of the data gap to the first and fifth quintile, to see the impact of the data allocation method;
3) Parametric data allocation method – modelling the Pareto tail: the data gap is allocated in the top 5% according to the Pareto distribution.

The distribution of the total income data gap was performed exemplary for three countries: Czechia, Finland, and Germany. Table 1 summarises sensitivity tests using the GINI coefficient. In this test the household income is not equalized.

Country	Coverage rate between EU- SILC and NA- total income	GINI EU-SILC data	1.GINI proportional allocation by total income	1.GINI proportional allocation by income component	2.GINI total gap allocated to 1st income quintile	2.GINI total gap allocated to 5th income quintile	3.GINI parametric data allocation
Czechia	72%	33.6	33.6	39.3	30.4	48.0	37.4
Finland	85%	34.8	34.8	37.1	27.6	42.0	39.7
Germany	75%	37.2	37.2	46.8	31.8	48.6	44.4

Table 1: Sensitivity tests using GINI coefficient

This experimental sensitivity test shows that the Gini coefficient varies largely for the selected countries depending on the data gap allocation method. We may conclude that the data gap allocation method changes the GINI coefficient and distribution of income. Further analysis is needed to choose appropriate data gap allocation methods and Eurostat is currently developing this work further.

3.2. Joint distribution of ICW

The comparison of probability density functions of the original distribution of total consumption in the HBS data set and the matched ICW data shows good results for most countries. Probability density functions of total assets in the matched ICW data differ slightly more from the original HFCS data, but do still reflect well the original distribution (see for example Belgium, Figure 3).

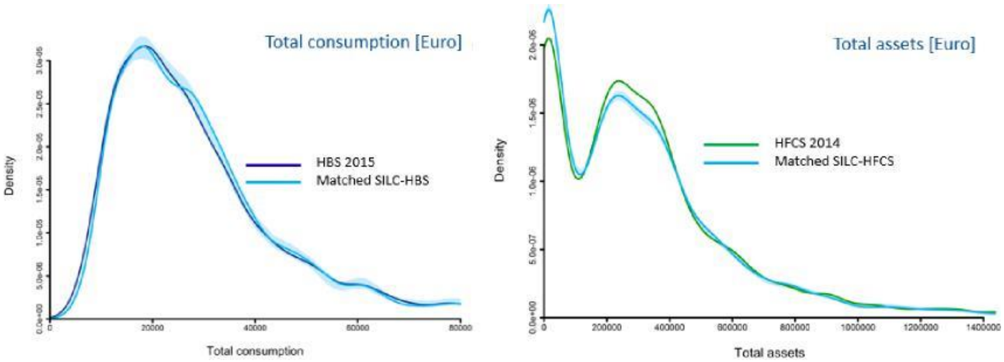


Figure 3: Belgium - probability density functions of total consumption and total assets from original HBS 2015 and HFCS 2014 data and the matched data set. The light blue area shows the 97.5 – 2.5 percentile range of the 100 repetitions of the matching.

Having income and consumption data in a joint micro data set allows to analyse the capacity of households to save. We compute saving rates as the difference of total disposable income minus total consumption expenditure divided by disposable income. The median saving rates of most countries remain unchanged between 2010 and 2015. However, there is a positive trend of increasing median saving rates over all income quintiles in some countries whereas others face an increasing inequality in the capacity to save between low and high income quintiles (Figure 4).

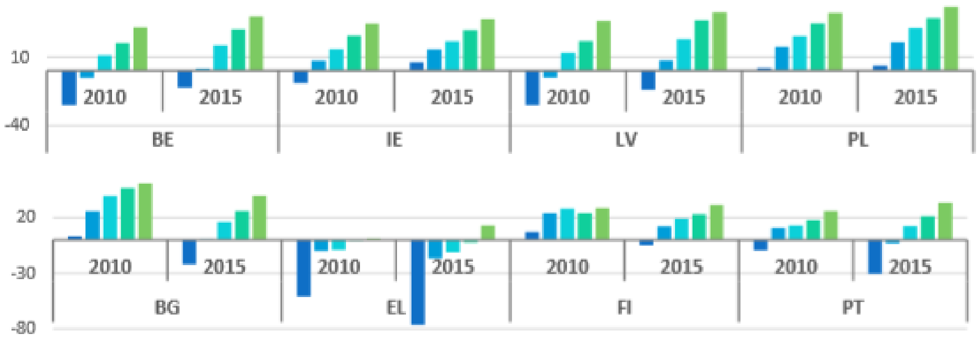


Figure 4: Median saving rates (%) by income quintile, 2010 & 2015. Belgium (BE), Ireland (IE), Latvia (LV, Poland (PL), Bulgaria (BG), Greece (EL), Finland (FI) and Portugal (PT).

This is partly reflected as well in the gini coefficients for savings. In the example shown for Belgium and Greece (Figure 5), the Gini coefficients for consumption expenditures and income remain unchanged whereas a small change is observed in the Gini coefficient for savings and net wealth. Again, these indicators are purely experimental at this stage given the strong assumptions that the joint ICW data set relies on.

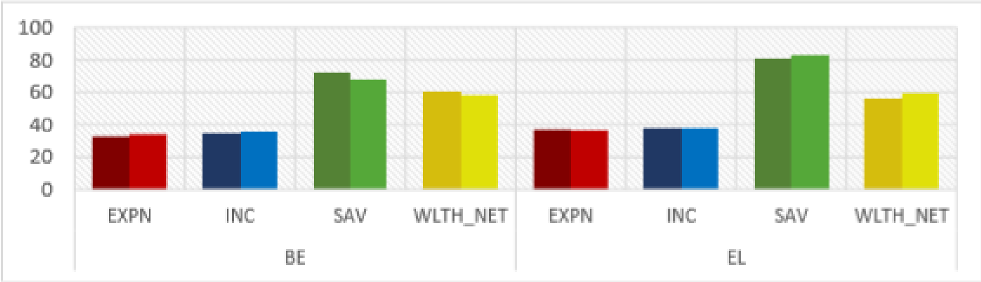


Figure 5: Gini coefficients2010 and 2015 for total consumption expenditure (expn), total disposable income (inc), savings (sav) and net wealth (wlth_net) for Belgium (BE) and Greece (EL). Dark colours: 2010, lighter colours: 2015.

4. Discussion and Conclusion

Eurostat is performing experimental data compilation and analysis on joint income, consumption and wealth data distributions as well as micro-macro comparison and reconciliation using the same methodology for all EU countries. The EU countries are encouraged to repeat these exercises at the national level, possibly using more detailed data or additional data sources.

First Eurostat results include:

- In 2015, the income data gap for the EU-28 between EU-SILC and national accounts was 27%. In general, conceptual and data comparability is high for the following income components: employee cash or near-cash income (excluding the employer's imputed social contributions), social benefits other than social transfers in kind received, and social contributions and taxes on income paid (excluding the employer's imputed social contributions). Income from self-employment shows medium comparability. For property income, comparability is medium/low. Income components with low conceptual comparability and low relevance in terms of GDI are taxes on wealth paid and current transfers received and paid.
- The household consumption comparison between sources was carried out for reference year 2010; the exercise will be repeated once the 2015 data become available. The average data gap between the HBS and national accounts for household consumption is around 27 %; the smallest differences and disparities among the countries are for food and non-alcoholic beverages.
- Eurostat is working on the distribution of the national accounts based on household surveys, the further developments are needed for quality framework and detailed work on the methods how to distribute data, including sensitivity analysis using distributional measures.
- Non-parametric hot-deck methods for statistical matching used to join

income data from EU-SILC with consumption data from HBS and wealth data from HFCS produce fair results with regard to reproducing the original distribution of key variables. From the resulting three-dimensional ICW micro data set a variety of indicators can be produced, notably saving rates for different groups of households. These indicators have to be interpreted with care though due to the unproven assumptions that the method is based on.

References

1. D'Orazio, M., Di Zio, M. and Scanu M. (2006), "Statistical matching: Theory and practice", John Wiley & Sons.
2. Törmälehto V.-M. (2018), "Reconciliation of EU-SILC data with national accounts", 2018 Net-SILC3 International Conference.
3. UNECE 2011, Canberra Group Handbook on Household Income Statistics
4. Zwijnenburg J., (OECD) (2016), "Further Enhancing The Work On Household Distributional Data –Techniques For Bridging Gaps Between Micro And Macro Results And Nowcasting Methodologies For Compiling More Timely Results".

Measurement of wealth inequality with distributional financial accounts indicators

Juha Honkkila, Ilja Kristian Kavonius
European Central Bank, Frankfurt am Main, Germany

Abstract

In view of the potential role that distributional data can play in explaining macroeconomic developments, the European System of Central Banks' Statistics Committee established an Expert Group on Linking Macro and Micro Data for the Household Sector (EG-LMM) in December 2015. The aim of this group is to understand, quantify and explain the main differences between the Household Finance and Consumption Survey (HFCS) and the Financial Accounts (FA). The systematic comparison of the two sources distinguishes between two main explanatory factors: generic differences, which potentially affect all or various components of wealth, and detailed instrument-specific differences in definitions and data collection. Ultimately, the EG-LMM aims to develop Distributional Financial Accounts (DFA) indicators.

The EG-LMM reported its provisional results in the first half of 2019 and this paper summarises the outcome. As a first step, this paper briefly describes the most important conceptual differences between micro and macro data on household wealth. Compared to pure survey data, DFA indicators provide information on household wealth that is coherent with macro aggregates. Inequality indicators between these two sources may differ, since the coverage of survey data is different for various kinds of assets and across various household groups. The aim of the group was to improve the comparability and DFA indicators also in long run. Therefore, it has provided number of issues, which need to be further worked in the medium-term. In the last part of the paper, these improvements are discussed.

Keywords

distributional financial accounts; wealth distribution; micro-macro linkage; household wealth

1. Introduction

The financial crisis of 2008 and changes in the economic environment have increased demand for timely, coherent and consistent distributional information for the household sector. These new data requirements are reflected in the G20 data gap initiative which encourages the production and dissemination of distributional information on income, consumption, saving, and wealth for the household sector. As a result of this, the OECD together with Eurostat has established an expert group investigating how to add distributional information in the National Accounts (NA). Eurostat and the European Statistical System have agreed in the "Vienna Memorandum" in 2016 to work towards this objective. Both of these initiatives are focusing on consumption, income and saving. The European Central Bank (ECB) coordinates the Household Finance and Consumption Survey (HFCS), conducted in the euro area as well as in several other EU countries, and produces the quarterly Financial Accounts (FA), both of which provide information on wealth.

The Statistics Committee (STC) of the ECB established in December 2015 the Expert Group on Linking Macro and Micro Data for the Household Sector (EG-LMM). Its first mandate focused on comparing and bridging the FA and HFCS. The EG-LMM submitted in March 2017 its final report covering the work conducted during the first mandate to the STC and it was agreed that the group will continue by further closing the gaps between the two statistics, by developing further the FA breakdowns which could be estimated by using this link and additionally, consider methods how to estimate time series for these breakdowns. This group delivered its final report by summer 2019.

This paper describes the work of the EG-LMM under its two mandates. In the next section the methodology of this work is summarised. This work has two dimensions. The first dimension is the results calculated at the ECB which are based on the linkage between FA and HFCS data and simple proportional allocation. The second dimension is the methodological work in which the EG provided a methodology and software code for the countries which allow estimating the distributional results and making corrections for instance for under-reporting and the under-coverage of rich households in the survey data. This approach also allows adjusting the estimation method depending on the available auxiliary data. The third section briefly shows preliminary results of the proportional method and the impact of the corrections for under-reporting and under-coverage of wealthy households. Finally, the fourth section concludes and presents further work recommended on the topic.

2. Methodology

First, the project focused on analysing generic differences between macro and micro statistics and then to adjust those to the extent possible. The following clear generic differences were identified: (1.) aim and set up of the statistics; (2.) definition of household; (3.) periodicity, timeliness and reference period; and (4.) valuation.

The aim and set up refers mainly to the fact that the FA is made to cover sectoral interlinkages and the balance sheet interlinkages between economic sectors while the HFCS is mainly focused on producing information on the wealth distribution between households. This appears in the collection of data. The FA data are often reported by counterparts, e.g. banks, as the HFCS data are typically surveyed directly from the households. This can also lead to different interpretation of economic concepts. The populations in the two statistics differ slightly, i.e. persons living in institutions[1] are excluded from the survey population. Concerning the periodicity, timeliness and reference period, the FA are quarterly statistics which are available within four months of the reference period (last day of the quarter for balance sheet items). The HFCS is conducted every three years in most countries and there is typically a long lag between the data collection and data availability. Although the HFCS has gradually synchronised its data collection, the fieldwork periods are still varying from country to country. Finally, concerning the valuation, the FA follow in principle market valuation, or a proxy of market valuation where there is no active market and therefore no easily measurable market price for some assets. Unlisted shares and other equity can be mentioned as an example of these types of assets. The valuation in the HFCS is based on self-assessment of households. This is supposed to be broadly similar to market valuation but particularly in the case of less liquid assets households may not be able to report market prices.

The work of the EG LMM aims at minimising the differences, i.e. to adjust the data when it is possible. Adjustments are made for the definition of household by using source-specific population figures of the two statistics in the calculation of per capita values, i.e. assuming that the portfolio of households living in institutions correspond with the average portfolio of the whole population. Regarding adjustments for periodicity, the closest quarter of the FA to the HFCS reference period – varying from country to country – was chosen for the analyses.

Additionally, related HFCS and FA specific issues and potential errors have been analysed. However, these are typically issues which cannot be corrected in the short-term. In the case of the HFCS – and households surveys overall – these are predominantly related to reporting and sampling bias. Particularly, the sampling biases vary from country to country, as different countries are dealing with those by means of different oversampling strategies, largely depending on the availability of auxiliary data sources applicable for oversampling. These different strategies affect the comparability of the results between the countries. Additionally, the way of collecting data vary from country to country. The majority of countries collect most of their data with traditional surveys via CAPI (Computer Assisted Personal Interviews), but the Netherlands collects their data through a web-survey and Finland uses a

[1] For instance: hospitals, old people's homes, residential homes, prisons, military barracks, religious institutions, boarding houses and workers' hostels, etc

combination of telephone interviews, registers and register-based estimations. In addition to Finland several other countries collect balance sheet or income variables from administrative sources. These different data collection practices affect also the reporting biases. In the case of Finland reporting bias does not exist for most balance sheet variables collected directly from administrative sources. In the case of the Netherlands the reporting error is different from the "traditional survey countries" because of self-administered reporting of data.

Main measurement issues in the FA data are related to the source statistics. The household sector data are typically based on the counterpart reporting, i.e. the reporting of banks and other financial institutions. This does not have an implication on the data quality but the counterparts typically cannot report all the relevant items. There are often weaknesses in the valuation of assets which are not traded on an active market and therefore do not necessarily have an easily measurable market price. Additionally, as the FA are a balanced system covering all economic sectors, some sectors need to be adjusted. In the case of households, other equity and other accounts payable/receivable are typically items which are adjusted. This means that these are items which are considered to be less reliable than other parts of the accounts – and therefore the inconsistencies are typically allocated to these less reliable items.

After identifying generic and source specific differences, the EG-LMM assessed the comparability of financial wealth and its components. The concepts and definitions of items included in household wealth in the HFCS and FA are different. In the FA the definitions of instruments, sectors and concepts such as valuation are given by the European System of Accounts (ESA 2010) and are mandatory in the all EU countries. The HFCS data collection is based on a set of common variables with predefined definitions according to an output-oriented approach. The definition of household wealth in the FA is the entire balance sheet, while the HFCS is able to measure only items that can be reliably collected during an interview. In particular, due to sensitivity issues, the value of cash held by households is usually not collected in household surveys. In addition, the collection of public pension wealth has proven to be difficult in both sources.

As a conclusion of this exercise a bridging table between the HFCS and FA was constructed and is included in appendix 1. It is essentially an update of similar tables presented in previous research[2] , the main difference being that the previous versions were based on the European Systems of Accounts 1995 while the current version was updated to correspond to the European System of Accounts 2010. Additionally, comparing to the previous work, the linkages

[2] The linkage has been presented before in: Kavonius and Törmälehto 2010. Kavonius and Honkkila 2013.

between different assets types has been assessed, i.e. whether the concepts had low, medium or high comparability. It is important to notice that this assessment is based only on the conceptual comparability and does not reflect actual measured differences.

The linkage describe so far in the paper has been used to estimate results by using the proportional method. The second part of this work has been to create a method for the countries which they can apply themselves and for which they should apply auxiliary national level data e.g. on the share of wealthy households above a certain wealth threshold. The EG LMM and ECB have developed a methodology based on iterative Pareto estimation in order to improve the estimations of wealth held by rich households. The adjustments to cover wealthy households that have a very low probability of being interviewed are essential in the estimation of the macro-micro gap. The fundamental assumption is that after a specified threshold the wealth distribution follows a Pareto distribution, the shape parameters of which can be estimated empirically.

However, under-reporting of non-rich household can play a significant role in some countries, and consequently the pure iterative Pareto method may overestimate the amount of wealth held by rich households in the economy, i.e. over-compensating for the lower wealth indicated by respondents. Therefore, as a second method adjustment to non-rich households were conducted. The idea of this applied "hurdle method" is in principle that some households owning certain assets do not report them and unreliable zeros are imputed based on observations from similar households. Both of these methods require country-specific adjustments. Access to auxiliary national level administrative data or rich lists would enrich these estimations.

The EG also concluded that the missing wealth of rich is an issue for the financial accounts. The main issues are related to the (1.) financial wealth abroad which is not captured by tax authorities; and (2.) non-financial wealth, e.g. holiday houses abroad. Concerning the first issue, it was identified that some additional information is available in the BIS locational banking statistics. Concerning the second issue, the ECB agreed to have a voluntary data exchange between NCBs and the EG-LMM. The new data should help to include housing wealth abroad and potentially, clean the other equity assets of the households from the non-financial wealth abroad (notional units).

3. Results

This chapter shows both the preliminary results obtained with the proportional allocation method and the impact that adjustments for under-reporting and under-coverage of the wealthy have on inequality. Figure 1 shows an example of distributional financial accounts (DFA) indicators for

France, Italy, Germany and Finland. The results are shown for the 'broad adjusted' wealth concept developed by the EG LMM, including only financial assets with high or medium conceptual comparability between macro and micro statistics, as well as non-financial assets. Calculations are based on the proportional allocation method, i.e. applying the HFCS distributions of each instrument as such and expanding the levels of wealth and liabilities – again at the instrument level – to the FA totals. Given that the HFCS/FA coverage of more unequally distributed asset types, such as shares and other equity, is typically lower than the coverage of more equally distributed assets, such as deposits and housing wealth, the DFA indicators usually show a higher degree of inequality than unadjusted HFCS data.

Figure 1. Distribution of assets (orange) and liabilities (blue) by net wealth quintile in France, Italy, Germany and Finland, EUR billions

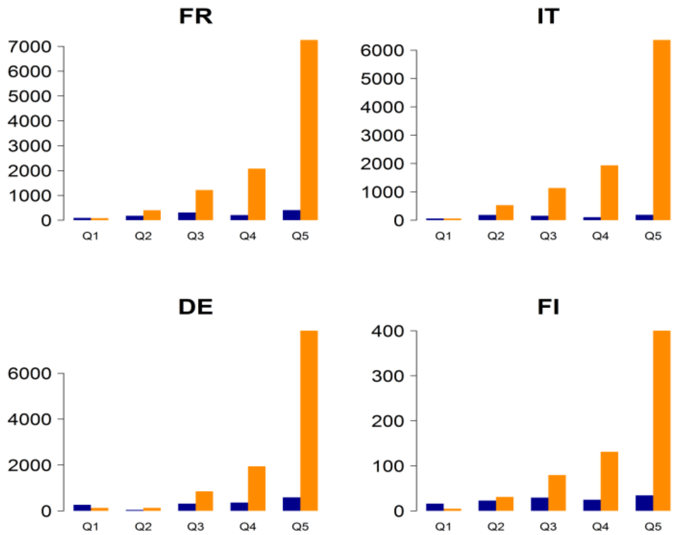


Table 1 illustrates the impact of selected adjustment methods on the distribution of wealth in the four above mentioned countries by comparing the top 10% wealth shares. The first row shows the result derived directly from the HFCS data. The second row indicates the result from pure proportional allocation, which assumes under-reporting by all surveyed households to the HFCS and does not make any correction for very rich households. The third row shows this indicator for pure iterative Pareto method, assuming no under-reporting by the surveyed households, and adjusting only the weights of very wealthy households (i.e. those with gross wealth above a certain threshold, fixed at EUR 1 million). The last column shows the result from the method combining the Pareto adjustment and the hurdle method, assuming possible under-reporting by all surveyed households, but also adjusting the number of very wealthy individuals.

The adjustments made to the coverage of wealthy households, as expected, produce higher estimates of inequality. The impact is particularly high in Italy which is also not applying any oversampling in the HFCS. In France, Italy and Germany accounting for under-reporting of non-rich households indicates that a pure Pareto method may overestimate wealth inequality, and emphasizes the need for adjustments in all parts of the distribution. Counter-intuitively, the hybrid method leads to highest inequality estimates in Finland. However, in Finland ownership of most assets is derived from administrative sources and the hurdle method may actually bias the results, given that Finnish administrative data sources produce reliable information on whether households own certain assets or not.

Table 1. The share of wealth held by the richest 10% of households according to various methods

	France	Italy	Germany	Finland
Base survey	49.4%	42.3%	57.5%	41.4%
Proportional allocation method	50.7%	49.7%	56.0%	54.1%
Pareto method	60.7%	63.4%	64.8%	45.3%
Pareto method combined with hurdle method	53.4%	52.5%	59.3%	50.8%

4. Conclusions and next steps

The work done on the macro-micro linkage of household wealth has established a linkage between the two statistics covering most financial and non-financial balance sheet items. Methodologies created to account for the wealth of rich households severely underreported in survey data has enabled the production of experimental distributional national accounts indicators.

The EG LMM has recognised needs for further development in the area. First, the coverage of non-financial assets in national accounts should as much as possible delineate dwellings and land underlying dwellings from other non-financial assets and land not-underlying dwellings, as well as exclude the holdings of NPISHs. Second, a drawback of the HFCS is that business wealth is a net concept, and current data does not allow a separation between financial assets, non-financial assets and liabilities of such businesses. Third, current comparisons do not include all instruments in FA balance sheets. To cover the entire balance sheet estimation methods for the remaining assets would need to be created. Finally, the possibility to use administrative data sources should be continuously pursued. These data sources are vital particularly in the estimation of wealth held by rich households.

References

1. Kavonius, I. K. and J. Honkkila (2013): "Reconciling Micro and Macro Data on Household Wealth: A Test Based on Three Euro Area Countries", Journal of Economic and Social Policy, Volume 15/2013, Issue 2, Article 3, Southern Cross University/Bepress.
2. Kavonius, I.K. and V.-M. Törmälehto (2010): "Integrating Micro and Macro Accounts – The Linkages between Euro Area Household Wealth Survey and Aggregate Balance Sheets for Households". the 31st General Conference of the International Association for Research in Income and Wealth (IARIW) in St. Gallen, Switzerland, August 2010.

Appendix 1: Bridging table Financial/National accounts vs. HFCS

ESA 2010 Code	National Accounts Instrument	HFCS Variable(s)	HFCS Wave 2 Variable and details	Conceptual Comparability
		Assets		
F.22	Transferable deposits	DA21011	Sight accounts	HIGH
F.29	Other deposits	DA21012	Savings accounts	HIGH
F.3	Debt securities	DA2103	Bonds	HIGH
F.4	Loans	DA2107	Amount owed to household	HIGH
F.511	Listed shares	DA2105	Publicly traded shares	HIGH
F.52	Investment fund shares	DA2102	Mutual fund shares	HIGH
F.512 + F.519	Unlisted shares + Other equity	DA2104 + Part of DA1200	Non-self-employment not publicly traded businesses + Value of Self-Employment Businesses, depending on legal form of business and number of employees.	MEDIUM
F.62	Life insurance and annuity entitlements	DA2109	Voluntary pension/whole life insurance schemes	MEDIUM
N.111 + N.2111	Dwellings + Land underlying buildings	DA1110 + DA1120	Household main residence + Other real estate properties	MEDIUM
(N.11 – N.111) + N.12 + (N.2 – N.2111)	(Fixed assets – dwellings) + Inventories + (Non-produced non-financial assets – land underlying dwellings)	Part of DA1200	Value of Self-Employment Businesses, depending on legal form of business and number of employees.	MEDIUM

ESA 2010 Code	National Accounts Instrument	HFCS Variable(s)	HFCS Wave 2 Variable and details	Conceptual Comparability
		Liabilities		
F.4	Loans	DL1000	Total outstanding balance of household's liabilities	HIGH

Inclusive growth: Does economic growth benefit all?

Peter van de Ven, Jorrit Zwijnenburg
Organisation for Economic Co-operation and Development (OECD)

Abstract

Government policy is very much driven by economic growth. Almost unconditionally, hardly giving any consideration to which population groups capture the gains from the additional income generated, growth of GDP is considered as the primary policy target. However, especially in developed economies, trends in recent decades, possibly reinforced by globalisation and digitalisation, have shown declining shares of compensation of employees, increasing inequalities in income and wealth, and median incomes hardly improving. Not only the resulting political tensions, but also the growing focus on trying to capture (material) well-being, have led to an increased policy attention for arriving at inclusive growth, i.e. economic growth that benefits the whole population and not only a happy few.

As a consequence of the above, user demands for data which link distributional information to macro¬economic statistics has increased tremendously. A problem in this respect is that data on the micro-level which describe income, consumption and wealth at the level of households and individuals, show, sometime quite significant, inconsistencies with data for similar indicators at the macro-level as can be derived from the system of national accounts. This hampers policy analysis and related advice.

This paper describes the work that has been done in the context of the OECD-Eurostat Expert Group on Disparities within National Accounts (EG DNA). The main objective of this group is to arrive at a consistent set of data on the distribution of income, consumption, saving and wealth, by bridging the gaps between the micro- and the macro-data. An additional objective is to develop methodologies for estimating more recent distributional information by combining more recent macro-data with the less timely micro-data. The paper will focus on the methodologies applied, the initial results, and the main problems encountered.

Keywords

Distribution of Income and Consumption; GDP; Household Disposable Income; National Accounts

1. Introduction

1. Growth of GDP is often considered as the primary policy target. However, such a focus on a single indicator ignores recent trends in declining shares of compensation of employees, increasing inequalities in income and wealth, and hardly improving median incomes. If anything, it has become clear that the assumption of "trickle down" is flawed. The resulting political tensions have led to a growing policy attention for arriving at inclusive growth, i.e. economic growth that benefits the whole population and not only a happy few. One of the most influential initiatives in this respect is the 2009 Stiglitz-Sen-Fitoussi Report on the Measurement of Economic Performance and Social Progress.

2. In response to this report, and also motivated by the OECD Inclusive Growth Agenda (among which the work on broader measures of well-being; see e.g. the OECD Better Life Index), several initiatives have also been taken in the area of national accounts, with the objective to put more focus on (the distribution of) household disposable income, instead of GDP:

- The dissemination of a dashboard on households' economic well-being; see http://www.oecd.org/sdd/na/household-dashboard.htm.
- The dissemination of a quarterly news release on "growth and economic well-being"; see e.g. http://www.oecd.org/sdd/na/Growth-and-economic-well-being-oecd-02-2019.pdf.
- A working paper on the decomposition of differences between GDP growth and growth in real household disposable income.
- A working paper on the impact of valuing unpaid household activities.
- Aligning micro data on the distribution of income and consumption to national accounts, to arrive at distributional information that is consistent with macro-economic indicators.

3. In respect of the last goal mentioned in the above, an Expert Group on Disparities in a National Accounts Framework (EG DNA) was launched in 2011, with the participation of a large number of countries and international organisations. The objective of this group was to carry out a feasibility study on the compilation of distributional measures of income, consumption and wealth across household groups consistent with national accounts data. The group developed a methodology on the basis of which first experimental results on income, consumption and saving for five income quintiles, four household groups based on main source of income, and eight groups based on the composition of the household were compiled and published in 2013. In 2015, the expert group engaged in a second exercise focusing on a more recent year and taking into account a number of adjustments to the methodology used in the previous exercise.

4. The results of the exercise show that in general all countries are able to comply with the methodology. Furthermore, countries have micro data available for most of the national accounts items and in case of lacking data, imputations lead to comparable results. However, the results also show that in some cases gaps between the micro aggregates and the national accounts totals can be quite substantial, possibly affecting the overall distributional results. In the remainder of this short paper, the methodology to arrive at distributional measures consistent with national accounts are discussed in Section 2, while some results are presented in Section 3. Session 4 discusses the way forward.

2. Methodology

5. The main objective of the OECD-Eurostat Expert Group on Disparities within National Accounts (EG DNA) is to arrive at a consistent set of data on the distribution of income, consumption, saving and wealth, by bridging the gaps between the micro-data and the macro-data from national accounts. An additional objective is to develop methodologies for estimating more timely distributional information, by combining more recent macro-data with the less timely micro-data. This section focuses on the main methodologies. It does not pay attention to the nowcasting techniques that are in the process of being developed; for more information on this topic, reference is made to Zwijnenburg (2016).

6. The expert group elaborated a preliminary methodology, on the basis of which first experimental results were compiled in 2013 (see Fesseau and Mattonetti, 2013). Subsequently, the group further elaborated and improved the methodology, and published results in a second exercise conducted in 2015 (see Zwijnenburg et al., 2017). The more general goal of the work is to break down the results for the household sector, as included in the national accounts, into more detailed household groups. Currently the project aims at a breakdown according to so-called equivalized disposable income (see below) into five income quintiles; main source of income into four groups; and household composition into eight groups. However, other and more granular breakdowns can also be envisaged, depending on the level of detail and quality of the underlying data, and the robustness of the methodology in combining micro data with national accounts totals. In breaking down the household sector, the project currently focuses on household income, consumption and saving, but in the future the aim is to also include breakdowns of wealth, in order to have a full and consistent set of accounts for the various household groups.

7. In line with the international standards for compiling national accounts, the 2008 System of National Accounts (2008 SNA), the EG DNA project

focuses on households, under the assumption that income is pooled and shared within the household, and that certain types of goods and services (mainly housing and food) are consumed collectively by the household, benefiting from economies of scale. Only private households are taken into account as these are assumed to have autonomy of decision in economic matters, whereas this is often not the case for institutional households (e.g., people living in prison, nursing homes or boarding schools). To adjust for differences in consumption needs for households of different size and composition, the project looks at so-called equivalized results. As the needs of a household increase with each additional household member, although not in a proportional way (due to economies of scale), equivalence scales assign a value to each household member in proportion to its needs, which leads to a number of consumption units for each household[1]. Household income and consumption are divided by this number of consumption units to arrive at comparable measures across households, i.e., *equivalized* income and consumption results.

8. The income measure used to analyse income inequality is household adjusted disposable income as defined in the 2008 SNA. This concerns the income of a household after re-distribution, mainly as a result of government policy, including social transfers in kind. Although the latter item is often excluded from micro distributional analyses, these in-kind transfers can be regarded as a direct alternative to providing households with a cash benefit to purchase the associated goods and services themselves. Their inclusion in income distribution analysis thus leads to a more comprehensive assessment of income inequality and of the impact of re-distributional policies. It also leads to an increased international comparability, because the way in which education and health, the most important components of social transfers in kind, are organised can differ significantly across countries. Consumption inequality is measured on the basis of actual final consumption, as defined in the 2008 SNA, so also including the consumption of goods and services provided in kind by government. As the work of the EG DNA concerns both household income and consumption expenditure, it also provides estimates on the saving of the various household groups.

9. The methodology for compiling distributional results uses a step-by-step approach combining data from national accounts with micro data. Figure 1 on the next page provides a schematic overview of this approach. In the first step, national accounts totals for the household sector, either

[1] In the EGDNA project, the Oxford-modified equivalence scale has been chosen as reference method. Accordingly, the first adult counts as 1.0 consumption unit, any additional persons aged 14 and over count as 0.5, while all children under 14 count as 0.3.

or not including non-profit institutions serving households (NPISHs), are "adjusted" to exclude items that do not concern private households resident in a country, such as NPISHs if combined with results of the household sector, and the income and consumption of institutional households. In the second step, at the most detailed level of available transaction and consumption categories, micro variables are identified that best match the various national accounts items. Different data sources may be selected in this step, depending on which sources provide the best link for the various income and consumption items. In the third step, the micro data are scaled to match the "adjusted" national accounts totals from step 1. Furthermore, imputations are made for items that typically fall outside the scope of micro data. This may relate to items that are specific to the system of national accounts (e.g., imputed items such as financial intermediation services indirectly measured (FISIM), investment income disbursements, and social transfers in kind), but also to items that are likely to be underreported or completely missing from the micro data (e.g., income from the underground economy and illegal activities). Finally, households are clustered into the various household groupings distinguished, and results are derived for the main aggregates and distributional measures.

Figure 1: Step by step approach for the estimation of distributional information

Step 1 - Adjust national accounts totals

Step 2 - Determine relevant variables from micro data sources in relation to the national accounts variables

Step 3 - Impute for missing elements and scale the micro data to the adjusted national accounts totals

Step 4 - Clustering households

Step 5 - Derive relevant indicators for the household groups

10. Some countries have already started to publish distributional results on the basis of the methodology developed by the EG DNA (Australia, Canada, France, the Netherlands, New Zealand, Slovenia and the United Kingdom). In the meantime, the expert group continues to work on further improvements in order to arrive at a more robust methodology thus motivating more countries to publish results. In this respect, apart from deciding on the methodology to allocate social transfers in kind to

households, a major issue concerns the way in which the statistical gaps between the micro-data and the equivalent national accounts aggregates are allocated to the various household groupings, as this will directly affect the distributional results. Table 1 presents the adjustment factors, i.e. the relevant national accounts aggregate divided by the aggregate from the micro data, for some income components from the 2015 exercise. It shows that the differences are particularly large for income from self-employed and unincorporated enterprises (mixed income) and the various property income items. On average, the distributed income of corporations received by households according to national accounts, for example, is more than 5 times as high as what can be found in the micro-data sources. Research is ongoing to improve the allocation of these gaps.

Table 1: Adjustment coefficients for a selection of income components

NA-Code	Item	Number of countries	Average	Minimum	Maximum
B2	Operating surplus	6	1.47	0.47	2.43
B3	Mixed income	9	2.69	1.30	5.24
D1R	Compensation of employees	9	1.15	1.01	1.38
D41R’	Interest received (not adjusted for FISIM)	8	2.08	0.66	6.40
D42R	Distributed income of corporations	7	5.06	0.70	17.76
D41P’	Interest paid (not adjusted for FISIM)	9	3.58	1.02	11.31
D5P	Current taxes on income and	10	1.18	0.78	1.54
D62R	Social benefits other than STiK	10	1.22	0.97	1.55

Source: Zwijnenburg (2016)

3. Results

11. Figure 2 presents some results from the 2015 exercise. The left panel shows the relative income position of each income quintile compared to the average, while the right-hand panel shows the saving rates for the five income quintiles. Of the six countries presented, Mexico has the highest income inequality, followed by the United States, while Slovenia has the lowest inequality in incomes. More surprising are the results for the saving rates, here presented as a percentage of disposable income. For a majority of the countries presented, households in the first and the second income quintile have (substantial) negative saving rates[2]. In Mexico, this is even

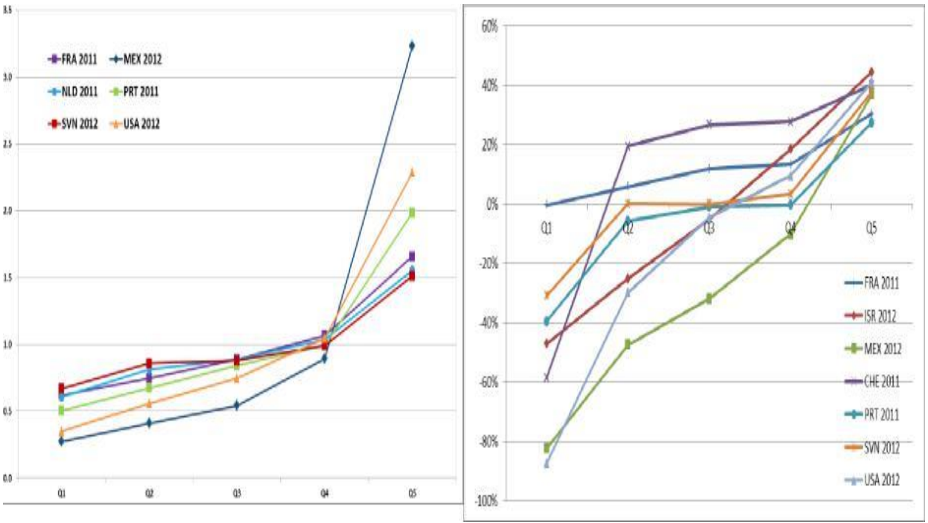
[2] It should be noted that the results for the saving rates in France are affected by the way in which the results from the micro-survey are being analysed and processed. In

true for all income quintiles, except the highest income group. More detailed results can be found in Zwijnenburg et al. (2017).

Figure 2: Some results from the 2015 exercise

Relative position of each household group compared to the average, for adjusted disposable income

Saving as a percentage of disposable income by equivalized disposable income quintile



12. The results for the saving rates have raised some eyebrows. Here, it should be noted upfront that the results are not caused by the methodology to align the micro-data to the national accounts aggregates; one can also generally observe such a divergence between income and consumption in the underlying micro-data. Further research is needed whether these negative rates constitute a statistical artefact, or actual numbers. There may be good reasons for large negative saving rates. Some households, for example farmers, may have very volatile income levels, switching from the lowest income quintile in years with negative income to the highest income quintile in other years. Similar but less extreme developments may occur for other households that become temporarily unemployed, while keeping up their level of consumption. Other plausible explanations may be related to the permanent income hypothesis and the life-cycle hypothesis. According to the latter theory, consumption patterns across age groups are more stable than income patterns, and especially young and old people may have consumption

case the level of consumption is exceeding that of income without households mentioning that they have to reduce their financial wealth or incur liabilities, the level of income is adjusted to bring it in line with that of consumption.

levels which do not match their income levels. For more information, reference is made to Van de Ven et al. (2017), chapter 9.

Whatever the case, having a large negative saving rate for a substantial period of time is not sustainable, and may indeed lead to liquidity and solvency problems, as a consequence of which the relevant households cannot manage their indebtedness anymore, need to look for debt negotiations, or even be declared bankrupt. Extending the distributional data for income, consumption and saving with similar data on wealth could be very helpful in analysing the results for the saving rates, as – disregarding investments in non-financial assets (mainly dwellings and assets for running an enterprise) – negative saving rates would have their counterpart in a run-down of financial assets or the build-up of debt.

4. Discussion and Way Forward

13. The above work of the EG DNA has much in common with the work of the Distributional National Accounts (DINA) team, led by Thomas Piketty. Apart from possible differences in the way micro-macro gaps are being allocated to groups of households, the main conceptual differences concern the target population (private households in EG DNA vs. adult individuals in DINA); the unit of analysis (equivalized household results vs. "equal-split" and "individualistic" results); and the income definition (household (adjusted) disposable income vs. national income). More details on the differences between the two approaches can be found in Zwijnenburg (2017).

14. A lot has already been achieved, and more and more countries start to disseminate distributional data according to the methodology of the EG DNA. However, much more still needs to be done. The main objectives of the EG DNA for the near future are as follows:

i. regular collection and publication of available results;
ii. further exploring ways to improve the allocation, to household groups, of the gaps between the totals from micro data and the national accounts totals;
iii. development of a centralised method for countries that do not compile own estimates;
iv. draft a handbook with detailed guidance on the methodology for compiling estimates;
v. further development of nowcasting techniques to arrive at more timely estimates; and
vi. include data on the distribution of wealth.

In relation to the last objective, the group intends to start looking into the issue of wealth distribution, most probably in close cooperation with the ECB and Eurostat. In addition to providing users with a more comprehensive overview of inequalities, providing users the opportunity to jointly analyse, at the level of household groups, results on income, consumption and wealth, it would also support the compilation of high quality distributional results, by adding another consistency layer to the compilation process, i.e. checking the consistency between saving rates and the accumulation of wealth.

References

1. European Commission, IMF, OECD, UN, World Bank (2009), "System of National Accounts 2008". https://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf
2. Fesseau, Maryse, and Maria L. Mattonetti (2013), "Distributional measures across household groups in a national accounts framework – Results from an experimental cross-country exercise on household income, consumption and saving", OECD Statistics Working Paper No. 2013/04. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=STD/CSTAT/WPNA(2013) 10/RD&docLanguage=En
3. Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi (2009), "Report by the Commission on the Measurement of Economic Performance and Social Progress". http://library.bsl.org.au/jspui/bitstream/1/1267/1/Measurement_of_economic_performance_and_socia l_progress.pdf
4. Van de Ven, Peter, and Daniele Fano (editors) (2017), "Understanding Financial Accounts", OECD Publishing, Paris. https://www.oecd.org/publications/understanding-financial-accounts-9789264281288-en.htm
5. Zwijnenburg, Jorrit (2016), "Further enhancing the work on household distributional data – techniques for bridging gaps between micro and macro results and nowcasting methodologies for compiling more timely results", paper prepared for the 34th IARIW General Conference, August 21-27 in Dresden. http://www.iariw.org/dresden/zwijnenburg.pdf
6. Zwijnenburg, Jorrit, Sophie Bournot and Federico Giovannelli (2017), "Expert Group on Disparities within a National Accounts framework – Results from a 2015 exercise", OECD Statistics Working Paper No. 76. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=STD/DOC(2016)10&docLa nguage=En
7. Zwijnenburg, Jorrit (2017), "Unequal distributions? A study on differences between the compilation of household distributional results according to DINA and EGDNA methodology", OECD. http://wid.world/wp-content/uploads/2017/11/054-DNA OECD.pdf

ISLP challenges alternative truths

Reija Helenius
Statistics Finland

Abstract
We live in a world awash with data. We are confronted with an uncontrollable flood of information daily – whether we want it or not. This stream of information contains a mix of statistical information, fake news and so-called 'alternative facts'. It has long been asserted that statistics mislead people - immortalized by Benjamin Disraeli's famous quip "lies, damn lies and statistics". But is it as simple as that? Perhaps some fault lies with our inability to interpret and use statistical information correctly? In the era of data deluge, how can we assist governments, citizens and businesses to filter the wheat from the growing volumes of chaff and use robust statistical information to support sound decision making? The International Statistical Literacy Project (ISLP) advocates for and promotes the correct use of statistics. The ability to critically analyze and evaluate statistical information is an essential civic skill. In today's data driven world, the ability to separate and discern reliable information from the unreliable is not optional but necessary. Equally important is the ability to compile and disseminate correct information properly. In the context of this new information age, this paper outlines the important contribution made by the ISLP in promoting statistical literacy all around the world and, particularly, supporting teachers to help educate young students - our future citizens and decision makers.

Keywords
Statistical literacy; International Statistical Literacy Project; Statistical Poster Competitions; Fake news; Media literacy

1. Introduction: Too much information - the challenges of information overload

The world is full of information. It has become a cliché to say that we are living in an information overflow. In fact, we don't live in it; we are about to drown in it. In the same way that the world's oceans are becoming polluted, our information channels are filling with both information and disinformation. We are ingesting disinformation in the same way that living sea creatures are ingesting microplastics. According to some estimates, 2,5 quintillion bytes of data is created each day (DOMO: Data Never Sleeps 5.0 [2017]). Google's vice president of engineering, Ben Gomes, estimated in 2017 that 0,25 percent of the Google's search engine results are clearly misleading or offensive (Google [2017]). Statistics are a part of this information overload: official statistics, statistical studies and unofficial statistics alike. Statistics always paint a picture of society – studies using the statistical methods can improve many things, ranging from the fundamental questions of society to highly specialised matters. However, statistics can also be used to mislead people. Statistics or its interpreters can also be sources for disinformation.

Information creators and researchers have the responsibility and obligation to promote the usage of trustworthy information, as well as the know-how of the methods used to create and analyse information. However, statistics are not an intrinsic value in themselves. It is more important to match the correct information with each need. Information needs to serve a purpose in decision making (Pullinger, J. [2013]). Statistics are a part of information data sets or - to use a fashionable term - information ecosystems. For functioning information ecosystems to form, the following things are required: collaboration between its' members, understanding of different fields of expertise and strengths, and an understanding of what actors and sources are not a part of this collaboration network. If we could filter disinformation out of the sea of information overflow, we would get closer to knowledge itself. When it comes to statistics, the right people to ask about the relevance of different sources in the information ecologies are statisticians and the representatives of statistical entities. This role is essentially a role of a fact checker – even a watchdog of sorts.

2. ISLP promotes statistical literacy everywhere in the world

The International Statistical Literacy Project (ISLP) has a history of over 30 years. The main idea the project was founded on is even more relevant today. The ISLP started off promoting the statistical skills of young people, now the promotion of statistical thinking and skills of citizens and other actors in society is also on the table (see more about ISLP history from Helenius, R. [2018]).

ISLP is a project initiated by the International Association for Statistical Education (IASE), which is the educational section of the International Statistical Institute (ISI). The project is run by a four-person executive team and supported by a five-strong steering group appointed by the IASE. The most important actors in the project are the 140 voluntary country coordinators who operate in 73 countries located in every continent.

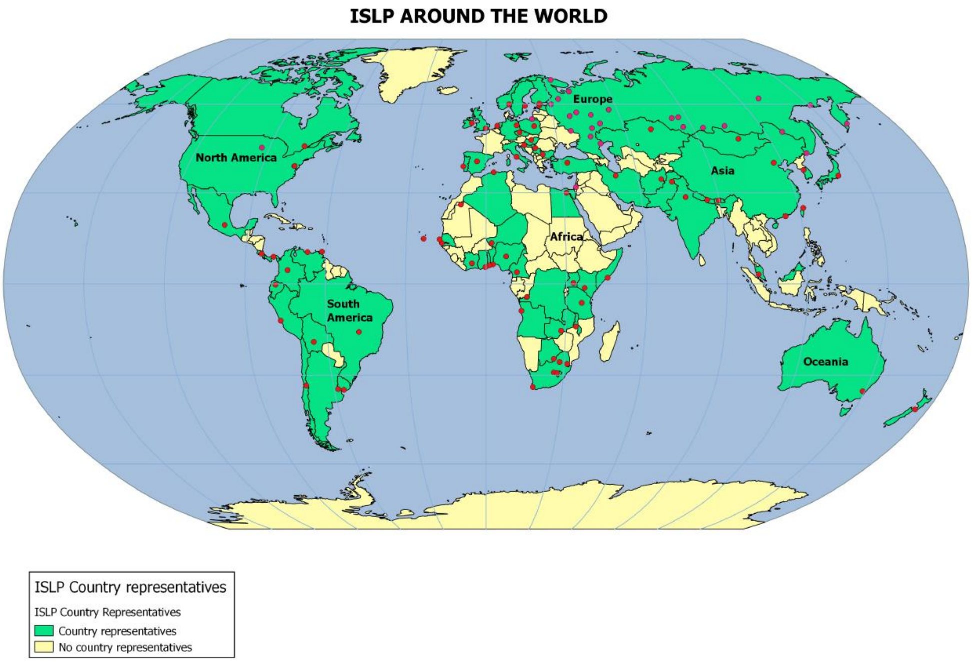


Fig. 1. ISLP network 2019

The project works in close cooperation with national statistical offices, statistical societies, universities and training institutes. All its contributors share a strong will, or even passion, to promote statistical literacy. This is not just for statistical know-how's sake, but for the instrumental value that statistics provide: statistics enable nations, citizens and communities understand and improve the society surrounding them. Official and trustworthy statistics are value creators in a society.

ISLP aims to share good practises to promote statistical literacy. The country coordinators have a significant role to play in their home countries, in the form of different events and trainings. ISLP publishes the ISLP newsletter (https://iase-web.org/islp/Publications.php), in which, amongst other things, the country coordinators share details about their work. ISLP also has a website (https://iase-web.org/islp/) where, amongst other things, essential links to pages regarding statistical literacy and numeracy are shared. ISLP organises a poster competition for young people (see chapter 4) and The Best Cooperative Award competition every other year.

The project is actively brought up in conferences via sessions, presentations and meetings. External funding is constantly sought, especially to support developing countries. Important supporters of the project have been Technology Industries of Finland Centennial Foundation, JMP/SAS and The ISI Tokyo Memorial Fund, Nokia and The Wakimato Memorial Fund.

3. What are the new literacies, how do they connect to statistics and why do we need them?

Literacy is traditionally thought of the ability to read and write. However, the large number of new medias and new technologies has revolutionised literacy. A new set of skills, approaches and abilities are needed to navigate in the digital world. Some say that the whole ethos of literacy is different in the new medias: instead of passively acquiring information, new media encourages participation and active discussion. Some of the new literacies are detailed and explained in this chapter.

Some examples of new literacies are **media literacy** and **critical literacy**. Media literacy "*builds an understanding of the role of media in society as well as essential skills of inquiry and self-expression necessary for citizens of a democracy*" (Center for Media Literacy). Media literate individuals can control the way they interpret things they see in the media – instead of letting the interpretations control them. Critical literacy comes close to this term. It is usually defined as an ability to analyse and evaluate different texts (or other media) and considering what bias, influence or discrimination it might have. Analysing Internet content, such as memes, critically, can increasingly reveal structures of social power and influence (Knobel, M. & Lankshear, C. [2007]).

The new medias also encourage exploring statistics in a new way. The statistics user can create new statistics and viewpoints to each data with different applications. Using apps requires a different skillset that is not necessarily related to traditional literacy or statistical understanding. While obtaining information is getting easier, using it requires new kind of skills. The term **data science skills** can be used; another new civic skill along the new literacies. Engel, J. (2017) illustrates the different components of data science skills: these are information technology and data handling skills, mathematical and statistical skills and skills related to phenomenon-based learning. The ProCivicStat -project and its final report (2018) emphasised the importance of social context and surrounding society when using and teaching statistics. The project yielded six key recommendations for this purpose. They are shown in the table below.

Table 1. Six key recommendations by ProCivicStat

#1	**Statistics education activities should promote engagement with social issues and develop leaner's critical understanding.**
#2	**Use relevant data and texts, and highlight the multivariate, dynamic and aggregated nature of social phenomena.**
#3	**Embrace technologies that enable rich visualizations and interactions with data about relevant social phenomena.**
#4	**Teaching methods should develop skills of critical interpretation of a wide variety of data and text sources.**
#5	**Assessments should examine the ability to investigate and critically understand data, statistics findings and messages about key social phenomena.**
#6	**Promoting the understanding of civic statistics requires a systemic change and collaboration by relevant stakeholders.**

Source: (ProCivicStat, 2018)

Media sources should be evaluated based on valid knowledge – not feelings or beliefs. This is where statistics form an integral part. Sashi Sharma outlines a few examples of where statistical literacy helps to make sense of the news media: for example, sensational news headlines ("*Kids who watch 'Sesame Street' do better in school*" (Sharma, S. [2017]) that are often based on singular studies, with small sample sizes, confounding variables and sampling error. Sharma puts it well: "*Indeed, citizens without statistical literacy may not be able to discriminate between credible and incredible information and will have difficulty in interpreting, critically evaluating and communicating reactions to such messages*" (Sharma, S. [2017]).

There are several formal definitions of statistical literacy available (See Gal, I. [2002]; Schield, M. [2010]; Wallman, K. [1993]). On a general level, the "traditional" view to literacy ('*literacy leads to development*') differs greatly from the new concept, which argues that literacy is rooted in social customs and has a social meaning. Social literacy implies training those who want to communicate something (UNESCO Institute for Lifelong Learning. [2013]). In recent years, ISLP has been concerned with responding to this more modern, and broad, concept of literacy. The definition of the term 'statistical literacy' by Gould, R. (2017) reveals the demands of a modern society, such as understanding issues of data privacy and ownership and the provenance of data and understanding how data are stored and how representations in computers can vary and why data must sometimes be altered before analysis.

Finally, statistical literacy leads us to the term **data literacy**. According to the Oceans of Data Institute (2015): "*The data literate individual understands, explains and documents the utility and limitations of data by becoming a* *critical consumer of data, controlling his/her personal data trail, finding meaning and taking action based on data. S/he can identify, collect, evaluate, analyze, interpret, present and protect data*." The outline of types of data literate citizen by Wolff, A. et al. (2016) is apt for the context of the ISLP project. Based on the situation where the data is used, they categorize the data literate citizens to communicators, readers, makers and scientists.

When promoting statistical or data literacy, we need to keep in mind that in global context we cannot begin an era of media, critical, or statistical literacy until the whole population of the Earth is able to read and write. This is particularly important in the ISLP project that wishes to focus its operations to developing countries. Currently around 14 % of adults (UNESCO [2016]) in the world's population (7,55 billion in the year 2017 [UNSD 2017]) are illiterate.

By educating people about statistical literacy, we are building a more informed community that is ready to think for themselves and critically interpret the news they see in the media. The purpose of the ISLP is to advance basic statistical literacy. In broad terms this means the aim is to promote statistical literacy among the widest possible audience. It also means advancing people's working life skills and social abilities. From an ISLP perspective, statistical literacy is a life skill required for today's information society. It is needed in ordering, conducting and understanding various surveys and reports, be they government, private sector or media. At a minimum, statistical literacy is an everyday skill for understanding news and media, and for making everyday decisions (see Figure 1.1) (Helenius, R., Campos, P., D`amelio, A. & MacFeely, S. [2019]).

Table 2. Dimensions of Statistical Literacy

Basic statistical literacy is a skill for people	Deeper usage skill is a growing requirement for people acting in information society
Understanding of basic concepts and key figures	Knowledge of concepts, figures, methods and presentations
Ability to use and understand numerical and statistical data in everyday situations	Ability to use and produce numerical and statistical data in work duties and decision making

Source: Helenius, R. (2016)

4. Awakening youth's interest in statistics

Alongside teaching statistical literacy, another goal of the ISLP project is to bring statistics as close to everyday life as possible – and thus demonstrate their necessity. Everyone gets educated: both the valedictorians of the class and those who struggle with mathematics or find it difficult to motivate themselves to study it. The latter is a key target group for promoting statistical literacy if we wish to extend statistical understanding as far as possible in society.

ISLP has been organising poster competitions since 2010. The aim is to teach students the whole statistical study process through creating a statistics poster: from defining the study question to collecting data, analysing it, presenting it and making conclusions. In the year 2018-2019, 37 countries and 14 721 students participated. The number of participants has grown every year. A statistical poster (see picture 2) is created as a team effort between 2-5 students, which makes it possible to combine different knowledge areas. One student might be gifted in the phenomenon researched, one in mathematics, third in written communication, fourth visually, and so on. Creating a poster also supports phenomenon-based learning, which has become common in several countries. As an example, Finnish schools had rarely used poster creation as a pedagogical method before. Through national poster competitions, creating a statistical poster has become one way of going through the basics of statistics in upper comprehensive or secondary school. Choosing their own study topic makes the creation process even more appealing to the students. The data used can either be self-collected or pre-existing reliable statistic data. Overall, after creating a statistics poster, one is well on their way to understanding the basics of statistics. The article by MacFeely, S., Campos, P. and Helenius, R. (2017) outlines the factors that create a successful poster competition.

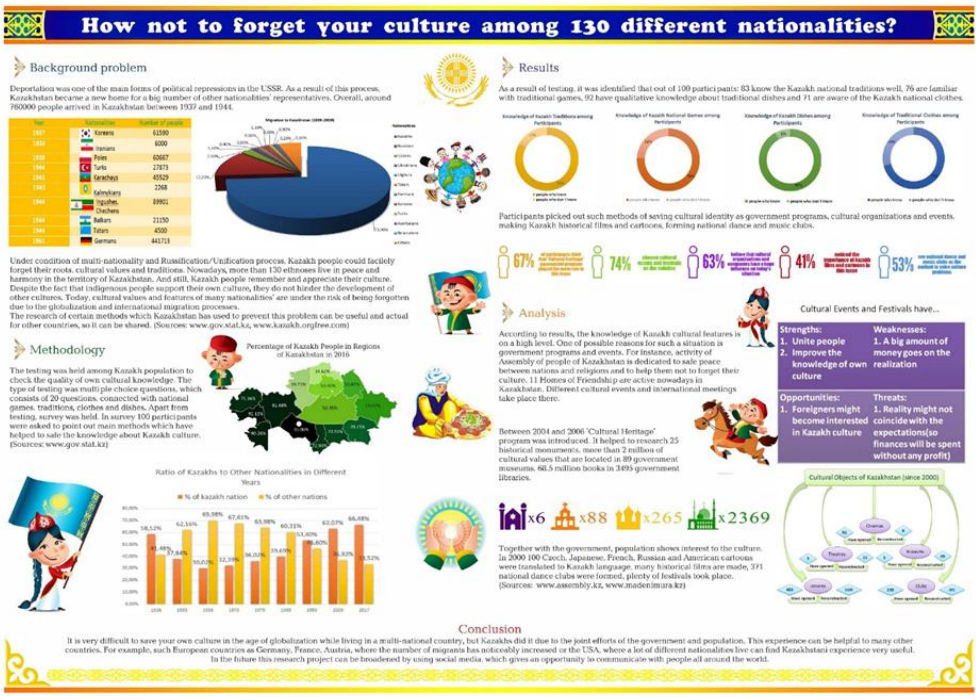


Fig. 2. An example of a statistical poster, which was awarded second place in the upper comprehensive school division of the 2016-2017 Poster Competition. Authors: Adiya Sadanova, Marina Tarassova and Botagoz Dauletbek, Kazakstan, (Source: ISLP Website, [2019]).

One way to spark the youth's interest in statistics for the first time is gamification. The Digicom-project by Eurostat has produced many excellent mobile- and computer-based games which pave the way to interest, and even addiction, to statistics (f. ex. Luhmann, S. et al. [2018]). ISLP awards The Best Cooperative Project every second year to promote great collaboration customs in the field of statistics worldwide. In the year 2015, the award was given to Exploristica – Adventures in Statistics by Statistics Portugal. It is a great example of how familiarisation with statistical thinking can happen through an adventure game. The game is geared towards 12-17-year olds and goes through the five important phases of the statistical process - select, collect, describe, estimate and interpret. The game consists of six parts, such as the Submarine game (players pilot a submarine through an underwater ecosystem and collect specimens of new species of reptiles for analysis), Quiz (teams compete in a rapid fire quiz to answer questions from Census 2011, in a "Who Wants to be a Millionaire" type setting) and the Giant Ape game (players help "Conga" the gorilla conduct a sample survey of voters in a small town). The game is a physical whole. (Campos, P. & Ribeiro, J. [2018].) There are also plans underway to make it into a mobile version.

Fig.3. Submarine game in Exploristica (source: Campos, P. & Ribeiro, J. [2018])

5. Discussion and Conclusion

Media literacy, statistical literacy and data literacy are increasingly important civic skills that affect all the operators of society in one way or another. ISLP, from its own part, promotes these aims globally through its collaboration network. The more comprehensive the ISLP network becomes, the more weight it has. In addition to statisticians and statistics professionals, it is even more important to network and build an ecosystem of statistical literacy with other agents in society, such as educational institutions and the media, who can act as partners in cooperation in the mission to raise critical information users and skilled information producers.

As we aim to develop the basic abilities of statistical literacy at length – which is the main goal especially for the ISLP project – it is especially important to examine information usage from the standing of each citizen and consider their life circumstances and skills. Information, or statistical information, is not a detached number or formula, but a part of each individual data user's information process. It can be a tool in both mundane and difficult decision-making situations, part of the learning process, a communicative method and so on. The poster competitions organised by ISLP for their part train youth to statistical thinking and using statistics in many ways and from different perspectives.

On a larger scale, reliable statistical information and its practical application is a part of democratic, egalitarian, and society developing capital. ISLP wishes to also be a part of this process and encourages more and more participants to join its network.

References

1. Campos, P. & Ribeiro, J.& (2018). *Exploristica – adventures in statistics with augmented reality*. . Retrieved from https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_9K3.pdf Invited paper.
2. Center for Media Literacy. *Media Literacy: A Definition and More*. Retrieved 9.7. from https://www.medialit.org/media-literacy-definition-and-more
3. Engel, J. (2017). *Statistical literacy for active citizenship: A call for data science education*. Statistics Education Research Journal, 16(1), 44-49.
4. Gal, I. (2002). *Adults' Statistical Literacy: Meanings, Components, and Responsibilities*. International Statistical Review, 70, 1, 1-25.
5. Gould. R., (2017). *Data Literacy is Statistical Literacy*. Statistics Education Research Journal, 16(1), 22-25.
6. Helenius, R. (2016). *Facing the Future with Knowledge* – Cooperation with Educational Institutions. Paper presented at Nordiskt statistikermöte, Sweden, August 22-24, 2016.
7. Helenius, R. (2018). *International Statistical Literacy Project past and now*. Retrieved from https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_7B3.pdf. ICOTS10 Invited paper.
8. Helenius, R., Campos, P. D`Amelio, A. & MacFeely, S. (2019). ISLP *country coordinators as ambassadors of statistical literacy and new innovations*. Unpublished manuscript, Statistics Education Research Journal.

9. International Statistical Literacy Project. *ISLP Poster Competition 2016-2017*. Retrieved from https://iase-web.org/islp/Poster_Competition_2016-2017.php.
10. Knobel, M., & Lankshear, C. (Eds.). (2007). *A new literacies sampler* (Vol. 29). Peter Lang.
11. Luhmann, S. et al. (2018). *A European effort to explore games and the gamification of official statistics*. Paper presented at CESS 2018 (Conference of European Statistics Stakeholders), Germany, October 18-19, 2018.
12. MacFeely, S., Campos, P. & Helenius, R. (2017). Key *success factors for statistical literacy poster* competition. Statistics Education Research Journal, 16(1), 44-49.
13. ProCivicStat Report (2018). Engaging Civic Statistics: A Call for Action and Recommendations. Retrieved from https://iase-web.org/islp/pcs/documents/ProCivicStat_Report.pdf
14. Pullinger, J. **(**2013). *Statistics make impact*. J. R. Statist. Soc. A (2013) 176, Part 4, 819–839.
15. Sharma, S. (2017). *Definitions and models of statistical literacy*: *a literature review*. Open Review of Educational Research, 4:1, 118-133, DOI: 10.1080/23265507.2017.1354313
16. Schield, M. (2010). *Assessing Statistical Literacy*: *Take CARE.* Assessment methods in statistical education: An international perspective, 133-152.
17. UNESCO Institute for Lifelong Learning. (2013). *2nd global report on adult learning and education.* Unesco Institute for Lifelong Learning. Feldbrunnenstrasse 58, 20148 Hamburg, Germany Alemania.
18. UNESCO Institute for Statistics. (2017). *Literacy Rates Continue to Rise from One Generation to the Next.* Fact Sheet No. 45, FS/2017/LIT/45.
19. United Nations Statistic Division. (2017). Demographic Yearbook 2017. Retrieved 10.7.2019 from https://unstats.un.org/unsd/demographic-social/products/dyb/dyb_2017/
20. Wallman, K. (1993). *Enhancing Statistical Literacy: Enriching Our Society.* Journal of the American Statistical Association, 88(421), 1-8.
21. Wolff, A., Gooch, D., Montaner, J., Rashid, U., & Kortuem, G. (2016). *Creating an understanding of data literacy for a data-driven society*. The Journal of Community Informatics 12(3), 9-26.

Experimenting datascience with the French SSP Lab

Dr. Elise COUDIN, Dr. Mathilde POULHES
INSEE, France

Abstract
INSEE created a new unit dedicated to innovation and R&D in terms of new data sources and new statistical methods for the production of official statistics. Part of the Directorate of Methodology, Statistical Coordination and International Relations, the SSP Lab (Official Statistics Service Lab) is a resource and animation centre for applied research, experimental development and new ways of working within the French official statistics service (Service de Statistique Publique; SSP). Sponsored by business units in charge of statistical production, the SSP Lab conducts experimental projects in partnership with these units. It also leads networks within the French official statistics service, in collaboration with external partners, including European peers and academics.

Keywords
Datascience; official statistics; datascience innovation

1. What is the background of the SSP Lab creation?

1.1. General context

The extraordinary development of new data sources, methods and processing capabilities of the digital revolution requires official statistics to produce ever more detailed information and to improve timeliness. INSEE, like other NSIs, therefore pays close attention to developing and federating innovation within the national official statistics service. Official statistics must absolutely maintain a high level of excellence in order to produce objective, relevant and meaningful statistical inputs in the public debate within a competitive context of new emerging data providers and actors. This level of excellence requires rare skills that must be mobilised and developed throughout the agents' careers, in a general context of reduction of public sector resources.

In this context, INSEE set up in May 2018 an innovation unit called the SSP Lab (Official Statistics Service Lab, SSP stands for *Service de Statistique Publique*, the official statistics service, i.e. INSEE and the Ministerial Statistical Services), which aims to monitor and disseminate innovative statistical methods and explore new data through experimental projects carried out in partnership with business units of the official statistics service.

The creation of this unit corresponds with the ESS vision 2020, which aims to explore new opportunities of the digital transformation and build organisations capable of working and collaborating with agility within the official statistics service and with European peers (e.g. through the Big Data task force and related ESSnets), but also with data producers and academics. It is inspired by the experience of other NSIs in the Netherlands, Italy, Canada and the U.-K. amongst others that have modernised their organisations to respond to these challenges.

1.2 Context in France

The SSP Lab was created following experiments using Big Data sources carried out in several of the INSEE and Ministerial Statistical Services units, e.g. the project incorporating mass retail scan data for producing the CPI (INSEE, started in 2011), the web scraping of job ads for estimating job vacancies (Ministry of Labour), and the use of administrative health data for statistical purposes (Ministry of Health). There has been also a positive regulatory context for exploring new sources of data since the adoption of the so called "Law for a Digital Republic" of 7 October 2016. At INSEE request, the latter makes mandatory the transmission of information from internal databases for companies concerned by a statistical survey. [1]

The SSP Lab was also created following the restructuring strategy of INSEE. INSEE set up in 2012 the Directorate of Methodology, Statistical Coordination and International relations (DMCSI) with the idea of pooling rare and strategic resources for possible synergies. Within the Department of Statistical Methods (within the DMCSI), the Division of Applied Econometrics and Evaluation (DMAEE) explored and disseminated innovative statistical and econometric methods within the official statistics service by providing support and advice to statisticians in charge of production. This division gradually integrated a role of coordination and animation of the work on Big Data within the official statistics service. It included two full-time data scientists in 2016. It joined the network of European peers through the Big Data task force and the ESSnet Big Data. It conducted several experimental projects on new sources, e.g. evaluating the interest of Internet sources for nowcasting economic indicators (Combes, Bortoli, Renault, 2015 and Combes, Bortoli, 2017), and started a collaboration with the Orange SenSE laboratory on mobile phone data. In response to requests from business units or Ministerial Statistical Services, investments were also conducted to acquire skills and experience on textual analysis and machine learning methods, and to disseminate practical instructions. These investments provided the opportunity to launch reflections

[1] It also cancels data transmission royalties between public administrations for statistical purposes

on working modes, allowing different profiles to work effectively together. In collaboration with the IT department, a team of two persons from the division and one from the IT department won the second prize of the Big Data Hackathon organised by Eurostat during the New Techniques and Technologies for Statistics conference in March 2017.

Datascience innovation for statistical production and studies is a crucial activity for maintaining high-quality, meaningful and relevant statistics, and is pledge of productivity gains. The visibility of its innovative works is also essential for the NSI to reinforce public confidence. In 2016, INSEE's medium-term strategy (INSEE 2025) recommended the creation of a unit dedicated to R&D for promoting reactive lab testing and experimentation, ensuring a technological watch, as well as leading and animating a network on related topics. Relying on this, a launch committee worked on the organisation and the objectives of such an entity from November 2016 to July 2017, and proposed to set up the SSP Lab unit within the DMCSI, capitalising on the division DMAEE resources and experience. The aim of the SSP lab would not be to concentrate all innovation within the official statistics service, but to be an expert resource and partner catalysing innovation. It would produce in partnership with business units innovative prototypes using new sources, new methods, new tools, even new angles of study, upstream of production projects. It would join innovation networks, within the official statistics service, but also with academics and European peers. All statisticians of the official statistics service are indeed trained at the master-level engineering *Grandes Ecoles* ENSAE and ENSAI, which are specialised in statistics, computational IT for statistics and economics, and have therefore already integrated datascience and technologies for Big Data in their curriculums. As such, a "branched" vision within the whole official statistics service of the datascience innovation can be adopted. The SSP Lab programme relies mainly on the business unit proposals, functions with reactivity (mainly through experimental projects that assume the risk of failure) and one of the important roles of the SSP Lab is the dissemination of a culture of innovation and knowledge, exchange of good practices, sharing feedback.

The French Decree of 10 April 2018 modified the INSEE organisation to set up this new unit. Closely related, the new Directorate of Information Technologies. This Directorate will include a unit dedicated to IT innovation called the Enterprise Architecture, Security and Innovation (EASI) that will foster technical innovation in the computing field (for example, by setting up distributed computing platforms and containerisation). The SSP Lab and the EASI will work in close collaboration.

2. How does the SSP Lab function in practice?

2.1. Composition, governance and resources

The SSP Lab team is made up of eight full-time datascientists, econometricians and IT specialists, including its chief and deputy. The idea is to group complementary profiles in terms of skills and experience (seniors / juniors).

The governance of the SSP Lab is ensured by the members of the executive committee of INSEE and the chiefs of the Ministerial Statistical Services, who review once a year the performance of the Lab experiments and provide global directions for future programmes. These directions are flexible enough to ensure reactivity and innovation on unexpected topics. Consequently, the investment decisions are made continuously.

The SSP Lab works in close collaboration with the IT innovation unit, notably through the use of the new IT platform that allows distributed computing and other treatments adapted to the needs of Big Data manipulation. The current version of this platform is unfortunately still incompatible with the security requirements of individual data. Consequently, only anonymised samples are used in the experiments.

2.2. Activities

The SSP Lab’s activities involve experimental projects, networking activities and dissemination.

a. Experimentation

The Lab conducts applied research and experimental developments involving statistical or datascience innovation in partnership with services in charge of production. The experimental projects are defined as follows. The subject is proposed and sponsored by a business unit. The project should last around six months. More generally, the idea is to define several stages that entail specific deliverables. The approach is exploratory, on a "small" scale, and similar to the lab testing approach. Consequently, the deliverables may be of different types: proofs of concept (POC), reports, experimental prototypes, etc., but cannot be directly integrated into a production process. Ultimately, they may not even be implemented. Experimental outcomes never being certain.

Several conditions must be fulfilled to launch an experimental project involving the SSP Lab and a business unit. As already mentioned, the needs are expressed by the business unit, should be in keeping with the orientations previously mentioned, and should concern (at least) one of the Lab's fields of action (new data sources, new statistical methods or new statistics). In addition, the SSP Lab should have the appropriate skills or should be able to acquire them rapidly.

If the previous conditions are fulfilled, then a mixed SSP Lab/Business Unit/IT team with the required skills (datascience, IT, etc.) is set up. Flexibility and real commitment to the project are required from the different participants. Several models of engagement, e.g. one day-per-week contractual commitments, customised projects and team descriptions are possible. Flexible and 'agile' ways of working are promoted. The mixed team is expected to work in steps and cycles, and to deliver regular outputs. The requirement is to broadly disseminate the results (final and intermediary) and to share experience through different media (intranet, newsletters, etc.).

b. Networking activities

For some experimental projects, the SSP Lab contributes to and benefits from European expertise. In particular, the SSP Lab takes part of the European exploration of the potential of Big Data to integrate it into the official statistics production (Big Data task force and related ESSnets). It participated in the ESSnet Big Data I Mobile Data Working Group, which aimed to 5 clarify the possibilities of accessing mobile phone data, lay the foundations for a methodology for their treatment and estimate population present within a given place and time and related indicators. The SSP Lab will continue to participate in the investments on mobile phone data in ESSnet BD II and will coordinate the French participation in other work packages, including those on satellite data and smart statistics.

For other experimental projects, the SSP Lab may also form academic partnerships in order to benefit from external expertise. For instance, it currently collaborates with the Institute of Public Policies (IPP) of the Paris School of Economics (PSE) to explore the modelling of professional careers for microsimulation purposes by using machine learning methods. Partnerships with private actors for experimentation on private data are also in the scope of potential activities.

c. Dissemination

The SSP Lab ensures the role of monitoring and dissemination of innovative statistical methods through training on datascience methods (e.g., machine learning, textmining and coding languages such as Python) for the statisticians of the official statistics service and the provision of technical documents. The SSP Lab animates networks on innovative topics within the SSP (dissemination of a Big Data newsletter and Big Data seminars). The SSP Lab and the IT innovation unit also work together to promote 'agile' ways of working, in particular via collaborative workshops and hackathons open to the members of official statistics service and close institutions.

3. Future plans

The short-term plans of the SSP Lab are first to conduct the new experimental projects, increase its visibility within the official statistical service and increase the dissemination of its outputs on appropriate medias (blog, intranet, experimental page on the Internet). A second important issue is to develop appropriate contractual frameworks to host in the SSP Lab external researchers, postdoctoral fellows and PhD students.

4. Some examples of experimental projects

The section details three examples of ongoing experiments. The entire list of activities planned 6 for 2018 is available in the appendix.

4.1. Employer identification in census survey

Sponsor: INSEE Social Studies Directorate (Census unit);
Team: SSP Lab (3 members), Census unit (2 members), IT (4 members), other units (2 members); Schedule: January 2018 (hackathon) and then from June to December 2018 (experimentation); Expected deliverables: training (hackathon), an experimental prototype and an experimentation report.

a. What opportunities do Big Data techniques offer?

Currently, respondents of the Census report the name of their employer, the activity of the legal unit and the address of their workplace. These response boxes are filled out in a non-standardised way, and frequently result into incorrect answers (spelling mistakes, imprecision and confusion between fields). In order to obtain a relevant industry code for each job, an automatic coding of employers is currently processed, but it is successful for only 45% of respondents. The remaining 65% are manually coded, requiring the work of around 70 INSEE agents for five months each year. Big Data techniques seem to offer great opportunities to improve this process.

b. Organisation of a Hackathon

The SSP Lab in collaboration with the IT department organised the first Hackathon of INSEE on this subject on 18 and 19 January 2018. It gathered more than 60 persons from the whole SSP (INSEE and Ministerial Statistical Services) and its partners (the Health Insurance InstituteCnam, the Central Bank, the Employment Agency-Pôle Emploi, etc.). Two days of training were organised before the Hackathon to present the subject and the approach. Different speakers presented the data involved (the Census and the business register, called SIRENE) and some techniques that could be useful during the Hackathon (web scraping, text mining, geocoding, etc.). This preparation phase was well received by the participants and the organising team received positive feedback.

c. Transforming this event into an experiment

Since the Hackathon, a small team (including Census department members, SSP Lab members, IT members and some participants in the Hackathon) has been working to the development of a prototype, implementing some ideas that emerged during the Hackathon and adding some new functionalities. This is still a work in progress.

4.2. Detecting wages/paid hours anomalies in employer payroll declaration statistical databases

Sponsor: INSEE Social Studies Directorate (Employment and Professional Income unit);

Team: SSP Lab (1 member), Statistical Methods Unit (1 member), Employment and professional income unit (2 members);

Schedule: from January to December 2018;

Expected deliverables: experimentation report, guidelines for implementation, and methodological and academic contributions.

a. A major change in the employer payroll and social contribution declaration format offering new opportunities

The Annual Declaration of Social Data ("déclaration annuelle de données sociales", DADS), mandatory fulfilled each year by each employer and to which reported individual wage-earner information is transmitted to fiscal and social services for payroll and tax purposes as well as for calculating social security wage-earners rights (e.g., pensions), has been replaced since 2016 by a monthly Nominative Social Declaration information. This change of sources completely modifies the national statistical service of information on employment and wages that relies on, but also provides the opportunity to rethink the automatic anomaly detection process implemented in the statistical production line, as the latter is deeply modified to integrate these new data. An adapted automatic detection of such anomalies would lead to productivity gains in the subsequent editing procedure.

b. Machine learning contributions to anomaly detection of wages/paid hours data

The experimental project carried out with the department of Employment and Professional income of the Social Studies Directorate tests different machine learning-based algorithms for anomaly detection of net and gross wages and related paid hours. The project has so far investigated unsupervised algorithms, such as fuzzy association rules, isolation forests and local 8 outlier factors on a small scale, with the intention to provide a probability score of an outlying position. The next steps will be to constitute a sample of observations labelled as anomalies or not to evaluate the performance of the methods tested in comparison to the current one (based on a classic regression) and to apply the methods selected on larger scales.

4.3. Mobile phone data

Sponsor: INSEE Regional Studies Directorate;
Team: SSP Lab (one permanent member and one intern);
Schedule: several experimentations since 2016;
Deliverables: institutional and academic contributions, data-processing techniques and experimental prototypes.

Mobile phone data has proven to form an exciting new data source for official statistics. The SSP Lab intends to explore the institutional, legal, technical and methodological challenges that come with the integration of mobile phone data in official statistics.

a. Data access at Orange Labs

Access to a pseudo-anonymised dataset collected by Orange for billing purposes has been made possible through an agreement between Orange Labs, Eurostat and INSEE. The dataset consists of Call Detail Records (CDR) describing information on each phone call and text message (SMS) sent or received by Orange users in the period from May to mid-October 2007.

b. First experiments

A number of experiments have been performed. The goal of the first experiment was to detect urban zones thanks to mobile phone data and application of supervised classifiers (Vanhoof, Combes, de Bellefon, 2017). The second experiment intended to measure the residential population by using the CDR during nights and advanced treatments of the data (de Bellefon, Givord, Sakarovitch, Vanhoff, 2018). The last experiment is still in progress and analyses the segregation by combining mobile data and fiscal data (Galiana, Sakarovitch, Smoreda, 2018 presented in this conference).

c. Future prospects

These experiments showed that mobile data are extremely rich, but could be unsuitable for some applications because of location imprecision or representative bias. In order to exploit the whole richness of information of these data, the following experiments will focus on the estimation of population present within a given place and time (as opposed to the residential population). Different time and geographical scales will be explored, potentially with signalling data. A new agreement for continuing the collaboration is ongoing.

5. Concluding remarks

Although the Lab is still in its early days, it is clear that the opportunities for new synergies with production units have been positively welcomed throughout the official statistics service ecosystem. The creation of the Lab is proving to be promising in terms of acquiring, maintaining and disseminating data science knowledge. What's more, it is already mobilising people within our organisation, and helping to inspire and motivate the next generation of our statisticians.

References

1. Combes, S and Bortoli, C: Apports de Google Trends pour prévoir la conjoncture : des pistes limitées, *Note de conjoncture*, March 2015
2. Combes, S., Bortoli, C. and Renault, T.:" Nowcasting payroll employment with traditional media content", New Techniques and Technologies for Statistics conference, March 2017.
3. de Bellefon, M.-P., Givord, P., Sakarovitch, B., Vanhoff, M. : "Allô, où es-tu ? Estimer la population résidente à partir de données de téléphonie mobile, une première exploration, in revision in *Economics and Statistics*, 2018
4. Galiana, L. and Sakarovitch, B. and Smoreda, Z.: "Understanding socio-spatial segregation in French cities with mobile phone data", 2018, presented at this conference.
5. Vanhoof M., Combes S., de Bellefon M.-P.: "Mining Mobile Phone to recognize Urban Areas" in A. Petrucci and R. Verde, eds., SIS 2017 Statistics and Data Science: New Challenges, New Generations. Proceedings of the Conference of the Italian Statistical Society (Florence: Firenze University Press, 2017) 1005-1012.

Annex: Ongoing projects

Exploring new data

- Student evaluation log files: more information to assess student response strategies [with the Education ministry statistical service, and IT department of INSEE]
- Mobile phone data - residential population, social segregation [with Orange Labs, ESSnet Big data] (cf 4.3) - Branch agreements: What protection for employees? A textmining approach [with master's students] - Satellite Data and city heat islands [with the INSEE geographic methods unit]
- New panel data about professional careers of both wage-earners and entrepreneurs [with INSEE social studies Directorate]

Applying new methods

- *Les champs de Sirene:* Automatic detection of employer in census [with IT department and hackathon participants] (cf 4.1)

- Detecting wage anomalies in employers' payroll declarations statistical databases [with INSEE Social studies Directorate] (cf 4.2)
- Machine learning for predicting careers and wages for microsimulation models [with INSEE Economic Studies Directorate and IPP-Paris School of Economics]
- Peer effects in Education [with the Education ministry statistical service]

Ongoing methodological reports

- Decomposition methods for inequality analysis
- Selection on observables: Propensity score in R (just released)
- Duration models in statistical studies

Other studies in collaboration

- Constructing control group for poor city districts (INSEE Regional studies)
- Differentials in peer effects in high school success (Education)
- Quantifying the effect of school avoidance on segregation (Education)
- Heterogeneity of the performance of high schools in France (INSEE Economic studies)
- Firm role in gender wage gap (INSEE Social studies)
- Wage discrimination against descendants of immigrants (INSEE Social studies)
- Evaluation of the 2014 Unemployment Insurance Agreement (INSEE Economic studies, Acoss)

Training sessions given in 2018

- Machine Learning (2d)
- Textual Analysis (1d)
- Python for the data science (in preparation)
- Decomposition methods for inequality analysis
- Evaluation of public policy

Dissemination and collaborative work practices

- Hackathons, collaborative workshops,
- Reading groups: Machine Learning and econometrics for career analysis with panel data (in preparation)

Dissemination

- Intranet, extranet SSM, Yammer, blog, Github
- Big Data newsletter, Big Data seminars

Networks

- Eurostat - Big Data task force, Essnet BigData I (and II)
- Eurostat - Grant City data (mobility and phone data)
- Academics - CREST, IPP-PSE, Dauphine

Role of ISI, and Statistical Societies in Measuring SDGs

Misha V. Belkindas[1], Alexey N. Ponomarenko[2]
[1]ODW Consulting, Rockville MD, USA
[2]Russian Statistical Association, Moscow, Russian Federation

Abstract

The data revolution and adoption of the Sustainable Development Goals (SDGs) gave a new impetus for the development of the national statistical systems. Currently the statistical institutions around the world are working on developing methodologies and conducting new surveys gearing up to better use of administrative records and private sector data for estimation of SDG indicators. This situation requires the larger efforts both from international community as well as from domestic institutions. External support may come from international organizations and other donors in the form of funds or, more importantly, technical assistance. International organizations are responsible for developing standard methodologies and they should provide assistance but this is not enough. Assistance may come from several alternative constituencies from the public sector (statistical agencies in developed countries) and the private sector (statistical consulting companies and individual consultants). As a part of statistical system, NGOs with statistical backgrounds also play their role. International NGOs, such as the International Statistical Institute (ISI), provide a network for professional communication for all interested participants using various events (conferences, seminars and webinars, etc.) as well as journals and websites for publication of related discussion papers. At the national level, the composition of statistical systems and functions of different participants in measuring SDGs may vary. It depends on each country's legislation, traditions, and other factors. In many countries, official statistical services play a key, if not an exclusive, role in SDGs measurement using support from the international agencies to solve the methodological problems and shortage of funds. In the case of the SDGs, this scheme may not work very effectively because even international agencies can't provide all necessary support due to ongoing methodological discussions. To solve the problem of measuring SDGs, members of national statistical systems need to act in the same way as their partners at the international level, that is, combine the efforts of official bodies, academia, private business and NGOs.

Keywords

data revolution; statistical capacity building; public private partnerships; NGO's; private sector

1. Introduction

From 2000-2015, the UN system was guided by the Millennium Development Goals (MDGs). The MDGs centered around eight goals, twenty one targets and 60 indicators. The goals were focused primarily on affecting change within developing countries and had ambitions such as halving extreme poverty. The indicators primarily focused on challenges faced by developing countries. The indicators came from well-established indicators produced through national statistical systems. Despite these indicators being well defined, there were still challenges around the accuracy, timeliness, and precision of the indicators. However, the data challenges were around data collected by national statistical offices, primarily using national household surveys or well-established tools.

Beginning in 2012, the UN began a process to identify a new agenda for 2015-2030 with an Open Working Group. The Open Working Group involved nearly sixty member states from all regions of the world. These groups identified a desire to keep some of the key issues embodied in the MDGs around social indicators but sought to expand the agenda to focus more on environmental indicators, peace and security, and inclusive development (Open Working Group 2012).

In parallel with the Open Working Group, the UN Secretary General formed a High-Level Panel of Eminent Persons to discuss the post-2015 agenda. This panel had senior officials from government, the private sector, think tanks and academia. They issued a call for "a data revolution for sustainable development." Such a data revolution would focus on filling gaps in data to improve monitoring of progress, the allocation of resources, and to boost accountability for governments and private sector actors. On data from the private sector, the panel wrote, "Only a few progressive, large businesses try to account for their social and environmental footprint. The Panel proposes that, in future – at latest by 2030 – all large businesses should be reporting on their environmental and social impact – or explain why if they are not doing so." (United Nations, 2013, p. 24). Such a framework for international statistics provides a significant shift from the MDG approach to data.

Following the recommendations of the High-Level Group, the UN Secretary General created the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. This group produced a report on the data revolution. The report makes several points about the role of private interests. The private sector can be one of the key sources of innovations and the governments can occupy a key role in balancing public and private interests (United Nations, 2015).

In response to the Independent Expert Advisory Group, the Global Partnership for Sustainable Development Data (GPSDD) was created. The GPSDD, currently hosted by the UN Foundation, boasts nearly 300 members, including several national statistical offices, think tanks and many companies. It is primarily a source for connecting the many stakeholders in the data revolution and encouraging them to work together. Some initiatives have included creating data roadmaps, developing inclusive data charters, and sharing innovations. Small amounts of funding have been awarded to projects involving more than one member, through financing from the World Bank.

The World Bank has committed to increase the frequency of poverty surveys in conjunction with the demands of the SDGs. They have also employed new methods and new approaches to estimating poverty in contexts where data are frequently missing.[1]

This paper is a case study based on practical experience of joint efforts of public sector and private sector institution as well as NGOs to solve the problems related to implementations of SDG estimations at the country level. Using this experience, we are going to demonstrate shortcomings and limitations which currently take place in cooperation of NGO's with other partners in statistical community.

2. Measuring the SDGs

No country in the world currently can provide data on all the SDG indicators as presented by the UNSD. Furthermore, methodology for some of the indicators are not even developed. This situation requires larger support to the national statistical systems particularly in developing countries.

The adoption of the 17 Sustainable Development Goals and the commitment to achieve them by 2030 puts enormous pressures on many countries mostly those with low level of development.

The 17 SDGs are measured by more than 240 indicators, divided in 3 tiers.

Tier 1 indicators are conceptually clear, have an internationally established methodology, all related technical standards are available, and estimated results are regularly produced by at least 50 percent of countries in every region where the indicator is relevant.

Tier 2 indicators are conceptually clear, have an internationally established methodology, all related standards are available, but estimated results are not regularly produced by countries.

Tier 3 indicators have no internationally accepted methodology or/and standards, but methodology/standards are being (or will be) developed and tested.

As of 13 February 2019: tier 1 contains 101 indicators, tier 2 contains 84 indicators and tier 3 contains 41 indicators. In addition to these, there are 6

[1] See for instance, World Bank https://www.worldbank.org/en/news/immersive-story/2017/07/26/the-tech-revolution-thats-changing-how-we-measure-poverty

indicators that have multiple tiers (different components of the indicator are classified into different tiers).

With the adoption of the SDGs comes the problem of their measurement. The methodology for SDG indicators is developed by Custodian agencies most of them are UN agencies or multilateral organization. These agencies have huge experience in such work and all necessary capacities. However, there is so much methodological and organizational work to be done that international agencies need the support of the entire statistical community. Academia and private institutions like Gallup Institute could provide additional ideas about possible methodological and technical solutions based on their experience and statistical NGOs may want to create a network for discussion for all interested parties.

ISI is a great example of such network. At the international level, ISI plays the unique role as the strongest and most reputable organization. Its membership has data scientists and statisticians who should work together to help combine the data coming from surveys, censuses, administrative sources and the "big data" into common data bases. These data bases should serve as a background to estimate the SDG indicators. The ISI through its Statistical Capacity Building committee design and roll out training courses and get involved in provision of methodological assistance.

But new methodologies are not the only problem.

What is more important is that the development of SDG indicators in many cases require new surveys or changes in current ones. In some cases, the questionnaires must be adjusted to get the disaggregation needed for the indicator, in some cases the sub-national data are not representative enough to allow for sub-regional estimates of SDG indicators. The countries are struggling to collect data to compile the SDG indicators as they need to introduce new surveys or adjust the ones which are in operation. These and other issues not only require a lot of efforts from country statisticians but also from those who are able to support capacity building in the statistical systems in developing countries.

Beside international agencies, there are two groups of organizations who support developing countries by providing technical assistance: statistical agencies of statistically developed countries (public sector) and consulting companies (private sector). The massive amount of interventions needs many actors to work together to help. Therefore partnerships between private and public sector entities in this endeavour are a key. They compete for projects in an open market. Often, they create partnerships in the forms of joint ventures, consortiums, twinning arrangement to compete. NGOs could provide support, participating in these projects or recommend to their members (including corporate members) to participate. For example, ISI, as a non-governmental, not-for-profit organization, can't compete in the market. If they have funds, provided by donors and they rely on volunteer labor force they cannot compete in the market as they would undermine the open competition. Therefore, it is imperative for the NGOs to find a way to join the partnership to be able to contribute to the capacity development in the field of statistics.

This paper is a case study based on practical experience of joint efforts of public sector and private sector institution as well as NGOs to solve the problems related to implementations of SDG estimation at the country level. Using this experience, we are going to demonstrate shortcomings and limitations which currently take place in cooperation of NGOs with other partners in statistical community.

3. Case Study

In 2018, the Bureau of Economic Analysis (BEA) Foundation, which is a non-profit entity which function is coordination of projects financed jointly by World Bank loans to Russia and Russian government funds for the Russian Statistical Service (Rosstat) has announced a project to support Rosstat in the compilation of a wide range of indicators for monitoring SDG.

The project calls:

- Propose criteria for compilation of 17 SDG indicators for further developing methodology for their calculation in order to enable a comprehensive monitoring of the SDGs in the Russian Federation and develop methodological recommendations for improving the empirical base that ensures regular calculation of these indicators. Compile these indicators using the proposed criteria.
- Review and systematize methods for the calculation of SDG indicators, included in the list presented in the methodological materials of the custodian agencies for SDG indicators by the spheres of their responsibility, and in the methodologies used by UN member states.
- Carry out test calculations of the SDG indicators for Russia as a whole and for the subjects of the Russian Federation using the available empirical base.
- Compare the calculation results for the SDG indicators with statistics from other countries and with data for Russia presented in the UN information sources. Make conclusions about the methodological reasons for possible differences between the calculated indicators and the results from other countries and those presented in UN information sources.
- Indicate additional variables needed to improve the coverage of SDG indicators and formulate recommendations for their consideration

within new or existing federal statistical observations, including suggestions on introducing new questions in the statistical surveys.

- Justify the optimal placing of the proposed above questions into the specific programs and statistical surveys taking into account their subject matter, frequency, and sample sizes. Estimate approximate budget for the calculation of SDG indicators in Russia.
- Develop guidelines for Rosstat to generate final statistics on SDG indicators, taking into account the need to supplement the empirical base, for submitting to the international organizations.
- Evaluate the possibilities for using alternative data sources (Big Data) for the calculation of SDG indicators.

US-based private consulting company ODW Consulting participated in competition. It has employed an international NGO Open Data Watch which works on different aspects of SDGs from calculation of indicators to estimating of costs needed to measure SDGs. In addition to work on the international methodology the project includes estimates using original Russian data as well as recommendations regarding improving of Russian official statistics. ODW Consulting employed several local experts to assist international experts, to collect initial data, and to make estimations. In this case ODW Consulting partnered with the Russian Association of Statisticians (RASt) to provide local experts with necessary background.

Currently this project is underway. In a recently conducted unpublished study, ODW Consulting looked at a sample of SDG indicators to understand the clarity of the write-ups by the custodian agencies placed on their websites. The study reveals that some indicators are not very clear, some write-ups are more than ten years old. Others have too many options for countries to choose from. It is known that no country in the world can deliver even on tier 1 indicators. We would expect that the champions can present around 90 indicators on their SDG web portals.

This concrete project may serve as a success story of cooperation between government, national statistical services, international agencies, private business and NGOs. This is because of clear distribution of functions between all these institutions. In many other cases, progress is not so visible.

The situation calls for strengthening the statistical systems worldwide to be able to deliver on indicators. The analysis of country and UN data portals demonstrate that the situation is not very promising. The call to support capacity development is not answered by those who finance statistical work – country governments, multilateral, bilateral, and private donors. Many estimates indicate significant gaps. Private and public sector entities should provide support to the least developed national statistical system. Ways of working together must be explored and agreements be reached rather sooner than later.

NGOs look like a weak point in this chain. NGOs have a high potential because in many countries their membership is very diverse. Their members are statisticians and data scientists working in statistical agencies, academia, private sector. In many countries, statistical societies are providing technical assistance and long- and short-term training. However, most of NGOs do not provide statistical training and do not conduct research themselves but work as a platform for communication for statisticians serving in both official and independent statistics and such communication is an important part of statistical system. This activity helps to keep dialogue open between the national statistical agencies, academia, universities and users of statistical data. Thus, NGOs should participate more actively in statistical capacity building process.

ISI is a flagship body of statistical community in the world. Other NGOs closely monitor its activity and try to use its experience at the national level. ISI should create the effective model for participation in statistical capacity building to provide example for the national NGOs.

4. Discussion and Conclusion

Some examples of possible interventions and organization of the technical assistance process within ISI should be considered.

International Statistical Institute is in a unique position to help countries to improve statistical systems due to its diverse membership. No other organization in the world has academics, experts in official statistics, mathematical statisticians, data scientists, economists and other specialists who together can help countries to improve their statistical systems to enable them to deliver on the SDG measurement. Through the Statistical Capacity Building Committee (SCB) the ISI is planning to support statistics in several areas.

The partial list follows:

- Provide technical assistance in official statistics by creating a cadre of high-level statisticians in different areas of statistics interested to donate some of their time to deliver TA in the field.
- Organize workshops, seminars on important topics for official statisticians, such as management, data revolution, measurement of SDGs, etc.
- Support training institutions in developing countries by providing curricula, liaising with professors and creating possibilities for professors from abroad to work with colleagues in developing countries to deliver lectures, master classes, etc.

- Support statisticians from developing countries (mostly young statisticians) to participate in conferences.

For these to happen ISI should first to obtain financing from donors to be able to conduct those activities. A serious fundraising campaign must be launched by ISI SCB. The candidates to be approached, are World Bank, IMF, African Development Bank and other multilaterals, UK, USA, Canada, Sweden and other bilaterals, Gates, Hewlett and other private foundations.

They should internally conduct a survey of members who are willing to donate their time for support the statisticians in developing countries. ISI should find the modus operandi how to work with partners in the field.

References

1. http://www.beafnd.org/en/projects/detail.php?id=475
2. https://web.archive.org/web/20180628015709/https://sustainabledevelopment.un.org/content/documents/1494sgreportsdgs.pdf
3. http://www.rusasstat.ru
4. https://www.un.org/sg/sites/www.un.org.sg/files/files/HLP_P2015_Report.pdf
5. http://www.undatarevolution.org/report/

Common approaches to capacity development in response to Agenda 2030

Hristina Aleksandrova[1]; Hans Reckhaus[2]
[1] GOPA Consultants, Frankfurt, Germany
[2] GOPA Consultants, Bad Nauheim, Germany

Abstract

The paper focuses on how the Agenda 2030 affects capacity development (CD) within national statistical systems (NSSs) and how private consultancies can contribute to this process. Achieving the 2030 Sustainable Development Agenda and new metrics caused by globalisation require robust and reliable data at national level. Today, national statistical authorities are expected to gradually increase the volume and quality of their data and the availability and comparability of statistics (OECD 2019). Developing new methods and extending capacity helps to take advantage of new technologies and new data sources, increase accuracy, reduce costs and cover emerging topics. The use of administrative data sources is important for the production of official statistics, in particular for the comparability of statistics required for policy use (ibid.). Common processes and quality standards are becoming increasingly important for the production of Sustainable Development Goals (SDGs), and the creation of interoperability between all statistical producers in a country is necessary to unify data exchange. The ever-growing availability of data and its strategic relevance for official statistics require common approaches to capacity development within an organization, implemented through private-public partnerships (PPPs[1]). Tailor-made long-term support is important, not only to improve several technical skills at the same time, but also to develop soft skills and deal with institutional constraints, where appropriate. This process can be sensitive as it is based on the willingness of the actors involved to take responsibility. We must also recognise the strengths and weaknesses of different partners who can assist and manage these developments, and adopt delivery models, which maximise efficiency, sustainability and impact.

Keywords

SDGs; statistical capacity development; private-public partnerships; national statistical systems

[1] A **private-public partnership** (**PPP**) is a cooperative arrangement between two or more public and private entities, typically of a long-term nature

1. Introduction

In light of the data requirements for the 2030 Agenda for Sustainable Development, with its 17 SDGs, 169 targets and over 230 global indicators, the national statistical systems are facing an unprecedented challenge in producing data that meets international standards and regulations. As a result, investing in the sustainable development of statistical capacity has been identified as one of the five cross-cutting thematic areas of the Transformative Agenda for Official Statistics (UNSD 2015). Furthermore, the United Nations Statistics Division (UNSD) recognised that besides investing in broadening technical skills through reorganisation and re-engineering production processes and adapting to a rapidly evolving technological environment, improvements have to be made with regard to change and project management (ibid.).

Thus, the challenge does not only relate to data production itself, but also to leadership and management skills. The path towards the SDGs set the objective to go beyond the traditional production-side approach. In the past, it has been criticised that statistics related interventions concentrate more on inputs and activities than on the utilisation of capacities.

Nowadays, capacity development has to be seen as more than only a one-off event, but as an intervention that focuses on the needs and priorities of the organisation/system as a whole. Formulating strategic aims and visions across the entire organisation becomes a prerequisite for successful development. Proper monitoring and reporting are indispensable parts of the process, where the focus should be on "celebrating progress and outcomes" rather than on outputs. This calls for a strategic and long-term thinking to be applied to the design of both the overall approach and the specific selection of methods.

The aim of this paper is to outline current trends in statistical capacity development and to assess where the concept of private-public partnerships can generate added value in order to achieve sustainable results in the overall intervention. To this end, the following chapter describes the methodology used in our research. Subsequently, a case study is presented, which gives a practical example of how the theoretical framework of capacity development can be put into practice. Afterwards, the results are summarised with a special focus on the added value of private-public partnerships. These findings are then presented in the last chapter and suggestions for further discussions are provided.

2. Methodology

In recent years, major international development organizations, governmental platforms and nongovernmental institutions (NGOs) have developed a variety of approaches and instruments to assess and promote the capacity within statistical systems. Capacity development has received additional recognition, particularly through the adaptation of the Agenda 2030 and its holistic approach to development[2].

A "whole system approach" is offered by the Capacity Development 4.0 (CD4.0) framework, in which statistical capacity is defined as "the ability of a **country's national statistical system**, its **organisations** and **individuals** to collect, produce, analyse and disseminate high quality and reliable statistics and data to meet users' needs" (PARIS21, 2018).

The CD4.0 framework presented by PARIS21 identifies three levels of intervention, as presented in Figure 1:

Figure 1: Different levels of intervention

Source: Adopted from PARIS21, 2019

The figure describes the different levels of intervention that can be addressed in order to develop capacity. It is important to remember that all different levels are interconnected and interdependent.

The individual level can be tackled through targeted training in order to improve **attitude, knowledge and skills** of staff members.

The organisational level is determined by the code of practices, resources, management skills, business and quality processes. These variables can be assessed through **in-depth analysis of business processes** and improved by targeted technical support. Formulating strategic aims and visions across the entire organisation becomes a prerequisite for successful development.

Finally, the system level is characterised by the national regulatory environment, **user demand** and **statistical laws**.

When it comes to capacity development, strategic and long-term thinking needs to be applied to the design of both the overall approach and the specific selection of methods. This is why we are supporting approaches such as PARIS 21 with its focus on sustainable growth.

A common approach to capacity development should not only address the question of "**How**" capacity is developed, but efforts should also be made

[2] See for instance: GIZ *Capacity WORKS* (2017); The World Bank *The Capacity Development Results Framework* (2009); World Vision *Organisational Capacity Development Manuals* (2008)

to identify best practices in response to “**Who**” implements the planned interventions.

The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action have recognised that “Aid is about building partnerships for development. Such partnerships are most effective when they fully harness the energy, skills and experience of all development actors” (s. OECD, 2005/2008 P. 17). The form in which these various types of skills are mobilized is a fundamental consideration for project design. Moreover, the challenge of meeting today’s increasing demands on NSSs, especially on data to monitor SDGs, in a resource environment that is decreasing in proportion to these demands, means that more and more in-depth consideration must be given to how to optimise efficiency, sustainability and impact. Traditional donor approaches, such as mobilising individual technical experts for smaller projects or ‘twinning’ arrangements between developing and developed NSOs, may not be optimal for maximising sustainable impact and efficiency, and may be more costly than necessary. In this context, private-public partnerships may provide the necessary capacities to successfully steer the development of NSOs in the long-run.

Private technical service providers specialised in official statistics organisations are the most common to efficiently run large and small-scale CD projects. They bring expertise in combining successful project management with a contextual understanding of the needs of official statistics. Their most important advantages are that they are able to:

- understand the practicalities of project implementation
- identify and mobilise the right subject matter experts
- improve cost-effectiveness in implementation processes
- have robust financial management systems
- coordinate partners and stakeholders across sectors, and
- form effective relationships between public and private actors where necessary (e.g. in gaining access to ‘Big Data’)

Case study

In the following we demonstrate how the theoretical approach to capacity development can be put into practice using the example of a project implemented by GOPA Consultants consortium together with the Lao Statistics Bureau (LAOSTAT). The specific objective of this project was to improve LAOSTAT’s capacity to produce and disseminate reliable data in accordance with international standards and in response to user needs.

Table 1 below illustrates how the project set-up has supported this objective by targeting all three dimensions of CD through an overall capacity plan:

Table 1: "Whole system approach" for LAOSTAT

Target/Level	Individual	Organisation	System
Resources	Recruitment engagement → process/employee HR strategy	IT system → ICT Plan, technical specifications and support to procurement process	Review of the Law and of the strategy for development of NSS (SDNSS)→recommendations
Skills and knowledge	Training plan/Capacity building plan Capacity building activities	LOSQAF + QMS Generic tools and methods Description of processes → Environmental statistics conceptual roadmap	Promotion of stats as management tool of public affairs → stats literacy Technical assistance with other ministries
Management	Behavioural framework/ performance management/ Middle management training/ Leadership training/ Project anagement training → HR strategy	Set of strategies and master plans Support to fundraising strategy ICP → change management/Middle management training/Leadership training/ Project management training	Support to the implementation of the Law and SDNSS → Implementation plan for the development of the NSS Support to SDGs data coordination role
Politics and power	Behavioural framework → HR strategy	Transparency Dissemination strategy, metadata strategy, website, LOSQAF	User satisfaction survey Support to stakeholders meeting WG
Incentives	Talent management/ Non-financial incentives (training and education opportunities, study visits, etc.) → HR strategy	ICP → vision, core values HR strategy → recommendation on compensation and benefits Dissemination strategy → reputation and trust	Promotion of stats as anagement tool of public affairs → Seminar for policy makers Support to the promotion of the law

3. Result

The private-public partnership model proves to be a good solution to support the global statistical system in meeting the challenges of Agenda 2030. Technical assistance can successfully be supported through PPPs to achieve sustainable results. The Laos experience has provided us with the following findings on how support for capacity development can or should be structured:

- Let the NSO take the ownership of their capacity development initiative.
- Focus on the needs and priorities of the organisation as a whole.
- Engage stakeholders in the capacity development and change management processes.
- Cultivate adequate political support and preserve autonomy.
- Establish an environment that is conducive to learning and change from the outset.

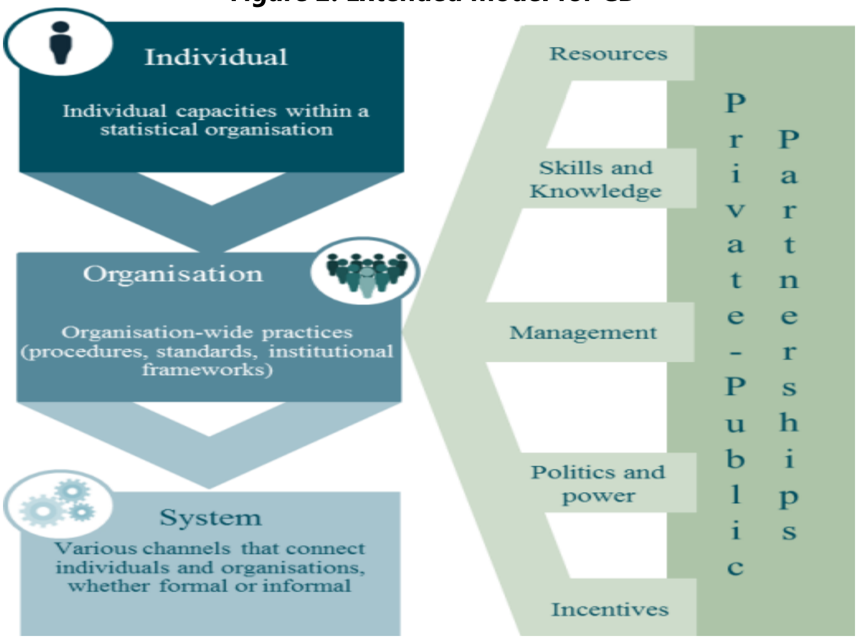
- Create and share a vision for change and communicate strategic aims and visions across the organisation.
- Achieve and disseminate quick wins
- Repeat!

In this regard, the added-value of involving private expertise providers is determined by the following advantages:

- Cost reduction with simultaneous provision of high-quality outputs
- Develop and manage cross-sectoral partnerships
- Innovation
- Invest where it makes economic sense
- Manage projects effectively

A final result of this paper is the extension of the initial CD approach through targeted assistance measures by PPPs, which is described in the following figure 2:

Figure 2: Extended model for CD



4. Discussion and Conclusion

The Agenda 2030 sets new objectives for official statistics, such as the promotion of a collaborative way for the transformation and modernisation of global, regional and national statistical systems. In addition, the data requirements for the 2030 Agenda are a major challenge both for national statistical systems and the international statistical community.

The paper has shown that PPPs can play an important role in the successful implementation of these changes, since they provide the stakeholders with an extended access to expertise, tools and measures. This holds especially true for activities that are related to enhancing the capacity of the NSSs, as shown in the example of the Laos case study.

CD projects should be based on the needs of specific NSSs. The models presented show that CD projects can become quite complex, as their work often has to cut across institutional units and/or institutions. Moreover, sustainability and efficiency often require that reforms are implemented on a large scale (over long periods of time).

For this reason, capacity building measures should ideally be engineered in two ways: covering the programming of the interventions, but also establishing successful models for implementation. Development partners, beneficiaries, and those implementing projects need faster and more transparent processes. However, the aspects that affect how knowledge and resources are used in practice are rarely addressed. A new and effective way of delivering capacity development "must explicitly acknowledge the real political economy challenges on the ground and aim to work within these constraints to improving data, and/or aim to alter the current incentives for producing and using good official statistics." (s. Krätke and Byiers, 2014: 1). The objectives of this approach go beyond the traditional production-side interventions and also include the strengthening of data use, literacy and results.

The paper at hand presents a theoretical approach to how CD can be promoted to prepare NSSs for the challenges that result from international agendas, such as the SDGs. Furthermore, using the example of the Laos project, a practical example of how this theoretical framework can be put into practice was given. From our perspective, the discussion on effective and efficient capacity development need to be enhanced at national and international level; with a preferred focus on the whole systems approach.

References

1. Gesellschaft für International Zusammenarbeit (GIZ) (2017): Cooperation Management for Practitioners. Managing Social Change with Capacity WORKS
2. Krätke, F., & Byiers, B. (2014). Implications for the data revolution in Sub-Saharan Africa. Link: http://ecdpm.org/wp-content/uploads/DP-170-Political-Economy-Official-Statistics-Africa-December-2014.pdf
3. OECD (2019): Recommendation of the OECD Council on Good Statistical Practice. Link: https://legalinstruments.oecd.org/public/doc/331/331.en.pdf

4. OECD (2005/2008): The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action
5. PARIS21 (2019): Why Capacity Development 4.0? Link: https://paris21.org/sites/default/files/inline-files/cd40-framework final.pdf
6. PARIS21 (2018): Proposing a Framework for Capacity Development 4.0, PARIS21, Paris
7. The World Bank (2009): The Capacity Development Results Framework. A strategic and results-oriented approach to learning for capacity development
8. UNSD (2015): Transformative Agenda for Official Statistics. A collaborative initiative for the modernisation and integration of statistics. Link:https://unstats.un.org/unsd/nationalaccount/workshops/2016/Ankara/Collaborative_Initiative_Moder nization EN.pdf
9. World Vision (2008): Organisational Capacity Development Manuals

The costs of gender-based violence against women in Spain

Rivera-Galicia, L.F.[1,3,4], Gallo-Rivera, M.T. [2,3,4]; Mañas-Alcón, E. [2,3,4], Montes-Pineda, O. [2,3,4]
[1]Department of Economics. University of Alcalá. Spain.
[2]Department of Economics and Business. University of Alcalá. Spain.
[3]Institute for Economic and Social Analysis (IAES). University of Alcalá. Spain
[4]Chair of Corporate Social Responsibility. University of Alcalá. Spain

Abstract

Violence against women (VAW) is one of the most severe forms of gender inequality. In 2013, the World Health Organization (WHO) estimated that one out of three women in the world has experienced physical and/or sexual violence throughout life. VAW causes thousands of deaths every year, and millions of victims who experience suffering and pain. Moreover, it has also significant economic costs in terms of expenditures on service provision and healthcare, lost income for women and their families, work absenteeism of victims and aggressors, police and legal costs, etc. This paper aims to estimate the global costs of VAW caused by intimate partner violence (IPV) in Spain during 2016. First, it starts describing the methodological aspects needed for the estimation of costs. Second, we depict the data on victims from the 2015 Spanish Macro survey of gender-based violence. Third, we classify the diverse impacts of IPV against women in three cost-assessment categories: employment, healthcare and legal sector. Finally, we estimate the tangible costs for each one of these categories, as well as for other concepts such as accommodation needs or other activities of third sector organizations. The main findings of this paper are that 2.7% of women aged 16 or more have suffered from physical and/or sexual violence from their intimate partners in Spain. However, this percentage increases up to 13.2% if we consider other types of intimate partner violence. Using the accounting model and different approaches for every category, we estimate the total tangible cost of VAW in Spain in 2016 between 1,281.0 and 8,540.9 million €, which represent between 0.11% and 0.76% of Spanish GDP.

Keywords

Gender-based violence; Intimate partner violence; Economic cost; Spain

1. Introduction

In 2013, the World Health Organization (WHO) estimated that one out of three women in the world has experienced physical and/or sexual violence throughout life. VAW causes thousands of deaths every year, and millions of victims who experience suffering and pain. Moreover, it has also significant economic costs in terms of expenditures on service provision and healthcare, lost income for women and their families, work absenteeism of victims and aggressors, police and legal costs, etc.

Calculating the cost of gender violence helps to know the economic resources lost by the agents affected by the violence: not only the victims, the people close to them and the aggressors, but also the companies of the private sector, the public administration, civil society, and future generations.

This type of studies has a double goal: to foresee the economic costs that represent not to reduce gender violence; and estimate the potential gains that would result from a significant reduction in the levels of gender violence. In addition, knowing the economic cost of gender violence helps to reduce the existing social acceptance of this problem and to improve the design of public policies to get rid of it.

In this paper, we estimate the total tangible costs of Intimate Partner Violence (IPV) against women in 2016 in Spain. In section 2, we present the methodology. In sections 3 to 5 we explain the estimates made in the following categories: employment, health and justice. Section 6 refers to other costs not included in the previous categories. Finally, section 7 shows the aggregate costs.

2. Methodology

The concept of gender-based violence in Spain is defined in the Integrated Law against Gender Violence (GREVIO, 2019, pp. 5-89). It refers to violence against women committed by men who are their partners or who have been a former partner. The reference year for the estimated costs is 2016.

We have used several data sources to estimate the prevalence and type of violence suffered by women. The Spanish Survey on Violence against Women 2015 has been very useful in determining the prevalence and types of violence, but we have also used administrative records for health and legal aspects.

We have estimated the tangible direct costs, referred to the monetary value of the goods and services consumed in the prevention and treatment of gender violence; and the tangible indirect costs, referred to the resources that are lost because of the reduction of the productive activity or the loss of income due to greater female inactivity or unemployment caused by gender violence.

The general methodology we have used is the Accounting Model, since it is one of the most common methodological options (Day, McKenna and Bowlus, 2005). It is well adapted to making estimates referred to a wide range of costs and affected agents, which are then added to provide the total cost. Component items can be costed using a ‘bottom-up’ or a ‘to-down’ approach (Chan and Cho, 2010). The analysis has been divided into categories and in some, more than one estimate has been made, due to the use of different methodological approaches and/or the hypotheses that support them. We offer, when possible, a conservative estimate that can be taken as the lower cost limit, which is accompanied by other estimates with less restrictive assumptions. This results in a range of values for the estimation of the cost of each category and, through the final aggregation, a range of values for the estimation of the total cost.

3. Employment costs

Gender-based violence generates costs on employment not only for the victims, but also for the aggressors, family and friends, companies, the public and the third sector. These costs are derived from absences, delays and distraction at work, or social and unemployment benefits perceived by the victims, or programs to improve employment levels, orientation or social reintegration. They affect employed women, but also those who perform unpaid domestic activities or care for children and relatives. Although a monetary transaction is not derived from their work, such activities or services may be diminished if the physical or mental health of the victim is affected, causing a reduction in the welfare of the household or the need to incur costs to contract them externally.

We have considered two methodological approaches for estimating the cost of this category. Both use the accounting method and determine the unit cost with a bottom-up approach.

Estimation 1 follows Zhang et al. (2012). Three estimates are made based on different groups: victims of physical and/or sexual violence with injuries or illness that limit their daily activity (option A), victims of any type of violence with injuries or illness that limit their daily activity (option B) and victims of physical and/or sexual violence, considering they have suffered moderate or severe violence (option C). We obtain (see Table 1) costs of € 354,750,623 (option A), € 710,882,100 (option B), and € 796,623,357 (option C).

Estimation 2 follows Dubourg et al. (2005). The unit costs are different for physical or sexual violence and they depend on the severity. In this case, the economic cost is € 3,639,749,999 (see Table 1).

Table 1: Estimates of Employment costs.

Estimation 1 (Zhang et al. 2012)			**Estimation 2 Dubourg et al. (2005)**
OPTION A Victims of physical and/or sexual violence with injuries or illness	**OPTION B** Victims of any type of violence with injuries or illness	**OPTION C** Victims of any type of physical and/or sexual violence (moderate or severe)	Victims of physical and/or sexual violence (moderate or severe)
€ 354,750,623	**€ 710,882,100**	**€ 796,623,357**	**€ 3,639,749,999**

Source: Adapted from Mañas-Alcón (2019).

4. Health costs

Gender-based violence is considered a public health problem because of its effects on women's physical and mental health (in the short, medium and long term). It has an important impact on health systems, due to the increase in the use of health services.

However, the quantification of the use of health services by women victims of gender violence is not complete. On the one hand, because the victims often do not communicate that they are suffering gender violence. On the other hand, because of the lack of resources and/or adequate training in the health system professionals for their effective detection and registration.

To estimate the costs of gender violence in the health sector, the methodology used is based on two major approaches to the accounting method, using a bottom-up approach. Only the tangible direct costs for the public sector are considered:

On the one hand, estimation 1 follows Zhang et al. (2012). It has been applied to two alternative options (A and B): In Option A, the most conservative, the total cost of gender violence in this category is estimated by multiplying the number of victims of physical and/or sexual violence in the last 12 months, who consult the medical or psychological services, by the estimated number of times they demand the service and its assigned unit cost. In this case, the costs of health care services are € 254,123,003, and represent the sum of the costs of healthcare itself (87% of the total cost), pharmacy costs (3.1%), and complementary services (psychosocial services, training activities and prevention policies, 9.9%).

In option B, the cost of gender violence is estimated as the product of the number of women victims of any type of gender violence in the last 12 months who consult the medical or psychological services, by the estimated number of times they demand the service and its assigned unit cost. According to this estimate, the costs of health care services are € 517,839,880 (91.8% healthcare, 3.3% pharmacy and 4.9% complementary services).

On the other hand, estimation 2 follows the methodology proposed by Dubourg et al. (2005). According to this estimate, the cost of health care services is € 2,483,646,332.

Table 2: Estimates of Health costs.

Estimation 1 (Zhang et al. 2012)		**Estimation 2 Dubourg et al. (2005)**
OPTION A Victims of physical and/or sexual violence who contacted healthcare services	**OPTION B** Victims of any type of violence who contacted healthcare services	Victims of physical and/or sexual violence (moderate or severe)
€ 254,123,003	**€ 517,839,880**	**€ 2,483,646,332**

Source: Adapted from Mañas-Alcón (2019).

5. Justice costs

This Category studies the cost of gender violence in Spain from a legal perspective, estimating the different types of tangible costs of such violence for the public sector. The legal dimension of gender violence is very relevant because the beginning of the legal process is usually the starting point to end a violent relationship and to escape from the aggressor.

However, it should be noted that women who enter the justice system represent a very small group compared to the prevalence of gender violence from the Survey on violence against women 2015. Institutional prevalence is well below true prevalence, then the cost is underestimated.

Unlike other categories, we use the institutional prevalence, that is, the one that refers to the number of victims who have accessed the goods and services provided by the Spanish institutions of justice in 2016.

The costs of legal, police, judicial and penitentiary assistance have been calculated. The costs of each category are obtained from the product of the institutional prevalence and the cost unit by type of service. These products are then added to get the total cost.

Two methodologies have been considered to estimate the legal costs, both using the accounting method. The two estimates (Table 3) present a similar disaggregation into concepts (information, prior advice and legal assistance, police, justice and penitentiary). In the first, based on Zhang et al. (2012), the unit costs are determined from a proportional top-down approach, which takes public budgets to determine the proportion that may be attributable to cases of gender violence. Two options are considered, A and B, which differ in the way they compute police cost.

Option A is based on the proportion of police officers. It is the most conservative and provides the lower cost limit of this itinerary at € 502,244,973. Option B is based on the number of reported incidents. The cost of the legal itinerary is 780,165,089 € in this case.

The second estimate uses the bottom-up approach. It is based on the unit costs (Dubourg et al., 2005) that are applied to the prevalence-year that is obtained for Spain from the Macro Survey 2015. According to this estimation, the costs of the legal itinerary are € 2,247,601,161.

Table 3: Estimates of Justice costs.

Estimation 1 (Zhang et al. 2012)		**Estimation 2 Dubourg et al. (2005)**
OPTION A Proportion of police officers in Gender Violence	**OPTION B** Proportional costs based on reported incidents	Murders Victims of physical and/or sexual violence
€ 502,244,973	**€ 780,165,089**	**€ 2,247,601,161**

Source: Adapted from Mañas-Alcón (2019).

6. Other costs

In this category we have considered other tangible costs not included in the previous categories. They are related to the situation of vulnerability of the victim, the housing needs, and the activities of organizations of the Third Sector.

The effects of gender violence on the situation of economic vulnerability of the victims and their families are important. However, due to the lack of information, it has not been possible to quantify them in this paper.

To estimate the costs of housing needs and the activities of Third Sector organizations, we have used the accounting method, combining the bottom-up approach and the prevalence for 2016.

The estimated cost of housing needs for victims of gender violence is € 167,171,335. It is the result of the sum of the costs of travel and moving expenses (€ 15,320,960) and the costs of shelter (€ 151,850,375). The first type of cost is paid by the victim and the second by the public sector.

The organizations and associations of the Third Sector focus their activity mainly on two areas: offering services and orientation to the victims; and improve awareness of this problem. Based on the Annual Reports of the activity of these organizations and information from their web pages and social networks, we analysed the activities and services provided to victims of gender violence. Using the estimated unit cost per user and considering the number of victims of gender violence who contacted some NGO/women's organization from the Survey on Violence against Women 2015, we estimated the economic cost of the services provided by these organizations in € 2,722,593.

The total cost of this category has been estimated in € 169,893,928.

7. Aggregate costs and comparison with other countries

The global estimation of the tangible economic costs of gender violence in Spain are in a wide range of values oscillating from € 1,281,012,528 to € 8,540,891,420 (see Table 4). The most conservative estimate is 0.11% of GDP and € 27.6 per capita per year, while the least conservative estimate is 0.76% of GDP and € 183.9 per capita per year.

Table 4: Aggregate costs in Spain. 2016

Categories	Estimation 1 (Zhang et al. 2012)			Estimation 2 Dubourg et al. (2005)
	Option A	Option B	Option C	
Employment costs	€ 354,750,624	€ 710,882,100	€ 796,623,357	€ 3,639,749,999
Health costs	€ 254,123,003	€ 517,839,880	€ 517,839,880	€ 2,483,646,332
Justice costs	€ 502,244,973	€ 780,165,089	€ 780,165,089	€ 2,247,601,161
Other costs	€ 169,893,928	€ 169,893,928	€ 169,893,928	€ 169,893,928
Aggregate costs	€1,281,012,528	€ 2,178,780,997	€ 2,264,522,254	€ 8,540,891,420
GDP and Population in Spain				
GDP in Spain	1.118.522.000.000€			
Population in Spain	46.440.099 inhabitants			
Tangible costs (as % of GDP)	0,11%	0,19%	0,20%	0,76%
Tangible costs (per person, in €)	€ 27,6	€ 46,9	€ 48,8	€ 183,9

Source: Adapted from Mañas-Alcón (2019).

The results of Table 4 are in line with the results of other studies, presented in Table 5. While Spain shows that the cost of the gender violence is between 0.11 and 0.76% of GDP, this percentage is 0.097% in Canada, 0.110 in France, 0.573 in UK and 0.604 in Australia.

Table 5: Comparison of the estimates of tangible costs for other countries.

	United Kingdom (2012) Walby and Olive (2014)	Canada (2009) Zhang et al. (2012)	Australia (2002- 03) Access Economics (2004)	France (2005-06) Nectoux et al. (2011)
Tangible costs (*)	6.455.502.531	1.522.986.450	4.557.000.000	1.937.015.203
GDP (current prices *)	1.126.463.000.000	1.567.365.000.000	753.943.000.000	1.765.905.000.000
Population	59.119.673	33.628.571	19.651.400	63.179.351
Tangible costs (as % of GDP)	0.573%	0.097%	0.604%	0.110%

Source: Adapted from Mañas-Alcón (2019). * National currency.

References

1. Chan, K. L., Cho. E. Y-N. (2010), A review of cost measures for the economic impact of domestic violence. *Trauma, Violence and Abuse*, 11(3): 129–143.
2. Day, T., McKenna, K., Bowlus, A. (2005): The Economic Costs of Violence Against Women: An Evaluation of the Literature, United Nations. Retrieved from: http://www.un.org/womenwatch/daw/vaw/expert%20brief%20costs.pdf.
3. Dubourg, R.; Hamed, J.; Thorns, J. (2005): *The economic and social costs of crime against individuals and households 2003/04*. Home Office Online Report 30/05.
4. GREVIO (2019): *Spanish Government Report on the Implementation of the Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention).* Group of Experts on Action against Violence against Women and Domestic Violence. Council of Europe. Retrieved from: https://rm.coe.int/state-report-from-spain/16809313e0.
5. Government Delegation for Gender-Based Violence (2015): *Survey on Violence Against Women 2015. Main Results.* Government Delegation for Gender-Based Violence. Ministry of Health, Social Services and Equality. Madrid, Spain. Retrieved from: http://www.violenciagenero.igualdad.mpr.gob.es/violenciaEnCifras/macroencuesta2015/pdf/S PAIN_SURVEY_ON_VAW_2015.pdf.
6. Mañas-Alcón, E.; Rivera-Galicia, L.F.; Gallo-Rivera, M.T.; Montes-Pineda, O.; Figueroa, C. & Castellano, M. (2019): *El impacto de la violencia de género en España: una valoración de suscostes en 2016.* Government Delegation for Gender-Based Violence. Ministry of the Presidency, Relations with Parliament and Equality. Madrid, Spain. Retrieved from: http://www.violenciagenero.igualdad.mpr.gob.es/violenciaEnCifras/estudios/investigaciones/ 2019/pdfs/El_impacto_de_la_VG_ES.pdf.
7. WHO (2013): Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva. World Health Organization. Retrieved from: https://www.who.int/iris/bitstream/10665/85239/1/9789241564625_eng.pdf?ua=1.
8. Zhang, T.; Hoddenbagh J.; McDonald, S.; Scrim K. (2012): *An Estimation of the Economic Impact of Spousal Violence in Canada, 2009.* Canada, Department of Justice.

The cost of victimization: Subjective and objective dimensions

Claudia Villante, Maria Giuseppina Muratore, Isabella Corazziari, Alessandra Capobianchi
ISTAT – National Institute of Statistics

Abstract

Public perception of crime is sometimes opposite to the official statistics data provided. Also citizens having not experienced crime victimization by their own, can be alerted and suffer fear of crime due to indirect victimization. Crime victimization consequences are not only economic: changes in ordinary life of victimized households and individuals, due to the direct and indirect crime experience, can be a heavy cost to be faced. Starting from a cross country analysis of victimization and literature, the paper, based on evidence of Citizen's Safety Survey, performed by Istat on a periodically every 5-7 years, household and individual costs due to the experience of victimization will be analyzed, according to different covariates as geographical area, the type of suffered crime, its dynamic, fear of crime among others. Costs will be estimated in terms of both material and not tangible changes: defense systems adopted for the house, individual strategies and behaviors changes adopted to prevent crime victimization, both in case of experienced than in case of fear of crime with not direct experience of victimization. Also the actual economic loss by type of crime on properties, both individual and household's ones, will be evaluated, using a model to analyze socio-economic impact of victimization.

Keywords

victimization; impact; fear of crime; crime; impact of crime

1. Introduction

Crime is a widespread activity that affects society and human living at all latitudes, but despite its pervasiveness the systematic measurement of its impact on society is far from being a major concern of policy makers. Fear of crime is a social phenomenon more complicate to analyze (Austin et al., 2002): it effects an high impact on the quality of life (Jackson and Gray, 2010), causing anxiety and other psychological issues on a personal level (Ruijsbroek et al., 2015) and, at a social level, affects prejudice and segregation. So that this insecurity has now become a major policy concern (Carro et al., 2010, Curiel & Bishop, 2017)[1].

[1] Rafael Prieto Curiel & Steven Richard Bishop, Fear of crime: the impact of different distributions of victimization, |DOI: 10.1057/s41599-018-0094-8 | www.nature.com/palcomms

Observing the mismatch between crime and its fear and the relevance it has at a social and political level, warrants further investigations of the aspects which might affect the personal perception of crime. For instance, demographic factors (such as age or gender), local factors (if it is a dark or crowded street) and others, such as the amount and style of media coverage of crime. How then should the mismatch between crime and its fear be explained and how does the fear of crime emerge as a social phenomenon? Furthermore, how can policies be designed to tackle this complex social issue if it is not clearly understood? Having data or observations to validate the analysis would be ideal, however, at an individual scale, it is almost impossible to measure the impact that suffering a crime has compared to the impact of other aspects of fear (for instance, hearing that a neighbour suffered a crime as opposed to being the actual victim of that crime). Observations at an individual scale are typically based on victimisation surveys, which frequently do not track the fears of the same individuals over time.

Fig.2. Social Decay or Incivilities rate (2002/2016)

In Italy, we observe a widespread fear of crime and risk of victimisation that has pushed crime to the top of the political agenda. In this context, it would be crucial to be able to answer, to name but a few, questions like 'What is the relative seriousness of different crimes?', 'How much do people spend for fear of crime?'. We need to consider the further source of distortions represented by the (incautious use of) official crime statistics, which being affected by a variable underreporting bias across offences can induce misperception about the effective crime rates and generate false myths as to what offence is on the rise. As is well known, reported crimes represent just the tip of a (sometimes big) iceberg. We said before that criminal activities have a strong impact on the community, but of course it is reasonable to expect that different types of crime produce different social costs.

Fig.3 – Concrete fear rate (2009/2016)

Fig.4 – Fear of crime rate[2]

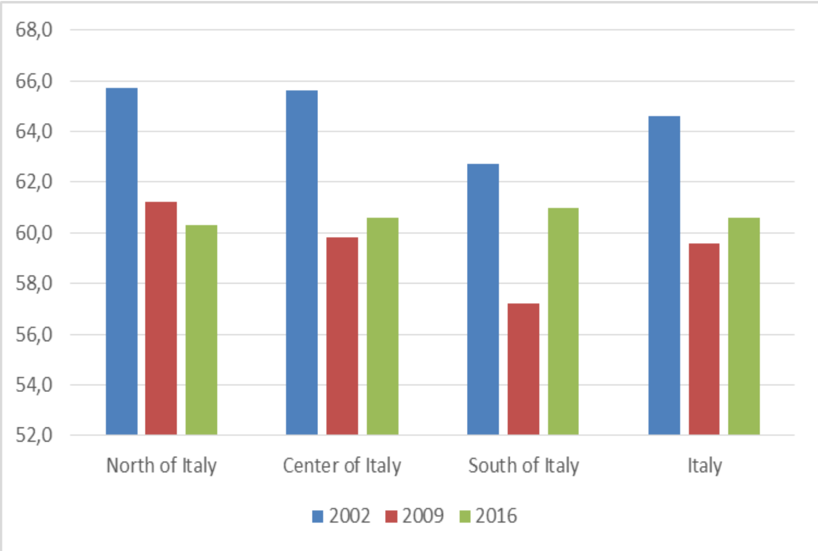
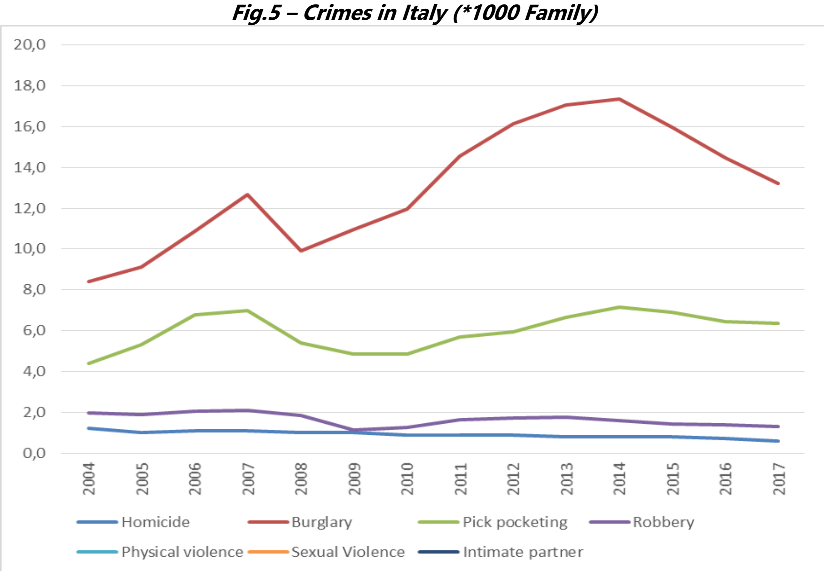


Fig.5 – Crimes in Italy (*1000 Family)



Recently an Europe cross country analysis shows that risk perception is positively correlated with fear of crime in all of the analysed countries, although the strength of the relationship differs. Contrary to expectations, countries with a higher victimization rate exhibit a weaker relationship between fear of crime and risk perception, while in countries with a lower victimization rate, risk perception plays a relatively important role in shaping

[2] People who feel safe walking in the dark alone in the area where they live

fear of crime. No effect of unemployment rate, size of migrant population or income inequality was found in the data.

2. Methodology

One issue that needs to be flagged about fear of crime and perception of risk is that, to some degree, 'worry is a good thing', a 'socially beneficial activity' that prompts us to care for ourselves, our children and others (Jackson et al 2009)[3]. Perception of risk is important, as it can be 'an accurate predictor of future victimisation' and it prompts us to adopt 'protective or risk-avoidant behaviours' (Brewer et al 2007 in Dichter and Gelles 2012)[4]. Accordingly, strategies to improve perceptions of safety should not be 'aimed naively at eliminating fear of crime', but at 'synchronizing fear with actual levels of crime'. (Cordner 2010)[5].

Ceccato (2012)[6] classifies three 'types' of fear of crime – 'individual', which is often associated with personal experience of crime, 'neighborhood' which is a result of what you experience where you live, and 'social macro', described as a 'social phenomenon' shaped by media and forming part of broader anxiety about global and social change (Ceccato, 2012).

Measuring the cost of violent crime victimization is difficult, and using that data to make comparisons across time and place can be even trickier. A significant amount of research has attempted to answer the question: *how much does it cost to be a victim?* [7]

Researchers have developed a variety of terms to characterize the numerous effects on costs of crime. While not all researchers use the same terminology to analyze costs, studies generally consider the following elements:

• *Tangible costs:* Tangible costs are those where the effect can be valued by observing product or service transactions that arise in response to or in anticipation of criminal activity. Such costs include the value of damaged

[3] Jackson J, Gray E and Farrall S (2009). 'Untangling the Fear of Crime'. Criminal Justice Matters, 75:1, pp 12-13. Centre for Crime and Justice Studies.

[4] Dichter M and Gelles R (2012). 'Women's Perceptions of Safety and Risk Following Police Intervention for Intimate Partner Violence'. Violence Against Women 2012, Volume 18 No 12.

[5] Cordner G (2010). 'Reducing Fear of Crime: Strategies for Police'. Office of Community Oriented Policing Services, U.S. Department of Justice, p.5.

[6] Vania Ceccato, (2011), The Urban Fabric of Crime and Fear, Springer Dordrecht Heidelberg New York London.

[7] Arthur Lurigio at Loyola University of Chicago notes that since the 1980s, the cost of crime victimization has been studied in three main ways: (1) drawing upon several data sources and applying a "basic costs calculus;" (2) surveying people on their willingness to pay for crime reduction programs; and (3) asking people to estimate "the effects of crime on housing prices and other community amenities" (Lurigio, 2014), or GAO, COSTS OF CRIME - Experts Report Challenges Estimating Costs and Suggest Improvements to Better Inform Policy Decisions, 2017

property, medical care to treat injuries, and costs to install alarm systems to avoid crime.

• *Intangible costs:* Intangible costs result from negative effects of crime, but the effects do not have directly observed market values. Such costs from crime include lost quality of life as a result of victimization, fear in the community, or avoidance behaviors to reduce the likelihood of being victimized.

• *Timing of the cost in relation to the crime:* While the most recognized costs of crime are those that occur as a direct consequence of the crime, such as the cost to recover from crime-related injuries, researchers consider secondary consequences of crime as well. Such consequences may occur before or after a particular crime occurs. For example, some costs of crime occur in anticipation of crime, for instance to avoid victimization. Other costs occur as a response to crime, such as the cost of judicial proceedings or the costs to incarcerate offenders[8]

• *Who bears the cost:* Crime costs can be borne by individuals, such as the victim or the victim's family, employers or businesses, and broader society, such as when public property is destroyed. Societal costs can also take the form of taxpayer funds for crime prevention or criminal justice system expenses. Cost estimates may also include not only victims' costs, but also costs to potential victims (such as costs associated with the fear of crime), future victims, offenders, and offenders' families[9].

Aim of the present work is to individuate different profiles of fear of crime perception, relating them to social decay indicators, behaviors and police work citizens evaluation in order to describe the intangible cost of crimes with available data. Another analysis will concentrate on economic loss of victimized people for the tangible costs.

International research according that the victimization rate is strictly correlate to the gender, age, social conditions, and that effects directly on tangible costs in anticipation of crime.

---

[8] Similar to this approach, researchers may categorize costs as ex ante or ex post, referring to if the cost occurred before or after the crime occurred. In addition, researchers may consider whether a cost is a direct result of a crime, including most costs borne by the victim. Or, researchers may consider whether a cost is an indirect cost of crime, including fear in the community or some criminal justice costs, without necessarily considering whether a cost is in anticipation of a crime or in response to crime. Studies may not necessarily agree on whether a cost should fit into a particular category or if the studies should consider these categories separately.

[9] Researchers may refer to costs as either external or social costs, which to some extent focus on who bears the cost. For instance, external costs are those imposed by one person onto another, where the affected person does not voluntarily accept the negative consequences. Costs imposed on the victim, such as medical bills, property damage, and lost quality of life are examples of external costs. Social costs are those that reduce the aggregate well-being of society and may include not only victim costs, but also costs considered to divert resources from socially productive uses. Researchers may not agree on which costs to consider or which costs may be external versus social costs.

In Italy the citizen security survey gives an overview of the crime phenomenon through the point of view of the victims. This survey allows having a "submerged" estimate for a large number of crimes not reported to the police and to identify risk population groups; moreover the survey provides other information on the nature of the crimes, the characteristics of the criminals and the relationship between victims and criminals. One of the major results of this survey, is the perception of the security and the fear and worry of suffering crimes), and the perceived risk of crime in the area where you live, of the relationship with the police and the strategies implemented by families and families to defend themselves.

The crimes considered are the theft, pickpocketing, theft of personal belongings, robbery, and the threat e aggression, credit card cloning, fraud, some computer crimes, vehicle theft and vehicle parts, theft of objects from vehicles, theft at home and unauthorized entry, the acts of vandalism, theft and mistreatment of animals. The estimate of violence against women who, requesting a particular methodology, it is detected with ad hoc surveys.

Starting from the micro-data of the Citizen Security survey, lead in 1997; 2002; 2008-2009 and 2015-2016, a dynamic factor analysis was carried out to draw a picture synthetic on the perception of security. Several indicators[10] were considered that characterize a phenomenon so complex, linked to individual subjective factors (personal vulnerability) and objective (concern of undergoing particular crimes, living in certain areas more or less at crime risk). The dynamic factor analysis was dedicated to the statistical treatment of data classified according to the three criteria (or modes, see Tucker 1966): statistical unit, quantitative variable and temporal opportunity for data collection. This type of data can be represented, by means of an array or cubic array of data X (Law and others, 1984) whose generic element is x_{ijt}, where $i = 1, \ldots, I$ is the unit index, $j = 1, \ldots, J$ is the variable index and $t = 1, \ldots, T$ is the index of times. The method can be applied when the same variables are observed on the same statistical units on different occasions. Several statistical models have been proposed to analyse this type of data: the Dynamic Factor Analysis (DFA) is the Italian proposal (Coppi, Zannella, 1979).

The factorial analysis has led to summarizing the phenomenon of safety perception with three main dimensions:

[10] The indicators analyzed are: feeling more or less well at night at home or on the street in your neighborhood; comes out at least once a week evening; bring at least something with him to defend himself; do not go out in the evening or at night alone because you are afraid (often or expected); puts the safe at the car doors when he is alone out of fear; the fear of crime is kept away from certain roads out of fear; is very or fairly worried about being subjected to certain crimes (robbery, theft, sexual violence, car theft or housing); the area where you live is quite a crime risk and crime has increased in the last period; the interviewee often sees tramps, people who take drugs or sell drugs, prostitutes, acts of vandalism, see degraded areas, and poor lighting; the interviewee don't seen police forces and they control little or for nothing and pass at least once a week in the street of residence. Each indicator was calculated as a percentage of 'yes' to the above indicators, age range, region and municipal typology in they live.

- the first relating to the insecurity motivated by crime and the degradation of the area in which we live and the poor control of the police;
- the second linked to the social behavior of individuals (it goes out every night for example);
- the third related to the degradation of the area in which one lives. The dimension of the defense strategy (individual and family) is also added to this analysis.

Alongside this methodology of analysis, a regression was conducted in parallel to understand which are the variables that influence the cost and therefore the economic losses related to the crime suffered in individual crimes (such as theft and theft of cars). to verify the damage actually suffered by the offenses considered.

A first descriptive analysis has been done on Italian Victimization data for the waves 2002, 2008-09 and 2015. The statistical model used is the Factorial Dynamic Analysis (first model, Corazziari et al. 2019), to obtain a picture over time of the complex phenomena of safety perception (intangible cost) related to subjective factors (individual vulnerability) and objective ones (cognitive fear of being victim of specific crimes, living in area social decayed, at risk of crime). The model combine regression and factorial tools to describe both the multidimensional characteristic of safety perception and its evolution over time.

Indicators from the data are: feeling unsafe at home or outside in the evening; bringing some object or locking the car for self-defense if outside alone; avoiding some streets or even to go out in evening due to fear of crime; influence of fear on self-behavior; worry about being victim of specific crimes as vehicle theft, burglary, bag-snatching, sexual harassment or rape, residence area perception in term of criminality, social decay (7 indicators); police ability to control and protect the residence area (if they pass at least one time a week, if people have been stopped for a control in the last 12 months). Other indicators regarding house defense strategies and systems have been discussed in the analysis

As regards the economic loss of victimized people for the tangible costs, a multiple regression analysis has been applied considering the reported loss of victims for two individual property crime; bag-snatching and robbery as response variables. As covariates and expletive variables have been considered some demographic variables about the victim (sex, age, residence Italian region and type of municipality), and variables about the place of the event, when it happens (at which hour in classes), and what has been stolen.

3. Result

Considering the analysis if intangible losses with FDA, in the observed period unsafe people when outside in the evening decreased over time, so as people using some individual strategy of defense; while unsafe people at home increased. No variation over time for cognitive fear of the above specified crimes, but social decay indicators have increased over time.

The multivariate analysis of the considered indicators over time allow to identify 6 profiles of individuals according to their safety feelings related to objective fear of specific crimes, to the social decay of the area where they live, to the presence of police, and are characterized by demographic characteristics as sex and age, by the geographical area and by the type of municipality where they live.

To defend themselves against crimes, families use a wide range of actions and behaviours: from traditional system as "solidarity among neighbours", to the most technological, for example the alarm system. In fact, 72.1% of families stated that their home is equipped with at least an instrumental protection system and 55.7% adopt at least one of the defence actions traditional. Overall, the most widespread are: the security door, by far the system most used defensive (51.3%), the request to neighbours to control (32.4%), the lights lit when they go out (27.3%), the lock on the windows (26.4%) and the bars (23.6%), the lights external with automatic ignition (22.3), the alarm device (20.8%), insurance against theft (14.5%) and the safe (14%). Only 10.8% of families claim to have adopted some defensive system for housing following a theft or a crime immediately while 43% had it installed because they was afraid to suffer it, also if they are never suffer it. In addition to adopting defensive strategies to protect housing, the reaction of the population the risk of victimization is also personal. While traveling, many people do protect them from the risks of suffering a crime:

- 28% of people, when they walk in the evening on the streets of your area, try to avoid situations considered at risk, such as some roads, places or people;
- 39.6% of those who use the car for their journeys are usually in safety to the doors and finally,
- 7.7% bring something with them to defend themselves or to ask for help in case of danger. All these attentions are more used by women, young people and adults and less by the elderly also in relation to the lower frequency of the latter's exit.

The perception of the risk of crime has an incisive influence on the sense of loss and discomfort that citizens feel.

We can considers that 33.9% of the population lives in an area that is very or fairly at risk of crime with more than 14 years, a decidedly higher figure than the previous survey (+11.9 points percentages in respect of the previous survey carried out during 2008-2009 years). 23.2% of respondents believe that the level of crime reached in the area where lives has increased, compared to a year before the interview, while 7% reported one decrease. People who live in areas considered to be at high risk of crime feel more insecure to go out at night (51.3% of cases against 27.6% of the average).

In the observed period unsafe people when outside in the evening decreased over time, so as people using some individual strategy of defense; while unsafe people at home increased. No variation over time for cognitive fear of the above specified crimes, but social decay indicators have increased over time.

The multivariate analysis of the considered indicators over time allow to identify 6 profiles of individuals according to their safety feelings related to objective fear of specific crimes, to the social decay of the area where they live, to the presence of police, and are characterized by demographic characteristics as sex and age, by the geographical area and by the type of municipality where they live.

The first group represent men of 35-64 years old, living in the center of metropolitan area or municipalities with more than 1000 inhabitants, in the Centre and North (Lazio, Toscana and Veneto, and at a lesser extent in Emilia Romagna and Lombardy) and in Sardinia: the live in decayed areas but with a good control by the police: such people feel safe and go out in the evening; notwithstanding the decay of their living area worsened over time.

The second group is similar to the previous one with the main exception of younger males (less than 44 years old) living in other regions in the North (Trentino Alto Adige; Friuli Venezia Giulia; Valle d'Aosta; Liguria; Piemonte), in the Centre (Toscana and Marche); in the South (Molise and Basilicata) and in the two main Islands. It differs from the previous group as decay decreased over time, so as worrying about crimes and the perceived risk of criminality.

Third cluster is of vulnerable people: women over 55 years old, living in small or medium municipalities in Friuli Venezia Giulia, Marche, Abruzzo, Toscana, Piemonte e Sicilia: they feel unsafe both at home than outside in the evening, in areas where social decay increased over time.

The fourth cluster represent unsafe people with no main reason: 70% are men, over 64 years old, living in small municipalities (less than 10000 inhabitants), mainly in the north (Trentino alto Adige, ; Valle d'Aosta; Friuli Venezia Giulia, Liguria) at a lesser extent in Umbria, Molise, Basilicata and Sardinia. Their fear looks difficult to be addressed as it remain constant notwithstanding the improvements of the area where they live in term of social decay.

The fifth and sixth clusters are mainly of women. The former represent young and adult women (half are 35-54 years old), unsafe and very worried and feared to be victims of crime, living in decayed areas: they live mainly in metropolitan areas in half the case, in Lombardy, Campania, Lazio and Puglia. Their situations worsened over time and they are very unsatisfied of the police and use defense systems for the house. The latter cluster represent younger women (more than 80% are less than 45 years old), less unsafe and worried about crimes than the previous cluster, living in better areas in term of social decay, in municipalities other than metropolitan areas, mainly in the South (Basilicata, Campania, Puglia, Sicilia) at a lesser extent in Liguria and Marche: their situation has improved over time.

As regards the economic loss of bag-snatching victims, the significant effects increasing the victim's economic loss are living in the center of a metropolitan area with respect to other municipalities' type, if the victim was within his/her car and at a lesser extent if he/she was within a shopping center, while the loss was lower than the mean loss if the victim was in a park when it happened, if money or jewelries have been stolen, and if it happened between 6 and 9 a.m.

As regard robbery the economic loss increases if the stolen objects, among other, are jewelry, luggage, watch, fur coats, silver ware, HiFi, Tv, furniture. Objects that suggest that the crime happen at home, the intangible lost is very high and there is a home violation.

4. Discussion and Conclusion

For the FDA models adopted (Lavit et al 1994) or Tucker models (Kroonenberg 1992) seems more suitable for cubic matrices of data as the third dimension is time, predicted by the peculiarity to be ordered: in FDA it is explicitly treated as an element of a different nature compared to the other two dimensions, unit and variable. The regression model gives an interesting result trying to highlight the effective cost of considered crimes.

The paper aims to validate a model of analysis of the economic impact of victimization on an individual and family level by comparing the subjective dimension with the objective dimension of the loss.

References

1. Aromaa K. and M.Heiskanen (eds.), (2008), "Victimisation Surveys in comparative Perspective", HEUNI -European Institute for Crime Prevention and Control, Publication Series No. 56, Helsinki.
2. Ceccato V., (2011), The Urban Fabric of Crime and Fear, Springer Dordrecht Heidelberg New York London.

3. Coppi R., Zannella F.(1979), *L'analisi fattoriale di una serie temporale multipla relativa allo stesso insieme di unità statistiche*, Atti della XXIX riunione della SIS.
4. Corazziari I., Dynamic Factor Analysis, in Vichi M., Opitz O., *Classification and Data Analysis, Theory and Application*, Springer, 1999, p.171-179.
5. Cordner G (2010). 'Reducing Fear of Crime: Strategies for Police'. Office of Community Oriented Policing Services, U.S. Department of Justice, p.5.
6. Dichter M and Gelles R (2012). 'Women's Perceptions of Safety and Risk Following Police Intervention for Intimate Partner Violence'. Violence Against Women 2012, Volume 18 No 12.
7. GAO, COSTS OF CRIME (2017) - Experts Report Challenges Estimating Costs and Suggest Improvements to Better Inform Policy Decision.
8. Jackson J, Gray E and Farrall S (2009). 'Untangling the Fear of Crime'. Criminal Justice Matters, 75:1, pp 12-13. Centre for Crime and Justice Studies.
9. Maggino F., Facioni C., Corazziari I. (2019) *Measuring Uncertainties: a Theoretical Approach*, DOI: 10.1504/IJCEE.2018.10011864, International Journal of Computational Economics and Econometrics, Inderscience Enterprises Ltd, vol. 9(1/2), pages 5-28
10. Prieto Curiel R., Bishop S.R., (2009) Fear of crime: the impact of different distributions of victimization, |DOI: 10.1057/s41599-018-0094-8 | www.nature.com/palcomms
11. Tucker R.L., (1966) *Some mathematical note on Three- mode factor analysis*, Psychometrika, Vol.31, n.3
12. van Dijk J., van Kesteren J, Smit P., (2007), Criminal Victimisation in International Perspective Key findings from the 2004-2005 ICVS and EU ICS, WOD

The integrated system on Violence Against Women (VAW) in Italy: Measuring, communicating and using VAW data

Maria Giuseppina Muratore, Lucilla Scarnicchia, Claudia Villante
ISTAT – National Institute of Statistics

Abstract

In 2017 an agreement between the Italian National Statistical Office (ISTAT) and the national Department for Equal Opportunity (DEO) at the Presidency of the Council of Ministers was signed in order to create an integrated data system on Violence against women (VAW) with the aim to support stakeholders in designing policies to prevent and combat VAW. The Council of Europe Convention on preventing and combating violence against women and domestic violence (2011), the so called Istanbul Convention , ratified in Italy in 2013, requires States to offer a holistic response to violence against women, through the "4 Ps approach": Prevention of violence through sustained measures that address its root causes and aim at changing attitudes, gender roles and stereotypes that make violence against women acceptable; Protecting women and girls who are known to be at risk and setting up specialist support services for victims and their children; Prosecuting the perpetrators; adopting and implementing state-wide integrated Policies that are effective, co-ordinated and comprehensive. Furthermore, based on article 11 "Data collection and research", the Convention recognizes for the first time the crucial importance and the role played by statistics in improving the awareness of gender based violence and its monitoring. With the aim to respond to the arisen needs the Italian integrated system was built up. It is a huge container fully accessible online which collect the main data and indicators, but also the existing legal frameworks, experiences and policies regarding violence against women. Data are being progressively organised in a dedicated data warehouse. The system collects official statistics from both kind of sources, survey data and register statistics, which are related to the following fields:

1. the prevention and more specifically data on the gender stereotypes and stereotypes towards violence against women, that are collected among the adult population and the student population, respectively using computer assisted telephone interviews (CATI) and computer assisted self-interview technique (CASI) ;

2. the size of the phenomenon in its several forms, as physical, sexual, psychological and economical violence inside and outside the family, as it is observed through the VAW surveys; sexual harassment and violence at work

investigated by the Citizens' safety surveys; administrative data on female homicides
3. the way out, with data about the calls to the to the national helpline against violence and stalking 1522 and the new census data collection on shelters and centers for victims of violence;
4. the paths in the justice system, as the analysis based on the reports to the police, the proceedings, the convictions and the prisoners data.
Most of these data are produced by ISTAT in agreement with NGOs and a number of National Ministries and Institutions as this is the only possible way to proceed producing good, accurate and meaningful data.
The paper will present the structure of the integrated system as well as the main key results from the new data collection on stereotypes and on specialized services for victims of VAW.

Keywords
Gender; multisource; Istanbul; police; justice

1. Introduction

In 2017 an agreement between the Italian National Statistical Institute (ISTAT) and the national Department for Equal Opportunity (DEO) at the Presidency of the Council of Ministers was signed in order to create an integrated data system on Violence against women (VAW) with the aim to support stakeholders in designing policies to prevent and combat VAW.

The Council of Europe Convention on preventing and combating violence against women and domestic violence (2011), the so called Istanbul Convention, ratified in Italy in 2013, requires States to offer a holistic response to violence against women. Chapter II of the Explanatory Report to the Convention describe the "3 Ps structure" of "Prevention", "Protection", and "Prosecution" of the victims: Prevention of violence through sustained measures that address its root causes and aim at changing attitudes, gender roles and stereotypes that make violence against women acceptable; Protecting women and girls who are known to be at risk and setting up specialist support services for victims and their children; Prosecuting the perpetrators. However, since an effective response to all forms of violence covered by the scope of the Convention requires more than measures in these three fields, the drafters considered as necessary to include an additional "P": adopting and implementing state-wide integrated Policies that are effective, co-ordinated and comprehensive.

Furthermore, according to article 11 "Data collection and research", the Convention recognizes for the first time the crucial importance and the role played by statistics in improving the awareness of gender based violence and its monitoring.

With the aim to respond to the arisen needs the Italian integrated system was built up. It is a huge container fully accessible online since the 25th of November 2017 which collect the main data and indicators, but also the existing legal frameworks, policies and experiences related to the violence against women. Data are being progressively organised in a dedicated data warehouse.

The system collects official statistics from both kind of sources, survey data and register statistics.

2. Methodology

According to the article 3 of the Istanbul Convention, as already defined in the 1993 UN Vienna Conference, the "violence against women" shall mean all acts of gender-based violence that result in, or are likely to result in, physical, sexual, psychological or economic harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life. The "gender-based violence against women" shall mean violence that is directed against a woman because she is a woman or that affects women disproportionately. The Convention provides that States have to consider in their penal code as forms of violence forced marriages, psychological violence, stalking, physical violence, sexual violence, including rape, female genital mutilation, forced abortion and forced sterilisation, sexual harassment.

However these definitions are not operational and using them for statistical purposes is a very complex issue. Looking for instance at most serious form of violence, the femicide has not a legal definition in Italy and in the EU countries since femicide does not represent a specific offence, as on the contrary happens in 16 countries of Latin America. Representing, however, a phenomenon of significant interest in the public debate, femicide is measured for statistical purposes on the basis of the relationship between the victim of the murder and its perpetrator.

This choice was endorsed at international level in May 2017, when the United Nations Office on Drugs and Crime (UNODC) Experts Group for the definition and implementation of the International Classification of Crimes for statistical purposes (ICCS) decided to adopt as a statistical definition for "femicide", the murder of women happened in the family context from partners, former partners or relatives.

The definition is important, in fact, looking at the Italian data, for instance, in 2017 123 women were killed, but how many of these murders are femicides? Considering the perpetrator/victim relationship, almost 3 out of 4 of 123 murders were committed in the family: 35.8% women were killed by partners, 8.1% by former partners and additional 28.5% women by relatives (data from

Ministry of Interior). On the contrary in the same year only the 15.8% of men were killed in the context of family relationships (37 out of 234 men).

Gender differences are evident and very stable. In Italy during the last decade the percentage of women homicides in the family context fluctuated from a minimum of 62.7% in 2010 to a maximum of 77% in 2014, decreasing to 72.4% in 2017.

Furthermore, while the murders of men constantly declined, women's murders are stable. Men victims of murder decreased from 4.0 to 0.9 per 100,000 males between 1992 and 2015 (according to the Causes of Death Data), while for women the rate fell from 0.6 to 0.4.

This means that for men although the incidence of murders still remains greater (about double) than for women, the progress was remarkable, especially because of the decline of deaths due to common criminality and to organized crime. While for women, who started from a much more favourable situation, the decrease followed a much slower rhythm in the same period.
Homicide data on women and gender related homicides tell a different story compared to homicides of men.

Concerning the other expressions of VAW, more common than homicide, ISTAT carried out a survey on Women's safety in 2006 and in 2014. Collected data show a sensitive gap between the number of victims of physical or sexual violence and the number of those who reported the suffered violence to the police and the competent authorities.

VAW occur mainly in the familiar environments, where a woman should feel safer. This characteristic explain why this phenomenon remains to a large extent underreported. The proximity with the perpetrator, who is a loved person, and the complex and opposing emotional and psychological reactions do not help the victims' disclosure.

It therefore is important to consider both sources of data, the police and judicial offices and the ad hoc surveys, designed to investigate directly the women experiences. The integrated information better represents the entity of the phenomenon and its characteristics.

In 2013 the Italian government established a task force on VAW measurement. The aim of the task force was to identify data useful to design policies to combat VAW, to protect victims, to help victims in escaping from violence and cope with suffered violence, caring them and giving help for re-integration.

ISTAT coordinated the task force that involved several actors and institutions: Ministry of Interior, Ministry of Justice, Ministry of Health, Ministry of Army, Ministry of Work and Ministry of Social Politics; NGOs (shelters, helpline 1522, VAW services); VAW experts, as academic and lawyers.
It was a long and difficult work, first of all to share the same language and the same aims and to achieve a wider view on the phenomenon.

The first step of the task force was identifying and understanding the data gaps depending on the type of agency providing data.

Data on VAW turned out to be collected in a very fragmented way. Some of the Italian Regions have developed local observatories on VAW. All shelters and specialized services collect data on their activities while only some hospitals and health services collect data on VAW. Local experiences differ in quality and have developed in their own way, without a common background in collecting data. Results are often of best quality but they stay as an isolated experience. The quality of data collection appeared very much linked to the individual sensitivity on the topic, the follow up of victims was lacking, there was no evidence of the most important data and no guarantee for periodicity and a rigorous methodology. Even in case of systematic collection of data, as in the case of administrative data from the police system and judicial offices, there was a low awareness about the importance of data to describe correctly the phenomenon.

The task force identified the best data suitable to represent the problem, looking at the best practices and at the possible solutions for each agency, including NGOs. The technical measures adopted by the task force became part of the VAW national Action Plan, approved and funded by the government. In this context, the task force identified the necessity to create an Informative system on VAW, as a comprehensive tool for storing data (macro and metadata) and analysis reports, periodically updated, developed and managed by ISTAT, in coordination with the VAW network.

Following the Istanbul Convention "3Ps structure" (Prevention, Protection, and Prosecution), the idea of ISTAT was to define a data warehouse built by connecting administrative sources and field research, as shown in the Table 1.
Table 1 – 3Ps structure of Istanbul Convention and mapping data in Italy

Table 1. Estimated coefficients, Odds ratios and p-values in weighted binary logistic model of drug use (Have ever used drugs)

Predictors	Coefficient	Odds Ratio
Single Male	1.7028	5.4894
Married Male	1.9523	7.0449
From Region 7	1.2486	3.4854
From Region 10	1.4388	4.2158
From Region 11	1.8654	6.4586
From Region 12	0.8888	2.4322
From ARMM	1.3380	3.8114
Often attends religious services (at least once a week)	-0.6080	0.5445
From birth up to age 18, 2nd person approves having a bf/gf before age 18	0.6342	1.8855
Family members frequently or almost always eat together at least one meal a day	-0.1911	0.8260
Sees gambling as wrong	-0.3696	0.6910
Rated Satisfied with life, all things considered	-0.2915	0.7471
Member of a fraternity/sorority	1.3429	3.8301
Had been suspended from attending classes	0.9211	2.5122
Have ever worked for pay, in cash or in kind	0.8068	2.2406
Had ever read pornographic or sexually explicit books/magazines/tabloids	0.9926	2.6982
constant	-5.8863	0.0028

**base category of the dependent variable: Never used drugs

3. Result

The Italian integrated system on VAW collect the main data and indicators, the existing legal frameworks, policies and experiences related to the violence against women. It is publicly accessible online, on the ISTAT website. Data are being progressively organised in a dedicated data warehouse.

The structure of the system reflects the Istanbul Convention "4Ps structure" Prevention, Protection, Prosecution and integrated Policies and is organised in six sections described as follows.

The context

The VAW is addressed giving references to national and international policies, existing legal frameworks including Italian regional laws, common definitions and internationally agreed indicators.

The phenomenon - Violence inside and outside the family

This section describes the size of the phenomenon in Italy in its several forms, as physical, sexual, psychological and economical violence inside and outside the family, as it is observed through the Women's safety surveys, carried out by ISTAT. Data refer to 2006 and 2014, while the next way of the survey will be held in 2020. Data are of high quality and show the prevalence of violence against women, its frequency, severity, causes, risk factors, dynamics, consequences and costs and the violence witnessing.

The phenomenon - Sexual harassment and violence at work

Sexual harassment and violence at work data are collected by ISTAT with the Citizens' safety surveys starting from 1997-1998 to 2015-2016 (4 editions) and the next wave is to be carried out in 2020.

The phenomenon - Female homicides

Female homicides are recorded in a historical series with national administrative data from 2002 to 2017 provided by ISTAT as Causes of Death Data and by the Ministry of Interior with information on femicides.

The phenomenon - Stereotypes

With the aim to give some information on the evolution of social relations between men and women, and the impact of women's stereotypes over time ISTAT collect data on gender stereotypes and stereotypes towards violence against women among the adult population and the student population. Individuals 18-74 years old were interviewed using computer assisted telephone interviews (CATI) with an ad hoc module in the Labour Force Survey in 2018, while students 11-19 years old will be interviewed by computer assisted self-interview technique (CASI) in 2020. The integrated system provides also data on surveys previously carried out by ISTAT which included some questions on gender stereotypes, namely Time use survey 2013-2014 and Survey on gender stereotypes and discrimination in 2011. ISTAT is also working on a "sentiment analyses" with the use of big data, as the twitter's text messages, which will focus on the moods and opinions towards gender stereotypes and violence against women.

The way out - Helpline against violence and stalking 1522

ISTAT analysed data on the calls to the national helpline against violence and stalking 1522 from 2012 to 2018 giving some information on the characteristics of the person calling, the type of violence suffered and the expressed needs.

The way out - Shelters and centres for victims of violence

The pathways out of the VAW are investigated by ISTAT with new data collections. A survey on services offered by specialized centres (280) to victims of violence was carried out in 2018 and is currently ongoing for its second edition. In 2018 there was also a survey on services offered to victims of violence by shelters (250). While in autumn will be launched a survey focusing on those women who used the specialized services. All these surveys will be annually updated and will provide a common set of information.

The way out - Programs for the perpetrators

The national Action Plan related to VAW call for programs for perpetrators. The aim is the rehabilitation of men in order to prevent the repetition of violence and to promote non-violent behaviors in the relationships. In 2018 the Italian National Research Council (CNR) collected data on centres which provide services to men who have been violent against women.

The path in the system of Justice

The section shows data related to the reports to the police system, the proceedings, the convictions and the prisoners. Data are collected on annual basis by the Ministry of Interior and by the Ministry of Justice and local Offices of the judicial system. They allow describing the type of violence and how often this has become acquainted with the law enforcement and the judicial system, the number of convictions and the characteristics of the sentenced persons. They also allow analysing the type and incidence of concomitant offences, some characteristics relating to the context and dynamics of the phenomenon and some information on the outcome of the proceedings.

In addition to the problem of the unreported episodes of violence, in Italy the statistics of the justice sector are affected by the lack of information on the sex of victims and the information regarding the victim-perpetrator relationship. For this reason, data collected in the integrated system on VAW are those related to some crimes, namely stalking, sexual assault, maltreatment in the family, female genital mutilation, which can be considered crimes to majority vocation of gender.

The prevention

The integrated system on VAW gives an overview of the awareness campaigns supported by the Italian Regions and by the DEO and list the actions carried out by the National Authorities to combat the VAW.

International experiences

Data and analyses provided by international organisations, United Nations Offices, Agencies of the European Union, National Statistical Offices of the European Union Member States are made available in this section.

4. Discussion and Conclusion

Good policies need regular, high quality, standardized, speaking data and coordinated data, but these goals are not easily achieved. Regular data are data periodically collected and updated in order to monitor the phenomenon; high quality data are collected using a sound methodology according to the specific sources; standardized data are comparable across countries and over time; speaking data are meaningful data suitable to appropriately describe the phenomenon, using core variables and indicators; coordinated data are comprehensible and comprehensive data and are the results of the good cooperation of the data providers. In this respect Italy, with ISTAT as coordinator of the dedicated task force, is building up a very good framework on VAW and the integrated system is among one of the achieved results. The system is online since the 25th of November 2017 (https://www.istat.it/it/violenza-sulle-donne) and is progressively updated, however a number of further steps needs to be done.

ISTAT interacts with the main stakeholders, but only a concrete willingness to convert in statistical terms some information already available in the computerized registers may lead to a correct definition of the violence against women. Ministries of Justice and Interior were asked to collect the crucial information which combine the sex of the victims and the relation with the perpetrators. Required changes in their systems of data collection are expensive and have found difficulties in being implemented. However this is absolutely important as the current national legislation is not adequate to detect violence in intimate partner relationship and in the family context. Data from the police sector and from the justice section should also keep the information on each victim linked to the information on the related perpetrator, making possible a follow-up of the victims in the reporting and judicial itinerary. Further developments are also needed to follow the judicial proceedings and to know, for example, the number of acquitted perpetrators and not only the persons convicted or imprisoned.

Data collection on women utilizing shelters, crisis centre and VAW services (NGOs) has been harmonized, i.e. based on the same variables across all shelters, and is planned to be systematically repeated, at least annually. However the pathways out of violence would be known only with a continuous flow of data, keeping together information on the same woman who may contact more than one shelter, several services, emergency services, hospital and police and justice authorities.

Data from hospitals, emergency and health or social public services are currently not very useful because it is not recorded if the woman is a victim of a gender based crime.

According to the measures adopted by the task force on VAW, hospitals and emergency departments will adapt some new procedures in order to recognize this kind of victims, asking them the victim-perpetrator relationship.

Monitoring VAW includes also observing the evolution of social relations between men and women and the impact of gender stereotypes. It is important to understand how men and women behave in society, for example at school, at work, in the family, at different ages. It is also important to evaluate the undertaken policies, in terms of prevention campaigns and raising awareness, for instance about a proper gender relationship in the education system.

Recent studies on social investment (SROI-Social Return on Investment) have highlighted the importance of prevention and raising awareness to stop the cycle of violence, namely to disrupt the vicious circle of the intergenerational transmission of violence, a mechanism on which the intervention has hitherto been modest.

Read and analysed as a whole, the integrated system on WAV will provide the governing bodies and all public and private actors involved an accurate and complete picture of the phenomenon under several aspects. When finished and completely operating, the system will be an essential tool for evaluating the effectiveness of policies undertaken to protect and support the victim and to prevent the phenomenon, in line with the principles of the Istanbul Convention.

References

1. Council of Europe (2011). Council of Europe Convention on preventing and combating violence against women and domestic violence. https://www.coe.int/fr/web/conventions/full-list/-/conventions/rms/090000168008482e
2. Council of Europe (2011). CETS 210 Explanatory Report to the Council of Europe Convention on preventing and combating violence against women and domestic violence. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016800d383a
3. OCSE, Paris21, National Strategies for the Development of Statistics Guidelines, Guidelines for the elaboration of a National Strategy for the Development of Statistics. April 2018
4. WeWorld, "violenza sulle donne. non c'è più tempo. quanto vale investire in prevenzione e contrasto. Analisi SROI (social return on investment) delle politiche d'intervento", February, 2017

Gender-based and intimate partner violence against women in Spain

Rivera-Galicia, L.F.[1,3,4], Gallo-Rivera, M.T. [2,3,4], Mañas-Alcón, E. [2,3,4], Montes-Pineda, O.[2,3,4]

[1]Department of Economics. University of Alcalá. Spain.
[2]Department of Economics and Business. University of Alcalá. Spain.
[3]Institute for Economic and Social Analysis (IAES). University of Alcalá. Spain
[4]Chair of Corporate Social Responsibility. University of Alcalá. Spain.

Abstract

Gender-based violence (GBV) is one of the most severe forms of Violence Against Women, especially when it is caused by an intimate partner. In 2013, the World Health Organization (WHO) estimated that more than one in three women worldwide has experienced physical and/or sexual violence throughout life. Awareness of this problem by governments and international organizations has increased in recent years.

This paper aims to analyse the prevalence of IPV against women in Spain during 2016. Based on data from the 2015 Spanish Macro survey of gender-based violence, which included 10,171 women aged 16 and over, we analyse five types of Intimate Partner Violence (IPV): sexual, physical, psychological (controlling behaviours and emotional abuse) and economic violence. First, we describe the Spanish population as a framework to understand GBV in Spain. Second, we analyse the sample and we characterize those women who have suffered IPV. Third, we characterize the aggressors, based on the information provided by the victims. Finally, we provide the profiles of the women who have suffered from IPV as well as their aggressors. The key findings of this paper are striking: 13.2% of Spanish women aged 16 or more have suffered from GBV; only 2.7% have suffered sexual and/or physical violence, but 9.2% have felt fear or have been controlled. The perception about their own health is worse amongst women who have suffered from IPV.

Keywords

Gender-based violence; Intimate partner violence; Spain

1. Introduction

Violence against women (VAW) is a violation of human rights, is rooted in gender inequality, is a public health problem, and an impediment to sustainable development. Intimate partner violence is one of the most common forms of violence against women and includes physical, sexual and emotional abuse and controlling behaviors by an intimate partner. More than

1 in 3 (35%) women worldwide have experienced physical and/or sexual violence by an intimate partner (WHO, 2019).

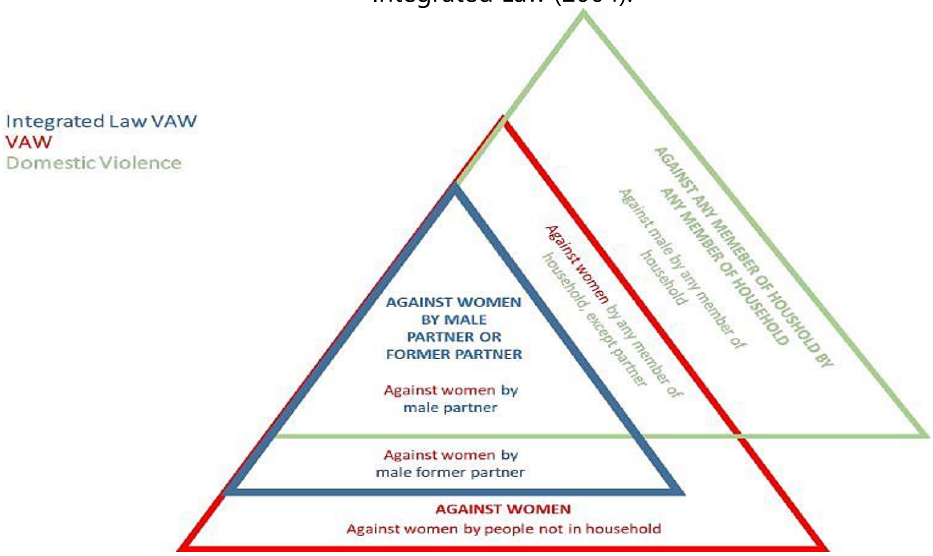
The Spanish Government is aware of the social problems that are caused by VAW since the 1990s (Cepeda, 2018). The 28th of December 2004, the Government approved the *Organic Law 1/2004 on Integrated Protection Measures against Gender Violence*, which considers this type of violence (directed against women for the fact of being women, this law is known as the Integrated Law against Gender Violence) as the most brutal symbol of inequality and one of the most flagrant attacks on fundamental rights such as freedom, equality, life, safety and non-discrimination (GREVIO, 2019, pp. 85-89).

2. The definition of Gender-Based Violence in Spain

The purpose of the Spanish Integrated Law against Gender Violence to combat the violence exercised against women by their present or former spouses or by men with whom they maintain or have maintained analogous affective relations, with or without cohabitation, as an expression of discrimination, the situation of inequality and the power relations prevailing between the sexes.

This Law establishes integrated protection measures whose goal is to prevent, punish and eradicate this violence and help its victims: women, their minor children and minors under their guardianship, care or custody. The gender violence to which this Law refers encompasses all acts of physical and psychological violence, including offences against sexual liberty, threats, coercion and the arbitrary deprivation of liberty.

Figure 1: Definitions of VAW, Domestic Violence and GBV according to Spanish Integrated Law (2004).



Source: Adapted from Mañas-Alcón et al. (2019).

3. The Survey on Violence against Women 2015

The *Survey on Violence against Women 2015* is the most relevant statistical operation produced in Spain on violence against women. It has been conducted every four years since 1999. Since 2011, the Government Delegation for Gender Based Violence manages it. It achieves the quality requirements introduced by the UN Statistics Committee (UN, 2014).

The questionnaire has been improved to better assess the violence suffered by women from their current or former partners, and a new battery of questions has been introduced in order to have, for the first time in Spain, data on the prevalence of non-partner physical and sexual violence against women. The reasons for the introduction of this battery were, on one hand, the recommendations of international organizations such as the United Nations, requesting the Member States to facilitate these data before the end of 2015, and, on the other hand, the need to comply with the provisions of the National Strategy for the Eradication of Violence against Women 2013-2016 in terms of raising the visibility of other forms of violence against women in addition to intimate partner violence.

The Survey on Violence against Women 2015 was conducted in partnership with the Center for Sociological Research to a sample of 10,171 women aged 16 or above, representative of the female population residing in Spain. The sample increased in more than 2,000 women with respect to the 2011 survey and women aged 16 and above have been interviewed, face to face, instead of just women over 18 as was done in previous surveys.

Following the international recommendations, 5 sets of questions were developed for the 5 different types of intimate partner violence, as shown in Table 1:

Table 1: Items included in the questionnaire to detect the different types of IPV.

Violence	Your partner*...
Physical violence	1- has ever slapped you or thrown something at you that could hurt you 2- has ever pushed you or shoved you or pulled your hair 3- has ever hit you with a fist or with anything else that could hurt you 4- has ever kicked you, dragged you or beat you up 5- has ever tried to choke or burn you on purpose 6- has ever threatened to use or has actually used a gun, knife or other weapon against you
Sexual violence	1- has ever forced you to have sexual intercourses** when you did not want to 2- did you ever have sexual intercourses you did not want to because you were afraid of what your partner could do if you refused 3- has ever forced you to do something else sexual that you did not want or that you found degrading or humiliating 4- has ever attempted to force you into sexual intercourse by holding you down or hurting you in some way 5-

Psychological violence: controlling behaviours***	1- tried to keep you from seeing your friends 2- tried to restrict contact with your family of birth 3- insisted on knowing where you are at all times 4- ignored you and treated you indifferently 5- got angry if you spoke with another man or woman 6- unjustifiably suspected that you were unfaithful 7- expected you to ask for permission before going by yourself to some places, for example a hospital or health center, a cultural or sports center, etc.
Psychological violence: emotional abuse	1- has ever insulted you or made you feel bad about yourself 2- has ever belittled or humiliated you in front of other people 3- has ever done things to scare or intimidate you on purpose (e.g. by yelling and smashing things, looking at you in a certain manner) 4- has ever verbally threatened to hurt you 5- has ever verbally threatened to hurt someone you care about
Economic violence***	1- refused to give you enough money for household expenses, even when your partner had money for other things 2- prevented you from making decisions about family finances and from shopping independently 3- forbid you to go to work or study outside the household
Fear	How often have you been or are you afraid of your (current/last/former) partner?
* Different questions were done to differentiate current/last/former parents. ** Sexual intercourse is defined as oral sex, forced anal or vaginal penetration. *** For controlling behaviours and economic violence, frequency was not measured. The questions refer to the general behaviour of the respondent's partner. This is because these behaviours refer to characteristics of the relationship and are often an ongoing pattern (course of conduct) rather than specific acts.	

Source: Government Delegation for Gender-Based Violence (2015).

In the Survey on Violence against Women 2015 information was collected about all partners a woman has had throughout her life while previous surveys only asked about the partners she has had in the last 10 years, thus resulting in a great increase of the percentage of ever partnered women with respect to the previous surveys. Therefore, the prevalence of lifetime intimate partner violence is more precisely measured.

4. Results

4.1. The Spanish Population

In Spain, there were 46,440,099 inhabitants in 2016. Their distribution, according to gender and age, is shown in Table 2:

Table 2: Population in Spain. 1st of January 2016.

Age Group	Males	Females	Total
0-13	3,387,554	3,185,739	6,573,293
14-15	465,293	438,346	903,639
16-64	15,203,861	15,065,141	30,269,001
65 and over	3,750,757	4,943,411	8,694,164
Total	22,807,464	23,632,635	46,440,099

Source: Mañas-Alcón et al. (2019).

49.1% of the population are males and 50.9% females. The Survey on Violence against Women 2015 considers only women aged more than 15, so the objective population is 84.7% of Spanish women. Women under 16 have been discarded due to legal reasons (parental permission is needed). Thus, sampling coverage is 43.1% of the Spanish population.

4.2. IPV against women

According to the data gathered from the Survey on Violence against Women 2015, 13.2% of Spanish women aged 16 or more suffered any type of IPV during the 12 months prior to the interview (1,339 women out of 10,171 women in the sample). The most common type of IPV suffered is psychological violence, with controlling behaviour (9.2%), followed by emotional abuse (7.9%) and physical and/or sexual violence (2.7%) (see Table 3):

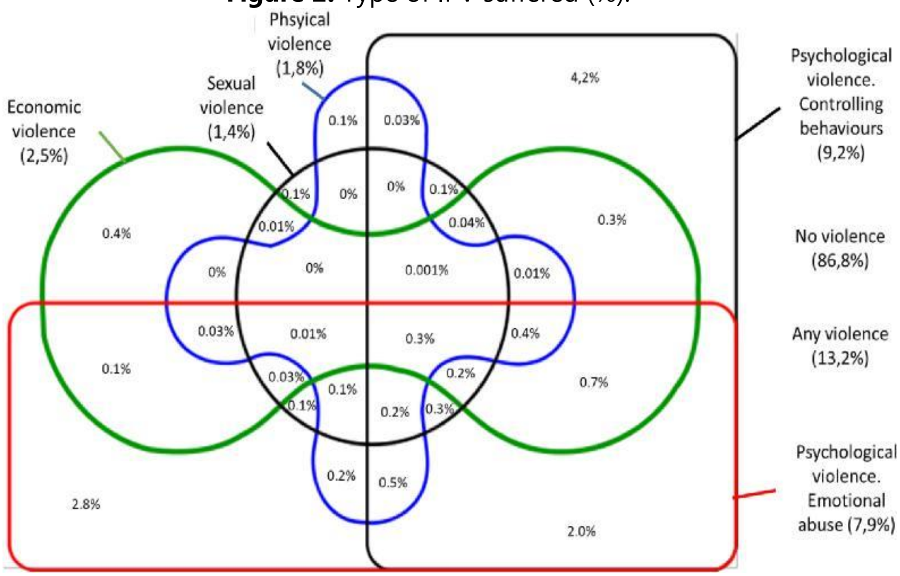
Table 3: IPV suffered in the last 12 months (%).

Types of Violence	Women 16 years old or
Any type of Violence	13.2%
Physical violence	1.8%
Sexual violence	1.4%
Physical and/or sexual violence	2.7%
Psychological violence: controlling	9.2%
Psychological violence: emotional abuse	7.9%
Economic violence	2.5%

Source: Mañas-Alcón et al. (2019).

Figure 2 shows all possible intersections between the different types of IPV that a woman can suffer. We can observe the percentages of Table 2. 0.03% of the women have suffered the 5 types of IPV (see the corresponding area in Figure 2).

Figure 2: Type of IPV suffered (%).



Source: Adapted from Mañas-Alcón et al. (2019).

4.3. The aggressors

Women who have suffered from IPV have been asked about their aggressors in the 12 months prior to the interview. The information in this section refers to a total of 1,248 men who infringed the different types of IPV according to the data provided in the sample.

Table 4: Proportion of aggressors who have infringed each type of violence:

Types of Violence	Aggressors
Physical violence	12.4%
Sexual violence	9.6%
Psychological violence: controlling	72.1%
Psychological violence: emotional abuse	58.4%
Economic violence	18.6%

Source: Adapted from Mañas-Alcón et al. (2019).

Table 4 presents the percentage of type of violence committed by the assailants. 72.1% of them have shown controlling behaviours and 58.4% have abused emotionally of their partners, infringing psychological violence to women.

5. The profile of victims and aggressors

Table 5 shows the distribution of the educational attainment of women and men to compare the differences that can be found within women (victims and no victims) and men (aggressors or general population). It is remarkable that the percentage of IPV is higher among women with secondary education, and lower among women with University studies. The proportion of aggressors is higher among individuals with a primary education level and lower among those who have university studies.

Table 5: Educational attainment. Women and men in Spain. Victims and aggressors. 12 months prior to the interview.

	Women		Men	
	No victims	Victims	All men	Aggressors
No education	6.7%	6.4%	6.6%	6.6%
Primary school	21.2%	20.0%	13.4%	23.6%
Partial secondary school	22.5%	23.4%	31.3%	25.2%
Complete secondary	14.0%	17,5%	14.0%	14.7%
Apprenticeship certificate	15.8%	16.4%	7.4%	13.7%
University studies	19.5%	16.0%	27.3%	13.8%
N.A.	0.3%	0.2%	0.0%	2.4%

Source: Adapted from Mañas-Alcón et al. (2019).

Table 6 presents the economic activity status for women and men and compares the distribution of this variable among victims and non-victims, and aggressors and general population. Most women work, although in the case of women who suffer IPV, the group of unemployed is bigger. In the case of men, it can be observed that the situation of unemployment increases the propensity to be an aggressor

Table 6: Economic activity status. Women and men in Spain. Victims and aggressors 12 months prior to the interview.

	Women		Men	
	No victims	**Victims**	**All men**	**Aggressors**
Work	38.6%	39.0%	53.3%	54.6%
Family business	0.8%	1.4%		
Retiree (previous work)	14.5%	9.3%	25.0%	24.1%
Retiree (no previous	5.4%	2.6%		0.1%
Unemployed	19.5%	26.5%	11.8%	14.9%
Student	6.1%	8.7%	6.7%	5.1%
Housework	14.6%	12.0%	0.0%	0.0%
N.A.	0.4%	0.4%	3.2%	3.2%

Source: Adapted from Mañas-Alcón et al. (2019).

Table 7 shows the self-perceived health as a measure of subjective health for women who have suffered from IPV and for those who have not. 58.7% of the victims state that their health status is good or very good, 11.1% say that it is bad or very bad, and 30.1% say it is fair. On the other hand, women who have never suffered IPV state to a greater extent than their health condition is good or very good (66.3%) and to a lesser extent bad or very bad (9.0%), or fair (24.8%)

Table 7: Women's self-perceived health status. 12 months prior to the interview.

	No victims	**Victims**
Very Good	16.4%	11.9%
Good	49.9%	46.8%
Fair	24.8%	30.1%
Bad	6.5%	7.1%
Very Bad	2.5%	4.0%
N.A.	0.0%	0,1%

Source: Adapted from Mañas-Alcón et al. (2019).

References

1. BOE (2004): *Ley Orgánica 1/2004, de 28 de diciembre, de medidas de protección integral contra la violencia de género*. Madrid, Spain. Retrieved from: https://www.boe.es/boe/dias/2004/12/29/pdfs/A42166-42197.pdf.
2. Cepeda, I. (2018): Fighting Prejudice: Campaigns on Gender Violence in Spain. *Journal of International Women's Studies*, 19(6), 17-34.
3. GREVIO (2019): *Spanish Government Report on the Implementation of the Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention)*. Group of Experts on Action against Violence against Women and Domestic Violence. Council of Europe. Retrieved from: https://rm.coe.int/state-report-from-spain/16809313e0.
4. Government Delegation for Gender-Based Violence (2015): *Survey on Violence Against Women 2015. Main Results*. Government Delegation for Gender-Based Violence. Ministry of Health, Social Services and Equality. Madrid, Spain. Retrieved from: http://www.violenciagenero.igualdad.mpr.gob.es/violenciaEnCifras/macroencuesta2015/pdf/S PAIN_SURVEY_ON_VAW_2015.pdf.
5. Mañas-Alcón, E.; Rivera-Galicia, L.F.; Gallo-Rivera, M.T.; Montes-Pineda, O.; Figueroa, C. & Castellano, M. (2019): *El impacto de la violencia de género en España: una valoración de sus costes en* 2016. Government Delegation for Gender-Based Violence. Ministry of the Presidency, Relations with Parliament and Equality. Madrid, Spain. Retrieved from: http://www.violenciagenero.igualdad.mpr.gob.es/violenciaEnCifras/estudios/investigaciones/ 2019/pdfs/El_impacto_de_la_VG_ES.pdf.
6. UN (2014): *Guidelines for Producing Statistics on Violence against Women*. New York. United Nations. Retrieved from: https://unstats.un.org/unsd/gender/docs/Guidelines_Statistics_VAW.pdf.
7. WHO (2013): Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva. World Health Organization. Retrieved from: https://www.who.int/iris/bitstream/10665/85239/1/9789241564625_eng.pdf?ua=1.
8. WHO (2019): *RESPECT Women: Preventing violence against women*. Geneva. World Health Organization. Retrieved from:https://apps.who.int/iris/bitstream/handle/10665/312261/WHO-RHR-18.19-eng.pdf?ua=1.

How to measure shelters and specialized services to support to victims of violence

Roberta Crialesi, Alessandra Battisti, Maria Giuseppina Muratore
Italian National Statistical Institute – ISTAT

Abstract

The Council of Europe Convention on preventing and combating violence against women and domestic violence (Istanbul Convention, 2011) requires Countries to provide for setting up "specialized services of immediate support, in the short and long term, for each victim of any act of violence that falls within the scope of the Convention". The types of support that these services provide are shelters and safe accommodations, telephone helplines, sexual violence services, legal support, counselling centres, and specific services for children as victims or witnesses. For long time in Italy the available information on this kind of services was fragmented, based only on partial investigations. The national Agreements between Government, Regions and Autonomous Provinces (signed in 2014) established the minimum requirements of anti-violence centres and shelter houses to guarantee homogeneous criteria at national level and supported the implementation of a more systematic and extensive data collection. In 2018 Istat carried out for the first time, in collaboration with the Department for Equal Opportunities (DPO) at the Presidency of the Council, the Regions and the National Research Council (Cnr - Irrps), a national survey on the services provided by the anti-violence centres for victims of violence. The survey aims to identify the different territorial models adopted by the centres describing the dissimilar opportunities available to support and accompany women and their children to go through the specific and individual path of exit from violence. The survey was carried out between June and July 2018, 281 anti-violence centres meeting the requirements of the 2014 Agreement were interviewed. The centres mainly provide listening and reception services, legal support, guidance and support for other services, psychological support, help in moving away from the violent partner, work orientation, support for autonomy. In some cases the service is provided directly from the centre, in others, from the centre in collaboration with the services in the area, in others, the centre performs an addressing function. This paper presents the results of a multivariate statistical analysis that took into consideration, on the one hand, the services provide to women, on the other, the organizational strategies adopted to respond to users' needs. The statistical analysis was carried out in two steps: a correspondence analysis and after a cluster analysis that have identified six clusters of anti-violence centres corresponding to six different organizational models.

Keywords
Domestic Violence; Specialized services; Istanbul Convention; Shelters; Cluster Analysis.

1. Introduction

The Council of Europe Convention on preventing and combating violence against women and domestic violence (Istanbul Convention, 2011) recognize a key role to the "specialized services" for a well-timed and productive support for each victim of any act of violence that falls within the scope of the Convention. The types of support that these services provide are shelters and safe accommodations, telephone helplines, sexual violence services, legal support, counselling centres, and specific services for children as victims or witnesses.

This document and the explanatory report (1) underline that the support for victims of violence against women should be rooted in a "gendered understanding" and focused on "the human rights and safety of the victim". Victims' human rights include dignity, privacy and bodily integrity. There are a number of further aims/principles for support, such as avoiding "secondary victimisation"– this should include not blaming victims or making them feel responsible for what has happened to them.

For long time in Italy the available information on this kind of services was fragmented, based only on partial investigations. The national Agreements between Government, Regions and Autonomous Provinces (signed in 2014) established the minimum requirements of anti-violence centres and shelter houses to guarantee homogeneous criteria at national level and supported the implementation of a more systematic and extensive data collection.

In 2018 Istat carried out for the first time, in collaboration with the Department for Equal Opportunities (DPO) at the Presidency of the Council and the regions and the National Research Council (Cnr - Irrps), a national survey on the services provided by the anti-violence centres to victims of violence.

This paper presents the results of a multivariate statistical analysis that took into consideration, on the one hand, the services provide to women, on the other, the organizational strategies adopted to respond to users' needs.

2. Methodology

The survey is aimed at providing a representation at national level of the services offered and the features of public and private anti-violence centres in order to guide policy interventions.

The survey was carried out in the months of June - July 2018 and 281 anti-violence centres (AVC) were found, meeting the requirements of the 2014 Agreement and active on 31/12/2017. Of these, 253 completed the questionnaire, 24 provided no information and 4 were no longer active at the time of the survey. The response rate is 90%. The study do not take into account the 106 anti-violence centres and services that do not belong to the Intesa Agreement.

The descriptive analysis have pointed out differences between centres in terms of organizational aspects and type of services provided. These differences seem to be relevant according to regions and years of experience of the single centre. For these reasons all the variables connected with the two dimensions were synthesized with two different correspondence analysis. The three main factors of each analysis were used to group the centres through a cluster analysis.

The organizational dimension was represented by: opening hours, presence of the centre in the list of the national number against violence 1522, participation of the centre and type of authority involved in the violence network, type of funding, training course of workers and volunteers and type of training, use of a scale for the Risk Assessment, monitoring activity, presence of local branch, workers classes, percentage of volunteers.

The services dimension was represented by: year of the centre opening, type and mission of the managing authority, type of services provided, classes of women users.

3. Result

The Italian ratification law of the Istanbul Convention (Law 27 June 2013, n. 77) identifies as an objective that of having an anti-violence centre every ten thousand inhabitants. At 31 December 2017, 281 anti-violence centres were active in our country, equal to 0.05 centres per 10 thousand inhabitants. Those who participated in the survey are 253, while the remaining 28 did not respond to the survey.

In 2017 women that have turned to the anti-violence centres are 49,152, of which 29,227 have started a specific and individual project of exit from violence.

The centres mainly provide listening and reception services, legal support, guidance and support for other services, psychological support, help in moving away from the violent partner, work orientation, support for autonomy. In some cases the service is provided directly from the specialized service, in others, from the service in collaboration with the no specialized services in the area, in others, the centre performs the addressing function.

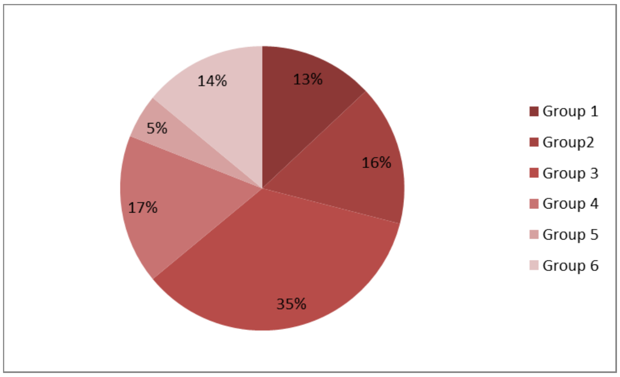
Most of the centres, 86%, work in a network with other territorial entities even though there are 11% of the centres that do not belong to any territorial anti-violence network. Almost all the anti-violence centres, 95%, adhere to the national helpline, the toll-free number 1522 against violence and stalking.

In 2017 there are approximately 4,400 operators that worked in the anti-violence centres, of which 56.1% were engaged exclusively in a voluntary manner. The professional figures that are more present in the centres, in line with the services provided, are lawyers, psychologists and reception operators.

93% of anti-violence centres provide for compulsory training for operators who are employed at the centre. In 85% of cases it is the centre itself that has organized training courses for staff.

In order to have a synthetic view of specialized services performance and organization, anti-violence centres were grouped on the basis of the results of a multidimensional analysis that took into consideration, on the one hand, their offer to women and, on the other, the organizational strategies adopted to respond to women's needs. Six types of centres emerged from the analysis.

Figure 1. The 6 groups of specialized services



Group 1	*The centres of medium-sized, integrated into the violence network*	12.6%
Group 2	*The medium-small centres, supported by the network*	*15.8%*
Group 3	*The centres with a strong and independent presence that also act together with the network*	*34.8%*
Group 4	*Small centres, but not isolated from the network*	*17.8%*
Group 5	*Small centres, which provide only basic services*	4.7%
Group 6	*The small centres, recently born, which provide few essential services, where the network does not exist*	*14.2%*

The centres of medium-sized, integrated into the violence network. This group includes 12.6% of the AVCs. These are centres that in 2017 followed between 100 and 200 women on their way out of violence. They are incorporated in the territorial anti-violence network - which also includes health services (local health authorities, hospitals) and social services (municipalities) - and are themselves providers - alone and together with other local authorities - of both the services envisaged by the agreement of 2014 and of the other services necessary for the realization of the personalized project of exit from violence.

The medium-small centres, supported by the network. The second group is composed of 15.8% of the AVCs; they are small centres with a territory of competence that is predominantly interprovincial. In 2017 they accompanied between 50 and 100 women, directing them, in most cases, towards the bodies that provide the services, since they are part of the territorial anti-violence network together with the Municipalities.

The centres with a strong and independent presence that also act together with the network. The third group is the most extensive (34.8%): it gathers fairly large centres in terms of both women followed and staff engaged in the centre. The operators are also specifically trained on the reception methodology and on foreign women while the training on the reception of women with disabilities is lacking. These centres offer a plurality of services (provided directly or by other territorial structures) and carry out prevention and information activities at the schools and training for law enforcement, lawyers and professional orders. They are the "historical" centres, whose promoter and manager is mainly private, who have been dealing exclusively with violence for more than thirteen years. Their rooting in the territory is also confirmed by the presence of a very articulated territorial anti-violence network, which includes the Municipality and health services but also law enforcement, prosecutors and courts. These centres receive both public and private funding and have almost total adherence to the requirements of the Agreement, both with respect to the definition, to the territorial network, and to the services offered.

Small centres, but not isolated from the network. The fourth group (17.8% of the AVCs) is composed of small centres, promoted and managed by private individuals, which do not deal exclusively with gender-based violence and do not have access to public or private funding. They are centres that do not focus on the training of their workers and, where they do, are supported by external figures. Also for the services provided they mainly have the role of accompanying towards the nodes of the territorial network to which they belong, often coordinated by the prefecture in which law enforcement agencies, prosecutors and / or courts, health services and social services participate. *Small centres, which provide only basic services.* The fifth group, the smallest one (4.7%), includes very small anti-violence centres, with a catchment area that does not exceed 40 women a year and not very accessible since they are open a few hours a day. They offer mainly listening and reception services, psychological and legal counselling, but they do not provide support activities to women's autonomy (job search, home search, removal from the perpetrator of the violence, emergency assistance, etc.), nor do they evaluate the risk of violence recidivism.

The small centres, recently born, which provide few essential services, where the network does not exist. The last group (14.2% of the AVCs) is composed of small-sized centres, both from number of users and employees point of view, born between 2014 and 2017. They are open less than 5 days a week but guarantee 24-hour telephone availability; they also set up a telephone number dedicated to the operators (law enforcement, first aid, social workers, shelter operators). These centres are mostly promoted by a public body, but the service is provided by a private entity that is still being trained in terms of gender-based violence; receive public funding; moreover, being of recent constitution, they are still little rooted in the territory and there is not an anti-violence network.

4. Discussion and Conclusion

The question "how do specialized centres perform?" is not an easy question. Data analysis, in fact, underlines that centres adopt different organizations models, that all can be suitable in order to achieve the goals of helping victims in getting out of violence.

The models are complex and have a geographical variability. Each region is autonomous and planned a different organizational model in order to respond to women's need. Regions have its own local legislation that affects shelters and anti-violence centres, their organization, funds management, requirements and monitoring tools.

The ISTAT survey is the first step towards harmonization, it represents for the first time the opportunity to have complete and standardized yearly data. The results of survey underline that shelters and specialized centres are not enough to answer to the victims' needs, especially in some area, and that their work, since it is based mainly on voluntary work, cannot be guarantee and continuative. The issue of the voluntary work is anyway very complex and discussed, since many NGOs believe that voluntary is the only way to guarantee the good quality of their work (mainly the relationship between women).

The analysis draw also attention to some best practices and to what play an important role such as the anti-violence network, the methodology adopted and the specific competencies of promoters and managers.

Firstly, it is very important to work in collaborative and synergic network with the other services, it is the best way to address the violence, to really help the victims to leave the partners or in general the violent situation and to gain the autonomy. Secondly the kind of adopted methodology is essential to help victims, a methodology based on the reciprocal relationship between women, according to the Istanbul Convention requirements, is indeed the successful strategy, as well as trained and capable workers. Thirdly, the characteristics of NGO that is responsible of the structure dedicated to women and experts is the key point to have good services.

References

1. AA.VV. (2005), Report of the expert group meeting, UN Division for the Advancement of Women, Expert Group Meeting Violence against women: a statistical overview, challenges and gaps in data collection and methodology and approaches for overcoming them, in collaboration with: Economic Commission for Europe (ECE) and World Health Organization (WHO) 11. 14 April 2005 Geneva Switzerland.
2. Bolasco, S., (1999) Analisi Multidimensionale dei Dati, Carocci Editore, Roma.
3. Council of Europe (2011). Council of Europe Convention on preventing and combating violence against women and domestic violence. https://www.coe.int/fr/web/conventions/full-list/-/conventions/rms/090000168008482e
4. Council of Europe (2011). CETS 210 Explanatory Report to the Council of Europe Convention on preventing and combating violence against women and domestic violence. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016800d383a
5. Liz Kelly, Mapping support services for victims of violence against women in line with the Istanbul convention standards. Methodology and tools, Strasbourg, November 2018
6. Holly Johnson, Bonnie S. Fisher, Veronique Jaquier (edited by), Critical Issues on Violence Against Women: International Perspectives and promising strategies, Routledge, New York, 2015
7. Sharmeen Farouk (2005), Violence against women: A statistical overview, challenges and gaps in data collection and methodology and approaches for overcoming them, paper presented at “Violence against women: a statistical overview, challenges and gaps in data collection and methodology and approaches for overcoming them”, UN Division for the Advancement of Women, Expert Group Meeting in collaboration with: Economic Commission for Europe (ECE) and World Health Organization (WHO) 11 . 14 April, 2005 Geneva Switzerland.

Future operational plan summary for census in The Kingdom of Saudi Arabia

Dr. Fahad S. Altukife[1], Fahad alfahead[2]
[1] General Authority for Statistics Saudi Arabia
[2] Mythologies & Operation of Population and Housing Census Project Saudi Arabia

Abstract
This presentation aims at demonstrating the programme of censuses that is conducted in Kingdom of Saudi Arabia, which includes Population and Dwellings Census, Establishment Census and Agriculture Census. However, in order to improve and further develop efficiency, as well as, prepare the Kingdom for an efficient move to an administrative based approach in 2030; the General Authority of Statistics (GASTAT) will use a Combined Census approach between traditional and register based census. This Combined approach will use administrative data supported by fieldwork. This new approach has required prior testing of administrative data such as the Saudi National Postal Address and data that is held by the National Information Centre (NIC). The presentation will further discuss the progress in the preparation and implementation of this new approach. This will include a thorough review of how the National Address system is being used as the framework for identifying buildings and households, as well as reviewing the results of the pilot studies initiated. The presentation will also discuss the implications for the fieldwork presented, and that includes, Computer Assisted Interviewing and the publicity and awareness censuses campaigns. Finally, the presentation will also discuss preparations for the release and publish of statistics from the 2020 round.

Keywords
GASTAT; Register Based Census; National Address System; National Information Centre

1. Introduction: Register-based Census

The Register-based Census is the dependency of administrative records in the different entities, The characteristics can be summarized in three main dimensions

- **Statistical dimension**
  - V' Using administrative records and collecting some data in the field.
  - V' Using self-enumeration and post-enumeration
  - V' Applying rules that directly detect errors while collecting data
  - V' Using addressing system
  - V' Applying the quality system through central operating rooms.

V' Sharing the form content with all clients.

V' The outputs of this census are used as a frame for surveys once the linkage processes are completed.

- **Technical dimension**

V' Use of buildings coordinates and addressing systems at the leveof cities and villages.

V' Data collection systems by using CAPI method "Computer-assisted personal interviewing". V' Synchronizing data from the National Information Centre and GASTAT databases while interviewing the head of the household by using tablets.

V' Following up the enumerator's progress in the field and calculating all the field quality indicators.

- **Data availability and Publication**

V' Instant results during the enumerating period and after completing the enumeration process.

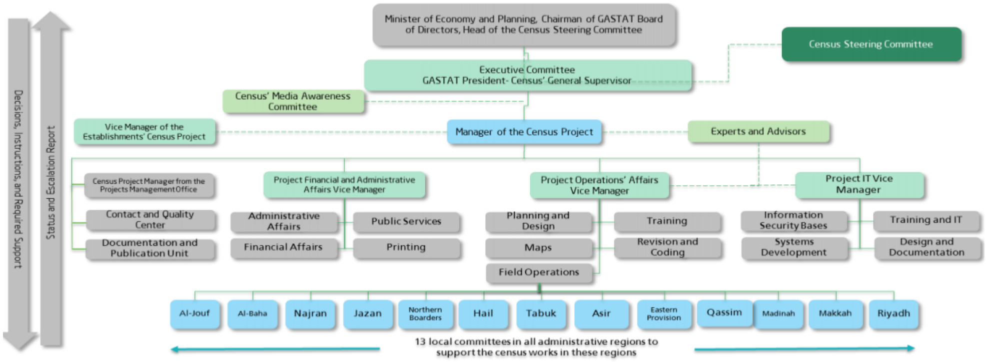
V' Flexibility of bulletin at any required geographical level.

V' Use of all means of bulletin (traditional and electronic), in addition to the use of business intelligence system.

2. Development of the Census Methodology

While conducting the Population and Housing Census in 1974, 1992, 2004, and 2010, Saudi Arabia relied on the method of traditional enumeration, in addition to the ICR technology. Portable devices were also used in the Establishments Census 2010, whereas tablets were used in the agricultural census, household and economic surveys, and Hajj statistics. Meanwhile, Saudi Arabia seeks to use the register-based method while conducting the Population, Housing, and Establishments Census 2020 to cope with the rapid development of the Information Technology field that facilitates the process of data flow and comprehension. Hence, it is necessary to change the method of enumeration from a traditional one into a more modern method, which is "the register-based method"

Governance



Linkage with the National Information Centre

What does linkage mean?

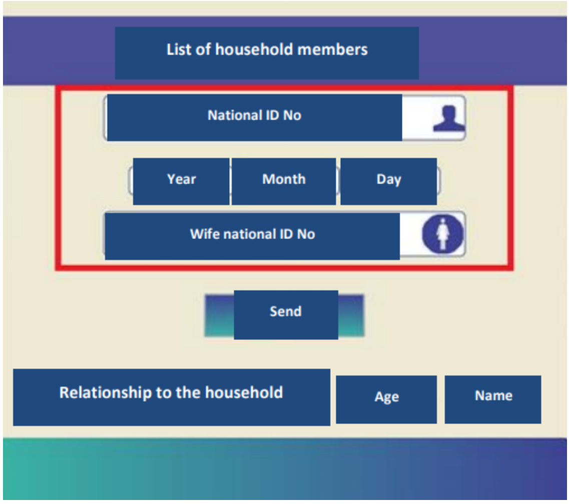
It is an electronic linkage between the General Authority for Statistics and the National Information Centre through private and secure network to exchange data.

Objectives of the linkage

- To access the records of individuals, households, housing, and establishments and link them with the individuals' national numbers.
- Unifying all the definitions and concepts, so that they can be transferred easily into statistical records that can be analysed and processed anytime.
- Use of register-based data in all household and establishments surveys.
- Founding a unified statistical number for establishments.

Household Register-based Data Request
Service inputs:
✓ Head of household ID
✓ Wife ID

Service outputs for all household members:
✓ Name
✓ Age
✓ Gender
✓ Nationality
✓ Relationship to the household head
✓ Marital status
✓ Educational status



Use of Addressing

To develop its coding system of buildings and housing units, GASTAT works on using a unified national coding system for addressing. This system includes more than (7 M) national addresses managed by Saudi Post. It relies on the Geographic Information Systems (GIS), whereas its structure was developed according to the international coordinates system with a 1-meter accuracy. Through this system, two layers might be built:

1. Postal codes that are based on geographical factors, population density, borders, and road networks.
2. A layer that provides an address for each location inside the postal codes.

Preparatory Stages of 2020 Census

Stage one: Updating populated areas' guidebook

- A guidebook was prepared for the populated areas according to Saudi Arabia's regions system; the data was processed and the administrative regions of Saudi Arabia was documented as well.

The number of the populated areas based on the administrative division is as follows:

Administrative regions	13	Municipalities	1527
Governorates	149	Populated areas	29591

- Estimate the number of dwellings for all populated areas and distribute them into statistical working areas.
- Confirm the guidebook connection to each national address in all populated areas and setting statistical boundaries through that connection.

Stage two: Updating cities

- Identify cities' boundaries clearly and accurately based on what the secretariats and municipalities have for the city.
- Estimate the number of dwellings and establishments for each national address.
- Distribute the cities, their neighbourhoods and blocks (national addresses) into statistical working areas.

Stage three: Coding of buildings, dwellings and establishments:

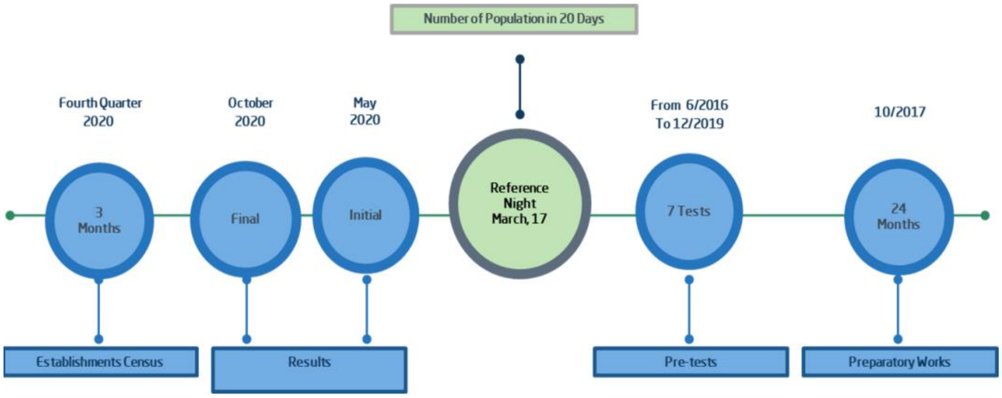
- Numbering the buildings, dwellings and establishments.
- Distribute observers' working areas into enumeration areas and identify the actual number of labour force.
- Prepare a complete framework for buildings, dwellings and households connected to the national addresses.

2020 Census's pilot-tests

2016: Testing the post address
2017: Testing the data of the National Information Centre.
2018: Testing the fulfilment of census's questionnaire.
2019 Q2: Testing the fulfilment of population and establishments census's questionnaire.
2019 Q4: Testing the fulfilment of population and establishments census's forms questionnaire.

Time Reference and Labour Force:

The time reference of the General Population and Housing Census was set on 17th of March, 2020. It will be implemented with the participation of more than (55000) supervisory and executive bodies and through (44000) field enumerators.



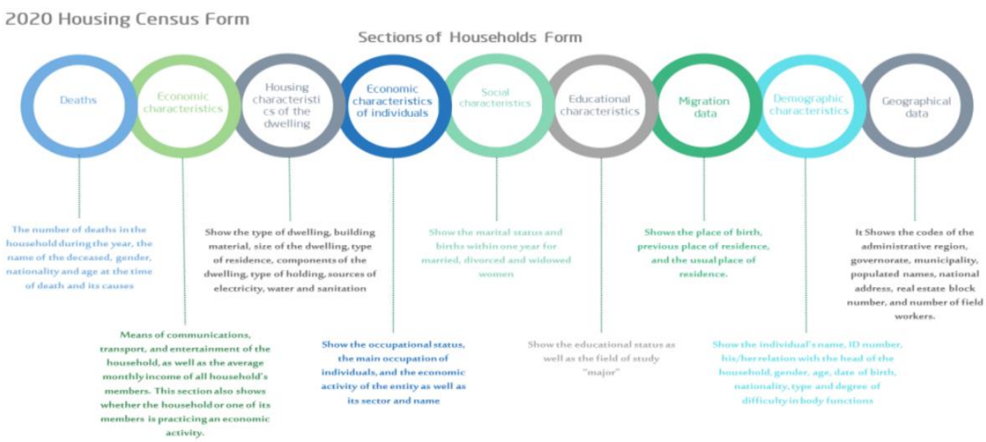
Data Collection

Enumerators will be using tablets to complete field data according to the household's post address available in the database of the National Information Centre after reviewing them with the head of the household and then sending these data to GASTAT's database.

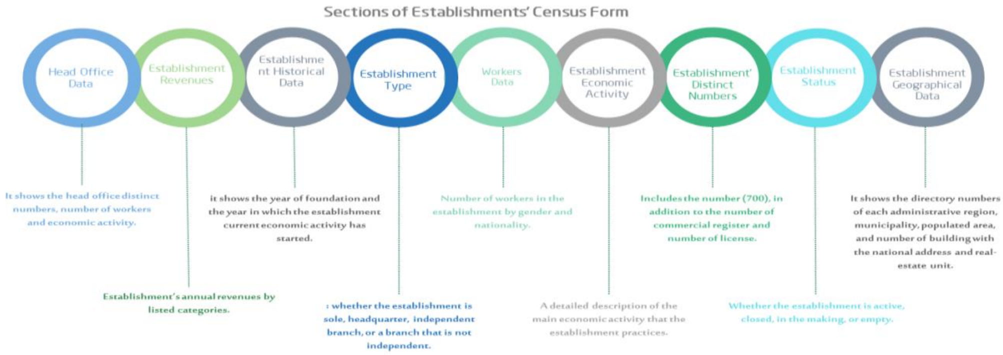
The census questionnaire will be filled out by

- Self-enumeration; the head of the household will be using an electronic link that is connected with the database of the National Information Centre.
- Transferring data of the field or electronic data from the database of the National Information Centre into GASTAT's database through sites visits.

Population and Housing Census questionnaire



Establishment Census questionnaire



3. Result

1- Conducting the Register-based Census through :

- Easy Data synchronizing.
- High Data security.
- Using geospatial technology in the enumeration stage.

Moving towards a Register Based Census in Spain

Jorge L. Vega Valle, Antonio Argüeso Jiménez, Marina Pérez Julián
Socio-Demographic Directorate, National Statistics Institute (INE), Madrid (Spain)

Abstract
Spain plans to carry out its first register-based census in the next 2021 round, becoming one of the biggest countries in the world using this approach. Since 1996, when the Population Register was created, census methodologies have evolved enormously and Spain has vastly increased the use of administrative registers in official statistics. In the last 2011 census, it was too soon to completely rely on administrative data, so a combined census was performed. Now, in 2018, 22 years after the creation of this register, INE faces this challenge with confidence. The availability of a variety of administrative registers and the fact that INE has access guaranteed by law to them, endorses this step forward. There are still some difficulties to overcome and a great effort is being made today, collecting any suitable administrative source, processing and integrating data, developing new IT tools, and of course, evaluating the quality of the whole statistical product. A complete population census test referenced to 2016 proved very satisfactory so we believe in the success of the 2021 census.

Keywords
Population and Housing Census; Administrative registers; Survey

1. Introduction

Once the 2011 Census was completed, 2021 Population and Housing Census project was proposed to be based primarily on the treatment of administrative records. In order to make a final decision about the methodology to be used in 2021, a work based on real data has been carried out. This exercise can be considered a pilot of the Population Census and it is referred to January 1st, 2016. Internally the information built is known as the 2016 preCensus File (2016-PCF). Additionally, the first viability study of the Housing Census, also based on registers, has also been conducted.

Regarding the Population Census, the results of the feasibility analysis of the 2016-PCF are conclusive: it is perfectly feasible a Census that is only based in the combination of administrative records. In certain and enclosed cases it would be necessary to complete the information with statistical procedures of imputation, but not to a greater extent than what is used in traditional censuses. The minimum level of requirement[1], which is the legal obligation established by the EU Regulation, is amply met for all the variables referred to population.

According to the Housing Census, developments are currently not so advanced. Anyway although it is proven that the EU regulation would also be complied with the strategy adopted, more efforts and time are needed to analyze the final quality of the product. In the least favorable scenario, the Housing Census would require a reduced and selective fieldwork operation in certain enumeration areas (those with poorest quality of information) of the national territory.

On the other hand, it is proposed to complete the Census project with a sociodemographic survey parallel to the census. The size of this survey could reach 1% of the total population and would allow to provide information useful for imputation of those variables that are not sufficiently well-covered by the Population and Housing Census. In addition, information of this survey has been largely demanded among users.

As a conclusion, 2021 Spanish Census can be considered “on the right track” and the main methodological doubts are cleared, but is not built yet. During the following years, an intense work of refinement in all the variables, especially in the Housing part, is required.

2. The Population Register (Padrón)

The main source as regards both population stocks and migration statistics in Spain is the population register, named Padrón in Spanish. Padrón is the official list of residents in each one of the 8,124 municipalities in Spain (as of 1st January 2018).

In Spain there are as many registers as municipalities. But there is a law, in force since 1996, integrating all these municipal lists into a single national database. There are also legal procedures to keep this database and the municipal files interconnected and updated on a monthly basis. This is made through the statistical office of Spain, INE. So, unlike other countries where the police or other administrative bodies are in charge of population registers, in the case of Spain, INE is the national institution that coordinates this single national population register.

Each month INE receives all the changes produced in every municipality. With this information INE performs validations and forwards these results to all the municipalities, to avoid duplications and also to include deaths, births or acquisition of Spanish citizenship that INE receives on a monthly basis from the Civil Register. Furthermore, all the consular offices of Spain throughout the world (around 250) are also connected to Padrón like the municipalities.

[1] More details in: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0712

According to the law, there is no restriction for registering in Spain in terms of legal situation. All people living or willing to live in Spain, regardless of their legal situation, have the right to be registered (this is actually an obligation) and they normally are, since being registered brings many advantages and no drawbacks.

For each person, Padrón contains these variables: gender, date of birth, place of birth (country, in the case of foreigners), nationality, educational attainment and national identity number. Foreigners in legal situation have the number of residence permit and people in irregular situation the passport number.

Because of all the details mentioned above and also because the quality of this source has improved every year since the beginning, Padron is the main element on which the Population Census is built. The 2016-PCF is created using Padron as of January 1st 2016 as starting point of information. This constitutes the backbone of the population file and dozens of administrative records of different nature have been linked to it in order to compose an individual information similar to the one that would be obtained with a questionnaire addressed to the households.

3. Two questions that must be answered by a Census

Every Census must give an answer to two questions. The first one, it should provide information about the number of people living in the country with a high level of geographical detail. But in addition, it must provide detailed characteristics (for example current activity status, year of arrival to the country or educational attainment) about all these people. From a graphic point of view, both questions can be seen as a large single file of two dimensions. The population figure will determine the number of records in the file (in vertical) and the number of variables will determine its amplitude (in horizontal).

	Identification	Sex	Age	Place of birth	Legal Marital Status	Current Activity Status	Educational Attainment	...	....	...
1										
2										
3										
...										
...										
...										

3.1 How many people live in Spain? The signs-of-life method

The original purpose of the censuses was to count the population. In Spain this main objective is not so important anymore, because due to the existence of Padrón, the uncertainty in knowing the population figures during the intercensal periods has disappeared.

In the same way that it has happened with the 2011 Census, 2021 Census takes Padrón as its basic element. The next step involves performing several additional controls:

- Verification that births and deaths are up-to-date as of the reference date.
- Analysis of the expiration date of Padron foreigners.
- Detection of "signs of life" of those people that appear in other files like Tax Agency or Social Security.
- Application of the 12-month criterion

Using all this information, a counting algorithm has been designed (still provisional and also susceptible of improvement) that provides a provisional population figure based on the signs-of-life method. As it can be seen in the following table, results obtained by the algorithm used in the 2016-PCF are very similar to those provided by Padrón:

	2016-PCF	**Padrón**	**Difference (PCF-Padrón)**
TOTAL	46,581,681	46,557,008	24,673
Spanish	41,972,389	41,938,427	33,962
Foreigners	4,609,292	4,618,581	-9,289

3.2 What are their characteristics?

If we take a look to the 2011 Spanish Population Census we can see that the individual questionnaire contained 22 questions, which can be grouped into eight blocks (basic demographic data, migrations, legal marital status and dwelling composition, studies, care provided, fertility, current activity status and mobility). Apart from that, the household questionnaire contained 7 questions about characteristics, facilities and equipment of the dwelling as well as the tenure status.

The main objective of the 2016-PCF was precisely to try to replicate the 22 individual questions; that is, to investigate to what extent a register-based census could at least contain the same level of detail as the 2011 census. It must be taken into consideration that not all the questions of the 2011 Population Census questionnaire have to be included in the 2021 census, in the same way that questions have been appearing and disappearing in each census along the years for many different reasons.

The first conclusion about the analysis of the 2016-PCF is that, regarding to the population, all the variables that are compulsory according to the EU regulation[2] can be obtained from registers. There are other variables[3] that are not compulsory but have a great census tradition or big interest among the community of users; in almost all the cases it is possible to obtain this information from registers.

In a very schematic way, below are the sources used in 2016-PCF for the most representative variables of the Census, as well as some of their most relevant characteristics:

Migration variables

Method and sources: Population register, but because it was not put into operation until 1996, it is also necessary to use the 2001 exhaustive Census.

General issues: In general, the quality of these variables is very high and all the information included in the 2011 Census questionnaire (years of arrival and previous places of residence) can be obtained from registers.

Relationship between household members (and other derived variables)

Method and sources: The starting point is the Padrón, which contains information about the people that live in the same household. In any case, information from the Tax Agency, Births and Marriages Bulletins, the Police Database (father's and mother's name are stored for every person) and previous Censuses are also used.

General issues: For each person it is investigated who is his/her father, mother, spouse and other relative. Next, derived variables related to the families, nuclei and structure of the household will be generated. In general, the search for father and mother (common surnames and certain restrictions of age and sex) is much simpler than the search for a couple. Regarding the couples, if they have a child living with them at home, they are much easier to detect. In other more complex cases (like cohabiting or same-sex couples without children), it will be necessary to establish a model of imputation that assigns couples to some situations of people living together.

Educational attainment

Method and sources: Padrón has information about this variable, but since its quality is not optimal it is necessary to use other sources like: several registers (diplomas, graduated people) from the Ministry of Education, information stored in the Unemployment Register, 2001 and 2011 Census.

General issues: Information of this variable will only be provided for people aged 15 or more. A person will have as value in this variable the highest educational level that has been observed in the different sources. Those residual cases that lack of educational level and also those ones that are kept in doubt among several values, are imputed taking into account the

[2] More details in:
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0543&rid=7

[3] For example some migration variables like year of arrival in the municipality or in the dwelling.

distributions according to sex, age and place of residence in an external source.

Current activity status

Method and sources: This is one of the variables that involves a greater number of sources: Social Security Registers, Unemployment Registers, Public Aids Database, Mutualities Registers, Register of Retired Civil Servants, Registers of Students, Tax Agency information, 2001 and 2011 Census.

General issues: Information of this variable will only be provided for people aged 15 or more. Using all these information, it is normal to have, in some cases, conflicts among data sources. In order to solve this issue, priority rules based on the recommendations of the United Nations and the European Regulation for Censuses have been used. It is normal that some people (for example women aged 55 or more) do not appear in any of the sources used and they will be classified as "others" inside the group Outside of the labour force. The results obtained for this variable in the 2016-PFC are very similar to those one of the LFS although it is still necessary to refine the category of the unemployed so that it resembles the ILO recommendation.

4. A new quality framework

One of the main novelties of the 2021 Census will be the inclusion of a new mechanism that enables users to evaluate the quality of each one of the census variables and that will also enable INE to make better decisions. The idea is to create for each variable (for example: legal marital status, educational level attained, etc.) a new one that would store information indicating the method or type of source used to provide the value for every person.

The procedure will consist in the creation (for each Census variable) of a new derived variable with several categories that will take into account various factors reflecting if it is a direct, indirect source or if the information has been imputed.

If we focus on the way we obtain the cell estimation, we will be able to quantify quality in a two dimensional basis: quality along a specific variable and quality in terms of each person.

An analysis by columns (variables) across people, allows us to detect for every variable involved what is the percentage of records provided by different sources or methods and the percentage of imputed records. This information helps us to detect the quality of the sources.

If we concentrate on rows (people) we can identify those records with the poorest quality level: those ones that have missing values or imputed information in several variables. It is very plausible to identify profiles of people with missing information that are difficult to estimate by administrative records, such as foreigners or people living in deprived areas.

With this information our users will have at their disposal more information available that will be useful to understand better the benefits of supporting the census information with administrative registers.

5. The Housing Census

The Housing Census consists, first of all, in an exhaustive quantification of all the dwellings (occupied and unoccupied), and secondly in a characterization of them and of the building where they are allocated. Analogously to what has been mentioned with the Population Census, it is planned to build an exhaustive microdata Housing file that contains all the dwellings and their characteristics, using the administrative sources available.

Likewise Padron is the main source of information for the Population Census, we could say that Cadastre will be the main source of data for the Housing Census. It has the great advantage that contains a unique identifier for all dwellings (cadastral reference) and that all the information is georeferenced. On the other hand, the fact that it is used for tax purposes at municipal level, has as one of the main consequences that the quality of the information stored is very high.

One of the main drawbacks of Cadastre is that it is not linked to the Population Register, so in some cases we do not know which household of Padrón corresponds to each housing unit of Cadastre (anyway some tests have been done with some provinces and the results seem to be quite encouraging). On the other hand, Cadastre contains information on certain census variables like period of construction or useful floor space, but not all, as for example the variable tenure status of households.

During the following months an intense job of data-linkage between sources will be done, but not any fieldwork operation is foreseen. In the end, it should be available an integrated system where people from the Population Census are assigned to housing units that belong to the Housing Census. The coherence between both products should not have any cracks.

In the same way as in the section of the population census, the situation for one of the most complicated variables of the next Housing Census is presented here schematically.

Tenure status of households

Method and sources: The main data sources are Tax Agency and Cadastre. In the Tax Agency all the declarants (around 65% of households) include the cadastral reference of their usual residence and also any other properties they own. On the other hand, Cadastre also contains information about the ownership of each housing unit.

General issues: Due to the fact that the proposed data sources are not totally exhaustive, it is necessary to carry out some imputation mechanism based on a survey. In addition, it is also pending to include certain improvements (for example information about rented dwellings that is stored in the declarations of the Tax Agency) in the data sources used that would help to increase the quality of this variable.

6. Other elements

6.1 Inclusion of other variables

The use of administrative sources opens the door to the incorporation of new variables, either because of the availability of new sources, or because of the possibility to exploit them in another way.

A traditional Population and Housing Census uses the questionnaire as the instrument for collecting information. In this situation, the main constraint when designing any census questionnaire is that it should be shorter and easier to understand better, because this will achieve a better quality of response. However, in the case of a register-based Census this limitation disappears; therefore, it can cover concepts that are difficult to include in a questionnaire and in many cases very demanded. Some examples of the variables that could be incorporated in the 2021 Spanish Census are:

- Information about property of vehicles, from Traffic institution
- Fertility, number of children born in the previous years
- Data on other owned dwellings, based on Tax Agency information and Cadastre

On the other hand, one of the strengths of a register-based census, which has not been sufficiently exploited yet, is the possibility of incorporating new context variables, that do not refer to a concrete individual (both because of legal restrictions or because of quality issues for individual data), but to the average or total of his/her enumeration area. Some examples of variables of this type that we could incorporate are: income level, electrical consumption or presence of green areas.

6.2 The survey

Maybe one of the main points of improvement of the Census is not the quality of the file that can be produced, but the shortcomings of information that it is not available in the different records that we use. The lack of certain information in relation to commuting (not mandatory but highly demanded) is an important lack. Furthermore the need for more demographic characteristics like second generation of immigrants, more reliable information about household members and better information about households is another handicap.

For all these reasons, and in order to complete the Census project, we are planning to conduct an ad-hoc survey that would target about 1% of households and would answer these and other questions.

7. Conclusions

A detailed analysis of the construction of variables for the 2016-PCF allows us to conclude that it is already almost a complete Population Census. If we consider the planned improvements that it will be included in the future, all the Regulation requirements would be amply met. 2016-PCF leaves out of doubt the viability of the registers as the main data source in 2021 Census. Spain would become one of the most populated countries (perhaps the largest) in the world with a register-based Census

Regarding the Housing Census, the situation is not so conclusive. The tasks for building the framework are still ongoing. Anyway it can be stated that a complete Housing Census based on registers will be also available with the same or higher quality than using traditional method.

Prospects for the development of higher education based on Big Data technology

Motoryn Ruslan[1,2], Motoryna Tetiana[1], Prykhodko Kateryna[2]
[1]Kyiv National University of Trade and Economics, Ukraine
[2]Taras Shevchenko National University of Kyiv, Ukraine

Abstract
The educational system continuously generates and accumulates a considerable amount of data. The issue of systematic work with these data by a wide range of subjects of education today can be called one of the most significant. Big Data technologies can be a powerful tool for transforming learning, to rethink approaches, to reduce costs and improve the efficiency of the educational system itself. Actual is the task of describing the technology of operating with big data, aimed at the development of educational systems through the identification of formed patterns in the education system. The authors determined the methodological aspects of the implementation of technology of the operating big data in higher education. Leading approach in this case serves simulation of the educational system with the inclusion in it of regulatory measures based on its analysis of the formed patterns. As a result of the research conducted, the authors of the article described the Big Data technology as a means of developing educational systems, including the distinguishing features of the Big Data technology, the system management processes, the direction of data collection in education are structured, and the properties of the collected Big Data base in higher education are defined.

Keywords
higher education, big data, data personalization, big data management.

1. Introduction

Over a long period, an enormous amount of information is accumulated in educational systems of educational institutions about various aspects of the educational process: students, their academic performance and attendance, teachers and their scientific, educational, as well as administrative activities, teaching materials (texts of lectures, workshops, audio, video recording, etc.), etc. These data should be stored, processed and analysed for the effective management of the learning process. Thanks to advances in information technology, big data can now be accumulated, analysed, managed in the field of education. To do this, it is necessary to study the existing data storage and processing technologies used by the leading countries of the world in education.

Research on the use of Big Data in the education system is quite fragmented.

An important aspect of Big Data research is the infrastructure of the data collected. Therefore, F. A. De Almeida Neto and A. Castro developed a model where the data created from the interaction between users and the platform itself is selected, collected and stored in local databases considering the online platforms that host educational events, [1]. Then local databases are collected and grouped into a global database.

Some aspects of this research area are issues related to the achievement of educational results. EDM (Educational Data Mining) is described as a means of increasing the effectiveness of e-learning. So, M. Nasiri, B. Minai, F. Wafai considered a model for predicting academic success through monitoring and supporting first-year students [2].

Other issues of the research are aspects of Big Data related to the interaction of subjects of education. G. Mobasher, A. Shavish, O. Ibrahim describe the structure of a large database in education, which contains demographic data of students, psychological characteristics of students, lecturers and parents [3]. V. Tem described the approach to the organization of collaborative learning, allowing to identify educational patterns based on a diverse set of educational online resources [4]. S. Dwivedi, V. S. K. Roshni, at the basis of analysts, describe the technology for students to select the most appropriate elective courses [5].

2. Methodology

The processing of large archives and large data streams requires new technologies, among which a special place is occupied by Big Data technologies.

The term Big Data refers to the large and complex data sets, which can be structured or unstructured, and occupy a very large amount of disk space.

Features Big Data can be described by «5V» rule:

- 1V (volume): the amount of physical data is significant. Large amount of data mean information about a large number of students and thousands of educational institutions. These data are accumulating and provide information that can be used to effectively manage the learning process.
- 2V (velocity): speed data collection and processing speed results фку relatively high; for example, data on marks for a lesson are entered no later than the end of the day they are received; the teacher after entering the data almost immediately can get acquainted with the analytics performance. The rate of change of big data allows to interactively monitor the learning process and respond to any changes in the learning process in a timely manner. The use of interactive tests allows teachers to identify students who give incorrect

answers to test questions and to provide them with the necessary content for the study and better assimilation of educational material in real time.

- 3V (variety): variability of processing algorithms for different types of collected results; for example, homework results can be presented by gender, age, health group, etc.
- 4V (veracity): high reliability of the data collected, allowing to formulate representative results; for example, after conducting a national study of the quality of education, it can be concluded that 4th year students have significantly higher marks than 1st year students.
- 5V (value): the value of the accumulated data should be based on the possibility of formulating useful multi-aspect dependences of the education system on their basis. For example, it can be noted that when students move from the first year to the second, the number of excellent and good students decreases, however, there is an equal change in the proportion of assessments, which may indicate a gradual complication of educational material; on the other hand, the number of “C students”. Big data in education allow teachers to receive a variety of information about the level of training students, the assimilation of educational information, the oversight and Labs. Another important problem of education is the identification of new, sometimes hidden, relationships in big data, new knowledge. For this data mining methods can be used to improve the organization of the educational process and improve its management efficiency.

Another important area of research are issues related to internal interaction. Predicting academic performance is one of the key themes of research in the field of Big Data in education. Assessing academic achievement is a difficult task, as student performance depends on various factors. The relationship between performance parameters and factors involved to predict performance in complex non-linear relationships, so the data collection areas should be inclusive. Big data management allows processing of information for the analysis of the key indicators of educational effectiveness. It is also important to note the benefits of using big data for administrative staff of higher education institutions. Academic performance, attendance, scholarships and other personal information about students is subject to continuous collection, processing and analysis. Working with this amount of data requires considerable effort. Automation of the already routine work will lead to savings of financial and human resources.

Methods of big data allow to form the relationship between types of education and assess the progress and potential of the student throughout his educational history. Such an approach can facilitate the formation of individual educational itinerary taking into account the characteristics of each student. There are five main types of data in the field of education:

- personal information;

- data on the interaction of students with electronic learning systems (electronic textbooks, online courses);
- data on the effectiveness of educational materials;
- administrative data;
- forecast data.

For structuring the management processes of Big Data in education five interrelated groups of processes can be allocated (fig. 1):

1) definition of the purpose and objectives of study;
2) selection of information sources, data acquisition procedures, information processing algorithms;
3) organization of data collection into single database;
4) analysis of the data, the formulation of ways to present the results;
5) development of practical regulatory measures.

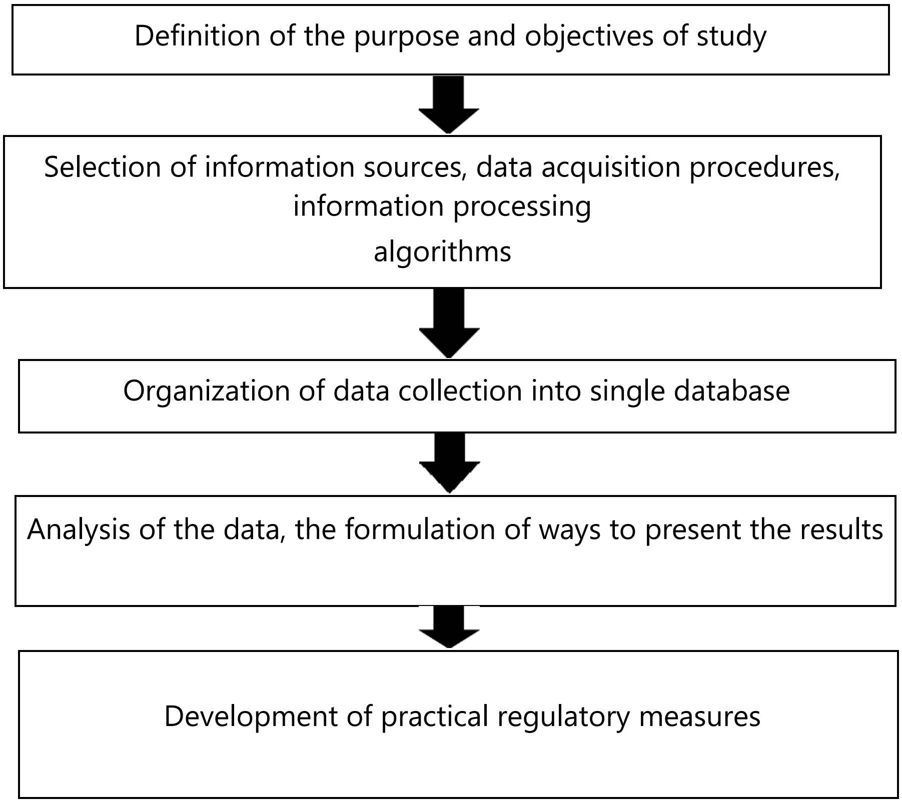


Fig.1. The sequence of the Big Data management process in higher education

How to evaluate the effectiveness of higher education? This requires a methodology for assessing its quality.

For each student, lecturer, education institution management - the assessment of the effectiveness of education is made up of different factors, with different influences and significance. Accordingly, our task is to create such a tool that will give an indicator sufficient and necessary for further decision making to each of the users separately and together.

The first, basic stage of solving the problem is to provide the user with coefficients brief and understandable, that is, estimated in specific quantities. Here is a difficult but important task - to find averaged indicators that will be important for everyone, for example, a lecturer, despite the fact that each lecturer has his own approach to learning and, accordingly, his own criteria for evaluating the effectiveness of achieving a particular result.

Therefore, first of all, it is necessary to determine the basic indicators on which it is further necessary to build models for assessing the current state and forecasts of further development.

The basic indicators are (fig.2):

- student assessment;
- truancy of student;
- the quality of the tools provided to achieve success (rating of lecturer, universities, educational materials).

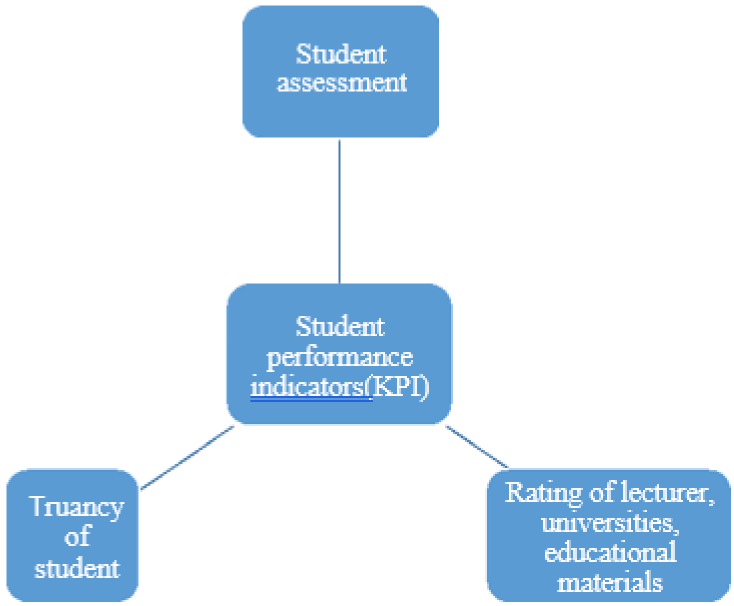


Fig.2 Student performance indicators

The main difficulty in the implementation that we have encountered is the number of systems that are sources of data and the volume of events and information in them. Relatively speaking, here were both sources of structured information (olympiad results; truancy; grades), and semi-completely unstructured information (training program; Internet events; names and structure of subjects; clicks on the electronic board).

Therefore, only the stage of data collection, integration of services (and often - their development from scratch) and all relational and non-leverage databases requires a large amount of time and, moreover, is continuous, because when a single result appears, a request for new developments immediately arrives.

3. Result

Let us give an example of the Big Data management process in higher education on the example of the Kiev National University of Trade and Economics (KNTEU).

1. Purpose - expert and analytical assessment of the human resources of KNTEU.

Tasks:

- identify the age characteristics of lecturers;
- determine the presence of objective vacancies or oversaturation of lecturers;
- assess the qualifications of existing lecturers.

2. Selection of information sources, data acquisition procedures, information-processing algorithms:

Selection of information sources:

- annex 1. Regulation of KNTEU "Rating assessment of the activity of educational research of employees of the university"
- the annual form of statistical observation 2-3NK "Report of the university at the beginning of the 2018-19 academic year";
- questionnaire form of student evaluation of lecturers.

The procedure for obtaining data:

- the selection of the necessary information contained in the Annex 1. Regulation of KNTEU "Rating assessment of the activity of educational research of employees of the university"
- selection of the necessary information contained in the form of statistical observation 2-3NK;
- conducting a questionnaire of student in the form developed by KNTEU.

Algorithm of data processing.

The analysis of existing personnel problems conducted in the context of the subject taught. The questionnaire of lecturer's assessment by students contained detailed information about lecturers.

According to the results of the study of the age structure, we construct the distribution of lecturers by age compared to the normal distribution. Normal distribution contributes to the stable functioning of the higher education system while maintaining a balance between the possibility of development and the preservation of tradition.

3. Collection and structuring of data

The stage involved the collection and structuring of data submitted to KNTEU, as well as the export of data collected in the form of 2-3NC.

4. Analysis of the data obtained the definition of ways to present the results, the fixation of patterns. Comparison of the distribution of lecturers by age compared with the normal distribution in the context of the subject taught. Construction of "anomalous" distribution graphs.
5. Development of practical measures to regulate processes

After the adoption of the final report - transmission of negative trends to the management bodies of

KNTEU, approval and implementation of an action plan for personnel policy for 2018-2021.

Now are implemented of student performance indicators:

- monitoring current academic progress and identification of deviations (got a bad mark; got a bad mark after an illness; got a bad mark, and the whole group got good marks);
- identifying features of training in the university (favourite subjects that work well, which does not skip);
  - identifying what types of activity are good and bad (written work, the answer is at the blackboard);
  - building a circle of interests, based on visits to classes in subjects.

4. Discussion and Conclusion

Methods of big data allows forming the relationship between types of education and assessing the progress and potential of the student throughout his educational history. Such an approach can facilitate the formation of an individual educational route, taking into account the characteristics of each student.

The site OnlineUniversities identified ten areas in which higher education will change under the influence of big data.

1. The method of working in groups will change: for example, at one of the courses at Harvard, students with different answers are paired so that they can come to a single decision, defending their point of view;
2. The learning experience will become more personal: technology allows to individually selecting not only courses, but also homework and careers;
3. Students will receive more recommendations: now the programs are able to predict how well the course will be completed, even before it has begun;
4. Data will play an important role in choosing a university: it assumed that applicants would not even have to submit applications, because the robots will select the best places for them;

5. Marketing will change: academic institutions will be able to learn in advance about promising candidates;
6. More students will get to the end of education: now technologies identify students at risk and help them;
7. Management of universities is optimized: institutions of different types will be able to receive more accurate recommendations;
8. Lecturer's will be able to help better lagging students: the programs will let you know exactly which areas have problems;
9. It will be easier to choose a career: digital portfolios will tell your whole story instead of you;
10. Data analysis will be a key element in the life of universities: using data analysis at all levels, the administration will be able to make decisions that are more effective.

References

1. De Almeida Neto F. A., Castro A. (2017) A reference architecture for educational data mining // Paper presented at the Proceedings Frontiers in Education Conference, FIE. October. - P. 1-8. doi: 10.1109/FIE.2017.8190728.
2. Nasiri M., Minaei B., Vafaei F. (2012) Predicting GPA and academic dismissal in LMS using educational data mining: A case mining // Paper presented at the 3rd International Conference on eLearning and eTeaching, ICeLeT. P. 53-58. doi: 10.1109/ICELET.2012.6333365.
3. Mobasher G., Shawish A., Ibrahim O. (2017) Educational data mining rule based recommender systems // Paper presented at the CSEDU 2017 - Proceedings of the 9th International Conference on Computer Supported Education. № 1. - P. 292-299.
4. V. Tam, E. Y. Lam, S. T. Fung, W. W. T. Fok, A. H. K. Yuen (2016) Enhancing educational data mining techniques on online educational resources with a semi-supervised learning approach // Paper presented at the Proceedings of 2015 IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2015. P. 203-206. doi: 10.1109/TALE.2015.7386044.
5. Dwivedi S., Roshni V. S. K. (2017) Recommender system for big data in education // Paper presented at the Proceedings - 5th National Conference on E-Learning and E-Learning Technologies, ELELTECH. doi: 10.1109/ELELTECH.2017.8074993

Improvements on a parallel/distributed algorithm for bootstrapping

Rub´en Carvajal-Schiaffino
Departamento de Matemática y Ciencia de la Computación
Universidad de Santiago de Chile

Abstract
Bootstrapping is a statistical technique that is very expensive from a computational point of view. For this reason, parallel versions have been developed, typically implemented on multicore architectures. However this approach can be limited by the amount of cores and the bottleneck produced by the simultaneous access to the memory. We present a comparison between parallel implementations in shared memory, private memory and one based on the use of massive parallel computing by means of the use of Graphics Processing Unit (GPU).

Keywords
Bootstrap; Parallel Algorithms

1. Introduction

In the resolution of any problem it is necessary to consider what the resources are available and therefore it is essential to choose the most appropriate tools. In the case of the implementation of algorithms, the resources are the processing time and the storage space. Considering these two resources, the tools are: the programming language in which the algorithm and the computation model are implemented. The programming languages can be classified according to their execution mode: compiled (C, FORTRAN), interpreted (R, MATHLAB) or running with virtual machine (JAVA). Regarding the computer model this can be sequential, parallel or distributed.

Since the bootstrap method requires a large number of operations, it is best to choose a language compiled as C. Regarding the computational model for bootstrap in addition to using the sequential model, it is possible to perform a parallel implementation. Parallel computing models can be classified into those that use shared memory where several processors can access a common memory such as *pthreads*[1] or *GPU's*[4] and those that use private memory where the processors collaborate with each other through message exchange [2,3].

In this paper we present a comparison between the sequential implementation of the boot-strap method to calculate the Pearson correlation coefficient, the parallel implementation using the cores of a computer and the parallel implementation using GPUs.

The rest of the article is organized as follows: in the methodology section the boostrap algorithm is presented to calculate the Pearson correlation and a calculation is made of the space required for its execution. Also two parallel implementations are presented. In the results section the execution times of the four implementations are shown and in the discussion section and conclusions the limitations and extensions of the implementations are presented.

2. Methodology

The number of operations required in the bootstrap algorithm for calculating the Pearson coefficient is $n\, timesb$ where n is the number of observations of each of the two variables and b is the number of iterations boostrap . We do not consider the number of operations required to calculate a Pearson coefficient since this calculation is done together with the reading of the elements of the sample generated. Regarding the space required, this is $2n + b$, where $2n$ is the space occupied by the observations and b is the space occupied by the Pearson coefficients calculated. This expression must be multiplied by 4 if the data is stored as real numbers of simple precision (float type in C language) to obtain the number of bytes required.

Table 1 shows the sequential version of the algorithm. Lines 2 and 3 can be combined in a single cycle for n operations. The parallel version can be

Table 1: Sequential Bootstrap

Input:	X, Y Set of n observations nbi Number of bootstrap iterations
Output:	P Pearson Coefficients
1	**for** $i \leftarrow 1$ **to** nbi **do**
2	$X', Y' \leftarrow$ **SampleFrom**(X, Y, n)
3	$Pi \leftarrow$ **CorrelationCoef**(X', Y')
4	**endfor**
5	**return** P

Table 2: Parallel Bootstrap

Input:	X, Y Set of n observations nbi Number of bootstrap iterations p Percent of the confidence interval cn Number of Cores
Output:	P Pearson Coefficients
1	**For each** core 1 . . . cn **do parallel**
2	**for** $i \leftarrow 1$ **to** $\left\lceil \frac{nbi}{cn} \right\rceil$ **do**
3	$X', Y' \leftarrow$ **SampleFrom**(X, Y, n)
4	$Pi \leftarrow$ **CorrelationCoef** (X', Y')
5	**endfor**
6	**endfor**
7	**return** P

easily implemented when executing each processor a bootstrap iteration. The ideal scenario would be to have as many processors as bootstrap iterations, in this case the algorithm make n operations. If the number of available processors is less than the number of bootstrap iterations, then they must be distributed among the processors. Let $i_1, i_2, dots, i_c$ be the number of iterations that each processor will perform, then $i_1 + i_2 + dots + i_c = b$. Then the number of operations of the parallel algorithm will be $n \times \boldsymbol{Max}\{i_1, i_2, \ldots, i_c\}$.

We describe below the implementation of the parallel algorithm using the pthreads library. In this model all the data are already stored in global variables or local variables of the functions that each processor executes are in the memory of the computer. This produces a bottleneck because the number of registers of the interface that communicates the processors with the memory of the computer is limited. As a way to reduce the impact of the bottleneck produced by the interface, a distributed version has been implemented using the library *mpich*[3] that runs on a cluster of computers each with several cores, in this case each process is executed in private memory.

In the case of implementation with GPU, the first feature to consider is that the card has its own memory and the code that its processors execute can only access this memory, so it is necessary to make data transfer options between the computer's memory and the memory of the card. The second feature is that the code executed on the card has different characteristics than the code executed on the computer. Since the functions executed on the card are not capable of generating pseudo-random numbers, it is necessary to build a sequence of these numbers with *n timesb* elements. A first alternative is to build the sequence on the computer and then copy it to the memory of the card, but fortunately there is a library called it curand that generates pseudo-random numbers directly on the card.

Input: X, Y Set of n observations
b Number of bootstrap iterations

Output: P Pearson Coefficients

GenPseudoInCard(*prand, n* x *b)*
CopyToCard(X, Y, n)
PearsonInCard <<< b >>> (X, Y, *n, prand,* P)
CopyToHost*(P, b)*
return P

Figure 1: GPU Parallel Algorithm

Figure 1 shows the algorithm of the parallel version that uses the graphics card. It is necessary to take into account that the memory space required in the card is $(2n + b + n \cdot b) \cdot 4$ bytes which can be a limitation.

```
__global__ void ComputeRhoInDev(float *v1, float *v2, float *prand,
                                float *rhodev, unsigned int n) {

    unsigned int i, j, index;
    float sx, sy, sx2, sy2, sxy, ro;

    sx = sy = sx2 = sy2 = sxy = 0.0;
```

```
for (i = 0, j = blockIdx.x * n; i < n; i = i + 1, j = j + 1) {
    index = prand[j] * n;
    sx = sx + v1[index];
    sy = sy + v2[index];
    sx2 = sx2 + v1[index] * v1[index];
    sy2 = sy2 + v2[index] * v2[index];
    sxy = sxy + v1[index] * v2[index];
}
ro = (sxy - sx * sy / (float) n) / sqrtf((sx2 - sx * sx / (float) n) * (sy2 - sy *
    sy / (float) n)); rhodev[blockIdx.x] = ro;
}
```

Figure 2: GPU Code

Figure 2 shows the code executed on the GPU. The variable *blockIdx.x* is the identifier of the process that executes the code and from this it is possible to access the corresponding pseudo-random number portion and the position of the array where the calculated Pearson coefficients are stored.

3. Result

In order to compare the different versions of the bootstrap implementation to calculate the Pearson correlation coefficient, the maximum size of the data to be analyzed that could affect the graphical card was determined, which corresponds to two variables with 49,000 observations each and 9,999 iterations were performed. boostrap On the computer with the graphics card, the sequential version was executed, the parallel version implemented with the library *pthread* and the parallel version with the GPU. The details of the hardware and software used are shown in Table 3. In the cluster with 24 computers, each with 4 cores, the distributed version implemented with the library *mpich* was executed. The hardware and software details of the cluster are shown in table 4.

Table 3: Hardware & Software of the Computer with GPU

Processor	intel Core i7-4790 CPU @ 3.60 Ghz x 8
Graphic Card	GEForce GTX 980 Ti – 2 Gib of Memory – 2048 cores
RAM Size	7,5 Gib
Operating System	Ubuntu 16.04 LTS - 64 bit
C Language	gcc 4.8.4

Table 4: Hardware & Software of the Cluster

Processor	(intel Core i5-3470 CPU @ 3.20 Ghz x 4) x 24
RAM Size	3,9 Gib
Operating System	Ubuntu 16.04 LTS - 64 bit
C Language	gcc 4.8.4

The execution times of the different implementations are shown in tables 5 and 6.

Table 5: Execution Time of Sequential Version and pthread Version

Version	Time (Seconds)
Secuential	4
Parallel - 2 Threads	7
Parallel - 3 Threads	8
Parallel - 4 Threads	8
Parallel - 5 Threads	7
Parallel - 6 Threads	8
Parallel - 7 Threads	8

Table 6: Execution Time of Distributed and GPU Version

Version	Time (Seconds)
Distributed - 96 Processes	0.047
GPU - 9,999 Processes V1	0.14
GPU - 9,999 Processes V2	0.000013

Version V1 of the implementation with GPU considers the transfer time of the Pearson coefficients from the card to the computer memory, whereas version V2 only considers the processing time on the card. In the distributed version, the maximum number of bootstrap iterations that some processors perform is 15 and the reported execution time is only local processing.

4. Discussion and Conclusion

We have presented the parallel and distributed implementations of the bootstrap method for the calculation of Pearson correlation coefficients and we observe that the execution times obtained depend strongly on the available infrastructure, which in this case is of accessible cost.

The implementation with the lowest performance is presented by the implementation with *pthread* due to the limitations of the number of available cores and in addition to the bottle-neck of the interface with the computer memory. In the case of distributed implementation, it provides adequate performance, and it can process much larger problems. The imputation with GPU presents the best performance, but has the limitation of the memory capacity of the graphics card.

Today it is possible to implement hybrid architectures with computers with more than one graphics card and/or computer clusters with these devices, so it is possible to solve instances of bigger problems with adequate execution times.

References

1. Posix Threads. https://computing.llnl.gov/tutorials/pthreads/
2. Bruck J., Dolev D., Ho C., Rosu M., and Strong R. (1997). Efficient Message Passing Interface (MPI) for Parallel Computing on Clusters of Workstations. Journal of Parallel and Distributed Computing. 40(1) 19:34.
3. MPICH homepage. https://www.mpich.org/
4. Nickolls J., Buck I., Garland M. and Skadron K. (2008). Scalable Parallel Programming with CUDA. ACM Queue, vol. 6 no. 2.

IMF Task Force final report on Special Purpose Entities (SPEs): Compilation issues and implementation in Luxembourg[1]

Paul Feuvrier
Banque centrale du Luxembourg.

Abstract

The final report of the IMF Task Force on Special Purpose Entities (SPEs) provides a comprehensive definition, which refers to both Business Statistics ("no or little employment") and Balance of Payments ("transactions with non-residents") concepts. Yet it also takes into account criteria related to the competitiveness of the jurisdictions hosting SPEs: The SPEs are institutional units established for specific purposes including for instance to "grant their owners access to capital markets or sophisticated financial services". How to identify institutions based on such non-standard criteria? The IMF report fortunately also includes a typology, which is an excellent starting point to go through the various Luxembourg SPEs business models. Luxembourg SPEs are mostly Captive Financial Institutions. The most commonly observed pattern is that Multinational Enterprises (MNEs) do not establish a single but instead a network of interrelated resident entities, characterized by complex ownership chains, to channel funds through Luxembourg. Some of these entities issue debt securities on behalf of the group, and most only have loans and equity (vis-à-vis affiliates of the group) on both sides of the balance sheet. As a consequence of this "channeling" function, the contribution of each MNE and therefore to SPEs in general, to Luxembourg's net international investment position (i.i.p.), should theoretically be rather small.

Keywords

External Statistics, Financial Account, Foreign Direct Investment, Portfolio Investment

1. Introduction

The second section of this note delineates which Luxembourg institutions are SPEs in accordance with the IMF Task Force's definition and the third section in which category they fall (typology). Section 4 presents BCL's survey strategy, editing system and the functional category of some financial

[1] This presentation should not be reported as representing the views of the BCL or the Eurosystem. The views expressed are those of the author and may not be shared with other research staff or policymakers in the BCL or the Eurosystem.

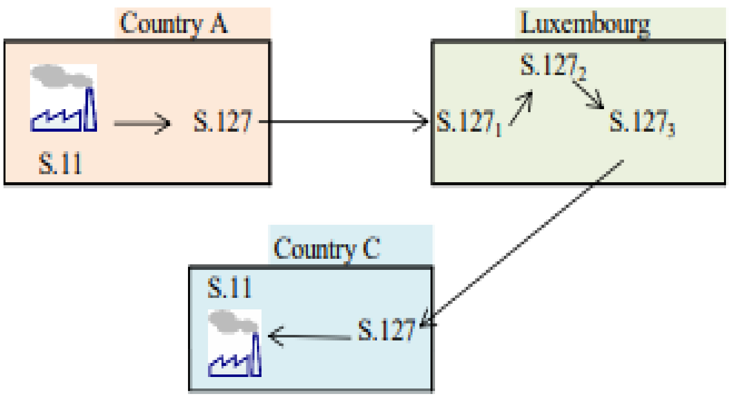
instruments issued by SPEs. Section 5 concludes with the need to carry on sharing data and the importance of SPEs typology[2].

2. IMF SPE's definition and its implementation in Luxembourg

An SPE, resident in an economy, is a formally registered and/or incorporated legal entity recognized as an institutional unit,
(...)
SPEs transact almost entirely with non-residents and a large part of their financial balance sheet typically consists of cross-border claims and liabilities.

All Luxembourg financial institutions are formally registered. Amongst them, which units looking like a captive financial institution would pass the institutional unit test? In the example below, a MNE makes an investment from country A to country C by setting up three entities in Luxembourg.
- S.1271, directly controlled by country A, is definitely an institutional unit[3].
- S.1272 theoretically may not pass the institutional unit test. However this is of little relevance for External Sector Statistics, because it only has domestic assets and liabilities.
- S.1273, is indirectly controlled by country A and has a claim on country C. It may not pass the test either but the claim of S.127 Luxembourg to country C would remain, even if the entities were consolidated.

Thus, a strict delineation of institutional units among captive entities does not prove essential for Luxembourg External Sector Statistics. The picture would change if, say, S.1271 would be a non-financial company S.11. Then all three Luxembourg units would need to be consolidated and the claim to country C would not be a claim of S.127 but a claim of S.11.



[2] This note does not touch upon the (important) question of SPEs output.
[3] SNA 2008 § 26.27 "Legal entities (...) are not combined if they are resident in different economies."

SPEs have no or little employment up to maximum of five employees, no or little physical presence and no or little physical production in the host economy.

Most Luxembourg captive financial institutions have no employment and very few Luxembourg SPEs have more than five employees. The assessment of SPEs employment is based on:

- Employment in the statistical business register,
- Non-zero staff cost in the profit & loss account, staff costs usually reflecting the compensation of directors. Some directors may participate in several boards.

However, not the companies themselves but providers of financial services (accounting, auditing, etc...) perform work on behalf of their SPEs clients. Those providers of financial services have indeed a physical presence in Luxembourg. It would therefore be a bit misleading to argue that SPEs "do not generate employment", without taking into account this indirectly generated economic activity[4].

SPEs are directly or indirectly controlled by non-residents.

Table below summarises the SPEs status of Luxembourg financial institutions Most Luxembourg captive financial institutions are directly or indirectly *owned* by non-residents. This ownership through a lasting interest (defining a Direct Investment) is the most common proxy for control between two institutional units. By contrast, Luxembourg Investment Funds are typically not held by direct investors "controlling" the entity but instead by minority investors, which makes them non-SPEs[5]. Last, Securitization Vehicles are usually held by direct investors, but the main part of their liability is made of debt securities (Portfolio Investment).

SPEs are established to obtain specific advantages provided by the host jurisdiction with an objective to (i) grant its owner(s) access to capital markets or sophisticated financial services; and/or (ii) isolate owner(s) from financial risks; and/or (iii) reduce regulatory and tax burden; and/or (iv) safeguard confidentiality of their transactions and owner(s).

This part of the definition does not refer to objective criteria such as the structure of a balance sheet or a concrete business model. It touches upon the competitiveness of each jurisdiction and is therefore much less

[4] The same goes for Investment Funds, which strictly have zero employment but still bring about labour intensive activity (custodian banks, transfer agents, fund administrations).

[5] In addition, a European convention (that the ECB suggested as a research topic in IMF External Sector Statistics) stipulates that all investment fund shares should be classified as Portfolio Investment.

straightforward. Do businesses invest in Luxembourg because of light regulatory burden, low taxes, skill of the workforce, political stability, or other criteria? Only a qualitative survey could theoretically answer those questions. As for (iii), captive financial institutions do not fall under the control of Luxembourg supervisory body. The same goes for most Luxembourg Securitization Vehicles and for some Investment Funds.

SPEs status of selected Luxembourg financial institutions

		Supervised		Controlled by direct investors/held by minority investors	SPE / Non SPE
		Yes	No		
S.124	Invest. Funds	Biggest part of the population	*Non-Regulated Alternative Investment Fund*	Minority investors/ Investment fund shares liability<-> Portfolio Investment	Non SPE
S.125	Overall Securitizatio n. Vehicles	*Vehicles Issuing debt on open markets*	Largest part of the population	Minority investors / Debt securities issued <->Portfolio Investment	Non SPE
S.127	Captive Financial Institutions		Complete population	Equity directly or indirectly held by a foreign institution or foundation <-> Direct investment	SPE

Source:BCL

3. Luxembourg SPE's typology (IMF report - Annex VI)

Luxembourg SPE aggregate balance sheet declines slightly in 2017 and more significantly in 2018. Following IMF report's typology, we distinguish between four main groups of SPEs: Conduits and pure holdings models are declining, while the intra group lending model is much more stable. The decline is mainly explained by two factors: First, Luxembourg signed the Multilateral Convention to Implement Tax Treaty Related Measures to Prevent BEPS (Base Erosion and Profit Shifting) in June 2017. The progressive phasing out of some financing arrangements incompatible with BEPS will probably continue in the next few years[6]. Second, US MNEs withdraw capital from Luxembourg further to the 2017 US tax reform.

logy of Luxembourg SPEs

Label and functional category (for RoW part)		**Total** 2016	**Total** 2018	3.1 - conduits 2016	3.1 - conduits 2018	3.2 - pure holdings 2016	3.2 - pure holdings 2018	3.3 - intragroup lending 2016	3.3 - intragroup lending 2018	3.4 - loan origination 2016	3.4 - loan origination 2018
Total assets		**9 599**	**8 577**	**891**	**696**	**3 415**	**2 707**	**5 204**	**5 096**	**90**	**78**
Debt securities held	PI	321	397	131	177	2	3	170	198	18	18
Participating interests	DI	6 819	6 152	342	221	3 283	2 616	3 167	3 293	27	22
Loans to affiliates*	DI**	2 276	1 846	380	278	120	86	1 736	1 450	40	32
Other assets	DI	183	182	38	20	10	1	131	155	5	5
Total liabilities		**9 600**	**8 577**	**891**	**696**	**3 415**	**2 707**	**5 204**	**5 096**	**90**	**78**
Debt securities issued	PI	1 115	973	720	515	16	17	360	419	19	21
Capital	DI	5 831	5 167	86	94	3 181	2 427	2 551	2 632	12	13
Loans from affiliates	DI	2 510	2 209	63	85	217	182	2 184	1 905	47	36
Other liabilities	OI	144	228	22	1	1	80	109	140	12	7

* to non affiliates for 3.4

** OI for 3.4

Source: BCL, Billions €

3.1 Entities issuing debt securities (other than securitization vehicles)

Those entities are further broken down into two subgroups:

- Some companies issue debt securities on open markets and lend the proceeds inside the group.
- Some other companies issue more sophisticated instruments (e.g. certificates), classified as debt securities, not necessarily on open markets.

This model has been declining since 2017.

3.2 Pure holdings

Some other entities have a basic wealth management structure, i.e. issue capital and hold participating interests. This model has been declining since 2017 as well.

3.3 Entities performing intragroup lending

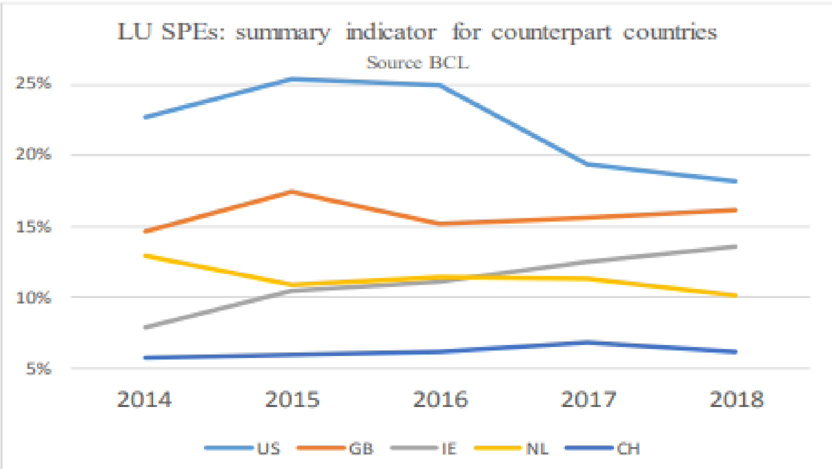
More diversified entities also take and grant loans inside the group. Debt instruments, overall loans, are classified Direct Investment. The model has remained stable since 2016.

[6] BEPS does not directly touch upon the most significant part of Luxembourg financial industry, i.e. deposit-taking corporations, Money Market Funds, Investment funds, Securitization Vehicles or Insurance Corporations.

3.4 Entities performing loan origination

Finally, some entities grant loan to non-affiliated entities. Their magnitude is much smaller than that of the three previous types. Loans outside the group are classified Other Investment.

In addition, main counterpart countries for operations of Luxembourg SPEs are US, GB, IE and NL. Operations with US have been declining since 2017, probably because of the tax reform.



4. Selected methodological issues on SPEs statistics

Luxembourg survey strategy on captive financial institutions

Captive Financial Institutions provide the BCL with a monthly security-by-security report and a quarterly full balance sheet. The BCL is therefore in a position to keep track of large and frequent balance sheet movements in an efficient manner. In a nutshell, the BCL collects positions and derives transactions. An ad hoc editing system spots large variations of positions (at a monthly basis on security items and on a quarterly basis for remaining items). There are indeed some blurred lines between transactions and price effects. A super dividend or a share premium would bring about a transaction, while an impairment would generate a price effect.

Editing system for SPEs operations

Luxembourg SPEs hold and issue illiquid financial instruments. The integration of those institutions to Luxembourg b.o.p./i.i.p. requires an ad hoc editing system, at semi aggregate and aggregate levels.

- *Semi aggregate level*

As has been previously emphasized, MNEs set up a network of financial institutions in Luxembourg. The network as a whole, takes external liabilities and holds approximately a similar amount of external assets (capital "in transit"), thus generates balanced positions (i.i.p.), and transactions (b.o.p.-financial accounts)[7]. How to identify the networks, ie how to group together Luxembourg affiliates of a single MNE?

The BCL's main source for this task is internal: i.e. companies' financial statements and survey data. Yet Eurostat Enterprise Group Register (EGR) also covers an increasing number of companies.

Number of Luxembourg financial institutions set up by MNEs
December 2018

	BCL survey		Covered by EGR	
	Number of groups	Number of institutions	Number of groups	Number of institutions
1 to 5 LU institutions in the group	1 310	2 319	691	1 227
6 to 10	81	600	41	314
11 to 15	31	389	15	187
More than 15	22	503	3	51
Total	1 444	3 811	750	1 779

Source BCL, Eurostat (EGR)

- Aggregate level

SPEs overall contribute to five "blocks", i.e. DI equities A/L, DI loans A/L and PI debt securities L. In 2018Q4, significant withdrawals of capital generate the decline of gross SPEs positions.

Flow / stock reconciliation of Luxembourg SPEs - External Statistics 2018 Q4

Billions of euros *Other = Price effect and Other Volume Change	Assets					Liabilities				
	Position 2018 3	**Trans.**	**FX**	**Other***	**Position 2018 4**	**Position 2018 3**	**Trans.**	**FX**	**Other***	**Position 2018 4**
Total- Captive fin. Inst. / RoW	**5 488**	**-254**	**32**	**-64**	**5202**	**5524**	**-291**	**25**	**-25**	**5233**
Direct investment	**5 137**	**-250**	**30**	**-45**	**4871**	**4506**	**-233**	**20**	**-22**	**4271**
Equity	3 880	-166	22	-43	3694	2954	-176	17	-17	2777
Loans	1 257	-85	8	-2	1178	1553	-57	3	-5	1493
Portfolio investment	**154**	**-11**	**1**	**-3**	**141**	**755**	**-53**	**5**	**-3**	**704**
Equity & investment fund shares	101	-11	1	-3	88					
ST debt securities	8	1	0	0	9	13	0	0	-1	12
LT debt securities	45	-1	0	0	44	714	-54	5	-3	662
Financial derivatives and Other Investment	198	7	1	-16	190	262	-4	0	0	258

[7] Along with balanced investment income (not described in this note).

Functional category of debt securities issued by SPEs

Let us come back to the above-mentioned second SPE type i.e. "Entities issuing debt securities (other than securitization vehicles)". Those conduits raise capital by issuing debt securities and then channel the funds to other affiliates of the MNE. They issue debt securities on organized financial markets. They also issue less standard products including so-called "hybrid" instruments, which are treated as debt securities in Luxembourg[8]. Those securities may not be transferred outside the group without the consent of the company. As for the functional category and from a theoretical perceptive, debt securities issued on open markets are definitely Portfolio Investment operations. Under certain circumstances[9], less liquid debt securities might theoretically ex ante qualify as Direct Investment. Mainly for pragmatic reasons (compilation purposes[10]), in Luxembourg both are classified as Portfolio Investment. In a similar way, securitisation vehicles, which are financial intermediaries (BPM6 § 4.77), issue debt securities on open markets. However, it can happen that, during a transitional period, securitisation makes use of intra-group financing[11], which, ex ante, also makes the line blurred between portfolio and direct investment. Again, for pragmatic reasons, the BCL classifies the operation as Portfolio Investment even during the transitional period.

In January 2019, the BCL sent a note to the IMF and suggested (research topic in External Sector Statistics) to remove Direct Investment from the list of functional categories associated to debt securities (BPM6 table 6.1, which gives a link between instrument and functional category).

5. Conclusion: SPEs coverage for reduction of asymmetries and BEPS implementation statistics

Reduction of asymmetries, data sharing and the Legal Entity Identifier (LEI)

The valuation and even the timing of SPEs equity transactions should be consistent with those of mirror country, which usually requires, at least at European level, bilateral discussions. International organizations have recently

[8] Those instruments are debt securities for statistical purposes. They may be treated as equity in other jurisdictions for fiscal purposes.

[9] BPM6 § 6.28. In addition, BPM6 § 6.56 details how Portfolio Investment is "not limited to securities traded on organized or other financial markets."

[10] Issuers are often not in a position to identify the holders of their debt securities. Thus the BCL uses the residual approach for PI liabilities (total liabilities net of resident holdings), which does not require the collection of counterpart country detail for debt securities issuances.

[11] The general rule, in the securitisation business, is that securitised assets are held by an undertaking different from that issuing the securities.

launched several such coordination initiatives, in which the BCL actively participates: the "FDI network" project and more broadly the documentation (metadata) of large individual SPEs transactions (microdata), made available to the ECB together with quarterly b.o.p./i.i.p. figures. Because of SPEs, micro data play an increasing role in the production of macroeconomic statistics.

Those data sharing operations would be more efficient if legal entities were identified by an international standard, like ISIN (or CUSIP) for financial instruments. In accordance with a 2017 European regulation, SPEs issuing debt securities on open markets are now identified by a Legal Entity Identifier (LEI) and securities issued have an ISIN code[12]. This paves the way for a better international standard identification of SPEs. Unfortunately, non-market oriented structures have had little incentives to register for a LEI so far. In addition, the LEI will bring a "logistic" help but factors of asymmetries as such will remain unchanged: coverage, timing of recording and overall valuation of unlisted equities from an accounting standard to another.

BEPS implementation and SPEs typology

The implementation of BEPS in Luxembourg started to bring about a reshaping of SPEs activities towards more "substance", bringing about some significant withdrawal of capital in 2017 and 2018. SPEs statistics collected by the IMF will also help to monitor this process at international level. Yet, business models are so complex and change so fast that some SNA 2008 and BPM6 standards (type of instrument functional category) may need to be completed with more targeted classifications in the medium run.

References

1. IMF BOP COM October 2018 Final Report of the Task Force on Special Purpose Entities.
2. BIS Proceedings of the ninth IFC Conference January 2019 "Are post-crisis statistical initiatives completed?"

[12] Regulation (EU) 2017/1129 of the European Parliament and of the Council of 14 June 2017 on the prospectus to be published when securities are offered to the public or admitted to trading on a regulated market

Special purpose entities and pass-through equity: A Micro-Analysis with BEA Data

Ryan Noonan
U.S. Bureau of Economic Analysis

Abstract
Special purpose entities (SPEs), legal entities with little or no employment or physical presence, play an important role in the global financial system. Financial transactions and direct investment positions recorded in macroeconomic statistics often reflect the flow of funds between business enterprises resident in one country and affiliated SPEs resident elsewhere. These statistics tend to record large direct investment equity positions in countries that host relatively more SPEs, irrespective of whether the equity passes through those countries to be used for production in other countries. Using firm-level survey data collected by the Bureau of Economic Analysis (BEA), this paper provides a summary of BEA's efforts to (i) identify SPEs in U.S. international economic statistics, (ii) measure the value of pass-through equity in affiliates, and (iii) understand the relationship between SPEs and pass-through equity. Preliminary results show that SPEs play a much larger role in U.S. direct investment abroad than in foreign direct investment in the United States and that SPEs account for a large amount of pass-through activity.

Keywords
Multinational enterprises, globalization, direct investment

1. Introduction

Special purpose entities (SPEs) play an important role in the global financial system. Increasingly complex multinational enterprise (MNE) structures use SPEs, legal entities with no or little employment or physical presence, to gain the benefits of varied legal and tax systems around the world. The impact of SPEs on economic statistics is difficult to assess and poses a challenge to national governments, international organizations, and other data users seeking to understand the global economy. Financial transactions and foreign direct investment (FDI) positions recorded in macroeconomic statistics often include the flow of funds between business enterprises resident in one country, affiliated SPEs resident elsewhere, and indirectly controlled affiliates in a third location. Because the purpose of many SPEs is to hold funds or securities that support production in other countries, these statistics tend to record large FDI equity positions in countries that host relatively more SPEs, though in most cases this equity simply passes through those countries to be used for production elsewhere.

To the extent that such "pass-through" equity is recorded in published FDI flow and position statistics it may create challenges for data users in at least two ways. First, total flows, including those to SPEs, may overstate the actual amount of FDI flows that stay in the host economy. For countries hosting many SPEs, these overstatements may be large, resulting in published FDI statistics that overstate the impact that FDI has on the host economy since much of the published flows are capital-in-transit. Second, for large economies with significant outward FDI flows to SPEs, such as the United States, recording FDI flows by immediate destination country obscures the ultimate destination and purpose of those flows.

In 2016, the International Monetary Fund's (IMF) Committee on Balance of Payments Statistics set up a Task Force on SPEs (TFSPE). At the conclusion of its work in 2018, the TFSPE proposed a definition of SPEs to be used for collection and analysis of macroeconomic statistics.[1] Under the definition, an institutional unit that meets all four of the following criteria is considered an SPE: (1) a maximum of five employees and little or no physical presence; (2) direct or indirect control by nonresidents; (3) established to obtain specific advantages, such as access to capital markets, financial services, risk mitigation, or tax or regulatory minimization; and (4) transactions almost entirely with nonresidents. The TFSPE recognized that one important activity of SPEs is holding assets for use elsewhere as pass-through equity.

This paper marks the first attempt to use data from the U.S. Bureau of Economic Analysis (BEA) to understand the prevalence of SPEs and their use of pass-through equity in U.S. FDI statistics. Pass-through equity is measured using company-level microdata according to a method proposed in a recent working paper by OECD economists.[2] Identifying SPEs and measuring pass-through equity enhances understanding of global macroeconomic statistics in three ways: (1) it illuminates the extent to which SPEs are used as pass-through vehicles by U.S. and foreign MNEs, (2) it provides evidence of whether SPEs are used for purposes other than passing equity to other destinations, such as for holding intellectual property, and (3) it provides information on the engagement of non-SPE firms in pass-through activity. Although the TFSPE focused on resident SPEs, this paper focuses primarily on nonresident SPEs as

---

[1] IMF Committee on Balance of Payments Statistics. "Final Report of the T ask Force on Special Purpose Entities." 31st Meeting: Washington, D.C., October 24-26, 2018. https://www.imf.org/external/pubs/ft/bop/2018/pdf/18-03.pdf.

[2] Borga, Maria, and Ceclia Caliandro. "Eliminating the Pass-Through: Towards FDI Statistics that Better Capture the Financial and Economic Linkages Between Countries." National Bureau of Economic Research Working Paper 25029: Cambridge, MA, September 2018. https://www.nber.org/papers/w25029.pdf.

those SPEs likely play a much larger role in U.S. FDI relationships, but it also discusses resident SPEs.

2. Methodology

BEA collects data on U.S. MNEs, their foreign affiliates, and U.S. affiliates of foreign MNEs using several surveys. Two of these surveys, the Annual Survey of U.S. Direct Investment Abroad (USDIA) and the Annual Survey of Foreign Direct Investment in the United States (FDIUS), collect an array of data items reflecting the activities of MNEs (AMNE). These include location, employment, sales, expenditures for fixed capital, and full income statements and balance sheets for the respondent firms.

While BEA surveys do not collect direct information about the purpose of an affiliate, USDIA and FDIUS data can be used to identify likely SPEs, both for SPEs in the United States (using FDIUS data) and foreign SPE affiliates of U.S. MNEs (using USDIA data). Based on the TFSPE guidelines, BEA identified an affiliate as an SPE using the following criteria: [3]

(1) Majority-owned by their parent(s).
(2) Zero to five employees.
(3) Less than $1 million of gross property, plant, and equipment (PPE).
(4) Less than 10 percent of total sales are to customers in the host country (for companies with sales).

This paper also uses data collected on the surveys to estimate the amount of pass-through equity moving through each affiliate. Because the USDIA data include both affiliates held directly by U.S. parents and those held indirectly through other foreign affiliates, any equity in one affiliate that passes through to another affiliate is, in effect, reported twice (or more) as it moves down the ownership chain. Removing this pass-through from intermediate affiliates creates a different picture of how equity relates to operations and provides a framework for recording FDI positions according to their ultimate destination countries.

Each affiliate in a multinational ownership chain can have “inward” and “outward” direct investment equity positions that correspond to its roles as both an affiliate of a foreign parent and a parent of foreign affiliates. An affiliate’s inward equity position is its total equity financing from its parent, which is included in owner’s equity in BEA AMNE publications. Correspondingly, an affiliate’s outward equity position is the affiliate’s equity in its foreign affiliates down the ownership chain, called equity in other foreign affiliates (hereafter, “equity in subsidiaries”) in BEA published data.

[3] BEA operationalized T FSPE’s “little or no physical presence” and “transact almost entirely with nonresidents” criteria with thresholds of less than $1 million of property, plant, and equipment and 10 percent of sales to local parties. These thresholds were tested for sensitivity and changes to them did not affect the results.

Following the process outlined by the OECD economists, the value of pass-through equity in an affiliate is defined as the smaller of its outward and inward equity positions—i.e., the value of equity invested in the affiliate by its parent that the affiliate then invests in its own foreign affiliates down the ownership chain.[4] In other words, inward equity positions that are large enough to account for outward equity positions are assumed to be pass-through equity. Subtracting this pass-through value from both the inward and outward positions provides a new estimate of owner's equity and equity in subsidiaries for each affiliate that seeks to remove double counting of equity that passes through the affiliate and identify where the equity is ultimately being invested.[5]

3. Result

III-A. Foreign (nonresident) SPE affiliates of U.S. MNEs

This section identifies and provides descriptive statistics about nonresident SPEs using the methodology described in the previous section. Nonresident SPEs, from the perspective of BEA, are foreign affiliates of U.S. MNEs.

In 2016, there were 78,413 majority-owned foreign affiliates (MOFAs) of U.S. MNEs. Of these MO-FAs, 16,021 (20.4 percent) met the SPE criteria and accounted for 39.7 percent of total affiliate (SPE and non-SPE) assets.[6] SPEs were identified in 182 of the 199 industries in which MOFAs operated in 2016. Table 1 shows the five industries with the largest share of assets and the number of affiliates in each industry. Holding companies, which are companies that only own the financial securities of other companies, accounted for 85 percent of SPE assets and nearly half of the number of SPE affiliates. By share of SPE assets, holding companies were followed by financial sector industries including financial investment activities, non-depository credit intermediation, banking, and securities and commodities.

[4] Borga and Caliandro.

[5] While these estimates better reflect the ultimate destination of the funds, there are at least three possible sources of mis-measurement. First, not all outward equity positions are necessarily funded by inward equity positions large enough to cover them. Second, there can be errors or omissions in the information about the chains of ownership among foreign affiliates. Third, there can be ambiguous relationships between inward and outward equity positions when directly-held affiliates have equity positions in multiple indirectly-held affiliates.

[6] Assets are used as a proxy for FDI position because position is only available for directly held affiliates, while assets are available for both directly and indirectly held affiliates. Assets are collected on company balance sheets and are not adjusted for ownership percentage.

Table 1. Nonresident SPEs and Assets by Selected Industry, 2016				
Industry of Affiliate	**Number**	**Share of SPEs (%)**	**SPE Assets ($B)**	**Share of SPE Assets (%)**
Total	*16,021*	*100.0*	*$10,032.4*	*100.0*
Holding companies	7,906	49.3	$8,533.0	85.1
Financial investment activities	910	5.7	$333.3	3.3
Non-depository credit intermediation	157	1.0	$285.7	2.8
Banking	49	0.3	$267.8	2.7
Securities and commodities	159	1.0	$117.6	1.2
Other industries	*6,840*	*42.7*	*$495.0*	*4.9*

The five countries hosting the largest share of SPE assets accounted for nearly 70 percent of total SPE assets, as shown in Table 2. The Netherlands, which hosted the largest share, and many of the other top host countries are commonly viewed as "countries of convenience"—that is, countries hosting large amounts of pass-through capital, intellectual property, or other SPE-related activities because of favorable tax rules and other regulations. These five countries reflect the concentration of SPE activity in a small number of locations.

Table 2. Nonresident SPEs and Assets by Selected Country, 2016				
Country	**SPEs**	**Share of Global SPEs (%)**	**Assets ($B)**	**Share of Global SPE Assets (%)**
Total	*16,02*	*100.0*	*$10,03*	*100.0*
Netherlands	1,578	9.8	$2,151	21.4
Luxembourg	1,141	7.1	$1,881	18.7
United Kingdom	1,880	11.7	$1,157	11.5
Bermuda	594	3.7	$969.4	9.7
UK Islands, Caribbean[7]	1,529	9.5	$762.8	7.6
Other countries	*9,299*	*58.0*	*$3,111*	*31.0*

<u>III-B. Pass-through equity in foreign (nonresident) affiliates of U.S. MNEs</u>

Following the definition outlined in Section II, pass-through equity in MOFAs of U.S. MNEs totaled $6.5 trillion in 2016. This amount accounted for just over 50 percent of all owner's equity in MOFAs and 83 percent of their equity in subsidiaries. Removing $6.5 trillion from BEA's statistics on U.S. direct

[7] The UK Islands, Caribbean comprises Anguilla, the British Virgin Islands, the Cayman Islands, and Montserrat.

investment abroad would decrease overall owner's equity in MOFAs from $12.9 trillion to $6.4 trillion. Equity in subsidiaries would decline from $7.9 trillion to $1.4 trillion. Reflecting the close connection between pass-through equity and SPEs, 81 percent of all pass-through equity moved through SPEs.

Table 3 shows the five countries that host the largest amount of MOFA pass-through equity. These five countries accounted for two-thirds of all pass-through equity. In "countries of convenience" such as the Netherlands, Luxembourg, and Bermuda, pass-through equity made up 70 percent or more of total owner's equity. MOFAs in countries where U.S. MNEs tend to be engaged in more economic production were less likely to have pass-through equity than MOFAs in countries of convenience.

Table 3. Nonresident Pass-Through Equity by Selected Country, 2016

Country	Pass Through Equity ($B)	Share of Global Pass-Through Equity (%)	Share of Country's Owner's Equity (%)	Share Accounted for by SPEs (%)
Total	*$6,511.4*	*100.0*	*50.4*	*81.5*
Netherlands	$1,483.7	22.8	72.9	85.2
Luxembourg	$1,028.8	15.8	71.6	93.4
United Kingdom	$792.4	12.2	46.9	78.7
Bermuda	$649.1	10.0	69.6	92.0
Canada	$363.7	5.6	47.8	73.7
Other countries	*$2,193.7*	*33.7*	*n.a.*	*n.a.*

While SPEs with pass-through equity made up approximately one-third of all SPEs, they accounted for a large majority of the total assets of SPEs. The 5,471 SPEs with pass-through equity had $8.8 trillion of assets, compared to a total of $1.2 trillion for the 10,550 SPEs without pass-through.

III-C. U.S. (resident) SPE affiliates of Foreign MNEs

Sections III-A and III-B focus on nonresident SPEs, as those SPEs likely play a much larger role in U.S. direct investment relationships partly due to the favorable tax and regulatory environments in some foreign countries of convenience. This section and the following section provide parallel statistics for resident SPEs.

In 2016, there were 22,054 majority-owned U.S. affiliates (MOUSAs) of foreign MNEs. Of these MOUSAs, 4,653 (21.1 percent) met all four SPE criteria. SPEs were identified in 180 of the 200 industries in which MOUSAs operated in 2016. Table 4 shows the five industries with the largest share of SPE assets and the number of SPEs in each industry. Holding companies, despite holding

95 percent of resident SPE assets, accounted for only 13 percent of the number of SPEs.[8] Resident SPEs were more likely to be found in the real estate industry than any other industry. While this industry includes real estate trusts, it also includes smaller-value properties that may meet the TFSPE's definition as applied here. Future research is needed to understand which of these affiliates are truly SPEs.

Table 4. Resident SPEs and Assets by Selected Industry, 2016				
Industry of Affiliate	**Number**	**Share of SPEs (%)**	**Assets ($B)**	**Share of SPE Assets (%)**
Total	*4,653*	*100.0*	*$159.4*	*100.0*
Holding companies	594	12.8	$151.7	95.2
Nondepository credit intermediation	48	1.0	$3.1	2.0
Real estate	1,571	33.8	$1.2	0.8
Financial investment activities	260	5.6	$1.2	0.8
Funds and trusts	135	2.9	$1.1	0.7
Other industries	*2,045*	*44.0*	*$1.1*	*0.7*

The five countries of ultimate beneficial owner (UBO) [9] with the largest shares of resident SPE assets accounted for more than three-quarters of total SPE assets. Most of these countries are longstanding trade and investment partners of the United States. The largest share of SPE assets in the United States was ultimately owned by United Kingdom-based MNEs, but the second largest share was ultimately owned by U.S. persons (commonly referred to as "roundtrip FDI").

Table 5. Resident SPEs and Assets by Selected Country of UBO, 2016				
Country of UBO	**SPEs**	**Share of U.S. SPEs (%)**	**Assets ($B)**	**Share of U.S. SPE**
Total	*4,653*	*100.0*	*$159.4*	*100.0*
United Kingdom	905	19.4	$48.5	30.4
United States	520	11.2	$43.7	27.4
Netherlands	499	10.7	$12.5	7.8
Canada	309	6.6	$9.2	5.8
Mexico	263	5.7	$8.2	5.1
Other countries	*2,157*	*46.4*	*$37.3*	*23.4*

[8] U.S. affiliates report on a consolidated domestic U.S. basis. The industry of a U.S. affiliate is determined by the industry from which the consolidated U.S. company derives the majority of its sales. Consolidated U.S. affiliates may include holding companies within the consolidation. This differs from the U.S. direct investment abroad data in which, according to survey instructions, foreign affiliates cannot be consolidated across industries or countries.

[9] The UBO is that person, proceeding up a U.S. affiliate's ownership chain, beginning with and including the foreign parent, that is not owned more than 50 percent by another person. Unlike the foreign parent, the UBO of an affiliate may be located in the United States.

III-D. Pass-through in U.S. (resident) affiliates of foreign MNEs

In 2016, pass-through equity in MOUSAs of foreign MNEs totaled $613 billion. This represented less than 10 percent of the value of pass-through found in MOFAs of U.S. MNEs, indicating that pass-through activity is more common for U.S. outward direct investment than for U.S. inward direct investment. Pass-through equity made up 24 percent of owner's equity in MOUSAs, compared with 50 percent of owner's equity in MOFAs. Pass-through equity accounted for 80 percent of MOUSA equity in subsidiaries, similar to the 83 percent observed for MOFAs. Subtracting $613 billion from BEA's inward statistics on foreign direct investment in the United States would decrease owner's equity in MOUSAs from $2.5 trillion to $1.9 trillion, while equity in subsidiaries would decline from $762 billion to $149 billion.

Another important difference between pass-through in MOUSAs and MOFAs is that MOUSA pass-through equity is much more distributed across SPEs and non-SPEs. While 82 percent of MOFA pass-through moves through SPEs, that figure is only 17 percent for MOUSAs.[10]

Table 6 shows the five largest sources of MOUSA pass-through equity by country of UBO. These five countries accounted for 57 percent of all pass-through equity in MOUSAs. Pass-through equity made up 40 percent of owner's equity in MOUSAs of UK-headquartered MNEs, though those U.S. affiliates were generally not SPEs. The largest share of pass-through equity in MOUSAs originating in the United Kingdom moves through financial investment affiliates.

Table 6. Resident Pass-Through Equity by Selected Country of UBO, 2016

Country of UBO	Pass Through Equity ($B)	Share of U.S. Pass-Through Equity (%)	Share of Owner's Equity of Country's Affiliates (%)	Share Accounted for by SPEs (%)
Total	*$612.5*	*100.0*	*24.1*	*17.4*
United Kingdom	$169.6	27.7	40.2	17.9
Canada	$62.0	10.1	22.3	5.1
Japan	$41.8	6.8	12.1	8.1
France	$40.5	6.5	18.8	2.1
Ireland	$38.3	6.2	24.1	7.5
Other countries	*$260.9*	*42.6*	*n.a.*	*n.a.*

[10] As with holding companies discussed in footnote 6, the lower share of SPEs in MOUSA pass-through may reflect the FDIUS survey's consolidation rules.

While most holding company MOUSAs (489 of 652, or 75 percent) host pass-through equity, SPEs in the United States, in general, are unlikely to do so. Of 4,653 SPEs identified, only 437 have pass-through equity. This absence of pass-through equity may indicate that SPEs in the United States are used for purposes other than pass-through equity and that because the United States is not a "country of convenience," foreign MNEs are likely using other countries for pass-through activities.

4. Discussion and Conclusion

This paper represents a first attempt to identify and analyze SPEs in U.S. direct investment statistics using the TFSPE's definition. This paper shows that SPEs play a much larger role in U.S. direct investment abroad than in foreign direct investment in the United States. SPEs—particularly nonresident SPEs—host a large amount of pass-through capital that moves from the parent company through an intermediary country before reaching the country where economic production occurs. The large overlap between SPEs and pass-through equity in the U.S. direct investment data is an indication that the SPE criteria developed by TFSPE will be useful in isolating SPEs in international investment statistics and that by identifying SPEs—particularly in "countries of convenience"—most pass-through activity will be identified as well. Additional work to fit the TFSPE's definition of SPEs to BEA microdata will allow BEA to further refine this identification. In particular, understanding which real estate affiliates in the United States are truly SPEs will provide clarity about the role of U.S.-resident SPEs in inward direct investment.

In addition to gaining a clearer understanding of the role of SPEs in such pass-through relationships, this paper outlines progress toward a framework for "pushing down" the U.S. direct investment position past SPEs to operating units and providing a clearer picture of how U.S. direct investment abroad relates to economic production in foreign countries. Future research will be able to study the distribution of U.S. direct investment after the elimination of pass-through.

Special purpose entities: Relevance and impact of harmonised concepts and measurement challenges – UK contribution

Sami Hamroush
Head of Balance of Payments, Office for National Statistics, United Kingdom

Abstract
Globalisation has led to a rapid increase in the importance of multinational enterprises to national economies, whose activities and corporate structures often span many countries with varied degrees of complexity. The role of Special Purpose Entities (SPEs) has also seen a rapid increase, as multinationals attempt to manage their access to capital markets, financial risks, and regulatory or tax burdens. As a member of the IMF Task Force on Special Purpose Entities, ONS has been utilising available data sources and the Task Force's definition of SPEs to develop a decision tree for identifying resident SPEs in the UK's FDI statistics. This paper summarises ONS's analysis of resident SPEs and outlines some of the compilation challenges faced when producing these estimates.

Keywords
FDI; SPE; MNE; UK; Globalisation

1. Introduction

Globalisation has led to an increase in the importance of multinational enterprises to national economies, whose activities and corporate structures often span many countries with varied degrees of complexity. The role of Special Purpose Entities (SPEs) has also seen a rapid increase, as multinationals attempt to manage their access to capital markets, financial risks, and regulatory or tax burdens.

The measurement of SPEs is important for ensuring Foreign Direct Investment (FDI) statistics accurately capture financial flows and corporate structures, while the ability to isolate SPEs from headline statistics is increasingly required by users who wish to analyse the true beneficiaries and risk takers of cross-border investment.

The Office for National Statistics (ONS) uses quarterly and annual FDI surveys as the main source for producing UK FDI statistics. As a member of the International Monetary Fund's Task Force on Special Purpose Entities (TFSPE), ONS has explored available data sources along with the Task Force's definition of SPEs to develop a decision tree for identifying resident SPEs in the UK's FDI statistics.

This paper outlines the agreed international definition of an SPE, summarises key findings of the TFSPE, and presents experimental estimates of UK inward FDI attributable to resident SPEs. The paper also outlines considerations and challenges faced when producing these estimates, which the author hopes will inform future discussions and work.

2. Methodology

SPE background and concepts

FDI refers to cross-border investment made by investors with the objective of establishing a lasting interest in an enterprise in an economy other than that of the investor.

An increasing globalised world with reduced capital controls, fewer barriers to trade and investment, deregulation of markets and technological progress have resulted in multinational enterprises becoming increasing important for national and international economic stability and growth. The increasing importance of multinationals have also led to many governments adopting policies to attract direct investment into their economies for benefits such as increasing productive capacity, job creation, knowledge and technological transfers, and productivity.

Growth in multinationals' international integration has also resulted in a rapid increase in the complexity of their corporate structures in terms of how they control investments, manage their risks, and exploit country specific advantages. Country specific advantages may be economic – such as an educated workforce or developed infrastructure – or financial, such as efficient capital markets or preferable tax regimes.

Managing these increasingly complex corporate structures and asymmetric financial, regulatory and tax regimes has led an increase in the use of special purpose entities – legal entities set up with the specific purpose of financial investments pass-through or holding overseas assets and liabilities on behalf of a wider international group. These entities have little to no physical presence in the economy they are based.

International manuals including the European System of Accounts, 2010; the OECD's Benchmark Definition of Foreign Direct Investment, fourth edition; and the IMF's Balance of Payments and International Investment Position Manual, sixth edition all include definitions on SPEs and recommend national statistical institutes separately identify FDI transactions and positions undertaken by resident SPEs in their statistics. The motive behind such recommendations is clear when FDI statistics are examined by partner country, where financial centres and offshore islands often have substantial FDI values relative to the size of their economies – reflecting the prevalence of SPEs.

IMF Task Force & decision tree

Although international manuals and organisations have emphasised the importance of separating out transactions and positions related to SPEs in FDI statistics, there has been an outstanding requirement for an internationally agreed definition of SPEs. With this in mind, the IMF Committee on Balance of Payments Statistics created the TFSPE in 2016 with a two-year mandate. The TFSPE's final report can be found here.

The TFSPE conducted a survey on the legal and national definitions of SPEs, their measurement and activity. Using this information, a SPE definition was proposed that included an upper limit of up to five employees, while no specific numerical threshold is recommended to account for physical presence and/or physical production. SPEs were also defined to be directly or indirectly controlled by non-residents, who set them up with the objective of: accessing capital market of financial services, isolating financial risk, reducing regulatory or tax burden, and/or safeguarding the confidentiality of transactions.

Using this definition and information gathered from the survey, the TFSPE produced a decision tree and typology. The decision tree was produced to allow national statistical institutes to classify entities as SPEs, while the typology was produced to provide context around the type of activities undertaken by SPEs – but was not designed to be an exhaustive list of activities undertaken by SPEs.

Data sources

To estimate the proportion of FDI attributable to SPEs, data linking was undertaking at using microdata from the FDI population and the Inter-Departmental Business Register (IDBR). FDI estimates used in this paper are for the 2016 reference period, consistent with results published December 2017.

The Inter-Departmental Business Register

The IDBR is a comprehensive list of 2.6 million businesses in all parts of the UK economy, other than the very small businesses and some non-profit making organisations. The two main sources of input are Value Added Tax PAYE income tax from HM Revenue and Customs. Additional input comes from Companies House, Dun and Bradstreet, business surveys and businesses profiling.

The Foreign Direct Investment population and survey

The FDI population is produced from four sources:

- Combining IDBR and Dun and Bradstreet (D&B) data to identify cross-border business relationships between UK businesses and foreign parents or subsidiaries (majority relationships).

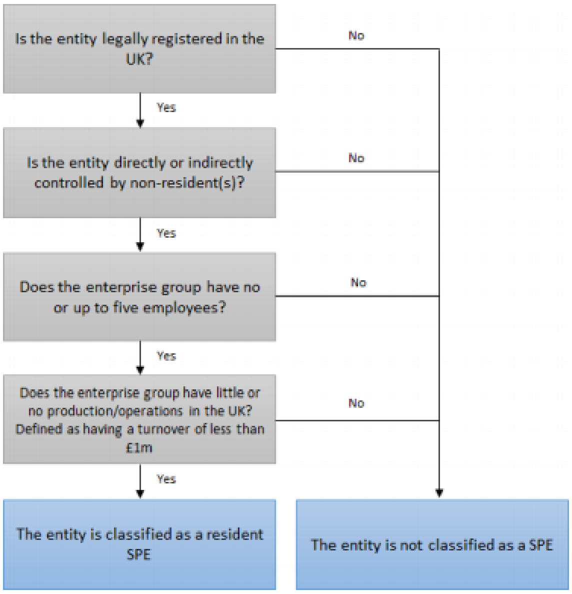
- A separately maintained database for enterprises that have previously been identified through the FDI or Mergers and Acquisitions surveys. This database includes minority share relationships that are not included in the IDBR and D&B.
- The FDI population is also updated regularly using information from the ONS Mergers and Acquisitions Survey, which is conducted on a quarterly basis.
- FDI estimation methods predict values for all non-sampled businesses, allowing for analysis of sampled and non-sampled businesses at a microdata level.

Decision tree

The FDI population and IDBR were matched to produce a dataset of cross-border FDI variables along with economic variables and UK employment and turnover for enterprise groups in the inward FDI population. More detail on the methods used to match the two dataset is available here.

It was also not possible to match all the inward FDI population to the IDBR, as some entities did not have enterprise identifiers, such as banks and bank holding companies (aggregates provided by the Bank of England), residential property, and public corporations. There were also some enterprises in the remainder of the FDI population that could not be linked to the IDBR – which were part of the separately maintained FDI population. Domestic balance sheet data are also not available via the IDBR; therefore, the decision tree utilised turnover as a proxy for physical UK operations.

Utilising the data available, the below decision tree was produced utilising the definition from the IMF Task Force on SPEs with data available at the time of analysis.



3. Result

High level results

Utilising the definition developed by the TFSPE and data sources available at the time of analysis, this paper estimates that £74 billion, or 8.4%, of the UK's inward FDI position in 2016 was attributable to SPEs.

As shown in Table 1, the UK's inward FDI position was £1,200 billion in 2016, of which £323 billion related to entities that did not have employment and turnover data available at the time of analysis – see previous section for more information on why.

Over two-thirds of the groups in the FDI population were matched to the IDBR, who in turn accounted for £877 billion of UK inward FDI in 2016. Of these, £82 billion of inward FDI positions were held by groups with five or fewer employees, which fell to £74 billion when the threshold of £1 million was applied.

Table 1: High level estimates of resident SPEs in the inward FDI population, £ billion, 2016

	Positions	*Income*
Total inward population	1,200	52
Banks, bank holding companies, property, public corporations, and unmatched businesses	323	15
Analysed population	877	38
Those meeting employment threshold	82	2
Those meeting employment and turnover threshold	74	1

Characteristics

Comparing the characteristics of enterprise groups classified as resident SPEs to non-SPEs can help provide more context to their structure and use.

Chart 1[1] presents the geographical composition of resident SPEs and non-SPEs. There is a clear distinction between the two, with the immediate parents of SPEs largely residing in "remainder" – a grouping comprising North America and non-EU Europe that has been supressed to mitigate disclosure. In contrast, a large proportion of inward FDI positions belonging to non-SPEs is attributable to immediate parents based in the EU.

[1] Note that figures presented in this chart and thereafter exclude banks, bank holding companies, property, public corporations, and entities with no employment or turnover information unless otherwise specified.

Chart 1: Geographical composition of resident SPEs and non-SPEs, 2016

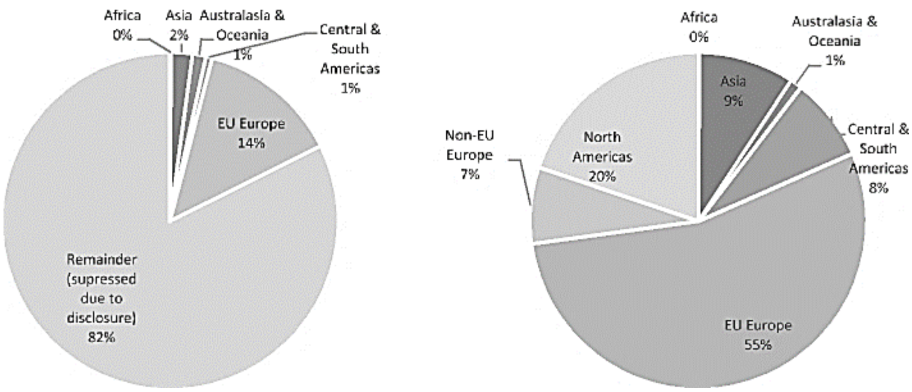
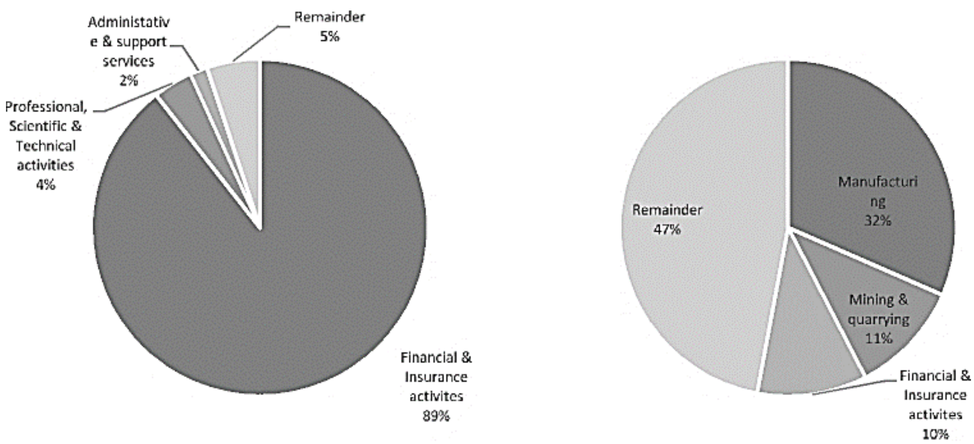


Figure 4 considers the industry composition of resident SPEs and non-SPEs. There are clear distinctions between the two classifications, with a notably larger proportion of enterprise groups classified as resident SPEs sitting within the financial & insurance services industries (89%). In contrast, non SPEs appear to be more diverse, with the majority sitting in manufacturing (32%), mining & quarrying (11%), and financial & insurance services (10%).

Chart 2: Industry composition of resident SPEs and non-SPEs, 2016



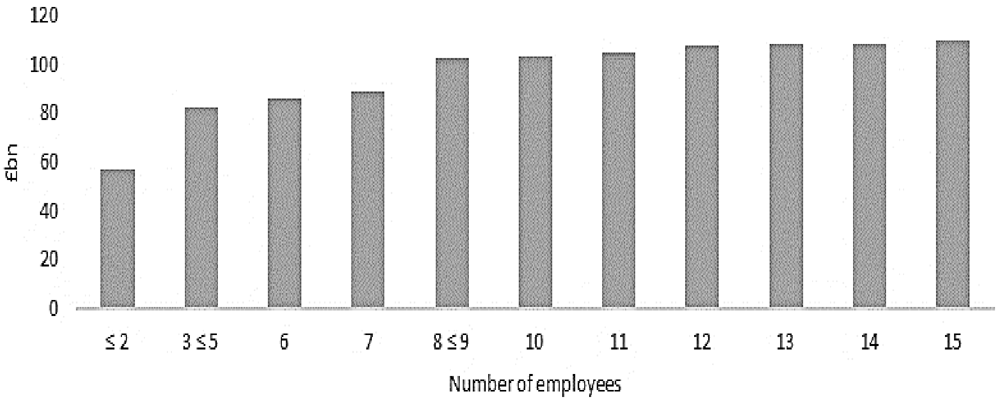
4. Discussion and Conclusion

Considerations

Results presented in this paper are experimental estimates produced through utilising the definition produced by the TFSPE and data that was available at the time of research. Given these are initial estimates, there are a number of areas that could be investigated in future to improve the accuracy of estimates.

An area for future consideration is the sensitivity of the thresholds used. The TFSPE classifies entities as SPEs based on assumptions regarding their characteristics – particularly employment and physical presence. To illustrate the sensitivity of these assumptions, Chart 5 presents the cumulative value of UK inward FDI positions attributable to businesses by employment size. As shown, enterprise groups meeting the employment threshold of up to two employees accounted for £57 billion of the UK's inward FDI position in 2016. Increasing the threshold to five employees raises this to £82 billion, to £86 billion for six employees, and £103 billion when the threshold is set to ten.

Chart 5: Sensitivity of employment thresholds, cumulative £ billion, 2016



A further consideration was that, using the data sources available at the time the research, it was not possible to separate enterprises that met the definition of a SPE but were also part of a wider enterprise group that had operations in the UK through other enterprises. For example, an enterprise group may control a special purpose entity in the UK to manage their international finances; however, they may also control a number of other enterprises undertaking economic output. Of approximately 19,000 groups analysed, slightly over 1,400 had an aggregate employment level that exceeded five when combining all their enterprises, but also had at least one enterprise meeting the employment threshold. Inclusion of these other enterprises would likely increase the value of inward FDI positions; however, it is unclear whether they should be considered SPEs.

A further area is the use of balance sheet information as part of the decision tree. Balance sheet data is not currently available via the IDBR, therefore the decision tree developed was arguably less sophisticated than that developed by the Task Force. Utilising wider data sources in future would allow for better measurement of UK physical presence, rather than the use of turnover as a proxy.

A further consideration was that, given resident SPEs are not currently identified, the FDI sample design and estimation methods do not currently treat resident SPEs separately. As such, businesses with low turnover can be grouped in smaller stratums unless previously identified through random sampling. Table 2 presents the sampling proportions for enterprise groups classified as resident SPEs and non-SPEs in this analysis. As can be seen, a smaller proportion of SPEs are sampled compared to non-SPEs – possibly due to fewer sitting in sampling stratums reserved for larger businesses according to turnover. Furthermore, while fewer are sampled, those that are account for a larger proportion of the overall value estimated for groups classified as SPEs.

Table 2: Sampling and estimation of resident SPEs and non-SPEs

	SPE	*NON-SPE*
Percentage of groups sampled	16%	19%
Percentage of inward positions sampled	95%	90%

Finally, a key area for future research is to classify all entities within the FDI population. As noted in the paper, not all entities in the FDI population were classified as it was not possible to link them all to the IDBR. A future review of the data sources used for these estimates may help improve the results.

5. Conclusions

Accurate coverage and separation of SPEs has become ever more important for ensuring FDI statistics are accurate and remain useful to users – who increasingly need to separate SPEs from headline statistics to understand the true beneficiaries and risk takers of FDI.

While international manuals and organisations have emphasised the importance of measuring SPEs in FDI statistics, there has been an outstanding requirement for an internationally agreed definition. As such, the TFSPE proposed definition, decision tree and typology are a welcome development that provides a clear framework for classifying SPEs.

Results presented in this paper have been produced through utilising definitions produced by the TFSPE and available data sources. The experimental results estimate that £74 billion of the UK's inward FDI positions were attributable to enterprise groups classified as SPEs, or 8.4% of the positions analysed. The results also suggested these groups' economic activity and their immediate parents are distinct from groups not classified as SPEs.

While these results are a positive step towards being able to separate resident SPEs from UK FDI statistics, the paper also highlighted a range of areas for future development to help improve these results. These largely focus on increasing the availability of data regarding multinationals conducting FDI in the UK, particularly on their UK operations and intra-group transactions.

References

1. Benchmark Definition of Foreign Direct Investment, 4th Ed, OECD (2008)
2. Balance of Payments and International Investment Position Manual, 6th Ed, IMF (2009)
3. Final Report of the Task Force on Special Purpose Entities, IMF (2018)
4. European System of Accounts, Eurostat (2010)
5. Foreign Direct Investment Involving UK Companies: 2016, ONS (2017)
6. UK foreign direct investment, trends and analysis: July 2018, ONS (2018)

Irish Special Purpose Entitites (Spes): Concepts and challenges of the implementation of the results of the IMF Task Force

Giovanna Bua, Brian Golden
Central Bank of Ireland

Abstract

Ireland is home to a large number of entities, many of which are potentially Special Purpose Entities. There is a wide variety in both the business models and objectives of the vehicles, which span from traditional securitisation to investment vehicles, funding and loan origination vehicles, and often form part of cross border, multi-entity corporate structures. The Central Bank of Ireland has been collecting data from these entities since 2015 using a locally developed definition. This paper aims to present an overview of some of the different structures active in Ireland. The authors will then assess some of these structures against the guidance provided by the IMF BoPCom TFSPE in order to underline strengths and weaknesses.

Keywords

ownership, sponsor, institutional units, IMF BOPCOM TFSPE, pass-through

1. Introduction

The statistical treatment of SPEs has been widely discussed at a global level for more than a decade. This work advanced significantly with the publication of the Final Report of the Task Force on Special Purpose Entities by the IMF (TFSPE) in October 2018. The TFSPE proposed an international definition of SPEs in the context of cross-border statistics as well as a data collection framework for cross-country comparable SPE data.

The Central Bank of Ireland (CBI) extended granular balance sheet reporting to non-securitisation SPEs in Q3 2015, on the same basis as those already applied to securitisation SPEs since Q4 2009 under European legislation. The Statistics Division at the CBI collects this data for multiple purposes, namely statistical purposes, regulatory monitoring and financial stability analysis. The definition for data collection was simply availing of a particular tax provision in Irish legislation, covering SPEs that engage in an extensive range of financial and leasing transactions.

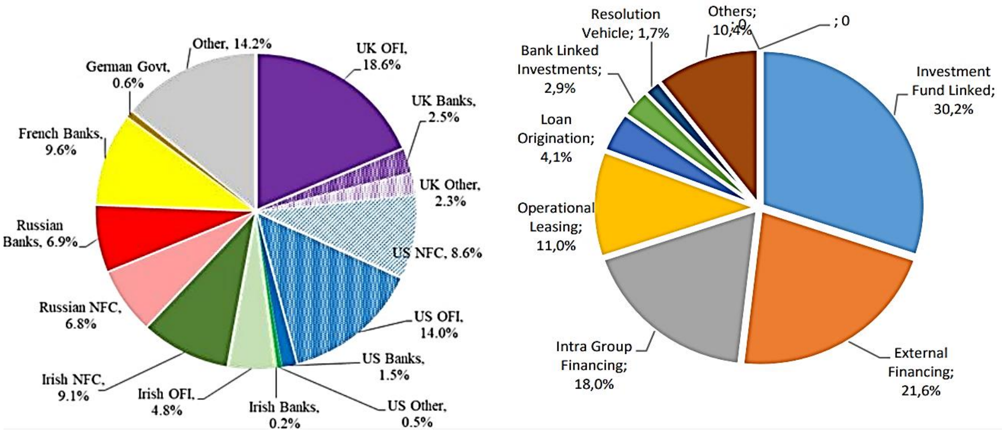
2. Special purpose entities in Ireland

Each SPE reports its entire balance sheet on a quarterly basis, mostly on a security-by-security basis, and profit and loss information annually. In addition, each SPE fills out a registration form, containing, among other items, the name, country and economic sector of the sponsor (the entity on whose behalf an SPE was established), the originator (the entity or entities that originally issued the SPE loans/assets) and the ultimate parent if the SPE is consolidated on an accounting basis. Noted interest, bankruptcy remoteness and orphan structure characteristics are also reported.

Total assets in the SPE sector amounted to €727 billion in Q4 2018. Within this, securitisation SPEs account for €447 billion in assets within 1,132 SPEs and non-securitisation SPEs, €281 billion of assets within 1,181 SPEs. Over half of all non-securitisation SPEs are consolidated into other entities compared to around one quarter of securitisation SPEs. Within both categories, most consolidation takes place into foreign entities while most SPEs are also sponsored by foreign entities. However, there is not complete alignment between both concepts, i.e. SPEs can be consolidated into a domestic entity but sponsored by a foreign entity and, less commonly, vice-versa. We also find that our SPE population reports fees and commissions payable for professional services but observe no evidence of employees and this view is supported by our market intelligence activities.

The non-securitisation SPE population is diverse in terms of both country and sector links and types of activity. These SPEs are often part of complex, cross-border corporate structures, often facilitating intermediation activity elsewhere in the chain. Looking at sponsors of SPEs (Figure 1a), these straddle the financial and non-financial sectors and a number of regions with domestic sponsors comprising 14 per cent of the total, by asset. In fact, disaggregating the other category provides almost 60 combinations of country and sector. Looking at activities (Figure 1b), non-securitisation SPEs span 14 pre-defined types of activity and a small residual category. Nevertheless, four categories comprise over 80 per cent of the sector in terms of total assets, namely investment fund linked, external financing, intergroup financing, and operational leasing. The two largest categories are presented below as business models to draw out some BoP relevant issues. For a full list of activities in our population, see Golden and Hughes (2018).

Figure 1a and 1b: Non-securitisation SPEs by sponsor and by activity (total assets)



Focusing on domestic links, Irish sponsors (€44 billion) are mostly multinational Non Financial Corporations (NFCs), including some companies that re-domiciled their headquarters to Ireland and are engaged in intra-group financing, external financing and operational leasing activities. Fund-linked investments linked to Irish-resident investment funds are also significant though these are largely sponsored by foreign fund managers. Most Irish sponsored SPEs are consolidated into the accounts of Irish entities (€36 billion) though consolidation into the accounts foreign entities also occurs (€2 billion). Within the non-securitisation SPE population as a whole, over half of total assets are consolidated into the accounts of other entities.

Based on this data, and in liaison with regulatory colleagues, Golden and Hughes (2018) developed the concept of SPEs (securitisation and other) as follows:

> “a legal entity, with little or no physical presence and narrow, specific, and/or ring-fenced, objectives, such as the segregation of risks, assets and/or liabilities, or as a cash conduit. The directors of an SPE typically have limited or no discretionary powers; rather activities are strictly defined by the terms of the SPE contract or arrangement. An SPE is often, though not exclusively, a satellite company of another financial entity and forms an ancillary part of the associate entity’s business by warehousing particular assets or risks.”

These SPE contracts are drawn up in accordance with the wishes of the SPE sponsor. The activities of the SPE are conducted in accordance with the specific terms of the contract. The SPE directors cannot act outside the terms of the contracts. This could be seen as a measure of control by the sponsor. Risks are distributed between SPE counterparties according to the provisions of each contract.

3. Challenges in the context of business models

In this section, we aim to assess some of the structures of Irish SPEs against the guidance provided by the TFSPE. The two business models outline the typical activities undertaken by Irish-resident SPEs. While these models do not cover the whole range of activities, they provide good guidance for assessing the criteria proposed by the TFSPE definition. For each model, we try to highlight those characteristics that challenge the identification of fundamental concepts such as institutional units, and economic and legal ownership. In order to tackle the relevance for External Statistics and to shed some light on their behaviour, we also look at the geographical composition of their portfolios. With this in mind, we first look at the average foreign portfolio share for both the assets and liabilities side and then we look at their distribution. Finally, we set thresholds on their portfolio composition and highlight the number of vehicles that have the following characteristics:

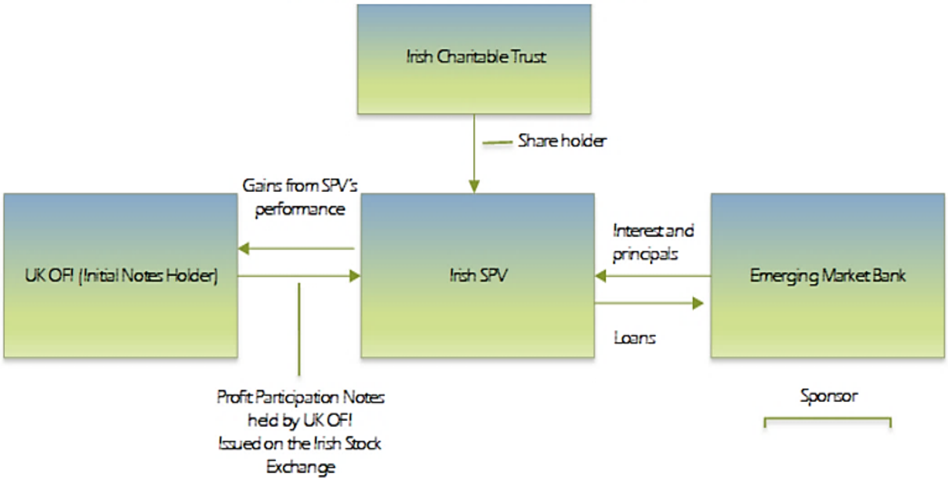
- Foreign assets and liabilities over total assets > 80 percent (pass-through);
- Foreign assets and domestic liabilities over total assets > 80 percent (mixed ass_liab);
- Domestic assets and foreign liabilities over total assets > 80 percent (mixed liab_ass);
- Domestic asset and liabilities over total assets > 80 percent (domestic)

Particular attention is given to the breakdown of domestic versus foreign sponsored entities, in order to understand whether the nature of the sponsor matters for the cross-border activities of the SPEs.

<u>External financing model</u>

This case study outlines a structure in which an Irish SPE is set up by a foreign company (e.g. emerging market bank) to obtain funding from third parties. The Irish SPE obtains financing from the issuance of profit participation notes backed by the sponsor. The proceeds of these notes are passed from the Irish SPE to the sponsor as a loan. The SPE is set up as a stand-alone entity outside of the sponsor’s group structure, with a charity or other body technically owning the SPE so that assets are not accessible by the bank.

Figure 2: External financing model



This framework outlines clearly the challenges in identifying control, its foreign and domestic nature and thus the concept of an institutional unit. The domestic legal owner does not exert control. Benefits are likely to be gained abroad and shared between the sponsor and the note holders.

However, the analysis of the sub-sample of SPEs classified as External Financing, suggests that in a number of cases the sponsor is a domestic NFC and its balance sheet is exposed to foreign entities.

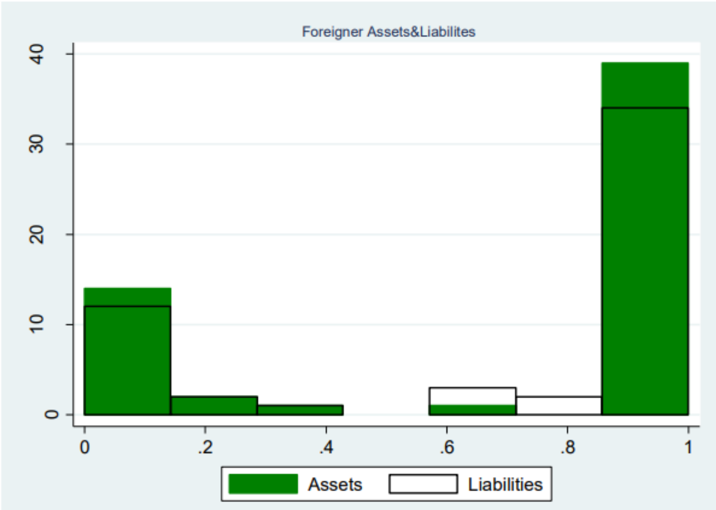
This would suggest that distinguishing between foreign and domestic controlled entities and consolidating domestic ones may not only be challenging but may also prevent distinguishing flows with a nature different from the functional category in which they are incorporated.

In Q4 2018, 64 SPEs (€56.6 billion) representing more than 34 percent of the total asset value of non-securitisation SPEs, were classified as External Financing[1]. The bulk of them were sponsored by foreign banks and NFCs. NFCs resident in Luxemburg and Ireland were also important players. 29 entities (€33.6 billion) – mostly NFCs - reported to be consolidated in a group and provided information on the ultimate parent. The ultimate parent and the sponsor were likely to be the same entity. Looking at the external financing SPEs portfolio composition, we find that on average 71 percent of their assets and 69 percent of their liabilities are foreign.[2] Most importantly, if we look at the distribution of foreign assets and liabilities over total assets, it is clear that their behaviour is extremely polarised and that the average is quite

[1] In the Central Bank survey, *external financing vehicles* are defined as *funding obtained from external sources, furthered as a loan to the parent.*

[2] As mall of asset and liability, positions are not disaggregated by country on the reporting form. Therefore the sum of foreign and domestic position as a share of total assets/liabilities may not sum to 100 per cent.

uninformative (see Figure 3)[3]. Looking at the thresholds on the portfolio composition suggests that the most common strategy is *pass-through* (26 vehicles), followed by *mixed_ass_liab* (8 vehicles). Irish sponsored vehicles are mainly *mixed_liab_ass*, investing the resources domestically and raising resources abroad.

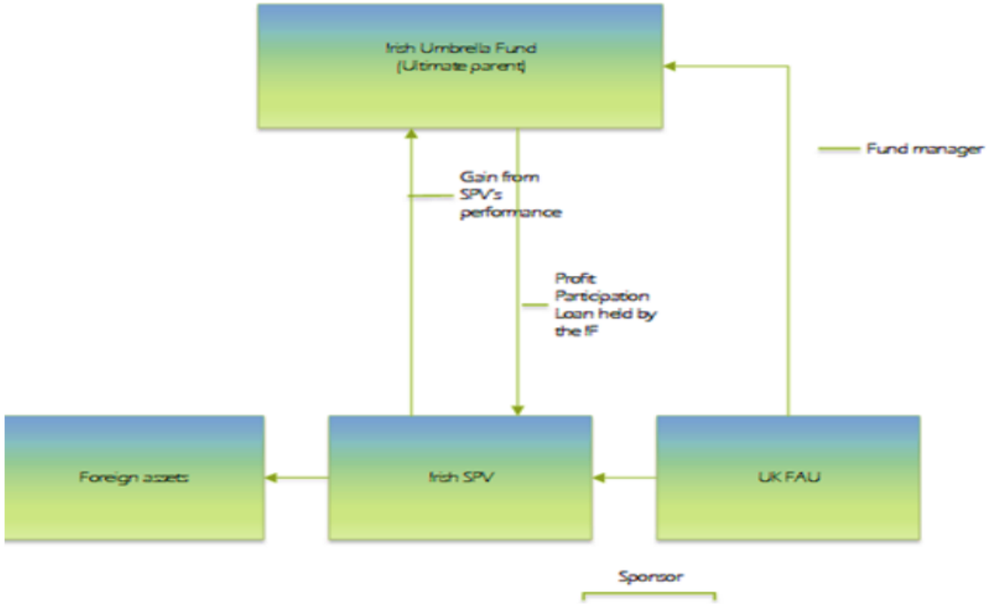


Investment fund linked

This case study outlines a structure in which an Irish SPE is set up for tax efficiency. The sponsor of the Irish SPE is a foreign investment manager of an Irish-resident investment fund. These fund units are owned by foreign investors. The Irish SPE receives a profit participating loan from the Irish investment fund and uses the proceeds to purchase foreign debt securities. In return, the fund receives interest dependent on the performance of the SPE's assets. The Irish SPE is wholly owned by the Irish resident fund and declares to be consolidated into its accounts.

[3] Assets and liabilities are analysed separately. Columns that overlap at the higher end, do not translate into pass-through entities.

Figure 4: Investment fund linked



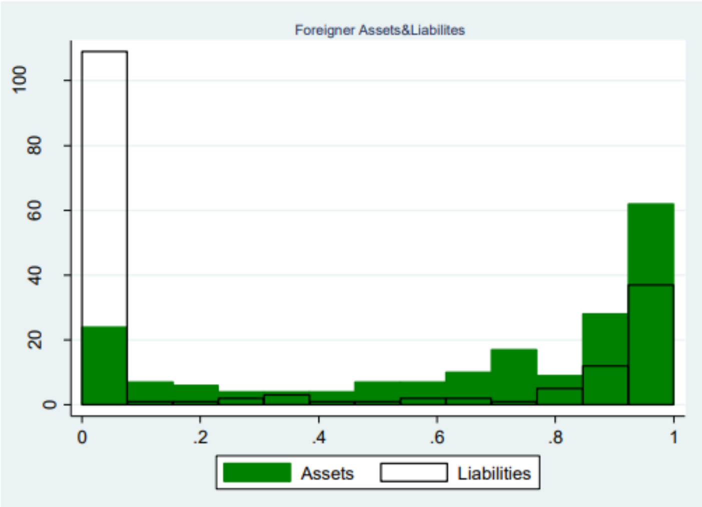
This example highlights the challenges that SPE structures pose to compilers of international accounts for identifying the residency of the controller. For this specific investment-fund linked structure, the sponsor (i.e. the foreign fund manager) is the entity that exerts control. However, should we consider the Irish investment fund as the owner, the SPE would be classified as domestically controlled and thus consolidated into the fund. As in the previous case, the analysis of the data for the entire sub-set of investment fund linked SPEs shows no clear difference in the balance sheet composition of foreign and domestically owned entities, suggesting that capturing this feature may limit the scope of the definition.

In Q4 2018, 233 SPVs (€83.5 billion) representing more than 34 percent of the total asset value, were classified as investment fund linked[4]. Only 32 SPEs were set up by Irish entities (€6.8 billion). Sponsors tend to be UK and US fund managers. 103 SPEs (€39 billion) reported as being consolidated into a group and provided information on the ultimate parent. In contrast to the previous case, the ultimate parent and the sponsor are very likely to be two different entities. Foreign fund managers are the sponsors, while Irish investment funds are the ultimate parent. Looking at their portfolio composition, we find that on average 66 percent of their assets and 32 percent of their liabilities are foreign. Figure 5 shows the distribution. In this business model, foreign and domestic sponsored SPVs behave similarly. Most of them have a very small portion of foreign liabilities, and a mix of domestic and foreign assets

[4] In the Central Bank survey, *Investment fund linked* are defined as *Vehicles linked to investment funds, which hold debt, equity, loans, or other financial assets with the goal of capital appreciation or dividend income*.

composition. A common strategy does not stand out. Many of them do not fall into any of the extreme categories we created (threshold > 0.8). It is worth noting that the frequency of pass-through is similar for foreign and domestic sponsored entities.

Figure 5: Distribution of foreign assets and abilities



4. Conclusion

Control and residency are key concepts for BoP. Due to their complex structure, however, SPEs pose challenges for the BoP community and the implementation of the TF-SPE definition. In this paper, we have shown that sponsor and parent are not always the same entity. In almost half of the reporting population, the legal and economic owner differ.

With this in mind, a key challenge going forward is to identify the proper proxy for control. If we believe that the sponsor is the best option, then we would need to clarify to which extent domestic sponsored entities have autonomy of decision. If they are not intuitional units, we would need to clarify whether their account should be consolidated into the sponsor or the parent. In this regard, it is also worth mentioning that the analysis did not identify a clear difference in the cross-border activities of domestic and foreign sponsored entities. This may suggest that consolidating SPEs into the account of a domestic entity would undermine the purpose of the definition itself, blurring the concept of FDI and financial intermediation. Finally, the analysis of the thresholds on their portfolios shows very diverse strategies, suggesting that considering only SPEs who transact almost entirely with non-residents may greatly reduce the number of entities included.

References

1. Barrett D., Godfrey B. & Brian Golden. "*New Data Collection on Special Purpose Vehicles in Ireland: Initial Findings and Measuring Shadow Banking*". Quarterly bulletin. Central Bank of Ireland.
2. Golden Brian & Hughes Patrick (2018). "*Shining a Light on Special Purpose Entities in Ireland*". Economic Letter. Central Bank of Ireland.
3. IMF (2018). "*Final Report of the Task Force on Special Purpose Entities*"

Linked administrative sources for the New Zealand census and population statistics

Christine Bycroft, Pubudu Senanayake
Stats NZ Christchurch, New Zealand

Abstract

Statistics NZ is beginning to see the benefits of a programme of research into greater use of administrative data for the population census and the wider population statistics system. The research effort has been driven by a need to modernise and reduce costs of the current full-enumeration census model, and the opportunities afforded by a rich and growing set of integrated administrative data. Development of new statistical models is integral to maximising the use of administrative sources. Included in this data mix are travel journey data for almost all border crossings – a rare advantage afforded by being an island country. A new measure of external migration has been developed based on actual time spent in NZ, rather than the previous approach that used intentions stated on arrival and departure cards. While NZ does not have an administrative population register, or any mandatory requirement to register an address with authorities, we have developed a statistical admin resident population based on a 'signs of life' approach. This work has proved to be a significant advantage given an unanticipated level of non-response in the most recent full enumeration census in 2018.

Keywords

population estimation; linked administrative data; external migration; census

1. Introduction

Administrative data have always been part of the production of Stats NZ's official population statistics, however, the ways in which administrative data are being used is changing. These changes are taking advantage of opportunities afforded by the integration of administrative sources which were formerly only available separately, and the development of new statistical methodologies that are needed to take full advantage of the linked data. As well as these opportunities, change has been driven by concerns about cost pressures and the sustainability of the current full enumeration census.

Population statistics in New Zealand are based on a five-yearly full field enumeration census. The census data are released in their own right and provide a rich array of social and economic information for small sub-groups and small geographies. A separate official Estimated Resident Population (ERP) series is the best measure of population living in New Zealand at a given time

(Statistics NZ, 2014a). The ERP adjusts for net undercount in the census, as measured by a coverage survey (Statistics NZ, 2014b). Population change from census day applies a cohort component method using high quality Birth and Death registrations, and measures of external migration.

New Zealand is fortunate in having excellent information sources on external migration. We are an island country with a record of border crossings that includes almost all travel journeys into and out of the country. Measures of external migration have until recently used passengers' stated intentions on arrival or departure to differentiate short-term travellers from long-term movements (migrants). In practice, intentions do not always correspond to actual length of stay in New Zealand or overseas.

Stats NZ committed to exploring the use of administrative data for the census in its Census Transformation Strategy (Statistics NZ, 2012). The strategy consists of two parallel strands. The first is a short to medium term focus on modernising the current full enumeration census, with the first implementation in the 2018 Census of Population and Dwellings. The second strand is investigating the feasibility of a census largely based on administrative data in the long-term. The two strands are inter-dependent and the 2018 Census includes a goal to increase the use of administrative data. The main use of administrative data in 2018 was to construct an address frame from administrative sources, which has developed some of the organisational infrastructure required for long-term change. The 2018 Census strategy also included use of alternative sources to improve data quality in the context of missing data for census questions, based on the findings of the longer-term administrative data research. The value of this joint strategy has become clear in the light of lower than anticipated response rates in 2018, which has considerably extended the role that administrative data now plays in the 2018 Census.

2. Census Transformation programme

Stats NZ's Census Transformation programme is exploring the feasibility of a census based largely on administrative data, and supported by sample surveys. Bycroft (2015) describes the administrative data landscape in New Zealand, and progress producing an administrative resident population to that point. Stats NZ reports to government periodically, outlining progress and seeking decisions on future direction. In a 2015 Cabinet paper government agreed that Stats NZ actively work towards a future census based primarily on government administrative data, supported by redevelopment of its household surveys. The linked administrative data sources available in Stats NZ's Integrated Data Infrastructure (IDI) have been the basis for this research. The IDI provides safe access to anonymised linked microdata for research and statistics in the public interest. The basic structure of the IDI consists of a central population list or 'spine' to which a series of data collections are linked. The IDI spine forms the conceptual centre of the IDI; other datasets are linked to it through an anonymised identifier.

For an administrative-based census, we aim to derive a list of people who are resident within New Zealand at a given point in time, without relying on a full enumeration census. We also need to determine where in New Zealand these individuals live. Gibb, Bycroft, and Matheson-Dunning (2016), and Stats NZ (2016, 2017a) describe the progressive development of a 'signs of life' approach. Activity in New Zealand as reflected in administrative data sources during a two-year window is used to indicate an individual's presence in New Zealand. Anyone who had died or migrated overseas before the reference date is removed. Geographic location is derived from address information sourced from multiple agencies. The *resulting admin NZ resident population* derived from the IDI is called the IDI-ERP, and is currently implemented in the IDI for use by researchers. This IDI-ERP population is also the source of administrative records for inclusion in the 2018 Census file.

In 2016, we released an experimental data series of national-level administrative population estimates, with further releases for a subnational geography time-series in 2017, and inclusion of ethnic groups in 2018. The experimental data series include estimates at 30 June from 2006 to 2016. Online tables compare the IDI-ERP with official population estimates over the same period. These comparisons are largely encouraging. Often there is close agreement with official figures, and consistency has increased steadily over time. However, there are still marked differences for some age groups and local areas.

While the IDI-ERP is a good approximation of the NZ resident population, it includes an unknown group of erroneous inclusions, and also misses some people. These coverage errors make it more challenging to derive an admin-based population estimate sufficiently accurate for official statistics, to low levels of geography. We are developing new population estimation models combined with a single coverage survey that will adjust for over-coverage in administrative sources as well as for under-coverage. We are also developing methods to adjust for the mis-classification of admin-based location information.

As well, a structured quality assessment of census variables derived from linked admin data sources has shown where administrative data has most to offer the census. The Census Transformation programme continues to publish results of investigations that compare 2013 Census variables with their counterparts derived from administrative sources. These include papers on income, educational qualifications, ethnicity, Maori population identifiers, households and families, and housing variables.

3. The 2018 Census

The 2018 Census strategy (Statistics NZ, 2016) set out an ambitious modernisation programme across all components of census taking. This followed several censuses of minimal content change, and limited innovation. Improvements to adjust for missing census characteristics data were signalled in the 2018 Census strategy. These included the use of alternative sources (the previous 2013 Census and good quality administrative sources), and more extensive use of statistical imputation.

While some aspects of the 2018 modernisation have been successful, major challenges were faced when implementing the new collection model, and the overall level of response was lower than expected. Once these collection issues became clear in mid-2018, Stats NZ embarked on development of new methods for including people in the 2018 Census dataset based on administrative data. The new methods have built on the research undertaken by the longer-term census transformation investigations, and fast-forwarded their development for use in the context of a full enumeration census.

The use of alternative sources for census characteristics has taken on greater significance than originally anticipated given an increased number of partial census responses and the contribution of administrative records to the final census dataset.

4. Methods for administrative enumerations in the 2018 Census

The admin NZ resident population, the IDI-ERP described above, is the source for including administrative records in the 2018 Census dataset. The IDI-ERP provides an initial admin list of people who can be considered for inclusion in the census dataset. We first remove NZ residents temporarily overseas on census night, as they are not included in the NZ census population definition, and remove those for whom we already have a census response. We also recognise limitations in the quality of administrative sources when determining who should be placed into the census file. When faced with choices of how to implement methods and where to set model parameters, decisions have been conservative.

Admin enumerations have been added to the census dataset when they will improve the overall quality of the census data. Administrative households are formed when have good evidence for improving census household information. Otherwise, admin enumerations are not placed in dwellings, but are placed within a small geographic area (a meshblock, up to 100 people), when we can do so while improving the quality of census small area data. We now outline the statistical methodologies used to include high quality admin enumerations in the census file.

Linking the census and administrative data

We link 2018 Census respondents to the same person in the IDI spine, so that we can remove them from the IDI-ERP, leaving only those who did not respond. This requires a high linkage rate and accurate linkages. Since New Zealand does not have a common identifier, probabilistic linkage methods are applied. The overall linkage rate of 97.7 percent is high in the NZ context. Census respondents who have not been linked to the IDI spine are a mix of those who:

- should have been matched to the IDI spine but were not (a missed or 'false negative' match)
- are not in the IDI spine (and therefore the non-match is correct)

The rate of missed matches has been estimated as 1.4 percent. False positive matches (when different people are incorrectly linked) are estimated as being less than 1 percent of the links made.

Admin enumerations in dwellings and households

The first and most demanding use of administrative data is the placement of groups of people within a dwelling to form households. For private dwellings where no census responses have been received, a statistical model has been developed to predict which households constructed from admin records are likely to have reliable data. The approach is based on methodology developed by the US Bureau of the Census who have a planned strategy to use admin enumerations in the non-response follow-up phase for their 2020 Census (US Bureau of the Census, 2017). The model produces a score that represents how reliable the administrative data is for representing the entire household in a given dwelling. 2018 Census responding households are assumed to represent the truth when training and assessing the model. The cut-off has been set as a balance between strict criteria of obtaining exactly the same people in the household as we observe in the census, and including admin households that reflect similar adult-child patterns as the census, even if we cannot guarantee that all household members are the same. Making the trade-off in this way means we include relatively more large or complex households than if we had set a more conservative cut-off, and makes some allowance for errors in census responding households.

While households where we have received some census responses may still be missing people, we have not developed a model to predict when admin records ostensibly for the same address should be placed within those responding households.

Admin enumerations in meshblocks

After deriving admin households for non-responding dwellings where possible, the next phase places admin people into the smallest New Zealand geographic area (meshblocks). The admin address ID provides an x,y location coordinate, but the person will be excluded from census households.

The remaining IDI-ERP population eligible for inclusion in meshblocks are those who have not been linked to a census respondent, and are not already included at a dwelling. We then explicitly adjust for potential coverage errors that could result from over-coverage in the IDI-ERP, and for duplication caused by missing linkages between the census and the IDI. We draw on new methods Stats NZ have developed to remove over-coverage and adjust for missed matches in the context of a new Dual System Estimation (DSE) population estimate based on the same linked census and IDI-ERP administrative population sources. DSE is a well-established methodology used in population estimation, often applied in the context of a full enumeration census and a coverage survey. Over-coverage is typically avoided by design of the coverage sample, and may be removed from the census, and extensive checks are made to meet the DSE assumption of no linkage error. Neither of these approaches are possible in this situation.

Removing over-coverage in the IDI-ERP: Our approach is designed to remove over-coverage caused by people selected in the IDI-ERP who are not NZ residents at the time of the census. As we cannot directly find over-coverage, we will also remove valid records. The goal is to effectively remove as much over-coverage as possible, while also minimising the amount of under-coverage being introduced. We apply a more rigorous selection criteria for inclusion, whereby people must have activity in at least two data sources: tax or health, plus one other. Tax and health datasets have high coverage of the population, and requiring an additional activity provides stronger evidence that people are in fact residents. This approach targets all age-sex groups, combines factors found to predict over-coverage in previous research, and is easy to apply and explain.

Accounting for missed linkages: We also adjust for duplicates caused by missed matches when we link the census file and the IDI spine. If we have not linked a census record to the IDI spine when we should have, it will still be in the eligible IDI-ERP dataset, and would be counted again. We are matching around 4 million census records to an IDI spine consisting of nearly 10 million records. We do not at present have the time or resources for checking even a sample to estimate missed matches.

We estimate missed matches on the basis of responses to census questions that closely match the criteria for inclusion in the IDI spine. A subset of the census, M*, is created through applying strong requirements of membership in two or more of the datasets that make up the IDI spine (Choi, 2019). Those in the M* census subset should all have been linked to the IDI spine, and those who have not been linked are missed matches. We obtain an estimate of 1.4 percent of the M* subset who were incorrectly not linked to the IDI spine, with rates available by strata (age, sex, geographic areas, and ethnic groups). We then assume that same rate of missed matches applies to the remainder of the census file. The appropriate number of people are removed through random selection within strata. This adjustment does not remove exactly the people whose census record should have matched to the IDI. Rather it removes a random selection of the right kind of people to a fine demographic breakdown.

Accounting for the quality of admin location data: After these adjustments we can be confident that the remaining eligible IDI-ERP people should have been counted by the census. However, we also consider the accuracy of admin location data, which decreases as geographies get smaller. To limit the errors for small sub-national geographies, we remove people who have a low probability of a correct meshblock location. The trade-off here is between including more individuals in the census dataset, and protecting the integrity of small area geographies. We include people for whom there is at least a 50 percent probability that we have their correct meshblock. This meshblock cut-off is the main driver of which admin people are included or excluded from the census file.

Previous censuses have applied statistical imputation methods to count some of the missed respondents in the final census dataset. The benefits of admin enumerations over imputation methods are evident in the 2018 results, and in comparisons with the 2013 Census. The IDI-ERP admin population does include people who are traditionally hard to count in a census. Including high quality admin enumerations does better at improving the census distributions than imputation methods that relied on local area missing at random assumptions, given the disproportionate nature of census non-respondents. In addition, the admin enumerations are real people for whom we may have associated characteristics from alternative sources.

5. New outcome-based measures of external migration

In New Zealand, the largest component of population change since the 2013 Census has been external migration. A reliable measure of migration is critical in ensuring a consistent and credible resident population estimate. A unit record level estimation of migration allows for the removal of people present within the administrative sources who have departed New Zealand, and addition of those who arrived into New Zealand as migrants. Therefore, with an individual level measure of migration, the largest component of population change can be coupled directly to the source data of population estimation.

Until recently, migration flows were largely based on the responses provided by travellers on their passenger cards when crossing the border. The intention to stay in NZ (or leave NZ) for 12 months or more, as reported by the traveller, was used to determine whether the border-crossing was a migrant crossing. This measure allowed New Zealand to produce some of the most timely migration statistics in the world. However inter-censal discrepancies, and analysis involving total traveller in-flows and out-flows showed that the measure was generating inaccuracies in the migration estimates, particularly between 2001 and 2006, where net migration was likely under-estimated by approximately 50,000 (Stats NZ 2017b) over the five years.

This led to the development of an outcomes-based measure of migration, where, rather than a traveller's stated intentions, the amount of time they spent in or out of New Zealand was used to determine migration status. Using passport data acquired as travellers cross the New Zealand border, a travel history of almost every traveller is created. This gives us a longitudinal 'register' of travel histories for individuals. By applying a classification rule to these travel histories, the migrant status of any given border-crossing can be classified. In New Zealand, the particular classification method is called the '12/16 month rule' (Stats NZ 2017b). For example, an arrival who has been out of the country for at least 12 months in the 16 months prior to their border-crossing, is classified as a migrant arrival if they spend 12 months or more in New Zealand over the 16 months *following* the border-crossing. This method estimates migration more accurately, as we are no longer reliant on travellers' self-reported intentions.

As with any administrative data, identity resolution is a challenge, especially given the number of border-crossings that we must resolve. Currently, our longitudinal administrative dataset extends back to 2013, with over 65 million border-crossings that require identity resolution to ascertain border-crossings by the same traveller. On average, 40,000 border-crossings are added each day to this dataset. Within these border-crossings, we estimate over 17 million unique travellers, based on data available on passport records since 2013. The resolution of these individuals provides the basis of a register of travellers.

While the outcomes-based measure provides more accurate measures of migration, it requires a wait of 16 months before migration levels are known with certainty. Official external migration estimates have been published monthly, within a month of the reference date. To maintain timeliness with the 12/16 method, we have recently developed a predictive classification model, which provides provisional estimates of migration for the latest 16 months.

This is a difficult prediction problem because there is a large class imbalance between overall traveller numbers and migrants (who make up less than 2% of all border-crossings). We have employed a machine learning approach, learning at the unit record level, combined with a multiple imputation prediction method, also at a unit record level, to produce migration statistics with associated uncertainties. These new migration estimates are revised each month to include the latest information.

Estimation at unit record level is beneficial as it allows us to build up any required aggregations. Importantly, it also allows the model to easily run across linked administrative datasets, such as those contained within the IDI. This means we can then estimate migration patterns of a subset of people with a given set of attributes available in the linked data.

6. Discussion

The 2018 Census dataset consists of people counted through responses to census forms plus those counted through administrative records. Stats NZ is now confident that it has compiled a census dataset that will provide census usually resident population counts and electoral counts of acceptable quality. The improved accuracy of measures of external migration will provide greater accuracy for population measurement in between censuses. The value of the census questionnaire component is evident for some census variables where no alternative sources are available. These variables consequently have higher levels of missing data in 2018.

Our experience with the 2018 Census where we faced unacceptably low response rates has highlighted the value of undertaking the longer-term research into an admin-based census. Being well-prepared to use administrative data to supplement responses, if necessary, provides a welcome risk management strategy.

It has become increasingly clear that a transition towards more effective use of administrative data in the census is best implemented as a step-wise approach based on a well-designed programme of research. Gaining the benefits of that research does not need to wait until conditions are ready to implement a full admin-based census, but can add considerable value through supporting the current census.

A statistical approach is essential to effective use of administrative data for official statistics. Requirements include the ability to safely access linked data sources, and statistical models developed for various estimation processes. The New Zealand outcome-based external migration measures are a combination of deterministic calculated migration estimates and modelled estimates with associated uncertainty measures. Switching to modelled migration estimates for the most recent periods (and away from intentions-based measures) replaces a perceived precise, but potentially biased measure, with a more variable measure where the uncertainties can be transparently defined. From an end-user perspective, this is a big change, especially dealing with substantial revisions for the most recent month when uncertainty is greatest. However, these approaches prepare us for a world where increasingly administrative data, and models that sit atop such datasets, will be required for demographic estimation, and projection.

The 2018 Census example is an intermediate step that combines field enumeration and administrative data. However it does not estimate the remaining missing data, nor provide uncertainty measures, which are produced through a coverage survey. We are working towards a model-based framework that would produce a full census population unit record file, along with measures of uncertainty.

The direction that the NZ Census will take over the next decade is unclear. We expect the next census in 2023 will continue the transition towards greater use of administrative data in support of the full field enumeration census. No decisions have been made on what will happen after 2023. However, administrative data will play an increasingly important role in future censuses and population statistics.

References

1. Bycroft C. (2015) Census transformation in New Zealand: Using administrative data without a population register. *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics*. 2015;31(3):401-411.
2. Choi, H. (2019) Adjusting for linkage errors to analyse coverage of the administrative population *Statistical Journal of the IAOS* vol. Pre-press, no. Pre-press, pp. 1-7, 2019
3. Gibb, S., C. Bycroft, and N. Matheson-Dunning. 2016. *Identifying the New Zealand Resident Population in the Integrated Data Infrastructure (IDI)*. Retrieved from www.stats.govt.nz
4. Statistics New Zealand (2012). *Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy.* Wellington: Statistics New Zealand, 2012.
5. Statistics New Zealand (2014a). *Estimated resident population 2013: Data sources and methods.* Available from www.archive.stats.govt.nz
6. Statistics New Zealand (2014b). *Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration Survey.* Available from www.archive.stats.govt.nz
7. Statistics New Zealand (2016). *2018 Census strategy.* Retrieved from www.stats.govt.nz
8. Stats NZ (2017a). *Experimental population estimates from linked administrative data: 2017 release*. Retrieved from www.stats.govt.nz
9. Stats NZ (2017b). *Outcomes versus intentions: Measuring migration based on travel histories.* Retrieved from www.stats.govt.nz
10. Stats NZ (2017c). *Defining migrants using travel histories and the '12/16-month rule'.* Retrieved from www.stats.govt.nz

Transforming population and migration statistics using admin data: Progress on the England and Wales Transformation Journey

Jay Lindop, Becky Tinsley
Office for National Statistics, England and Wales

Abstract
The Office for National Statistics (ONS) has been using integrated data to pursue the UK Government ambition that 'censuses after 2021 will be based on alternative sources of data.' ONS is working with other National Statistics Institutes (NSIs) which are also looking at the use of administrative data for their censuses. The Administrative Data Census project has made progress in: producing a range of Research Outputs that are typically produced by the ten-yearly census; comparing these outputs with official statistics; and seeking feedback from users. A key milestone in this ambitious programme of work is to put administrative data at the core of population and migration statistics for England and Wales by Spring 2020. This paper summarises progress towards this milestone, where the aim is to produce a coherent set of estimates for both population size (stocks) and components of population change (flows), highlighting methodological challenges and our approach to tackle them in order to deliver this transformational change.

Keywords
Population statistics; Migration statistics; Administrative data; Data integration; Transformation

1. Introduction

The Office for National Statistics (ONS) is transforming the way we produce population and migration statistics, to better meet the needs of our users. Working in partnership across the Government Statistical Service (GSS), we are progressing a programme of work to put administrative data[1] at the core of our evidence on international migration (UK) and on population (England and Wales) by 2020[2]. This ambition is based on our current plans for acquiring access to the further administrative data sources we need to deliver this. Our work programme is also an integral part of the work over the next four years to make a recommendation to the UK Government in 2023 about the future of population and housing censuses in England and Wales.

[1] Administrative data refers to data collected by other organisations (such as government departments) to support the delivery of services or for other operational purposes.

[2] International migration is a reserved policy area whereas population is devolved so ONS official statistics cover migration at a UK level and population for England and Wales. See section 2 which explains the different systems for measuring the population in Northern Ireland and Scotland.

The published report provides an update on our GSS Migration statistics transformation programme and builds on the previous research delivered through this and the ONS Administrative Data Census project. Previous publications have set out our progress in linking multiple data sources together to produce estimates of the size of the population. We have also carried out targeted work to better understand international migration, including reports into student migration and work to compare the International Passenger Survey (IPS) and Home Office visa data for non-EU migrants.

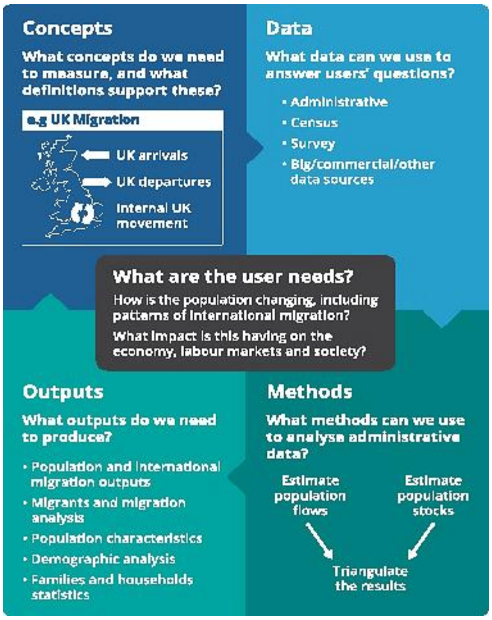
ONS has long acknowledged that the International Passenger Survey (IPS), which is currently used to estimate migration to and from the UK, has been stretched beyond its original purpose and that we need to consider all available sources to fully understand international migration. At the same time, our previous research clearly demonstrated that no single source of information can tell us everything our users want to know, or fully reflect the complexity of our changing population. Instead it has shown the value that can be gained from using linked administrative data, while highlighting the challenges of using this data to measure traditional definitions such as short- and long-term migration, and usual residence.

A report, published in January 2019, brings together our transformation work on population and migration, building on our knowledge and investigating how the administrative data sources now available to ONS[3] can be used, alongside surveys, to improve the way we measure population stocks and flows in the future.

We are still developing our future system and are in the process of acquiring the further administrative data sources needed to deliver this – for example, to address coverage gaps for EU migration. As such, this report does not directly compare administrative data with our existing IPS-based migration statistics or make any overall assessments of their statistical quality. Instead, it provides an update on our approach towards building an administrative data-based system that will, over time, give us new insights on the quality of the IPS and our official international migration estimates for the UK. It also provides evidence of how different administrative sources can help us to better understand aspects of migration and reveals some of the different travel patterns that migrants make, such as circular patterns of movement.

[3] The Digital Economy Act 2017 provides new opportunities for ONS to access existing data held by other government departments, for the purpose of producing research and statistics.

Figure 1: Our draft framework for transforming population and migration statistics



As shown in Figure 1, we have also developed a new draft framework that puts users at the heart of what we do and describes the important questions we need to answer to meet their needs: What concepts do we need to measure and what definitions support these? What data can we use to answer users' questions? What methods can we use to analyse administrative data? What outputs do we need to produce? The published report is structured around this framework and is summarised in this paper.

2. Methodology

As set out in our framework, once we have identified the right data sources to deliver a future administrative-data based system, we need to develop the right methods for using them to measure what users need. Our aim is to integrate administrative data sources in a way that creates a flexible system for producing the range of insights our users need – aligned with ONS principles for how we manage and secure data for public benefit. This is both in terms of putting administrative data at the core of our official statistics for population and migration in future, but also strengthening of the evidence base on important areas such as the impact of international migration on society and the economy.

Using administrative data to develop a stocks-based approach

Our previous Administrative Data Census research focused on our progress in developing a future **stocks-based approach** where we combine linked administrative data and applied a set of rules to produce a Statistical Population Dataset (SPD). This work demonstrated the potential for producing estimates of the usually resident population using administrative data, but early research showed the need for further refinements to produce estimates about the flows of the population between two points in time. It has also highlighted the importance of using a greater range of data sources and the need for a survey which can measure and adjust for coverage issues seen for different areas and different groups in the population.

Using the knowledge that we have developed from analysing our SPDv2 and understanding key data sources, we have developed a set of data-driven rules that we can use as part of a future system for determining which administrative records are part of the usually resident population. This approach focuses on identifying the data source that provides the best coverage for a given age group ('first hierarchy'). We then supplement any gaps in coverage, or limitations of that source by using other sources, to create a 'hierarchy' of rules ('second hierarchy' etc.). Developing rules in this way will enable us to be flexible as new administrative sources become available, or as sources change over time.

We have developed an initial approach for specific age groups based on our understanding of the coverage and quality of each data source. As an example, our work on the 5-15 age group illustrates this approach. We have used the England and Wales School Census as our 'single source' as this provides high coverage of children attending state schools, which represents a large proportion of this age group. We then supplement this with 'activity' from Department for Work and Pensions (DWP) Child Benefit data and have included records from NHS Patient Register (PR) to account for gaps in coverage (for example, children attending private schools and home-schooled children who are not covered by the School Census data).

Using administrative data to develop a flows-based approach

In the existing system for producing official estimates of the England and Wales population, we use a cohort component method. Our starting point is the ten-yearly census. Each year, we age everyone on, use administrative data to add births, remove deaths and make adjustments for internal migration. We use the IPS to estimate international migration flows (people immigrating into the UK, and those emigrating from the UK), and distribute that to local levels using administrative data. Once these flows have been added to the previous year's stock total, we are able to produce a stock total for the current year. This can be thought of as a **flows-based approach**. Our ambition is to make far greater use of administrative data to produce these national estimates of international migration in future, so we have carried out new research to explore how we can develop a flows-based approach based on wider data sources.

Our published research report outlined the challenges in measuring and understanding how the population is changing over time, particularly for immigration and emigration (flows, both nationally and locally).

We are confident that births and deaths are well recorded in administrative sources. For long-term immigrants and emigrants, this is more challenging as no single data source captures the patterns of movement for all types of migrants. Like the data-driven rules described in the stocks-based approach, we are developing a similar inclusion approach for long-term immigrants - we are calling this confidence-based rules.

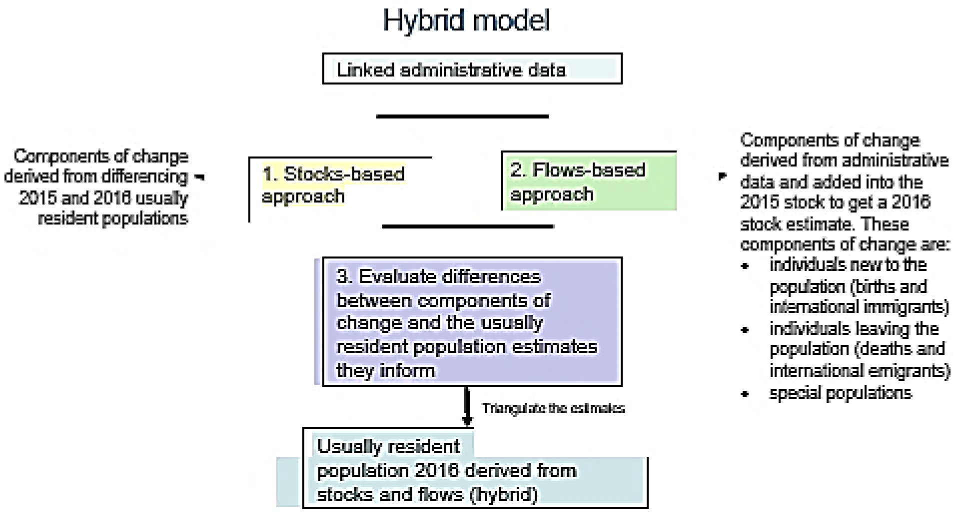
Our aim is to bring together the multiple sources to build a comprehensive and granular evidence base for migration to (and eventually from) the UK.

Bringing it together: our proposed hybrid model

Focusing on these approaches separately offers us the opportunity to produce the best-possible estimates for stocks and flows in future, using the best data and methods available to us. However, to produce a coherent set of statistics for our users requires us to develop an approach that brings the two methods together.

Figure 2 sets out our hybrid model for delivering a transformed population and migration statistics system. We call this a hybrid model to reflect that we are aiming for an approach that produces the best possible stocks and flows of the population.

Figure 2: Our proposed hybrid model



Once we've produced a set of statistics from both approaches, we need to evaluate the stocks and flows that we have produced and triangulate them to enable us to produce a coherent set of population and migration statistics. This is likely to involve the need for a survey to measure and adjust for coverage patterns seen in the data (we call this a 'Population Coverage Survey' (PCS)).

We have updated on our progress about developing a PCS in our previous publications. As part of our wider survey transformation work, we are looking at how we can integrate the PCS with the Labour Market Survey and other residual data requirements into an Integrated Survey Framework. This model would provide other vital characteristics of the population, which, along with information from administrative sources, will help us shed light into the impact of different groups of population on society and the economy.

We will ensure that we develop the hybrid model, described above, to be as flexible as possible, to enable new sources and methods to be used as they become available, or as they change over time. This approach also opens up the potential for providing a longitudinal aspect into better understanding the dynamics of population change. This could give us the chance to offer more insights into key policy and research areas, such as the economic outcomes of international migrants depending on length of stay and age at arrival.

3. Result

As noted above, we have previously produced admin-based estimates of the size of the population by linking four key data sources[4]. We applied a set of rules to include records into a Statistical Population Dataset (SPD) to represent the 'usually resident' population. Since then, we have shown that our previous SPD rule of requiring records to appear on at least two of our four key data sources results in varying coverage patterns for certain age groups or geographies.

We have now made progress towards a new approach for producing population stocks and flows using administrative data, by bringing more sources together to fill gaps in coverage. We have linked immigration, education, health and income records, and have explored how we can use these sources to determine the usually resident population of England and Wales and immigration flows to the UK. This includes developing data driven rules, based on registrations and 'signs of activity' we can identify from each data source.

[4] NHS Patient Register, DWP/HMRC Customer Information System, England and Wales School Censuses, Higher Education Statistics Agency student data

We have also produced a series of case studies that put a spotlight on what different administrative data sources tell us about international migration.

This latest research has clearly demonstrated the benefits of combining multiple sources to provide new insights into migration. The key points are:

- Our new analysis of circular patterns of movement using Home Office Exit Checks data clearly demonstrates the complexity of the travel patterns we can see in the data. Looking at individuals who arrived in the UK on a non-visit visa and their travel patterns for the following two-year period, we were able to identify a range of circular journeys into and out of the country, categorise these into groups and look at their characteristics. For example, those we defined as having a low or medium number of journeys tended to be here for around 2-5 months and travelled for the purposes of study or family. There is therefore potential to produce statistics on circular migration in future, so we will explore how we can do this based on feedback from our users about what aspects are most important to them.
- Our improved understanding of what administrative data can tell us about migration from the EU, building on previous work which focused mainly on non-EU citizens. For example, when we linked NHS Personal Demographic Service (PDS) data to the Migrant Worker Scan (MWS), we found median lags between arrival and NHS registration of 276 days for EU nationals and 60 days for non-EU nationals in our linked dataset. This indicates that those registering for a National Insurance Number (NINo) do not tend to access health services immediately. For EU nationals it also emphasises that wider data sources will be particularly important for identifying migration into the UK, given the time lags in the health data.
- New insights into international student migration by linking Higher Education Statistics Agency (HESA) data with Home Office Exit Checks and Her Majesty's Revenue and Customs (HMRC) Pay as You Earn (PAYE) data. We found that a greater proportion of EU students in our linked HESA and PAYE cohort were employed compared to non-EU students – which likely reflects immigration rules and the economic background of students – and also provides evidence that our move to bring more sources into our SPD will help improve our coverage for non-EU students. By linking HESA and Exit Checks data, we also found that almost half of non-EU students in our linked data spend between 300-400 days in England and Wales during their first 14 months of study within a 16-month period. Our existing definition of long-term migration counts those who spend 365 days in England and Wales as resident, so this work shows the need to explore what new or

additional concepts and definitions would help our users better understand migration patterns.

- Our latest research linking the MWS to both benefits and annual PAYE data showed clear potential for earnings data to be an important source for identifying and measuring migration patterns. Our analysis illustrated that 4 in 5 non-UK nationals in our linked dataset had signs of activity in income and benefits data following arrival in the country, with more identified in PAYE data. Further linkage work, using more comprehensive PAYE Real Time Information (RTI) data, will therefore be instrumental in helping us to identify patterns of migration to and from the UK.

As outlined in section 2, we are using our knowledge of the different data sources to produce a series of data-driven rules with the aim of better representing the usually resident population. We are also developing a set of confidence-based rules aimed at improving the coverage of international migrants.

Further analysis and a series of case studies that explore how we can use administrative data to identify "activity" for the migrant population can be found in Annex B of the published report, and an accompanying slide pack.

4. Discussion and Conclusion

We will continue to collaborate closely across the GSS and beyond to develop our approach for putting administrative data at the core of our statistics. The case study research in this report focuses on the findings from linking individual data sources together to explore specific topic areas. To take this further, our next steps will be to link across a fuller range of data sources available to ONS, to continue to develop our data-driven rules and build an integrated system for measuring population and migration.

Important data sources which we plan to focus on in our next research phase are; further Home Office administrative data, PAYE RTI (and Self-Assessment), further healthcare data, further linked education data, Council Tax and others such as Electoral Register and DVLA registrations. This will improve the coverage of our data, particularly for groups such as EU migrants. Whilst we have improved our knowledge of what administrative sources such as health and income data can tell us about both EU and non-EU migration, our existing evidence base is much stronger for non-EU migration. Our next steps will focus on how we can use these further data sources to improve our coverage and address known challenges such as using administrative data to measure emigration from the UK. As we progress, we will also continue to consider the role of surveys in our future system, alongside developing our methods for producing improved statistics at a regional and local level.

A key next step is to put together all of what we have learned so far to produce administrative data-based estimates about the stocks and flows of the population. We will continue to publish the findings from our research on an iterative basis to demonstrate our progress towards our ambition to put administrative data at the core of our evidence on migration and population statistics by 2020. We plan to publish our next update on this work in spring 2019.

Alongside this, we will also carry out further work to compare what existing survey sources tell us about population and migration, including the IPS, the Labour Force Survey (LFS) and Annual Population Survey (APS). We will publish an update on our findings in May 2019 alongside the latest Migration Statistics Quarterly Report, with conclusions following later this year. We also plan to publish further work in 2019 investigating the impact of migration on the health and education sector, as set out in the timetable we published in our previous transformation update.

Estimation of gross population flows when SPDs are incomplete

John Dunne
Central Statistics Office, Ireland

Abstract

Gross population flows are made up of population inflows and population out¬flows in the population stocks between two time points or over a period of time. Population inflows comprise of immigration and births while population outflows comprise of emigration and deaths. A Statistical Population Dataset (SPD) is a dataset that typically contains one record for each member in the population. The compilation of an SPD is typically the first step in the compilation of population estimates from microdata. If a traditional Census based population estimate, the final census dataset can be considered an SPD. With a register-based population estimate, the SPD is based on the population register. SPDs can also be compiled directly from administrative data as in the Irish PECADO (Population Estimates Compiled from Administrative Data Only) project. When SPDs are complete and assuming perfect linkage, gross population flows can simply be derived by compar¬ing the SPDs between two consecutive timepoints or periods. Population estimates are simply obtained by taking counts from the SPD. If the SPD contains only undercount and not overcount then the SPD (list A) can be adjusted using a second list (list B) in a Dual System Estimation (DSE) setup to obtain population estimates. The PECADO project proposes a system of population estimates where both lists are compiled from administrative data sources. Reliable population estimates will provide for reliable estimates of net population flows but not gross population flows. We review the system of population estimates in the PECADO project before proposing an extension of DSE methods for the estimation of gross population flows that relies on the same lists used to compile population estimates. While not perfect, the proposed method has the property that as the coverage of the SPDs becomes complete the bias and variance of the gross population flow estimates goes to zero. We also consider alternative options and argue against them in favour of the proposed method. Finally, we present estimates of gross population flows using the proposed method and compare them to the official estimates. The practical application in the PECADO project focuses on two consecutive periods (calendar years) rather than timepoints.

Keywords

administrative data; DSE; migration; undercoverage; SPD; PECADO

1. Introduction

For countries that do not have a Central Population Register (CPR) on which demographic statistics can be compiled, the production of coherent statistics on population stocks and flows can prove challenging. This is particularly true for those countries that have relatively high and variable migration flows that are difficult to estimate.

The traditional approach to population estimates in these scenarios is an application of the demographic component or cohort component method to postcensal population estimates and then a recalibration of population estimates for intercensal estimates. In a Eurostat review of 31 countries (EUROSTAT, 2003) published in 2003, 19 countries were identified as using the component method for population estimates. This approach can be labelled a *flows first* based approach in that the population flows are first estimated and then used to derive the stocks based estimate. The weakness with this approach is that any errors or bias in estimating the components of population change (births, deaths, immigration, emigration) will be carried forward from timepoint to timepoint. These concerns, amplified in the presence of high migration flows, are one of the reasons why some countries, such as Ireland, undertake a census at 5 yearly intervals. The census provides a benchmark to recalibrate the population estimates at regular intervals. In a subsequent review in 2015 (EUROSTAT, 2015), Eurostat found that 31 of 44 countries depended on the census for annual population estimates and of these 31 countries only 9 supplemented their population estimates with information from registers.

For many countries, the census is the backbone to the system of population estimates. Conducting a traditional census is a major logistical exercise presenting many challenges, including those of cost, logistics and timeliness. Given the associated challenges, these countries are seeking to modernise and use administrative data as the basis of a modern system of population statistics.

In the 2000 round of censuses only 4 of 44 countries conducted a register-based census where the enumeration was based on a Central Population Register (CPR) (UN-ECE, 2008). In a register based census persons and households are simply counted using the CPR. The CPR, available in some national administrations, is the backbone to the co-ordination and delivery of public services to individuals and households.

The ability to compile reliable population estimates from administrative data sources is a significant milestone in transitioning from a traditional system of population statistics (with the traditional census at its core) to a modern system of social statistics with administrative registers and administrative data at its core. A modern system also holds the promise of being able to conduct a census on an annual basis at a fraction of the cost of what a traditional census would cost. Reliable annual population estimates then makes redundant the requirement to estimate the demographic components of change to obtain the following years population estimates. All that remains is to obtain an estimate of gross population flows in a manner that is coherent with the net population flows to ensure coherence in the system of population estimates. We label this type of approach a *stocks first* based approach in that population estimates are first estimated before deriving the flows using the stocks as an input.

This paper presents the methodological approach proposed in the PECADO project to estimate population flows when SPDs are incomplete.

2. Methodology

2.1. PECADO Project Overview

The project proposes a new system of population estimates based on DSE methodologies. The system first estimates population stocks before considering estimation of gross population flows. There is no CPR in Ireland to underpin public administration systems. However, there is a system of official Person Identification Numbers (PIN) used by persons when engaging with public administration systems.

2.2. Estimating Stocks - population size

The PECADO project approach to estimating population size can be summarised as follows:

Compile an Statistical Population Dataset (SPD) based on Signs of Life (SoL) An SPD is a dataset that, in theory, contains one record with attributes for each person in the population. In practice, an SPD may have under coverage and over coverage with respect to the population. In the PECADO project, the SPD is compiled with respect to each calendar year. The SPD contains a record for each person where there is significant evidence that the person is resident in the State in the calendar year. This evidence is typically of the form of a transaction, e.g., paying tax, receipt of child benefit, enrolled in education, receiving public healthcare etc.. Administrative data sources are chosen to cover the different life stages from the *cradle to the grave*; births, child benefit (universal payment), education (primary, secondary and tertiary), employment, self employed, public health care, pensions and deaths. A Protected Identifier Key (PIK) is used to link and summarise the records such that there is only one record for each persons where one or more Signs of Life are identified. The PIK is an encrypted form of the PIN and preserves linkage capabilities across data sources and over time while minimising privacy risk by eliminating the need to use identifiable information. Linkage error is considered minimised as it is based on a high quality PIN. By design, the SoL approach eliminates over coverage from the SPD, leaving only undercoverage to be dealt with. In the Irish PECADO project the SPD is called the Person Activity Register (*PAR*).

Estimate population size using DSE methods The PECADO project identifies a list B derived from another administrative data source to use in the DSE setup with the SPD as list A. List B is comprised of all those applying for or renewing their driving licence in the calendar year. We call this list the Driver Licence Dataset (DLD). The DLD is excluded from the SPD.

The methodology underpinning this approach is documented by Zhang and Dunne (2018). The population size estimator $\widehat{N} = nx/m$ where n, x and m are the respective sizes of list *B*, list *A* and the match between list *A* and list *B*, is underpinned by the following 3 assumptions

1. *No erroneous records:* A closed population ensures no records from outside the population are included but we also suppose there are no duplicate records or incorrectly identified records in either list *A* or list *B*.

2. *Matching assumption:* There is no linkage error when matching records between list *A* and list *B*.

3. *Homogeneous capture with respect to list B:* Every unit *i* in the population *U* has an equal chance π of being captured in list *B*.

An additional assumption of *independent capture in List B*, that is the event of any person is captured in list *B* has no impact on the likelihood of any other person being captured in list *B*, provides the variance estimator $\widehat{V}\left(\widehat{N}\right) = \left(n(n-m)x(x-m)\right)/(m^3)$.

The DSE methods described here (Zhang and Dunne, 2018) provide for a broader application of DSE than the more traditional approach as described by Wolter (1986). In particular, we can now consider a DSE model when list *A* is compiled from administrative data sources where it can be difficult to justify that the traditional assumptions hold.

This system produces stock population estimates on an annual basis on a population concept similar to the annual resident population (Lanzieri, 2013). The annual resident population concept used here is based on a calendar year rather than a point in time. The demographic accounting framework is adjusted slightly to reflect this. The accounting identity becomes: population resident in year (t) is equal to the population in year (t-1) less outflows in year (t-1) plus inflows in year (t). Reliable stock estimates will provide reliable estimates of net flows, that is the difference between inflows and outflows, but won't provide estimates of gross population flows. We propose an extension to the DSE methodology above to estimate gross population flows building on the stock estimates.

2.3. Estimating Gross Flows: Inflows and Outflows

The PECADO project reuses the underlying data sources from two consecutive years, 1 and 2, to estimate gross population flows. Gross population flows are made up of Inflows (Births plus Immigration) and Outflows (Deaths plus Emigration). The PECADO system to estimate gross population flows can be summarised as follows:

Estimate Stayers in SPD between two consecutive years 1 and 2 The same DSE methodology is used to estimate Stayers in the SPD between two years where lists A and B are formed from the *PAR* and *DLD*. Stayers corresponds to the group of persons present in the population in both years 1 and 2.

Gross estimate to obtain estimate of Stayers in population The next step is to simply gross the estimate of Stayers in the SPD to obtain an estimate of Stayers in the population. The factor used for grossing is obtained by dividing the population estimate by the size of the *PAR* for either year 1 or 2. We make the assumption that the proportion of Stayers in the population not included in the *PAR* is equal to the proportion of Stayers in the population included in the *PAR* for the respective year. This is the only additional assumption made. In making this assumption, we note the impact of any violation of this assumption will reduce as the coverage of the *PAR* with respect to the population goes to 100%. Noting that we can estimate Stayers using the population in either year 1 or year 2, we can also consider combining the two estimates to improve precision (reduce variance).

Estimate flows by subtracting Stayers from population estimates The gross flows can now be obtained using the Demographic Accounting Framework as follows:

Outflows, year 1, can be estimated by subtracting the estimate of Stayers from the population size estimate in year 1, Inflows, year 2, can be estimated by subtracting the estimate of Stayers from the population size estimate in year 2.

To estimate gross population flows, we consider the population in year 1, U_1, the population in year 2, U_2, the population common to both years, $U_S = U_1 \cap U_2$, the population of Outflows (those in year 1 but not in year 2), $U_O = U_1 \setminus U_S$ and the population of Inflows (those in year 2 but not in year 1), as $U_I = U_2 \setminus U_S$. We denote the respective population sizes as N_1, N_2, N_S, $N_O = N_1 - N_S$ and $N_I = N_2 - N_S$. DSE methods, as previously described, are used to estimate N_1 and N_2 as $\widehat{N}_1 = x_1 n_1/m_1$ and $\widehat{N}_2 = x_2 n_2/m_2$ where x_i is the size of list A in year i ($i = 1$ or 2), n_i is the size of list B and m_i is the size of the match ($PAR_i \cap DLD_i$) between list A and list B for year i. If we can estimate N_S we can obtain estimates of N_O and N_I, the gross population flows.=

To estimate N_S, we first estimate K_i, the number of Stayers in PAR_i using DSE methods and multiply by $\widehat{N}_i / x_i$, we can use i=1 or 2. We get

$$\widehat{K}_i = \frac{x_{i:j} n_{i:j}}{m_{i:j}} \qquad \text{for } i, j = 1,2 \text{ and } i \neq j$$

$$\widehat{N}_s = \frac{\widehat{N}_i}{x_i} K_i \qquad\qquad \text{for } \ i = 1 \text{ or } 2$$

Where $x_{i:j}$ is the size of the set $PAR_i \cap PAR_j$, $n_{i:j}$ is the size of the set $PAR_i \cap DLD_j$ and $m_{i:j}$ is the size of the set $(PAR_i \cap PAR_j) \cap (PAR_i \cap DLD_j)$.

The DSE setup to estimate K_i can described as follows

(i) List A is the set $PAR_i \cap PAR_j$.

(ii) List B is the set $PAR_i \cap DLD_j$, , i = 1, 2, j = 1, 2 and $i \neq j$.

(iii) Both list A and list B are subsets of $PAR_i \cap U_s$ the population of Stayers in PAR_i

(iv) Linkage based on official identification numbers ensures no linkage error.

(v) DLD_j satisfies the *homogeneous capture assumption* for population U_j, and as $(PAR_i \cap U_s) \in U_j$ then $PAR_i \cap DLD_j$ will satisfy the *homogeneous capture assumption* when estimating K_i the size of the population $PAR_i \cap U_s$.

This approach gives two estimators for N_s depending on which of K_1 or K_2 is taken as the starting point which could be combined if desired.

3. Discussion and concluding remarks

A theme running through this paper is adherence to the *demographic accounting principle* that the change in stocks (population estimates) from one period or timepoint to the next is equal to the difference in the associated flows, inflows less outflows (Inflows = Births + Immigration, Outflows = Deaths + Emigration). This principle underpins the traditional demographics cohort component approach , a flows first based approach, used to derive post-censal population estimates and is the basis on which a coherent, census based system of population estimates has traditionally been built. However, this system is prone to significant revisions when post-censal estimates are replaced with higher equality inter-censal estimates. The PECADO project proposes a stocks first based approach to a modern system of population estimates. Population estimates for the years 2011 to 2016 have to date been released by CSO as research outputs[1]. This paper describes an extension of these methods to estimate gross population flows in an overall coherent system of population estimates.

The aspiration or goal of census modernisation is to generate a high quality SPD with one record per person and no erroneous records. Once this goal is achieved, the methodology becomes less important. When SPD's are

[1] http://www.cso.ie/shorturl/480

complete (and free of erroneous records) estimating stocks and flows becomes simply a matter of counting records contained in the SPD in two time periods (timepoints), including those records that are common to the two time periods (timepoints). In the journey to achieving these goals, the methods developed can provide coherent estimates and when the goal is achieved can be provide reassurance around that achievement. As the coverage of the SPDs improve so too does the quality of the estimates of stocks and flows.

In summary, key characteristics of the proposed methodology to estimate gross population flows include:

Only one additional assumption is required beyond those required for estimating the population size, that the SPD is considered *representative* of the population.

Violation of this assumption may only have a low impact, as $x_i \rightarrow N_i$, the impact any violation goes to 0. In practice the size of the SPD is going to be close to that of the population (> 90%)..

A coherent set of estimates for stocks and flows through reusing the same data sources and using the stock estimates as a basis for estimating flows.

In terms of simplicity, the same data sources are used for estimating stocks as well as flows in a coherent system using simple DSE methods. Combining simple DSE methods with appropriate set intersection diagrams will aid explanation of these methods to users.

Further work in the PECADO project includes investigation of methods to detailed disaggregate state estimates by geography. If this is achievable, it may also be feasible to further extend the methods to estimate internal migration flows.

The author notes, the modernisation of systems of population estimates requires new and innovative methods and applications to fully exploit administrative (and other) data sources. There is a lot to be gained by the international statistical community in sharing ideas and practices and hope that this paper can make a contribution in this regard.

References

1. EUROSTAT (2003). *Demographic statistics: Definitions and methods of collection in 31 European Countries*. European Communities.
2. EUROSTAT (2015). *Demographic Statistics: A review of definitions and methods of collection in 44 European countries*. Eurostat.
3. Lanzieri, G. (2013). On a New Population Definition for Statistical Purposes Note. In *CES Group of Experts on Population and Housing Censuses*. UNECE.

4. UNECE (2008). *Measuring Population and Housing, Practices of UNECE countries in the 2000 round of censuses*. United Nations Publication.
5. Wolter, K. M. (1986). Some Coverage Error Models for Census Data. *Journal of the American Statistical Association*, 81(394):338–346.
6. Zhang, L. and Dunne, J. (2018). Trimmed dual system estimation. In Bohning, D., van der Heijden, P. G., and Bunge, J., editors, *Capture-recapture methods for the Social and Medical Sciences*, chapter 18, pages 237–258. CRC press.

An introduction to INEXDA's metadata schema[1]

Stefan Bender[1], Brigitte Hausstein[2], Christian Hirsch[3]
[1] Head of Research Data and Service Centre, Deutsche Bundesbank.
[2]GESIS Leibniz Institute for the Social Sciences
[3] Research Data and Service Centre, Deutsche Bundesbank.

Abstract
This paper introduces the metadata schema used by the international network INEXDA to describe granular datasets from different countries. The schema, agreed on by all members, facilitates a comprehensive inventory of existing granular datasets conducted in the member institutions. This inventory, in turn, will foster harmonisation activities between INEXDA members, broaden metadata sharing and potentially future data sharing between institutions represented in the network, and pave the way for metadata on publicly available granular datasets to be shared with external researchers. The INEXDA metadata schema was developed to be easily adaptable for non-INEXDA institutions.

Keywords
Metadata, Microdata, International Network

1. Introduction

Metadata is essential for documenting data, citing it and finding it in catalogues. Metadata can be defined simply as information about data, ie a description of data. It is "...structured or semi-structured information which enables the creation, management, and use of records [i.e. data] through time and within and across domains in which they are created. Recordkeeping metadata can be used to identify, authenticate, and contextualize records; and the people, processes and systems that create, manage and maintain and use them" (Wallace, 2001, p.255). Metadata provides a means for visibility and presentation of data as well as discovery. It supports the re-use, management, exchange and long-term preservation of data. "Data without metadata is just stuff. Nobody needs more stuff today" (Recker, 2014).

The use of a metadata schema goes through various standardisation phases, from which more or less binding metadata standards develop. If such a standard is specified, documented and legally recognised, it is referred to as a standard. In addition, there are de facto or quasi standards based on

[1] The views expressed here are those of the authors and do not necessarily reflect those of the Deutsche Bundesbank, GESIS, or the INEXDA Network.

discipline-specific practical experience and recognised rules of a community. Meanwhile, there is a large array of metadata standards focusing on a particular subject domain, content type, function or application, such as:

- Data Documentation Initiative (DDI)
- Dublin Core Metadata Initiative (DCMI)
- DataCite Metadata Schema
- da|ra Metadata Schema
- Metadata Encoding and Transmission Standard (METS)
- Preservation Metadata Maintenance Activity (PREMIS)
- Statistical Data and Metadata Exchange (SDMX)

The significance of a metadata standard lies in the uniform language, which makes it possible to record information about data in an understandable and structured manner. In the case of domain-specific standards, metadata often constitutes very granular – semantically rich – statements about an object. In order to describe data uniformly and unambiguously, standardised terminologies (eg controlled vocabulary such as thesaurus or keywords) are used for certain metadata elements to describe particular metadata with uniform information. This should be distinguished from semi-structured or unstructured metadata, which describe important contextual information on the creation of the data.

Since its foundation in January 2017, the members of the International Network for Exchanging Experiences on Statistical Handling of Granular Data (INEXDA) have, amongst other things, collaborated to harmonise metadata structures and complete an extensive stock-take of available datasets in member institutions. This stock-taking exercise has three aims.

- First, it is designed to give member institutions an overview of the available and possible comparable granular datasets.
- Second, it makes it easier for data users to discover and use datasets appropriate for research and analysis by using harmonised metadata.
- Third, it provides a framework to facilitate a possible harmonisation of datasets in the (near) future.

Because the description of the data should be comparable, all members agreed on a metadata schema for the granular data. This paper presents the metadata schema underlying the stock-taking exercise by INEXDA members. We also explain practical considerations when implementing a metadata schema for microdata, which can easily be adopted by other institutions. Naturally, the INEXDA metadata schema was designed with certain goals in mind. Therefore, some choices may not translate to other situations. For example, the INEXDA schema allows for information about the relation of different datasets to each other, which is essential for the possible harmonising of datasets.

2. The INEXDA metadata schema

2.1 Scope

The INEXDA metadata schema is designed to provide metadata for microdata on the dataset level. Since microdata are data at the level of individual reporters, the data may allow re-identification of individual reporting units. Because of its sensitive nature, microdata is therefore always subject to protection of confidentiality for individual observations. Metadata about microdata also has to adhere to the same high standards when it comes to protecting confidentiality.

The INEXDA metadata schema is based on the da|ra metadata schema (Version 4.0) which was jointly developed by GESIS – Leibniz Institute for the Social Sciences and ZBW – Leibniz Information Centre for Economics (Koch 2017). da|ra operates a registration service for social science and economic data which allocates a Digital Object Identifiers (DOI) to different types of datasets, allowing a unique identification of resources (https://www.da-ra.de). da|ra maintains a cooperation with DataCite (https://www.DataCite.org).

Adapting an existing metadata schema to fit the purpose of INEXDA provides a level of standardisation for microdata coming from different countries, institutions, and collection purposes. In addition, the interoperability with the da|ra schema allows for a seamless transition between the INEXDA and da|ra databases, which makes it easier to obtain a DOI identifiers in the future. However, ensuring interoperability may also result in restrictions on altering metadata items (see section 2.2).

2.2 Metadata items

All INEXDA members agreed on a metadata schema, which, on the one hand, describes the datasets in a sufficient way for the purposes mentioned above while, on the other hand, being easy to handle for potential users or data producers. Therefore, the INEXDA network came up with 21 items (see Table 1), which we describe below in more detail. For each metadata item, we also provide an example drawn from Bundesbank datasets to illustrate its use. Please note that you will find the example after the description of the item. These examples are written in italics.

Resource Type: A description of the underlying resource for which metadata is compiled. In case of INEXDA, the resource type is always a dataset. This item is included in the INEXDA metadata schema to facilitate interoperability with the da|ra schema.

Resource Identifier: The "Resource Identifier" is a unique value to disambiguate between resources. A version number can be provided as a reference that changes have been made between versions.

Name of Dataset: This item displays the name of the dataset.

Example: Securities Holdings Statistics – Base (name of one Bundesbank dataset)

Creator: The item "Creator" refers to the name(s) of the institution, and/or division, and/or department responsible for developing, collecting and/or managing the dataset. Names of individual persons should not be provided.

Example: Deutsche Bundesbank, Statistics Department

DOI Proposal: The schema allows suggesting a DOI name for the dataset. A Digital Object Identifier (DOI) is a permanent, persistent identifier used for citing and linking electronic resources (texts, research data or other content).

Example: 10.12757/Bbk.SHSBase.05121603 (for one version of the Securities Holding Statistics)

URL: URL is the address of a webpage which displays information about the dataset, if available. Alternatively, the landing page of the research data centre or the institution could be used if no specific webpage exists for the dataset.

Example:

https://www.bundesbank.de/Redaktion/EN/Standardartikel/Bundesbank/Research_Centre/research_data_micro_data_securities_holdings_statistics.html?nsc=true&https=1

Language of Resource: This item refers not the language in which the metadata elements are expressed but instead refers to the language in which the dataset is available. Default is English. List a language other than English only if dataset is not available in English.

Example: English

Publication Date: The date on which the dataset was released internally or made publicly available needs to be entered here in ISO 8601 format. If a DOI has been assigned, the convention in the INEXDA network is to use the date of DOI registration.

Example: 2016-08-01

Availability: Providing information on procedures under which data are being made available to data users may help to better understand which datasets they might be able to access and how to access those datasets. Availability procedures are best described by controlled vocabulary, i.e. by a pre-defined list detailing possible procedures which avoids confusion of data users due to the same content being described by different language. This item lists the procedures under which data is being made available to researchers. Entries allowed are restricted to the following predefined list:

- Download – researchers can directly download data from the website;
- Delivery – researchers can receive the data set or access data from any location via remote access or send codes to the data owner and receive controlled output;

- On-site – researchers have to come to the premises of the data owner to see or work with the data;
- Not available – researchers cannot access / use the data by any means;
- Unknown – other type of availability, please specify in free text.

Example: On-site[2]

Sampled Universe: The item "sampled universe" provides information on elementary units about which inferences are to be drawn and to which analytic results refer when analysing the dataset.

Example: The SHS-Base is a full census (no reporting thresholds apply), i. e. all financial institutions domiciled in Germany report any securities they are holding for domestic and foreign customers. In addition, domestic banks provide information about their own holdings, irrespective of where the securities are held. The financial institutions who are obliged to report comprise domestic banks (monetary financial institutions excluding money market funds), domestic investment companies and "other" domestic financial companies.

The data collection involves holdings of debt securities, shares and investment fund shares or units 3, irrespective where the securities were issued, in what currency they are denominated or if they are listed or not. Only securities which are in circulation and which can be assigned to an investor are included in the Securities Holdings Statistics.

A basic set of information is required to be reported on a security-by-security level. This includes the International Securities Identification Number (ISIN), the nominal amount or number of units held, and the sectoral classification and residency of the holder. For securities quoted as a percentage (eg bonds and debt securities), the nominal value is stated in the relevant nominal currency and for securities quoted as a number of units (eg equities, mutual fund shares) the number of units held is reported. As the significance of securities repurchase and securities lending transactions has increased strongly in recent years, securities holdings which are passed on or acquired as part of such contracts are to be flagged separately. This information is mandatory for the own holdings of domestic banks only. Since reference month January 2014 the monetary financial institutions have to report the book values of their own securities holdings. Securities that are attributed to the trading portfolio must be labelled.

[2] Most Bundesbank data are available for non-commercial research via the Research Data and Service Center (https://www.bundesbank.de/Navigation/EN/Bundesbank/Research/RDSC/rdsc.html).

Sampling: This item describes the type of the sample and sample design used to select the observations to represent the population. As an INEXDA convention, the value of the item will be set to "Total Population" if dataset contains no sampling, i.e. dataset is the population.
Example: Total Population
Temporal Coverage: The item "Temporal Coverage" refers to the sample period of the dataset, which is the time period during which the data was collected or observations were made. The time period should be expressed in ISO 8601 format. As an INEXDA convention, information on month and quarter may be omitted if frequency of the dataset is annual.
Example: 2005-12 to 2016-03
Time Dimension: Describe the time dimension of the data collection. In the INEXDA schema, the following three time dimensions are allowed,

- "Panel" datasets contains information collected from the same (or almost the same) set of entities over time.
- "Time Series" data is collected repeatedly over time to study changes in observations. The "Time Series" category is appropriate for almost all macro datasets, e.g. data on GDP or unemployment rates.
- "Cross-section" dataset are data about a population collected only once.

These items are further broken down by the frequency (e.g. monthly, quarterly) with which data about entities is collected in the dataset.

Furthermore, the following INEXDA conventions apply to this metadata item.

- In case time intervals between consecutive data collections are not equally-spaced (e.g. datasets collected from event-driven reporting), the highest available frequency in the dataset is selected, that allows for a meaningful analysis of the data.
- In case the frequency of data collection changes over time (e.g. because of structural breaks following a change in regulation governing the collection of the dataset) the most recent frequency available in the dataset should be used.

Example: Panel: monthly
Collection mode: This item provides information on the method used to collect the data.
Example: Reporting agents file their reports electronically to Deutsche Bundesbank.
Unit descriptions: Microdata are data on the level of the individual observation. Microdata regularly contain numbers or strings attached to individual observations with the aim to make identification easier. These identifiers may exist on the company level (e.g. country tax number, LEI) or

security level (e.g. ISIN). For analysis and research purposes it is often necessary to combine microdata from different sources or policy domains using these identifiers.

The item "Unit Descriptions" provides information about available identifier(s) in the dataset. More than one
identifier is allowed, separated by a comma. Identifiers from external data vendors are allowed.
Example: Bank ID, ISIN

Descriptions: The purpose of the metadata item "Descriptions" in the INEXDA schema is twofold.

1. It provides a short description of the dataset.
2. It is used to share information about the scope of structural breaks in the dataset, where structural breaks are defined as major events and revisions that have impacted the dataset.

Designing a metadata schema for microdata needs to employ a procedure documenting significant changes made to a dataset over time stemming from a change in underlying reporting requirement rules. These structural breaks, if they occur, affect metadata items in that the appropriate documentation becomes time dependent. For example, the reporting frequency of a dataset changes in 2012 from quarterly to monthly. In the INEXDA metadata schema the item "Temporal Coverage" will now depend on the time period: before 2012 the appropriate value is quarterly, after the structural break in 2012 the value is monthly.

Besides changes to the time frequency with which data is collected, other examples of structural breaks include:

- Changes to the set of collected variables.
- Changes in the population or sampling.

Example:
The SHS-Base is the core module of the Deutsche Bundesbank's Securities Holdings Statistics (WpInvest). The motive behind creating the SHS-Base is to be able to answer the question: "Who holds what from whom and how much?". Financial institutions domiciled in Germany report securities which they hold for domestic or foreign customers ("custodian-approach"). In addition, domestic banks provide information about their own holdings, irrespective of where the securities are held. Reporting agents are domestic banks (monetary financial institutions excluding money market funds), domestic investment companies and "other" domestic investment companies. The SHS-Base is collected by means of a full census, i.e. every reporting agent has to send a report (if no securities are held in safe custody, a nil report is filed). The reporting agents provide information on securities

holdings broken down by the holder's economic sector and country of origin on a security-by-security basis. The reporting scheme comprises the holdings of debt securities, shares and investment fund shares or units. Holders are classified according to the ESA sectoral classification scheme and the amounts held in safe custody are transmitted to the Deutsche Bundesbank. Securities data are indispensable for monetary analysis as any shifts in financing between the banking system and the securities markets may affect the transmission of monetary policy. In addition, data are essential to monitor the development of amounts held and the distribution among different (groups of) investors. Financial stability analysis makes use of the data to measure the risks associated with different types of instruments and exposures to individual counterparties. Information on the composition and quality of the portfolios of holders is used to better understand investor behavior.

- *From 200512 to 201212: end-of-quarter collection*
- *From 201301 to 201603: end-of-month collection*

Geographical Coverage: The item "Geographical Coverage" provides information on the region where the data was gathered or on which the data is focused.
Example: DE (for Germany)
Keywords: Please enter the keywords describing the general content of the dataset. You may choose up to 10 keywords. Entries allowed are restricted to a predefined list.
Example: Banks, Securities, Financial markets, Debt securities, Equity and investment fund shares or units, Financial derivatives
Alternative identifier: This metadata item is used to indicate that the information about the dataset provided here belongs to the metadata collection of INEXDA.
Example: INEXDA
Relations: The item "Relations" will be used in the future to describe relations between datasets and databases in the INEXDA metadata database. Examples of relations will include:

- Different versions of a dataset.
- Relation between dataset and database (in a given country)
- Relation between datasets containing similar units (in different countries)
- Dataset feeds into an ECB dataset

Publications: Allowing descriptions of publications be associated with related datasets provides additional information which is complementary to existing metadata. Publications include information on how a dataset is being used (e.g. variable transformation), what a dataset is being used for (e.g. topics or methodology), and who has used the dataset. All this information is useful to help data user better understand and discover datasets relevant for their research.

Therefore, the INEXDA metadata schema contains an item that provides information on scientific publications relating to the registered dataset. Publications listed here may also include descriptions of datasets (e.g. technical reports, data reports, or user guides) which are publicly available. *Example: Bade, M., Flory, J. and T. Schönberg (2016). SHS-Base, Data Report 2016-02 - Metadata Version 1-1. Deutsche Bundesbank Data and Service Centre.*

3. Conclusion

The relevance of metadata also becomes clear in connection with the re-use and corresponding citation of the data. Bibliometrical methods, which should make the performance of data production measurable similar to the impact factors in text publications, are based on metadata. In addition to the bibliographic and content information, these metadata should also include a persistent identifier (so called PID) that enables the identification and localization of the used data uniquely and permanently. There are now a number of services that offer the allocation of PIDs in the form of Digital Object Identifiers (DOI) for research data. The German Leibniz Institutes GESIS and ZBW are offering a special service for the assignment of DOI names for social and economic data.

References

1. Day, M., (2005). Metadata, DCC Digital Curation Manual. In: S. Ross, M. Day (eds), Retrieved 10 July 2018, from http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata
2. Harzenetter, K., Borschewski, K.: da|ra: Solutions to the Challenges of Data Registration, Access & Exchange. CNI Spring 2016 Membership Meeting; San Antonio (Texas); April 4-5, 2016. Retrieved 10 July 2018, from https://www.cni.org/topics/information-access-retrieval/dara-solutions-to-the-challenges-of-data-registration-access-and-exchange
3. Jensen, U., Wasner, C., Zenk-Möltgen, W. (2018): Metadatenstandards im Kontext sozialwissenschaftlicher Daten. In: U. Jensen, S. Netscher, K. Weller (eds.), Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Opladen. (in print)
4. Koch, Ute and Akdeniz, Esra and Meichsner, Jana and Hausstein, Brigitte and Harzenetter, Karoline: da|ra Metadata Schema:

Documentation for the Publication and Citation of Social and Economic Data. Version 4.0. GESIS Papers 2017/25.)
5. Recker, A.: Was sind Daten | Forschungsdaten | Metadaten? Forschungsdatenmanagement in den Geowissenschaften. Vortrag. Berlin, 3-4 July 2014. Retrieved 10 July 2018, from https://www.slideshare.net/GESISarchivetraining/was-sind-daten-forschungsdaten-metadaten
6. Wallace, D. (2001). Archiving metadata forum: report from the Recordkeeping Metadata Working Meeting, June 2000. Archival Science, 1(3), p 253-269.

INEXDA - The granular data network[1]

Emily Witt[1], Stefan Bender[2], Olympia Bover[3], Giovanni D'Alessio[4], Luís Teles Dias[5], Robert Kirchner[2], Renaud Lacroix[6], Paulo Guimarães[5], Manuel Ortega[3], Lyon Michael[7], Christian Hirsch[2]
[1]European Central Bank
[2]Deutsche Bundesbank
[3]Banco de España
[4] Banca d'Italia
[5]Banco de Portugal
[6]Banque de France
[7]Bank of England

Abstract

The financial crisis of 2007-08 has highlighted the need for using granular data on financial institutions and markets to detect risks and imbalances in the financial sector. Data producers such as central banks and national statistical institutes are witnessing a growing need to improve granular-data access and sharing. When making granular data available, data producers face significant legal and technical challenges related to, among others, safeguarding statistical confidentiality. This paper introduces the INEXDA international network, which provides a platform for data producers to exchange practical experiences on the accessibility of granular data, metadata as well as techniques for statistical analysis and data protection.

Keywords

Microdata, International Network, Data Access

1. The motivation for INEXDA

In 2009, the finance ministers and central bank governors of the G20 endorsed the first phase of the Data Gaps Initiative (DGI-1) to promote actions to close data gaps that had come to light in the wake of the global financial crisis that emerged in 2008. During the process of DGI-1, data users and data compilers increasingly expressed the need for improving data sharing, particularly of granular[2] data, in order to foster the understanding of global

[1] The views expressed here are those of the contributors and do not necessarily reflect those of the Banco de España, Banca d'Italia, Banco de Portugal, Banque de France, Bank of England, Deutsche Bundesbank, or European Central Bank.

[2] In this paper, granular data are defined as less aggregated data than traditional statistics (eg finer breakdowns of aggregates in traditional statistics) or microdata. Microdata are data at the

developments, for example with regard to risks and imbalances. Consequently, the second phase of this initiative (DGI-2) contains a new recommendation (II.20) promoting the exchange of (granular) data as well as metadata.[3]

To help meet data users' and data compilers' demand for (granular) data sharing within the legal framework of the individual jurisdictions and to facilitate the implementation of Recommendation II.20 of DGI-2, a group of central banks established the **I**nternational **N**etwork for **E**xchanging E**x**perience on Statistical Handling of Granular **Da**ta (INEXDA). In accordance with the objectives of INEXDA outlined below, participation is open to other central banks, national statistical institutes, and international organisations. Other examples of exchanging experiences in the context of data sharing include the Conference of European Statisticians Task Force on the Exchange of Economic Data, which focuses particularly on the activities of multinational enterprises (MNEs), as well as the work on data sharing by the Bank for International Settlements (BIS) Irving Fisher Committee (IFC).

INEXDA was explicitly mentioned in the report of the Inter-Agency Group on Economic and Financial Statistics: "Update on the Data Gaps Initiative and the Outcome of the Workshop on Data Sharing", March 2017. The paper was welcomed by the G20 Finance Ministers and Central Bank Governors in March 2017 and by the G20 leaders: "We welcome the recommendations of the Inter Agency Group on Economic and Financial Statistics (IAG) for sharing and accessibility of granular data." (p. 5, Communiqué of the G-20 FMCBG Meeting (2017)).

2. A brief history of INEXDA

On 6 January 2017, the Banca d'Italia, Banco de Portugal, Bank of England, Banque de France and Deutsche Bundesbank (see also figure 1) founded INEXDA during a meeting at the Banco de Portugal. In this meeting, the BIS – which participated as a guest – offered to support the work of INEXDA by providing access to the eBIS[4] platform. All INEXDA information is therefore stored and shared via the eBIS system.

The second INEXDA meeting took place at the Bank of England on 7 July 2017, where the Banco de España and European Central Bank (ECB) joined INEXDA as first-time guests. During this meeting, particular emphasis was placed on developing a metadata schema for the INEXDA network. In this regard, a presentation by the GESIS Leibniz Institute for the Social Sciences on "The da|ra Data Referencing System and its potential for the INEXDA Project"

level of individual reporters or at a low level of aggregation that may lead to the identification of individual reporting units.

[3] More information on DGI-1 and DGI-2 can be found at http://www.imf.org/external/np/g20/pdf/2015/6thprogressrep.pdf.

[4] https://www.ebis.org/auth/login.

was considered very useful by INEXDA members (see Bender, Hausstein and Hirsch (2018) for a more detailed description of the INEXDA metadata schema).

At the third INEXDA meeting on 11 January 2018 at the Banque de France, the INEXDA network welcomed the Banco de España and ECB as new INEXDA members, increasing the number of INEXDA members from five to seven. Furthermore, the Banco Central de Chile, Banco de México, Oesterreichische Nationalbank, Central Bank of the Republic of Turkey and – for the first time, a national statistical institute – Office for National Statistics UK attended the meeting as guests. One notable outcome of the meeting was the consideration of establishing working groups on different topics within the framework of INEXDA (see section 4).

The fourth INEXDA meeting was held on 27 August 2018 at the BIS in Basel, where Banco Central de Chile and Central Bank of the Republic of Turkey participated as new members. Alongside the guests in attendance at the third meeting, the Bank of Russia, Federal Statistical Office of Germany, Eurostat, and the Swiss National Bank were attending the meeting as first-time guests.

3. INEXDA's objective

INEXDA was established with the overall aim of facilitating the international use of granular data for analytical, research and policy purposes within the limits set by the applicable confidentiality regimes.[5] This overall aim can be further broken down into the following two, more specific objectives.

First, INEXDA will provide a basis for exchanging experiences on the statistical handling of granular data that are accessible to external users. Examples of "statistical handling" include the processes, methods, and tools for data and metadata access, techniques for the statistical analysis of granular data, procedures for data confidentiality and data security, and procedures for output control. Second, INEXDA will provide a framework for investigating possibilities to harmonise access procedures and metadata structures, to develop comparable structures for existing data, and to further foster the efficiency of statistical work with granular data.

The higher level of data disaggregation in the case of granular data is also associated with an increased need for data protection. European and national legal provisions regulate both the user group and the access channels to microdata and oblige data providers and data recipients to maintain data confidentiality at all times. Therefore, the overriding principle of the work of INEXDA is compliant with the respective statutory secrecy and data protection

[5] INEXDA's objectives are outlined in the Memorandum of Understanding (MoU), which must be signed by each member and is available on the websites of each member institution.

requirements, and thus maintaining the confidentiality of the information submitted by the reporting agent.

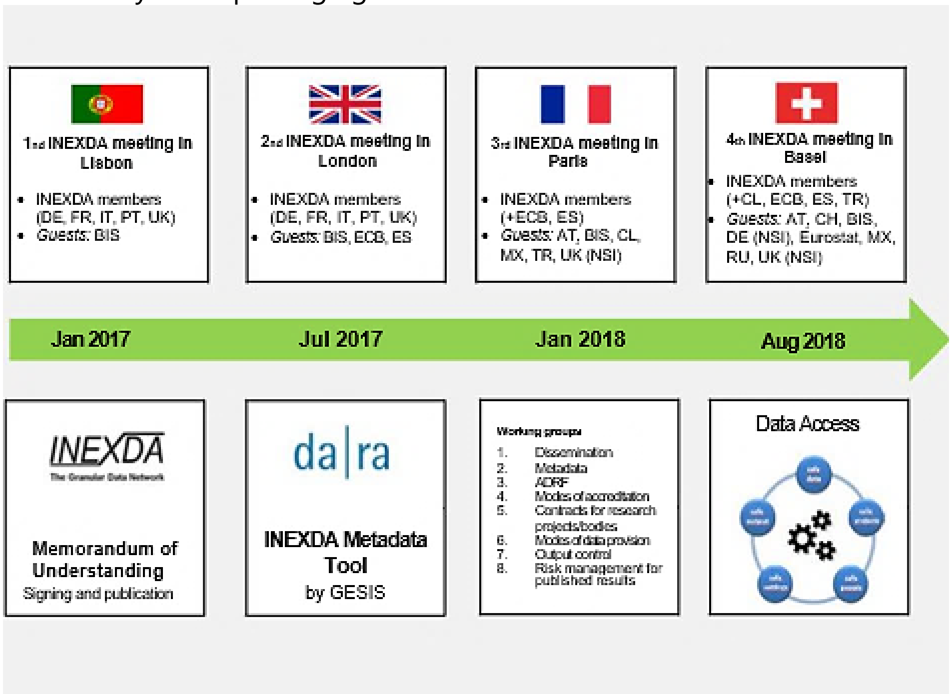


Figure 1: Overview of participants and important outcomes of the first four INEXDA meetings.

4. The current INEXDA work programme

For the current work programme, INEXDA members have decided to find a balance between keeping the momentum and not being overly ambitious. Therefore, INEXDA has identified eight potential topics for the work programme:

1. Dissemination (of granular data)
2. Metadata (see section 4.1 for a brief overview)
3. Tools for supporting the work of INEXDA members (ADRF, see section 4.2 for a brief overview)
4. Modes of accreditation (see section 4.3 for a brief overview of items 4, 5, and 6)
5. Contracts for research projects/bodies
6. Modes of data provision
7. Output control
8. Risk management for published results

INEXDA aims to have an agile structure, so the topics of the working programme should produce tangible results after six months as a minimum.

Besides these activities and the contribution of INEXDA to the 9th biennial IFC Conference, INEXDA will also make contributions to the 2018 Conference of European Statistics Stakeholders (CESS) in Bamberg, and the 62nd ISI World Statistical Congress in Kuala Lumpur in 2019.

4.1 Comprehensive inventory of data in member institutions

From the start, the INEXDA network has collaborated to harmonise metadata structures by conducting extensive stock-taking of available data sets in member institutions. The goals are:

1. to provide an overview of available and potentially comparable granular data sets from participating institutions;
2. to enable data users to discover and use appropriate data sets for their own research and analyses, which the participating institutions agree to share;
3. and to prepare a framework to facilitate a possible harmonisation of data sets in the (near) future.

Because the descriptions of the data should be comparable, an agreement on a metadata schema for the granular data was established between all members. To this end, the INEXDA metadata schema closely follows the da|ra metadata schema (version 4.0), which was jointly developed by the GESIS – Leibniz Institute for the Social Sciences and the ZBW – Leibniz Information Centre for Economics. The INEXDA metadata schema is designed to provide metadata for microdata at the data set level.

Adapting an existing metadata schema to fit the purpose of INEXDA provides a level of standardisation for microdata produced in different countries, institutions, and with different aims. Furthermore, the interoperability of the INEXDA metadata schema with the da|ra metadata schema allows for seamless transition between the INEXDA and da|ra databases, which makes it easier to obtain digital object identifier (DOI) for datasets in the future.

All INEXDA members agreed on a metadata schema, which, first, describes the data sets in a comprehensive way for the purposes mentioned above. Second, the schema is easy to use for potential users and data producers. It should be noted, that the metadata schema revolves around a "standardised data set", which is a snapshot of data produced in an institution (eg credit register) taken at a certain point in time (e. g. 1999-2017). To this end, INEXDA devised 21 items for its metadata schema (see table 1).

Furthermore, INEXDA has created a platform (see figure 2) for collecting and exchanging the metadata information produced during the inventory. This platform is available to all INEXDA member institutions. The platform is being developed jointly with GESIS.

Because of its sensitive nature, microdata are always subject to protection of confidentiality of individual observations. Metadata about microdata also have to adhere to the same high standards when it comes to protecting confidentiality. INEXDA's metadata system is designed to address these issues.

1	Resource Type
2	Resource Identifier
3	Name of Dataset
4	Creator
5	DOI Proposal
6	URL
7	Language of Resource
8	Publication Date
9	Availability
10	Sampled Universe
11	Sampling
12	Temporal Coverage
13	Time Dimension
14	Collection Mode
15	Unit Descriptions
16	Descriptions
17	Geographical Coverage
18	Keywords
19	Alternative Identifiers
20	Relations
21	Publications

Table 1: The INEXDA metadata scheme

4.2 Evaluating tools to support INEXDA's harmonisation process

While the highest priority is given to completing the inventory of available data described in 4.1, the investigation of harmonisation possibilities at other levels of the data lifecycle (eg access procedures and registration processes) remains an important task in the current INEXDA work programme. Standardised software applications could be a way forward, as these would not only facilitate communication between the INEXDA partners but also help to maintain common standards.

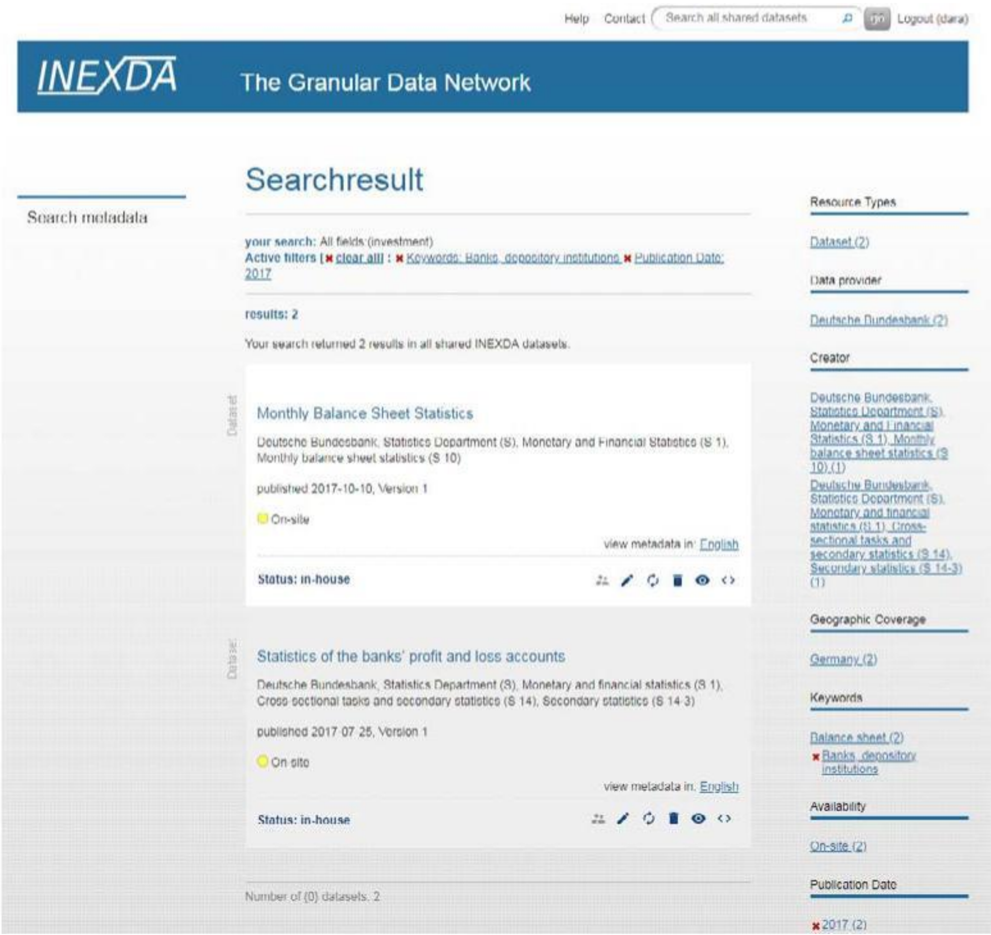


Figure 2: Hypothetical example of the INEXDA Metadata Platform

The New York University (NYU) has established, under the assignment of the Bureau of the Census, the Administrative Data Research Facility (ADRF), which provides a set of analytical tools, data storage and discovery services, and general computing resources based on cloud solutions for a diverse set of users, including government analysts and researchers. As the ADRF framework is considered to be potentially very useful for the harmonisation process, INEXDA will consider cooperation with NYU.

4.3 Taking stock of the access procedures and registration processes for researchers

One overarching goal of INEXDA is to provide a basis for exchanging experiences on the accessibility of data, procedures for data confidentiality, and security of data. Since access to microdata is in the scope of official statistics, INEXDA will benefit from national and international experiences to shape the outcome of this work stream. In the context of INEXDA, and following up on a survey of the Working Group of Statistical information Management (WGSIM) of the European System of Central Banks (ESCB) on national central banks' (NCBs) approaches to granting external researchers access to confidential data for research purposes, Emily Witt and Jannick Blaschke (ECB) conducted interviews with several central banks (Oesterreichische Nationalbank, Deutsche Bundesbank, Banco de España, Banque de France, Banca d'Italia, De Nederlandsche Bank, Banco de Portugal, Central Bank of the Republic of Turkey, Bank of England, European Central Bank) and Eurostat. The result is an overview[6] of selected NCBs' and Eurostat's approaches to providing access to non-published granular data for research purposes that complements other work in this area.[7]

Besides international experiences, national experiences are helpful in identifying the best practices with regard to access to microdata. For example, the Deutsche Bundesbank recently provided an overview of the microdata access procedures used, where three different user groups of microdata have been identified (internal analysts, internal researchers, and external researchers). The paper (Schönberg (2018)) described different access modes for each user group in detail. A unit called Internal Service for Micro Data Analysis handles internal analysts' data access requests following a multilevel approach (modelled after the European System of Central Banks (ESCB) standard approach).

At the end of August 2018, INEXDA will likely start a working group focusing on best practices on how data users could be allowed to access granular data once they have completed the accreditation process and have signed all relevant contracts. The task of this working group is to take stock of existing models of data provision used by INEXDA members. Possible topics may include:

- data access via secure access facility and/or remote access (eg technical design and specifications of limitations);
- anonymisation of methodologies and tools;
- provision of services to external researchers (eg provision of standard or ad hoc data sets, linkage of various data sets, upload of external data sets, access to licensed data sets);
- provision of analytical tools and allowing/facilitating code sharing.

[6] The participating interviewees agreed to share the results with INEXDA members and guests.
[7] For example, the "Guidelines for the assessment of research entities, research proposals and access facilities" (Luxembourg, November 2016) from the European Commission, Eurostat, Directorate B: Methodology; Corporate statistical and IT services, Unit B-1: Methodology and corporate architecture, or the results from the FP7 project "Data without Boundaries" (DwB, see https://www.facebook.com/dwbproject).

4.4 INEXDA web page

A web page for the network will be launched by the end of 2018. The website is intended to be independent of the signing parties' websites and, to this aim, the following domains were reserved: www.inexda.org; www.inexda.com.

5. INEXDA working arrangements

The members of INEXDA have implemented the following working arrangements.

- All decisions are made on a consensual basis.
- The work within INEXDA will be performed at the operational levels of the member institutions.
- INEXDA members convene at least once per year. Guests may be admitted to meetings. A pre-meeting will be organised prior to each INEXDA meeting for the purpose of inviting INEXDA guests to discuss the progress INEXDA has made so far.
- The chair of INEXDA is elected for a two-year term on a consensual basis. Responsibilities of the chair include co-organising the meetings in close collaboration with the local organiser, coordinating activities, and drafting a report at the end of the chairmanship, which has to be agreed on a consensual basis.
- The eBIS facility operated by the BIS provides the centralised location for exchanging documents and fostering collaborative activities among INEXDA members.

6. The INEXDA application process

The following procedure has been established for admitting new members. It is mandatory for institutions that want to join INEXDA to have a representative attending at least one INEXDA meeting in person before submitting a formal application. The application letter should be signed by the head of the statistical department of the respective institution (or, in the case of national statistical institutes or international organisations, by the head of the responsible department) and sent to the chair of INEXDA. Any application to join INEXDA from a new institution and the signing of the MoU must be agreed by all members. Furthermore, the applicant institution is invited to attend an INEXDA meeting to give a presentation on the current state of its granular data sharing and its motivation for becoming a member of INEXDA.

7. Conclusion

The International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) was founded to facilitate active dialogue on practical experiences – in particular on the accessibility of granular data, metadata, and techniques for statistical analysis and data protection. Until recently, the network was predominantly focused on establishing a metadata schema and conducting a comprehensive inventory of data in member institutions. In the next phase of the work programme, access procedures and registration processes for researchers will come to the forefront of INEXDA's activities.

The overall aim is to facilitate the international use of granular data for analytical, research, and comparative purposes without jeopardising and always subject to the respective applicable confidentiality regimes.

References

1. Bender, S., Hausstein, B., and Hirsch, C. (2018). An Introduction to INEXDA's Metadata Schema, Technical Report 2018-02, Deutsche Bundesbank, Research Data and Service Centre.
2. Communiqué of the G-20 FMCBG Meeting (2017). Retrieved 31 July 2018, from https://www.bundesfinanzministerium.de/Content/EN/Standardartikel/Topics/Featur ed/G20/g20communique.pdf
3. G20 Action Plan (2017). Retrieved 31 July 2018, from
4. https://www.g20.org/profiles/g20/modules/custom/g20_beverly/img/timeline/Germ any/2017-g20-hamburg-action-plan-en.pdf
5. Inter-Agency Group on Economic and Financial Statistics (2017). Update on the Data Gaps initiative and the outcome of the Workshop on Data Sharing. Retrieved 31 July 2018 from data.imf.org/api/document/download?key=61400076
6. Memorandum of Understanding on the Establishment of the International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) (2018). Retrieved 31 July 2018 from https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/inexda/html/index.en.html
7. Schönberg, T. (2018). Data Access to Micro Data of the Deutsche Bundesbank, Technical Report 2018-01, Deutsche Bundesbank, Research Data and Service Centre.

Appendix: List of INEXDA members and INEXDA guests[8]

INEXDA members are institutions that have signed the MoU. INEXDA members:

- Banca d'Italia
- Banco Central de Chile
- Banco de España

[8] as of 12 October 2018

- Banco de Portugal
- Bank of England
- Banque de France
- Central Bank of the Republic of Turkey
- Deutsche Bundesbank
- European Central Bank

INEXDA guests are institutions that have participated or have confirmed participation in at least one INEXDA meeting but have yet to sign the MoU. INEXDA guests:

- Banco de México
- Bank for International Settlements[9]
- Bank of Russia
- Federal Statistical Office of Germany
- Eurostat
- Oesterreichische Nationalbank
- Office for National Statistics UK
- Swiss National Bank

[9] The Bank for International Settlements supports the INEXDA initiative without being a full member.

How Banca d'Italia disseminates granular statistics to external researchers: The case of firms' and households' survey data

Laura Bartiloro
Banca d'Italia – DG Economics and Statistics – Statistical Analysis Directorate

Abstract
This note illustrates Banca d'Italia's long tradition in disseminating granular survey data to the public, the rationale for the choices done, pros and cons of the different options. After discussing the treatment of confidential data in Italy, the note describes the dissemination of households' and firms' survey data. In particular, Banca d'Italia Remote Data processing (BIRD) allows external researchers to perform economic analysis on granular data without any direct access to them. The note will also present Banca d'Italia's project for the future, namely the creation of a Research Data Center, including a laboratory, in order to facilitate the way internal and external users access microdata, to increase data availability, to better address related methodological issues and to facilitate data sharing with other institutions.

Keywords
Microdata access; Remote execution; Confidentiality

1. Introduction

Economists usually strive to have data at the maximum degree of detail possible in order to perform their econometric exercises and better understand economic phenomena. The use of granular data is increasing over time, thanks to micro-founded macro-economic models, becoming more and more popular, as well as the integration of macro data (aggregates) with micro information (distributions, etc.).

Elementary data can be disseminated to the public in different ways, depending on the confidentiality content of the information: a simple distinction is between anonymized data diffused on the producer's Internet web site via PUF (Public Use File) for information with loose or manageable confidentiality restrictions, and data with restricted access to protect the anonymity of the respondents (persons, organizations, firms), for which specific safeguards have to be implemented. Confidentiality means that in no way a user can trace back the name of the reference unit (household, firm, bank, etc.); it can be due to legal restrictions or to an agreement between data providers and respondents.

Banca d'Italia has started collecting information from households since 1962, followed, since 1974 by surveys on industrial and service firms. In due respect of the legal protection of personal data, ruled by national legislation before the European initiative (i.e. GDPR), the results of these surveys have been made available for the research community since the early Eighties, first in aggregated form and, more recently, with the provision of elementary data in various forms.

Similarly to what the ECB, Eurostat, other central banks and national statistical institutes collecting micro data do, Banca d'Italia has always granted access to microdata for scientific purposes only (i.e. the user must be part of an organization that has to be recognized by the data provider as a research entity).

This note describes how Banca d'Italia disseminates granular data to external researchers in the case of the Survey on Households Income and Wealth and some (but not all) of the business surveys run by Banca d'Italia, namely the Survey of Industrial and Service Firms, the Business Outlook Survey of Industrial and Service Firms, the Survey on Inflation and Growth Expectations and the Italian Housing Market Survey.

2. Confidentiality vs utility

The literature has made it clear that the core challenge in the dissemination of granular data is the balancing of the risk of re-identification with the utility associated with data analysis (see for example Schouten and Cigrang, 2003; Lane and Schur, 2010). The choice of the amount of risk that data providers accept in the name of utility depends on technical, organizational and legal issues. In general terms, the dissemination of Banca d'Italia's granular data faces two different confidentiality safeguards, the legal protection of personal data (GDPR UE/2016/679) and the professional secrecy of information collected by Banca d'Italia acting as banking supervisory and resolution authority (art.53 CRDIV and art.84 BRRD). The professional secrecy does not apply to households' and firms' data, where data are collected only for research purposes. As for the legal protection, GDPR does not forbid the processing of personal data "for scientific or historical research purposes or statistical purposes", while providing "appropriate safeguards for the rights and freedoms of the data subject [...]. Those safeguards should ensure that technical and organisational measures are in place in order to ensure, in particular, the principle of data minimisation." GDPR also foresees an additional protection of sensitive data, which does not concern households' and firms' survey data as this type of information is not usually included in Banca d'Italia's questionnaires.

In theory a pre-defined amount of risk or re-identification can be admitted, but for Banca d'Italia has always aimed at avoiding as much risk as possible. Still, data dissemination of granular data has increased over time, depending on the improvement of technological means that permit data usage by external researchers, while preserving confidentiality, where needed (Bruno, D'Aurizio and Tartaglia-Polcini, 2011). Households' survey data are under the general protection of the GDPR, while GDPR rules do not apply to companies, therefore no limit exists, in theory, for the dissemination of firms' elementary data. Contrarily, Banca d'Italia has chosen to not disseminate firms' granular data, because our business surveys include information that companies may not want to disclosure, as expected turnover, investment plans, etc.; therefore, in order to preserve a high response rate, we have chosen to not disseminate firms' elementary data. Indeed, Banca d'Italia makes it clear to her surveys' respondents that information provided through business surveys will not be diffused to the public in an elementary format. The promise of this additional safeguard is crucial for the collection of precious information that would not be elsewhere available.

3. Firms' survey data

Having voluntary chosen not to disseminate firms' elementary data, we had to find a way to safeguard confidentiality. Different options were available. Anonymization is the first, easy possibility. Still we have chosen not to go this way as we consider it an insufficient protection for firms' data, as researchers could always trace back a respondent firm by simultaneously using stratification dummies (region, size, sector). It could also then be possible to expunge these variables from the database, but by exactly matching a particular data (as for example turnover), re-identification of the respondent could in theory still be possible. Another possibility was the use of confounding techniques, but, again, we valued these techniques as not completely safe, because the algorithm used for confounding data could be identified. We neither chose the option usually adopted for data with the highest confidentiality restrictions, i.e. providing access to elementary data in a devoted laboratory (also called "data enclave"), where computers don't allow the user to take away any data if not in an aggregated form, and laboratory employees check the result of the elaborations and their non-violation of confidentiality restrictions. A laboratory significantly reduces utility, because the researcher has to go in person to the devoted location. On the contrary, willing to disseminate granular data to external economists and having adequate technical devices, we chose remote processing. Remote processing allows researchers to process elementary data, ensuring that individual information cannot be visualized.

Banca d'Italia introduced BIRD (Banca d'Italia Remote access to micro Data) in March 2008.[1] We are not the only central bank or data provider allowing for remote job submission (see for example: Germany – Gesis - Leibniz Institute for Social Sciences, with MISSY; Netherlands – CBS, with Microdata Services; Sweden – National Statistical Institute, with MONA).[2][3]

Instructions for using BIRD, information on available datasets and other useful documentation are published on Banca d'Italia web site, but BIRD does not work online. All communication between the researcher and BIRD happen via email. First interested researchers have to file for using BIRD. As explained on the dedicated web page, each researcher has to request a user license, filling in a form providing personal information (including a valid ID) as well some details on the research project for which data are needed. As said, this is an important step, as Banca d'Italia intends providing the use of elementary data only for research purposes and not for commercial use. In order to verify the scope of the requested access, the dedicated personnel of Banca d'Italia examines the form, mainly focusing on the current position of the researcher (whether or not she works for a university, research institute, etc. or for a private firm) and the research project. The user also signs a formal agreement with the privacy law and the deontological code. Once it is clear that there will be no commercial use of the data, Banca d'Italia admits the researcher to work with BIRD. Authorized users receive detailed instructions on how to submit programs for processing. Hence the researcher performs her statistical and econometric analyses in writing a program in a supported format (Stata, SAS, R) and submits it by email to a dedicated address.

First the program is subject to a legitimacy check: BIRD verifies that the program does not include any command comprised in a given list of commands (depending on the statistical package in use), namely those potentially able to disclose individual information (like "list" in Stata, or "proc print" in SAS). When the program includes one of these commands, execution is blocked and the system automatically sends an email to the researcher informing her about the formal rejection of the submitted elaboration. On the

[1] BIRD relies on the Lissy system of the Luxembourg Income Study, a project started in 1983 with the objective to open up microdata from a large number of countries for comparative research. Lissy consists of a series of software components connected through one or more networks, these components work together to receive, process and return statistical elaborations. All the components of the system are physically separated and at no moment is a user in direct contact with the data.

[2] See Schouten and Cigrang (2003) for a brief survey of other remote access systems worldwide.

[3] Available databases include: Survey of Industrial and Service Firms (since 1984); Business Outlook Survey of Industrial and Service Firms (since 1993); Survey of expectations of inflation and growth (since 1999); Italian housing market survey short-term outlook (since 2009); Survey on cross-border transactions in services by non-financial and insurance firms - direct reporting (since 2013); Selected items of banks' balance sheets (forthcoming).

other hand, when this initial check is passed, BIRD runs the program and a dedicated Banca d'Italia employee examines the output of the program in order to further verify that confidentiality is not breached by the submitted computations.[4] Once the output does not violate any confidentiality restriction, i.e. it does not identify information referred to any single or restricted group of firms (or banks, in the future), the output can be released to researcher. She hence receives an email with the cleared output. If, on the opposite, the manual check envisages a confidentiality violation, the researcher receives an email explaining the reason of the rejection. It could also be the case that the program is miswritten, and BIRD ends in error. Again the researcher receives an email reporting the misspelling of the program. In order to reduce the number of these occurrences and speed up the process, since 2016 a dataset with fake figures in semicolon-delimited ASCII format that replicates the internal structure of the original data from the Survey of Industrial and Service Firms (but contains randomly generated data) is available on Bank of Italy's web site, so that researchers can test the editing of their codes before submitting them to BIRD.

In order to prevent any violation to confidentiality restrictions there are several firewalls, at three different levels: user (users are identified, qualified and registered; registered mailboxes are whitelisted; outputs are monitored and archived; deontological code, privacy law, specific penalties); data (identifying variables are expunged from the datasets used for remote processing; extreme data are censored; stratification variables are collapsed); processing (forbidden to display individual data; keyword parser implemented with blacklist and greylist; particularly long and/or complex programmes are always reviewed manually; all submissions are reviewed manually).

Each institution allowing for remote processing of granular data provides a similar set of controls. Remote execution platforms are then considered reasonably safe and useful and thus remain an important tool for the dissemination of granular data for many data providers around the globe.

4. Households' survey data

The Survey on Household Income and Wealth (SHIW) was begun in the 1960s to gather data on the incomes and savings of Italian households. From the beginning to the publication of Banca d'Italia's Internet web site, dataset

[4] There's a clear trade-off between the length of the commands that are automatically forbidden, and the flexibility allowed to researchers in running computations and regressions. The higher the flexibility granted to the researcher, the higher the role played by Banca d'Italia employee in manually checking the output of the job submitted. As for the moment, the list of forbidden keywords is very limited, but should users and jobs submitted increase significantly and dedicated employees not balance this rising burden, the list of forbidden keywords could be enlarged.

were made available to the researchers (usually universities) that requested the dataset with a letter via ordinary mail. The dataset was sent in magnetic tape archives. Since the late Nineties, with the Banca d'Italia's web site, households' survey dataset have been made freely available on the web site, and lately also earlier datasets have been included: datasets are now out for every single survey in full (also including data found in the monographic sections of the individual surveys) from the 1989 wave on.

For households' survey data GDPR (as earlier the Italian law) states that individual data can be disseminated after data minimization. We therefore anonymize households' information and don't include identification variables (birth date and place, residency, day and hour of interview, etc.) in the datasets freely available in dedicated Public Use Files (PUF) on the Banca d'Italia web site. PUF is a common choice for data for which anonymization and the exclusion of some (indirect) identification variable are enough for safeguarding confidentiality.

No dataset request form is needed in order to download the PUF file. Many other institution (as for example the ECB and Banco di Espaňa) ask households' survey data users to fill in a request form online and data are then sent (or download is unblocked) after the request is validated by the data provider. We have deliberated for no request form and the PUF file is free for download without the identification of the user, also for the consideration that PUF can easily be transmitted between researchers. Maybe also thanks to this easy access, households' survey data are widely used, as witnessed by the numerous research papers that have been written over the years, and their use by some universities during exercising classes. Banca d'Italia just makes it clear to users that data are distributed for research purposes only and their use is under full and exclusive responsibility of the user. The drawback of this wide dissemination policy is that we can't track the users of households' survey data, as researchers usually fail to send us back their paper. We are therefore thinking about adding an online compiling form on the web page, in order to collect some basic information (the name of the researcher, institution or university affiliation, title of the research project, etc.) before unblocking the download of the PUF.

In the light of providing researchers with as much as useful information as possible, aside the archives for every single survey we deliver additional datasets with specific data adjustments. The first is an historical database containing data from 1977 but only including the subsets of variables considered useful for longitudinal analyses, after having performed a number of adjustments: as the way questions are posed (and therefore variables and/or variables options) may differ over time, variables are ex-post harmonized, so that time-series are continuous and longitudinal analyses can be easily performed by the researchers, without obliging them to prepare the dataset on their own before running the statistical computations. In addition, in the historical database some identifying variables (i.e. province of residency), previously expunged, have been added back for the period 1977-1986. We also deliver the Italian component of the Household Finance and Consumption Survey (coordinated by the ECB), adding some variables that are not included in the original SHIW dataset (as, for example, gross income). For all datasets, the information needed for data usage (questionnaires of the latest waves, variables names, instructions for data usage, etc.) are available in pdf format, while data are available in different formats (SAS, STATA, CSV).

5. Banca d'Italia experience so far and the way forward

As said we usually don't receive, and hence cannot count and archive, all research papers written by external researchers using households and firms survey data, we don't have a measure of the utility of Banca d'Italia's granular data dissemination. On the other hand, we can count the number of people submitting jobs through BIRD for business survey data and downloading the file with households' survey data. As for the first figure, we observe a huge volatility of users and jobs submitted: on average 7 jobs are run every week, even if the number of researchers is rather limited (around 10 new researchers per year file for using BIRD). We believe that these figures could easily increase as accessing data becomes more user-friendly. To this end Banca d'Italia has included in her 3-years strategic plan the goal to enrich the offer of granular data to the general public. We have hence started working on a brand new Research Data Center (RDC) in order to facilitate the way internal and external users access microdata, also increasing data availability and improving methodology. It will be a single entry point, notwithstanding differences in the permissions of use and access points allowed for every dataset, owing to the already discussed differences in households' and firms' survey data.[5] The RDC will therefore include all datasets already available as Public Use File (hence the households' survey data, and more) and in BIRD, but it will also provide two new important tools. The first one is a web tabulator, an easy to use device, available online, for performing tabulations from micro data in a personalized way, i.e. the researcher can choose categories and breakdowns to a certain extent. The web tabulator will have some built-in firewalls in order to prevent the identification of a single respondent or of a small number of respondents. In fact a simple, and commonly used, way to prevent re-identification, is imposing a minimum number of respondents

[5] This approach is common to other data providers. The Data without Boundaries programme for European data, ended in 2015, noticeably supported this approach, up to listing a set of principles for access to elementary data (Schiller and Welpton, 2013).

for any given variable and stratification variable (Freiman et als., 2016). Another possible solution has been recently proposed (Chetty and Friedman, 2019), which consists in adding noise to each statistics in proportion to its sensitivity to the addition or removal of a single observation from the data, in order to more efficiently cope with the trade-off between privacy loss and accuracy.

Another innovation of the Banca d'Italia's RDC will be the creation of a data enclave, a laboratory, situated in Banca d'Italia premises, where external researchers can receive assistance from a dedicated personnel and perform analysis on reserved datasets, today not available for external researchers (namely for datasets other than households' and firms' survey data). The creation of a RDC is also part of Banca d'Italia's commitment as participant of the International network of exchanging experiences on statistical handling of granular data (INEXDA), whose final goal is to facilitate the use by external economists of granular data produced by participating institutions for research and comparisons. The memorandum of understanding already signed by the 7 participating institutions foresees two steps: perform a pilot exercise consisting in a detailed inventory of available databases and existing procedures; explore harmonization for future integration of participating RDCs. This will possibly allow in the future to make comparative analysis using granular data on reserved datasets, thereby heavily contributing to understand international heterogeneities on firms' and households' behavior.

References

1. Bruno, G., D'Aurizio, L., Tartaglia Polcini, R. (2011). The Bank of Italy's experience with remote processing of business microdata. Mimeo.
2. Chetty, R., Friedman, J.N. (2019). A practical method to reduce privacy loss when disclosing statistics based on small samples. America Economic Review Papers and Proceedings. Forthcoming.
3. Freiman, M. H., Schar, B., Hasenstab, K., Lauger, A. (2016). Evaluating a remote access system. Research Report Series. Center for disclosure avoidance research. US Census Bureau.
4. Lane, J., Schur, C. (2010). Balancing Access to Health Data and Privacy: A Review of the Issues
   and Approaches for the Future. Health Services Research. Vol.45.
5. Schiller, D., Welpton, R. (2013). Providing remote access to European microdata. Conference paper. NTTS.
6. Schouten, B., Cigrang, M. (2003). Remote access systems for statistical analysis of microdata. Discussion paper n.03004. Statistics Netherlands

Sharing and using financial microdata

Alejandro Gaytán González, Manuel Sánchez Valadez, Mario Reyna Cerecero
Banco de Mexico

Abstract

Banco de Mexico has been collecting financial microdata of all market operations done by banks and brokerage houses for over 15 years. This data collection was possible because of a broad data sharing agreement between Mexican financial authorities after the 1995 financial crisis. More recently, new needs for monitoring financial institutions and markets, and the lessons learnt from the 2008 global financial crisis have made necessary improving data collection, data sharing and dissemination of microdata. Over the last years Banco de Mexico has: i) taken the task of a "Trade Repository Like" for derivative operations with new collection templates and new improvements on the quality of data and services; ii) implemented a new data sharing framework with other financial authorities, and new MoUs have been signed; iii) continued the revision and expansion of metadata as an undergoing process for improving the use of these microdata bases; and, iv) designed and developed a new dissemination portal for microdata.

Keywords

Microdata; data sharing; derivatives

1. Introduction

Banco de Mexico has been collecting financial microdata of all market operations done by banks and brokerage houses for over 15 years. The Mexican financial crisis in the mid 1990's unveiled several data gaps and potential improvements in the collection of financial information. As a response, Banco de Mexico started collecting financial microdata solutions to provide flexible solutions for different information needs in Banco de Mexico and other financial authorities. This decision transformed the financial information model from a traditional model to generate central bank statistics to a model heavily based on timely daily granular microdata, mainly with all market operations by all banks and brokerage houses (Gaytán and Sánchez, 2017).

Banco de Mexico concentrated the collection of microdata of financial market operations, regulatory regimes and consumer credit, while the Bank and Securities Supervisor collected microdata on mortgages and the commercial credit portfolio, in addition to a set of other regulatory reports.

This specialization was the result of an agreement to reduce the regulatory burden of information reporting and to share financial information between financial authorities. In 2000, different financial authorities signed this data sharing agreement: Banco de Mexico, the Ministry of Finance (SHCP), the Bank and Securities Supervisor (CNBV), the Financial Service protection Agency (CONDUSEF), the Deposit Insurance Agency (IPAB), and some years later, the Pension Funds Supervisor (CONSAR). Recently, the 2008 financial crisis and the international initiatives to improve financial stability, have implied several improvements in the acquisition, management and sharing of financial information, including the G-20 data gaps initiatives, the implications of Basel III, the initiative to mitigate risks in the Over the Counter (OTC) derivatives market, shadow banking, among other.

Financial information at the level of market operations microdata allows attending different users and information needs, such as open risk positions of an individual institution or the network of exposures in different markets and in the whole financial system. The increased complexity of the interlinkages, instruments and institutions, require increasing capacity to identify potential risks, even though the costs of a detailed model of microdata is high, specially "in times of financial turmoil, the advantages of having the precise information surpasses any maintenance costs associated with such a model, nonetheless, there are also great benefits in steady times" (Gaytán, 2014). The costs of such an information model are high, both for the authority that collects it and for the reporting institutions. Thus, to maximize the social value of this model, it is important to broaden its use by improving data sharing schemes among authorities and find ways to provide wider access to academic researchers, market analysts and the public.

The paper proceeds as follows: section 2 presents a description of current data sharing schemes at Banco de Mexico. In section 3 describes enhancements on the scope of information of the financial system managed by Banco de Mexico, focusing on information of derivative operations. Meanwhile, section 4 describes recent undertakings to expand data sharing with other Mexican financial authorities and the improvement of tools for data dissemination for diverse audiences. Finally, section 5 mentions some challenges ahead regarding data sharing.

2. Current Schemes for Data Sharing at Banco de Mexico

The 2000 agreement signed by financial authorities to coordinate actions to compile, store, share and disseminate the information received from financial intermediates, set the foundations of a more efficient system of financial reporting to authorities (Gaytán and Sánchez, 2017).

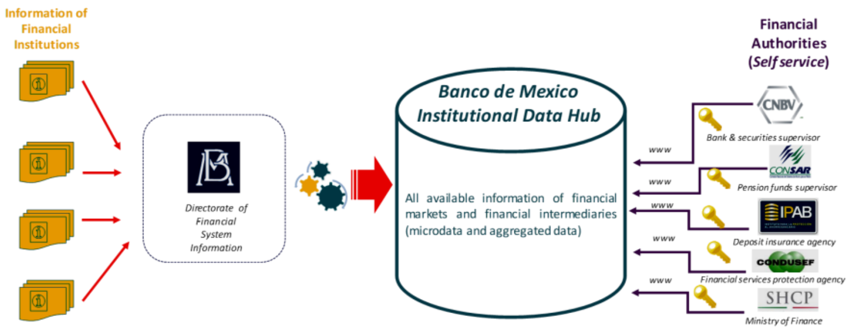
Currently, Banco de Mexico uses three main schemes for sharing data with other financial authorities and information users (Figure 1). First, a Central Data

Hub provides secure direct access both to granular and aggregated microdata of financial institutions reporting to Banco de Mexico. This hub provides querying tools to databases, reporting services and business intelligence tools. Second, a controlled service scheme to share very large volumes of a predetermined set of microdata delivered according to a calendar agreed with the user. Finally, public platforms to access time series and interactive graphics.

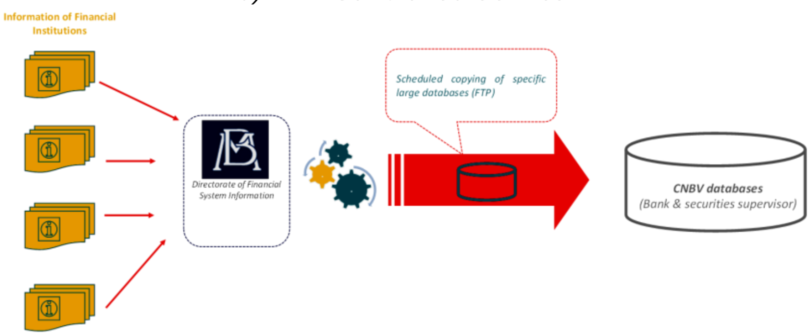
The data sharing schemes with financial authorities were significantly improved in 2014, when a financial reform included the basis of a new mandatory framework of data sharing among domestic financial authorities for specific purposes: preserve financial stability, avoiding disruptions in the functioning of the financial system and/or the payments system. In addition, Banco de Mexico was given the faculty to share information with foreign financial authorities after the signing of Memorandums of Understanding (MoU) that establish the conditions of the information exchange and include a reciprocity principle. These changes made improvements in the analysis possible, supervision and regulation functions of financial authorities (Gaytán and Sánchez, 2017).

Figure 1. Data Sharing Schemes

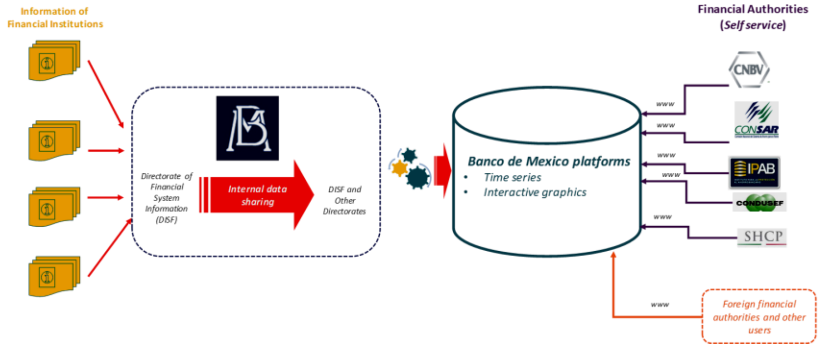
a) Central Data Hub (direct access)



b) Controlled Service



c) Self Service



Source: Gaytán and Sanchez (2017).

3. Recent Improvements in the Model of Financial Information

Over the last years, Banco de Mexico has been working on improving the scope of its financial system information model. With respect to the credit market, Mexican financial authorities have information loan by loan of banks and other regulated credit institutions (Sofomes E.R.), Banco de Mexico collects data on consumer credit portfolio (credit cards and other consumer loans) and CNBV collects the mortgage and commercial loans portfolio. In recent years, there have been important improvements. Banco de Mexico started requesting information to Credit Bureaus´ on loans to firms and households, which has improved the integration of the information and has increased the availability of data on loans provided by unregulated financial institutions. In addition, the bank supervisory improved the collection of commercial portfolio data with an improvement in the consistency and level of detail of information. On the other hand, Banco de Mexico has improved the collection of consumer credit by:

i. Improving individual credit risk information;
ii. Requesting information on consumer credit clients, that will help to improve both financial stability and financial inclusion analysis; and,
iii. Including the initial Total Annual Cost (CAT),[1] which facilities cost comparisons of credit products across institutions.

The global financial crisis also had important implications for the development of new precautionary regulation for financial institutions. To enhance the resilience of the financial system, Basel III established new standards for capital adequacy as well as new standards on liquidity, namely the Liquidity Coverage Ratio (LCR) and the Net Stable Funding Ratio (NSFR). In Mexico capital adequacy regulation is a responsibility of CNBV and liquidity is a joint regulation of Banco de Mexico and CNBV. However, Banco de Mexico collects and verifies the information of both regulatory regimes as it uses the

[1] The CAT is the domestic name of the Annual Percentage Rate (APR), which includes the annualized interest rate plus all fees attached to the loan.

microdata model to replicate a large proportion of these ratios using the microdata of operations. Currently the LCR is a daily requirement with a 10 days lag revision. These improvements have enhanced financial stability analysis and surveillance of financial institutions.

Banco de Mexico, has also implemented new information requirements to improve the surveillance of the payments system. In 2015, Banco de Mexico improved the template, timeliness and quality of the report of all credit and debit card transactions, increasing the fields of the information to include data on the location of each transaction and the activity of the business, among other.[2] The timeliness of the reporting is one day. More recently, this information was extended with a new requirement to include not only information on the transaction, but also, information that includes all the confirmation messages related to each transaction. In 2017 Banco de Mexico implemented a new requirement about all cross-border transactions made by financial intermediaries via correspondent banking.

Another important improvement has been the information model on derivatives. Banco de Mexico has been collecting information transaction by transaction on derivatives operations performed by banks and brokerage houses since 1999, and it was a natural step to operate with the functions of a "Trade Repository (T.R.) like" infrastructure for the international initiative of IOSCO-CPMI to reduce the risk in the OTC derivative market.

To better perform the functions of a T.R. like, an important update of the reports was made in 2015 to include new developments in markets and instruments, information to respond to new demands by users and information related to regulatory changes, particularly the standardization of OTC operations and the central counterparty. In this respect it was designed to fulfil a set of data elements consistent with the principles proposed by international institutions (IOSCO) and regional regulators (ESMA), which includes a better identification of underlying assets, products (ISDA taxonomy), netting agreements, among other.

Another improvement is to centralize in a single template, a catalogue of entities, the characteristics of counterparties of all operations, including derivatives, of the banks and brokerage houses in financial markets. These characteristics include the tax identifier as well as the Legal Entity Identifier (LEI), economic activity, residency, and relevant relations with the party, among other.[3] This new requirement allows a better identification of the different entities trading in the domestic financial markets and of foreign exposures. In addition, and linked to the counterparty information, the data on collateral of

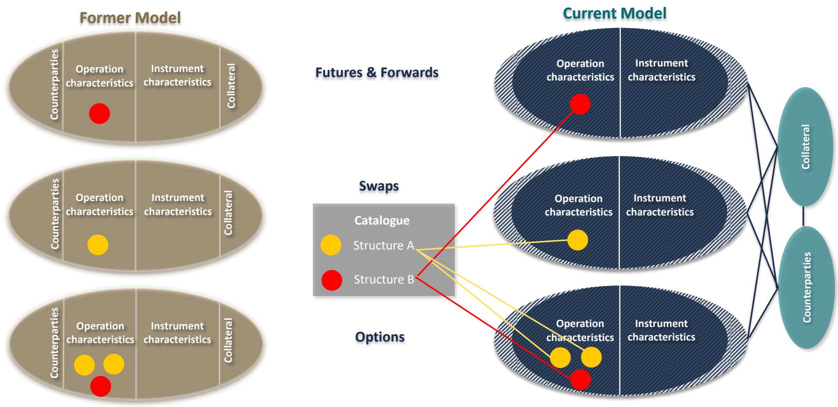
[2] To identify the kind of activity of merchant is used the Merchant Category Code (MCC).

[3] The economic activity identification is according to North American Industrial Classification System (NAICS).

derivative operations was separated from the individual transaction as collateral is generally related to a set of operations with the same counterparty (Figure 2).

Before 2015, the template had some limitations to account precisely of the “strategies” with derivatives operations, and these limitations sometimes hindered the complete identification of what transactions were part of the strategy. To close this data gap, it was created a precise way to link transactions that form part of a “structure” or “strategy”.

Figure 2. Changes in Derivatives Information Requirements



Source: Banco de Mexico.

Nevertheless, to operate as a T.R. like for derivatives it is not sufficient to improve the data model, it is also necessary to adjust the operation of processes of data and adjust the services provided to perform the functions of a Trade Repository in a central bank. IOSCO-CPMI define a series of best practices established in a set of Principles for Financial Market Infrastructures defined in the “Principles for financial market infrastructures” (BIS & IOSCO, 2012). To implement the necessary changes, it was defined a new operational model for the functioning of the T.R. like. The idea is to pass from a model centred on the service of the provision of information to another with three main services: Registration of derivative transactions, the provision of information and the design and development of information solutions.

One of the main functions of the initiative of Trade Repositories is to increase the transparency of the OTC derivatives market, which implies the establishment of mechanisms of data dissemination and data sharing. A T.R. should provide information to market participants, other domestic and foreign financial authorities and provide publications of the operation of the derivatives market. Currently, the scope of data dissemination is done using the schemes for data sharing at Banco de Mexico that were presented (Figure 1). The data hub and the direct access scheme to microdata can be used to attend data requirements of financial authorities and there is a broad set of statistics and graphs of the operation of the market published using the time series platform and the interactive graphic platform. Nonetheless, a broader definition and implementation of improvements of the data dissemination processes and tools is still work in progress.

4. Recent Improvements on Sharing and Dissemination of Microdata and some Potential Uses

As mentioned above, the financial reform strengthened the data-sharing scheme. In recent years as a complement to the 2000 general agreement, two additional MoUs were signed between Banco de Mexico and CONDUSEF in 2015, and with CONSAR in 2017. In addition, Banco de Mexico is currently working on MoUs with other domestic financial authorities.

Banco de Mexico shares information with CONDUSEF about banks' financial products (deposit products and credit products) and their characteristics and services, fees and interest rates, transactions in retail payment systems (checks, ATM transactions, TPS transactions, electronic transfers, etc.) and e-commerce payments.

CONSAR and Banco de Mexico share information on derivatives and securities transactions performed by banks with the Pension Funds (SIEFORES), daily information about the composition of the investment portfolio of these funds and data on savings, and the demographic characteristics of workers.

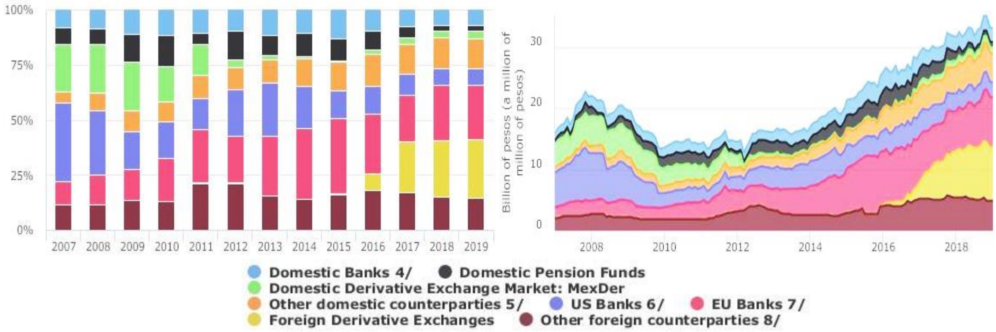
Data dissemination by publication is another important way of increasing the value of the information collected. Banco de Mexico has been expanding the publication of financial data, particularly on derivatives.

In 2016, Banco de Mexico published a broad set of statistics on derivatives (turnover and outstanding operations, and the forward exchange rate). In the same year, it was launched an interactive portal for financial information (PIIF). In this platform, it was included a graphic overview of the derivatives market in Mexico using some of the most relevant published derivatives statistics. More recently, graphic information was published on outstanding government securities sectored by holder (securities issued by the Federal Government, IPAB, and Banco de Mexico).

Although currently the PIIF has only been used to disseminate graphic information, this portal will be used to disseminate micro-data on loans and market operations. Figure 3 and Figure 4 are examples of derivate operations information included in the PIIF. Figure 3 shows outstanding derivatives by type of counterparty and, Figure 4 shows turnover and average exchange rates on forwards on US dollar vs Mexican peso.

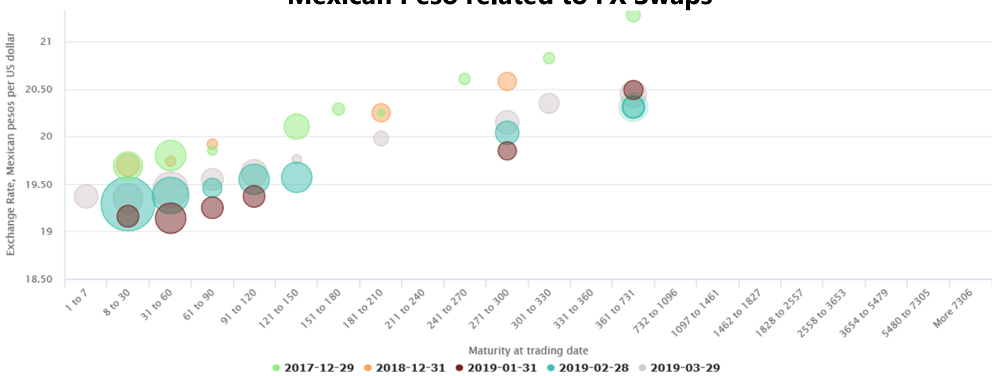
Figure 3. Outstanding Derivatives by Type of Counterparty

a)Structure 1/ 3/ b) Outstanding 2/ 3/



1/ Figures at end of year. Figures at the end of January 2019. 2/ Figures at end of month. 3/ Figures on Options and Warrants are preliminary from September 2015, due to significant revisions in course. 4/ Figures are only on domestic commercial banks transactions. 5/ Domestic development banks, brokerage houses, other financial entities, and non-financial entities other than private enterprises. 6/ Figures are only on commercial banks located in the United States. 7/ Figures are only on commercial banks located in the European Union. 8/ Figures are only on commercial banks located in Latin America and others foreign financial entities and non-financial entities. Source: Banco de Mexico.

Figure 4. Turnover and Average Exchange Rates on Forwards on US Dollar vs Mexican Peso related to FX Swaps



Source: Banco de Mexico.

Recently all financial information at Banco de Mexico went through a process of revision of its metadata to determine a set of characteristics that will help the identification of its characteristics and, thus, its use. These metadata forms the basis of a new financial data inventory. The metadata includes information about the collection (frequency, timeliness, granularity, etc.), content, and sources of information to help traceability, the reporting institutions, sensitivity information among other. It is worth mentioning that the revision and expansion of metadata is an undergoing process for improving the use of microdata managed by Banco de Mexico.

5. Concluding Remarks

One of the main objectives of enhancements in the model of information at Central Banks is to maximize the potential social value of data. According to the experience of Banco de Mexico, in order to maximize the social value of information it is necessary to improve its: i) efficiency in its generation, ii) accuracy, and iii) dissemination.

In the case of Banco de Mexico, the process of generation of derivative operations information is under revision and enhancement, which has implied adopting international best practices, improving information security and increasing the focus on solving user's needs. From this experience, these improvements will be adopted into the rest of the information process at Banco de Mexico.

As was mentioned above, having microdata allows improving the accuracy of information. In effect, having microdata allows the Banco de Mexico to check consistency between different information requirements, which in turns helps improve information quality. In particular, one potential enhancement is to join diverse data bases allowing an easy identification of parties and counterparties in different markets, which could be helpful to identify risk exposures.

Another key aspect is data sharing practices. Banco de Mexico has improved data sharing agreements and platforms. In this respect, the PIIF, a relatively new platform of data dissemination, has allowed to share new statistics under different formats.

These improvements notwithstanding, Banco de Mexico is still working on data dissemination, specifically in microdata dissemination. In this respect, a data room is an option to expand microdata dissemination to diverse audiences.

It is worth mentioning that our final objective of all improvements in the model of financial information and data sharing practices is to increase the amount and quality of information in order to do better analysis and to take better decisions, which as result will maximize the potential social value of information.

References

1. Banco de Mexico Web page, PIIF site:http://www.banxico.org.mx/IndicadoresGraficos/actions/portada?locale=en
2. Bank for International Settlements and International Organization of Securities Commissions (2012) Principles for financial market infrastructures.

3. Gaytán Gonzalez M.A. (2014) The use of micro-data in the financial system information model of Banco de Mexico. In: Integrated management of micro-databases. BIS, IFC Bulletin No 37.
4. Gaytán González M.A. and Sánchez Valadez M. (2017) Sharing information with financial authorities. The case of Banco de Mexico, Document presented at 61th World Statistics Congress of the International Statistical Institute (ISI), Marrakech.

Fake news and Big Data: An overview of solutions

Nuno Guimaraes[1], Kimmo Vehkalahti[2], Pedro Campos[1,3]
[1] INESC TEC, University of Porto
[2] University of Helsinki
[3] Statistics Portugal

Abstract
Rapid online spread of misinformation like fake news has become one of the top trends in modern societies. It has kept increasing, affecting elections and other policy-making processes, skewing public opinion, and diminishing trust in science. To tackle the problem, Facebook, Google, Twitter, and other companies are now using big data to identify websites that publish fake information. By applying methods of data science, fake news can be detected by scoring web pages, flagging sensationalistic words, weighing facts, and checking website domain names and Alexa rankings to gauge their reputation. Nevertheless, critical literacy is needed more than ever for reading and understanding of news, especially in the context of Online Social Networks. In this paper, we discuss how machine based actions and human based skills may help in fighting against fake news.

Keywords
Critical Literacy; Online Social Networks; Social Data Science

1. Introduction

The era of Online Social Networks (OSN) has revolutionized the way content is propagated on the web. Before the rise of Facebook and Twitter, social networks already existed but none of them achieved the status that these two OSN have achieved regarding the distribution and propagation of personal and newsworthy content. Consequently, it has forced well-established news media to adapt to a situation where information is diffused at a fast pace and where users' focus is limited to a few milliseconds per post. In addition, with the decay on the sales of newspapers and tv viewership, social media has provided the obvious path towards the survival of traditional news media.

In order to monetize on OSNs, news media have relied on online subscriptions and ad-revenue (generated through clicks and views of the source website). However, the easiness of creating accounts on social media and publishing content as well as the wider reach that these provide, has lead some users to spread dubious content on the platforms with the purpose of profiting through ad-advertising method or changing other users' perceptions.

In 2014, World Economic Forum identified the rapid spread of misinformation online as one of the top 10 trends in modern societies (World Economic Forum, 2014). The continuous increase of dubious information in OSN is indeed concerning as it affects real-world events like elections (Allcott & Gentzkow, 2017) and other policy-making processes. With approximately 67% of the Americans consuming news through social media (Gottfried & Shearer, 2017) and other similar results all over the globe (Reuters Institute, 2017) it becomes crucial to develop mechanisms that are capable of detecting this type of dubious information online.

This paper is structured as follows: We start in Section 2 by describing various types of dubious information, namely false information, biased news, and misleading statistics. In Section 3, we present possible solutions for the problems, ending up with Discussion and Conclusions in Section 4.

2. Dubious information

In this section, we describe three types of dubious information: a) false information, b) biased news, and c) misleading statistics.

a. False information

A common Online Social Network (OSN) user should learn to be critical regarding his/her personal feed, as there are many types of false information online and several techniques that are used to deceive users into believing such false information.

One of the most common techniques is URL mimic where false information websites have a similar address to real ones. A wellknown example is *bloomberg.ma* website which was responsible for publishing a news piece about an offer for the takeover of Twitter, resulting in an increase in the shares of the companies (de la Merced & Goldstein, 2018). So, users must be aware of the origin of the content. In order to tackle this problem, Facebook has implemented a new feature where it provides a context to external URLs shared in posts (Facebook, 2017). Similar tools in other social media are yet inexistent.

A more recent method to spread misinformation uses the hacking of a verified account to reach a high potential number of followers. The "Doubleswitch" method targets particular high reputation social media accounts (e.g., journalists, politicians, human rights activists) to spread dubious information and makes recovering the access to those accounts non-trivial (Security, 2017). Therefore, even when the sources of posts are trustworthy, its content should not be taken as facts.

Another problem that highlights the importance of journalists and publications in the spread of false information is the "false confirmation" or "circular reporting". This particular scenario occurs when information that seems to be checked and confirmed by several sources is indeed coming from a single one, with other sources reprinting the content without double checking it (Filkins & Fisher, 2003). However, this type of circular reporting also happens when crowdsourced wikis are used by journalists to gather information. For example, when a journalist copies false information from a Wikipedia page without checking other sources, then the Wikipedia page can cite the journalist's publication to increase the trustworthiness of the false information published. A real-world example occured when *The Independent* published the information that the actor Sacha Baron Cohen worked at Goldman Sachs (Techbug, 2009). Therefore it is important for journalists to be aware of this type of events and always validate their sources several times. Creating and spreading false news in mainstream sources is much depending on the journalists writing for them.

b. Biased News

Although false information makes up a large portion of the dubious content spread in OSNs, there are specific scenarios where the information spread is not false, but the way it is published can lead readers to perceive it in a biased way where one or several of the agents are diminished or criticized compared to others. In addition, a careful choice of words and the omission or decontextualization of some facts is used to tell a story that guides users' believes in a certain direction.

Several extremely biased websites were created during the 2016 U.S. Presidential election, with their content proliferating in social media. One of the most recent examples was the doctored Jim Acosta video released by an extreme-right website and later "retweeted" by the White House Press Secretary (Harwell, 2018). The video portraits a real situation. However, it had been deceptively edited, causing a serious bias in the users' perception.

Nevertheless, it is not only this type of news websites that share bias content. A common example in mainstream news media is referenced in Gentzkow & Shapiro (2006) regarding the battle in the city of Samara in Iraq. While *The New York Times* began their report of the event by stating "American commanders vowed on Monday that the killing of as many as 54 insurgents in this central Iraqi town would serve as a lesson to those fighting the United States, but Iraqis disputed the death toll and said anger against America would only rise" (Filkins & Fisher, 2003), Al Jazeera wrote "The US military has vowed to continue aggressive tactics after saying it killed 54 Iraqis following an ambush, but commanders admitted they had no proof to back up their claims." (AFP, 2003). This kind of biased information highlights the importance of critical literacy in today's society.

c. Misleading Statistics

Statistics have a crucial role everywhere in the modern society. They appear in the form of numbers, summaries, tables, and visualizations. Although there is an old saying about "lies, damned lies and statistics", the statistics may be perfectly reliable and accurate but still seriously misleading because of the way they are presented or because of the way they are interpreted or (mis)understood. Even simple numbers may sometimes be dramatically misleading. For an example, see Taub (2018) or Willis (2018).

In general, most people do not feel comfortable with statistics. For common OSN users but also for journalists, simple numbers or percentages may often be surprisingly demanding to cope or work with, let alone statistical summaries like means, standard deviations, and correlations, or some results of statistical analyses and tests (say, p-values, confidence intervals, and regression coefficients). Perhaps the most challenging thing here is that statistics are not just plain numbers (although that is what they typically look like). Instead, statistics are reflections of complex phenomena behind the numbers, and hence interpreting statistics requires at least some level of understanding of those phenomena, which might often be far from trivial. Understanding the uncertainties and pitfalls related to data gathering and measurement is also crucial for assessing the validity and reliability of statistical information.

The classic book *How to Lie with Statistics* by Darrell Huff (1954) does still the job of summarizing the typical ways and forms of presenting misleading statistics. However, the (at least apparently) easy ways of creating visualizations and extremely easy ways of spreading them rapidly online has multiplied the number of possible situations where presenting or interpreting statistics appears misleading in a way or another.

It is also noteworthy that in the current world of OSNs, (dis)information typically spreads via quick instinctive reactions instead of slow critical thinking. A recent book *Factfulness* (Rosling et al., 2018) goes through ten specific instincts that easily distort our views of the world. Many of those instincts are indeed closely related to interpreting statistics, especially in visual forms. The aim of the book is to encourage the reader to replace his/her instinctive reactions by critical thinking. In addition of old and new books, plenty of useful materials in various forms are openly available on the web, see for example, Blauw (2016), Liddell (2016), Rosling (2010), and Steinberg (2017). Excellent handbooks especially for (data) journalists have been compiled by Gray & Bounegru (2019) and Gray et al. (2012).

3. Solutions

All this calls for skills of critical statistical literacy. Following Weiland (2016), critical literacy is related to individuals being able to write both the word and world—transforming their lived realities through the power of literacy[1]. Based on the intersection between Statistical literacy[1] and Critical literacy, the author introduces the key elements of critical statistical literacy in terms of reading and writing information that are crucial to critical citizenship in today's data centric societies. Among these key elements, Weiland (2016) identifies the following: "evaluating the source, collection and reporting of statistical information and how they are influenced and shaped by the author's social position and sociopolitical and historical lens" and "Communicating one's social location, subjectivity, and political context to others and how it shapes one's meaning making of the world when reporting results of a statistical investigation."

The problem of misinformation online has reached a proportion where companies like Google, Facebook, and Twitter are forced to intervene. Solutions are needed both for machine based actions and human based skills. Facebook has implemented a system to give more information on the source where an external URL is propagated in a post (Facebook, 2017), Twitter has removed several accounts who are bots and who were spreading fake news (Timberg & Dwoskin, 2018), and Google is tackling the problem by improving media literacy (England, 2019). Several independent initiatives to develop tools for detecting misinformation have occurred with the "Fake News Challenge", a competition to developed machine learning algorithms for detecting a stance of a claim (Riedel, Augenstein, Spithourakis, & Riedel, 2017) and SemEval task 4 whose goal was to detect if a piece of news was hyperpartisan (Kiesel et al., 2018).

The research community has also been active on the topic. Several solutions have been proposed to tackle problems of misinformation online. More specifically, research has tackled the analysis of false news (Shao et al., 2018) and their propagation on the network (Vosoughi, Roy, & Aral, 2018), the detection of bot and spam accounts (Benevenuto, Magno, Rodrigues, & Almeida, 2010; Chu, Gianvecchio, Wang, & Jajodia, 2012) and other users that are responsible for the spreading of unreliable content (Guimaraes, Figueira, & Torgo, 2018). The targets of research also include the classification of misinformation using machine and deep learning (Ruchansky, Seo, & Liu, 2017; Tacchini, Ballarin, Della Vedova, Moret, & de Alfaro, 2017) and the development of systems that allow fact-checking of claims recurring to external knowledge databases (Ciampaglia et al., 2015; Shiralkar, Flammini, Menczer, & Ciampaglia, 2017).

---

[1] For a deeper understanding of the concept of statistical literacy, see, for example, Gal (2003) and Watson and Callingham (2003).

4. Discussion and Conclusion

The continuous increase of dubious information in Online Social Networks (such as false information, biased news, and misleading statistics) is indeed concerning as it affects real-world events like elections and other policy-making processes which critical citizens must critically make sense of and evaluate. It is crucial, as stated by Weiland (2016), for critical citizens to be able to use that "statistical" power to influence, shape, and transform the socially constructed discourses and structures around them in order to create a more just world. Modern methods using machine and deep learning and fact-checking of claims recurring to external knowledge databases may help us getting empowered.

Let us close this paper by quoting Samuel S. Wilks, who in his 1951 presidential address to the American Statistical Association paraphrased a note, originally written by H. G. Wells in his 1903 book *Mankind in the Making*, in a shortened and simplified form: "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

References

1. AFP. (2003, December). Samarra clash toll still a mystery. *News | Al Jazeera*. Al Jazeera. Retrieved from https://www.aljazeera.com/archive/2003/12/200849131528113385.html
2. Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2)211–236. https://doi.org/10.1257/jep.31.2.211
3. Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting Spammers on Twitter.*Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS 2010)*, 1–9.
4. Blauw, S. (2016). How to defend yourself against misleading statistics in the news, TEDx Maastricht. Retrieved from https://youtu.be/mJ63-bQc9Xg
5. Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*, *9*(6), 811–824. https://doi.org/10.1109/TDSC.2012.75
6. Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., & Flammini, A. (2015). Computational fact checking from knowledge networks. *PLoS ONE*, *10*(6), 1–13. https://doi.org/10.1371/journal.pone.0128193
7. de la Merced, M. J., & Goldstein, M. (2018, January). Twitter Shares Jump After Faked Bloomberg Report. *The New York Times*. The New York Times. Retrieved from

https://www.nytimes.com/2015/07/15/business/dealbook/twitter-shares-jump-after-fake-bloomberg-report.html

8. England, R. (2019). Google explains how it's fighting fake news. Retrieved from https://www.engadget.com/2019/02/19/google-explains-how-it-is-fighting-fake-news
9. Facebook. (2017). New Test to Provide Context About Articles. *Facebook Newsroom*. Retrieved from https://newsroom.fb.com/news/2017/10/news-feed-fyi-new-test-to-provide-context-about-articles/
10. Filkins, D., & Fisher, I. A. N. (2003, December). U.S. Sees Lesson for Insurgents in an Iraq Battle.
11. *The New York Times*. Retrieved from https://www.nytimes.com/2003/12/02/world/us-sees-lesson-for-insurgents-in-an-iraq-battle.html
12. Gal, I. (2003). Expanding Conceptions of Statistical Literacy: an Analysis of Products from Statistics Agencies. *Statistics Education Research Journal*, 2(1), 3–21.
13. Gentzkow, M., & Shapiro, J. M. (2006). Media Bias and Reputation. *Journal of Political Economy*, *114*(2), 280–316. https://doi.org/10.1086/499414
14. Gottfried, B. Y. J., & Shearer, E. (2017). News Use Across Social Media Platforms 2017. *Pew Research Center*, *Sept 2017* (News Use Across Social Media Platforms 2017), 17. Retrieved from http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/
15. Gray, J., Bounegru, L. & Chambers, L., eds. (2012). *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. Retrieved from https://datajournalismhandbook.org/
16. Gray, J. & Bounegru, L., eds. (2019). *The Data Journalism Handbook 2: Towards a Critical Data Practice*. Retrieved from https://datajournalismhandbook.org/
17. Guimaraes, N., Figueira, A., & Torgo, L. (2018). Contributions to the Detection of Unreliable Twitter Accounts through Analysis of Content and Behaviour. In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, {IC3K} 2018, Volume 1: KDIR, Seville, Spain, September 18-20, 2018.* (pp. 90– 99). https://doi.org/10.5220/0006932800900099
18. Harwell, D. (2018, November). White House shares doctored video to support punishment of journalist Jim Acosta. *The Washington Post*.

WP Company. Retrieved from https://www.washingtonpost.com/technology/2018/11/08/white-house-shares-doctored-video-support-punishment-journalist-jim-acosta

19. Huff, D. (1954). *How to Lie with Statistics.* Available online at https://archive.org/details/HowToLieWithStatistics
20. Kiesel, J., Potthast, M., Mestre, M., Shukla, R., Stein, B., Corney, D., ... Adineh, P. (2018). SemEval 2019 Task 4 - Hyperpartisan News Detection. https://doi.org/10.5281/ZENODO.1406208
21. Liddell, M. (2016). How statistics can be misleading, TEDEd. Retrieved from https://ed.ted.com/lessons/how-statistics-can-be-misleading-mark-liddell
22. Reuters Institute (2017). *Digital News Report*. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017 %20web_0.pdf
23. Riedel, B., Augenstein, I., Spithourakis, G. P., & Riedel, S. (2017). A simple but tough-to-beat baseline for the Fake News Challenge stance detection task, 1–6. Retrieved from http://arxiv.org/abs/1707.03264
24. Rosling, H. (2010). The Joy of Stats. Wingspan Productions for BBC. Retrieved from https://www.gapminder.org/videos/the-joy-of-stats/
25. Rosling, H., Rosling, O. & Rosling Rönnlund, A. (2018). *Factfulness: Ten Reasons We're Wrong About the World – and Why Things Are Better Than You Think.* London: Sceptre.
26. Ruchansky, N., Seo, S. & Liu, Y. (2017). CSI : A Hybrid Deep Model for Fake News Detection, Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 797–806. https://dl.acm.org/citation.cfm?id=3132877
27. Security, D. (2017, June). The "Doubleswitch" social media attack: a threat to advocates in Venezuela and worldwide. *Access Now*. Retrieved from https://www.accessnow.org/doubleswitch-attack/
28. Shao, C., Hui, P. M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018). Anatomy of an online misinformation network. *PLoS ONE*, *13*(4), 1–23. https://doi.org/10.1371/journal.pone.0196087
29. Shiralkar, P., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2017). Finding streams in knowledge graphs to support fact checking. *Proceedings - IEEE International Conference on Data Mining, ICDM*, *2017-November*, 859–864. https://doi.org/10.1109/ICDM.2017.105

30. Steinberg, L. (2017). Lies, Damned Lies and Statistics: A Data Literacy Primer. Retrieved from https://eavi.eu/lies-damned-lies-statistics-data-literacy-primer/
31. Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some Like it Hoax: Automated Fake News Detection in Social Networks, 1–12. Technical Report UCSC-SOE-17-05, School of Engineering, University of California, Santa Cruz. https://arxiv.org/pdf/1704.07506
32. Taub, A. (2018). Lies, Damned Lies, and One Very Misleading Statistic. The New York Times.Retrieved from https://www.nytimes.com/2018/02/28/world/americas/un-sexual-assaults.html
33. Techbug. (2009). Wikipedia Article creates Circular references. Retrieved from https://techdebug.com/blog/2008/04/19/wikipedia-article-creates-circular-references/
34. Timberg, C., & Dwoskin, E. (2018). Twitter is sweeping out fake accounts like never before, putting user growth at risk. Retrieved from https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk
35. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
36. Watson, J., & Callingham, R. (2003). Statistical Literacy: a complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46.
37. Weiland, T. (2016). Towards a Framework for a Critical Statistical Literacy in High School Mathematics. Wood, M. B., Turner, E. E., Civil, M., & Eli, J. A. (Eds.), *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Tucson, Arizona.
38. Willis, K. (2018). Statistic claiming there have been 18 school shootings in 2018 is misleading. Retrieved from https://www.ajc.com/news/national/statistic-claiming-there-have-been-school-shootings-2018-misleading/ODJ4SFcVfr1gX6LtOZ4olL/
39. World Economic Forum (2014). Top 10 trends of 2014: 10. The rapid spread of misinformation online. Retrieved from http://reports.weforum.org/outlook-14/top-ten-trends-category-page/10-the-rapid-spread-of-misinformation-online/

Birth registration: An essential element for the individual and for government

F. Sam Notzon
Director, International Statistics Program, National Center for Health Statistics, CDC

Abstract
Birth registration establishes the identity of the individual, is a requirement for many government services (education, health services, etc.), and provides data useful for researchers, government and others. The article describes the importance of birth registration for the individual and others, considers measures to improve birth registration completene ss in low- and middle-income countries, and cites various applications of birth registration data. The focus of the proposed improvement measures is on reasonable-cost, sustainable improvements for low-resource countries.

Keywords
Civil registration; birth registration; vital statistics

1. Introduction

According to the United Nations, birth registration is required to establish the identity of the individual and to safeguard individual rights to social status and benefits. Birth registration is best assured through a civil registration system, that is, the continuous, permanent and compulsory recording of the occurrence and characteristics of all vital events, in accordance with the laws of each country.[1] Once collected, this information on births is used to produce legally valid birth certificates and birth statistics. The characteristics of birth registration within a civil registration system – continuity, permanence and universality – make it extremely useful in addressing the individual and governance needs of a country. Consequently, all developed and most developing countries ensure registration of births via a civil registration system.

2. Advantages of birth registration

As noted, birth registration is important for both the individual and the state. Birth registration and the subsequent issuance of a birth certificate are crucial elements in establishing the identity of the individual. Birth certificates provide proof of place of birth and parentage, information essential in

[1] United Nations Department of Economic and Social Affairs, Statistics Division. Principles and Recommendations for a Vital Statistics System. *Statistical Papers*, Series M, No. 19/Rev 3. United Nations, New York, 2014.

establishing citizenship and inheritance rights. Proof of citizenship is required in many countries for access to services such as health and education, and for obtaining basic documents such as a national identity card, work permit, passport, driver license, marriage license and social benefits. In addition, proof of age is required for placement in school and other age-limited services and is essential in preventing exploitation of children, for example through underage marriage, child labor and underage induction into military service.

Statistics derived from births occurring during a given year and registered during the allowable time period are extremely useful to governments. Because birth registration is compulsory, birth statistics at any level from national to local are not subject to sampling error; because registration takes place at the time of birth the statistics are not subject to recall error. The fact that registration is continuous means data are available for any time period, without the need to organize and fund a periodic survey. These advantages make birth statistics ideal for government, academic and private use. Birth data are useful for identifying those in need of maternal and child health care services, immunization services and others. In addition to planning for health services, governments use birth statistics to plan for future needs for education, transportation and many other community services. Birth registration data are a key component of intercensal population estimates. Birth counts serve as denominators for a number of essential demographic and public health measures, such as the infant mortality rate, maternal mortality rate, and immunization coverage rate. Given the multiplicity of uses and applications of birth registration, the UN states that birth registration should be universal, covering all groups including births to unwed mothers and to ethnic minorities such as indigenous groups.

3. The challenge

Despite the acknowledged importance of birth registration, births of more than a third of children under age 5 are not registered worldwide, and that proportion is nearly two-thirds for children in the least developed countries.[2] The proportion of births registered within the first year of life, the requirement for accurate birth statistics, will be lower yet. In addition, civil registration systems, which include the registration of births, deaths and other vital events have shown very slow improvement in the developing world in recent years, or even recent decades.[2]

For countries with deficient civil registration systems, birth registration completeness is measured primarily via surveys, principally the Demographic and Health Surveys (DHS) sponsored by USAID, and the Multiple Indicator Cluster Surveys (MICS) sponsored by UNICEF. Using survey data for the years

[2] United Nations Children's Fund.

2005-2012, the lowest regional birth registration estimates were 39% for South Asia and 44% for Sub-Saharan Africa.[3] As low as these birth registration measures are, the true estimates may be lower yet due to measurement issues, principally respondent confusion about what constitutes actual registration. The registration process is complex, as described below, and may lead the respondent to report a birth has been registered when in fact only part of the process has been completed.

4. Birth Registration Process

The registration of a birth in a civil registration system is a legal process, generally defined according to the country's registration law and regulations. While birth registration is similar across countries, there are differences which are best described by a review of various possible procedures required to register a birth:

Declaration. Many countries require a family member to visit the local civil registration office to inform the registrar of the birth. Typically, the family member is known as the *declarant.*

Notification. To speed the registration process and improve the quality and details of information reported, country civil registration systems are increasingly relying on selected authorities to inform them of vital events. As delivery in hospital or clinic becomes more common, CR law in some countries now accepts notification of births and their characteristics from health facility authorities. Similarly, CR systems may accept information on deaths, including cause of death, from health facility authorities.

Registration. Once the reporting requirements for a birth have been fulfilled, the CR office can *register* the birth. This means recording the legal and other required information, either in a registration book, on an official birth registration form, or entering the information electronically into a birth registration database. Typically, a copy of the information, either on paper or electronically, is sent to the central CR office. The UN recommends that registration of births be provided free of charge for registration within the allowable time period.[1]

Certification. Once registered, the parents or other authorized individuals can request a copy of the birth certificate, which is *certified* by the registrar. In some cases, such as when the civil registration system places a registrar within a maternity hospital, the family member can be provided with a copy of the birth certificate at the time of registration. In countries where a declarant is not required to physically visit the CR office, certificates may be printed for all registered births and then distributed to families via health personnel. In other

---

[3] United Nations Children's Fund. *Every Child's Birth Right: Inequities and trends in birth registration,* UNICEF, 2013.

cases the certificates are only printed on demand. Most countries assess a fee for birth certificates.

Delayed registration. Many countries have a deadline for reporting births to the local CR office, often one month from the date of birth. To register a birth beyond that date, some countries require payment of a small fine, while in other cases delayed registration requires a judicial hearing. To ensure that annual birth statistics are as complete as possible, most countries include births registered during the early part of the following year, often up to the end of the first quarter of the year. It should be noted however that births registered outside the allowable period, while still vitally useful to the individual, will not be included in birth statistics.

There are many variations on the registration steps noted above. Countries may require health facilities to prepare birth notification forms but give the form to the family rather than send the form to the CR office. In this case a family member will be required to visit the CR office with the form to complete the registration process. Some countries may have very specific requirements concerning the family member authorized to report a vital event to the CR office, while other countries may accept any family member as a declarant. Some countries may not allow the registration of births to unmarried mothers. Other countries may not allow the registration of births to refugees/displaced persons.

5. Birth Registration Barriers

Despite the abundance of reasons to register a birth, barriers and other problems may prevent birth registration. The challenges may be due to the attitudes, perceptions or resources of the parents, they may be due to aspects of the registration system itself, or due to a combination of the two.[4] The table below lists a number of common barriers to birth registration. These barriers are well known in the civil registration field.

Barriers to Birth Registration
Barriers related to parents: 1. Parents may not be aware of the need to register a birth, or feel that a birth certificate may not be needed under their present circumstances. 2. Parents may not have the means to travel to a distant civil registration office to complete the registration process. 3. Cultural practices may require time before the first name of the child is selected, meaning that the birth cannot be registered during the accepted time period after birth. 4. Parents may assume that children that die shortly after birth do not need to be registered, or do not report the birth and infant death to avoid burial costs. 5. There may be alternative sources of identification for children, such as a health passport, that suffice for childhood.

[4] Pelowski MW, Wamaj RG, Wnagombe J, et al. Why don't you register your child? A study of attitudes and factors affecting birth registration in Kenya, and policy suggestions. *Journal of Development Studies,* 8:1-24, 2015.

Barriers to Birth Registration
6. Parents may opt to skip registration of the birth and wait until the child is old enough to apply for a national identification card. *Barriers related to the registration system:* 1. The family may be part of a marginalized population, such as an Indigenous group, that is not well served by the civil registration system. 2. Registration fees, may discourage parents from registering a birth. 3. The civil registration system may not allow the registration of a birth to an unmarried mother.. 4. Requirements for the delayed registration of a birth may be so onerous that the parents are not able or are unwilling to complete the registration process. 5. Registrars who are frequently absent from their office may prevent parents from registering a recent birth, particularly if the child dies during this interval and the parent assumes registration is no longer necessary. 6. Birth notification forms completed at the hospital may be lost during transfer to the local registration office. 7. Hospital staff may refuse to forward the notification form to the local civil registration office until the family pays the hospital bill for delivery care.

6. Solutions

Useful solutions are those that produce a positive result and are within the means of the country to sustain. Expensive, donor-driven activities may produce a short-term gain but will not be sustainable. There are a number of solutions available to address barriers to birth registration, none of them perfect and some requiring outside assistance to initiate. These solutions will be addressed in turn, with discussion of the pros, cons and potential impact of each on the birth registration system.

Comprehensive assessment of civil registration system. An approach that has proved successful for civil registration improvement is to begin with a comprehensive assessment of the system.[2] International organizations and other groups have published detailed guidelines on comprehensive assessments and many organizations are promoting these assessments in low- and medium-resource countries.[56] The assessments serve to identify major system problems or barriers, and by organizing workshops to discuss the results can assure agreement across government on the changes needed.

Amend registration law to facilitate birth registration. Changes to registration law could address a number of the barriers mentioned above: deputizing health personnel or others to provide information legally acceptable as the basis for birth registration; removing registration fees; removing onerous requirements for delayed registration of births; ensuring that births to unmarried mothers are registered; ensuring that alternative

---

[5] World Health Organization. *Improving the quality and use of birth, death and cause-of-death information: guidance for a standards-based review of country practices*. WHO, Geneva, 2010.

[6] United Nations Economic Commission for Africa. *Improving National Civil Registration and Vital Statistics Systems in Africa. Volume 1: Guideline for Conducting Comprehensive Assessments of National Systems.* UNECA, Addis Ababa, 2016.

identity documents such as health passports are not accepted as proof of identity; requiring a birth certificate for issuance of national ID card; etc. Changes to registration law can be addressed through a formal review of all laws and regulations affecting birth registration, and involving civil registration authorities and others in the review.[7] Donors may be best placed to finance this activity, typically bringing in legal experts to work with local legal experts on the review, followed by a workshop for local authorities to discuss the results. Amending registration law is a slow process however, requiring parliamentary approval. .

Eliminate requirement for family member to declare birth at civil registration office. There are various ways to accomplish this, including amending the registration law to accept notifications from health facility personnel as the basis for registration. This would have the greatest impact on birth registration, eliminating the need for the family to send a member to an often-distant civil registration office. This would ensure facility births are registered, and if enlarged to include other health personnel such as immunization personnel, this change could expand coverage of non-facility births as well. Immunization personnel could also ensure delivery of birth certificates, as schedules for certain immunizations require repeat visits to the immunization center. The use of immunization personnel would not be advisable under certain circumstances, such as during mass immunization campaigns. A disadvantage of involving immunization personnel in the registration process is the potential for fraudulent birth registration.

A related solution has been to assign volunteer community health workers or village leaders to report births to civil registration authorities. This approach has been tried in several countries, with mixed results. Community health workers often are overloaded with tasks related to assistance projects and as a result may fail to report recent births. Requiring village leaders to maintain paper village registers, as was done in Malawi, has not been successful.[8] A demonstration project in Kenya using community health workers with mobile devices succeeded in transmitting birth registration information to a central location. The project did not succeed in increasing birth registration however, as the reports were not accepted by the registration system as the legal basis for registration.[9] In addition, using community health workers or village leaders as informants also runs the risk of fraudulent birth registration.

---

[7] Schwid A, Federes A, Bronson G, et al. *Civil Registration and Vital Statistics. Legal and Regulatory Review: Tools and Methodology.* Vital Strategies, New York, 2017.

[8] Singogo E, Kanike E, van Lettow M, et al. Village registers for vital registration in rural Malawi. *Tropical Medicine and International Health,* 18(8):1021-4, 2013.

[9] World Health Organization. *Move it: Report on Monitoring of Vital Events using Information Technology*. WHO, Geneva, 2013.

To eliminate the possibility of fraudulent birth notification, a number of countries have stationed civil registration staff within health facilities. In addition to registering the births occurring in the facility, in some cases the registrar is able to issue a birth certificate on the spot. Despite the advantages of this approach it has been only a partial solution, as for cost reasons it has been limited to the largest maternity hospitals and clinics.

Education campaigns. Providing information to families about the importance of birth registration, through organized meetings, political or religious leaders, teachers or others can lead to increased birth registration. The increase is often limited to the most motivated parents however, and in any case requires regular repetitions of the campaign.

Mapping the flow of registration documents or information. Business process mapping is designed to document the flow of paper forms or information from the source to the final destination at the civil registration office. Organized as workshops, these activities bring together representatives of many groups involved in different aspects of birth registration. The workshops are useful in identifying barriers to document flow and in encouraging a spirit of cooperation across government organizations. As with legal reviews, donors are often best placed to organize and fund these workshops.

Requiring birth certificate for government services. This approach has the potential to incentivize parents who otherwise would not see the need for birth registration. This requirement should not be used to deny children access to education, child support payments or other government services. Instead, civil registration systems should make birth registration available in connection with these services, for example establishing temporary registration offices next to schools at the beginning of the school year.

Facilitating access to birth certificates. Requiring birth certificates for various services will only be workable if the civil registration system can assure individuals ready access to their birth certificates. Because birth registration in many countries has been a paper-based system, the original documents are generally stored in the civil registration office where the birth was registered. What this means in practice is that individuals who have moved from their place of birth are required to return to that location to request a copy of their birth certificate. In larger countries this could involve long-distance, expensive travel requiring many days to complete. Even if copies of the original documents have been forwarded to the central office of the civil registration system, the fact that carbon paper was used for the duplicates means that many of the duplicate forms will be blank, the result of inadequate supplies of carbon paper for local registration offices. Computerization of registration offices is an obvious solution, but is expensive and may be beyond the financial resources of many national civil registration systems.

7. Birth registration and the national identification system

National ID systems are touted as important for economic development and promoted by international organizations.[10] The spread of national identification systems in developing countries however, runs the risk that governments will divert funding from civil registration to develop the national identification system, believing it can substitute for birth registration. Instead, birth registration and a national identify system should work together, by assigning a national identification number at birth and including that number, or a link to that number, on the birth registration form. This will ensure that children possess proof of identity until they reach the age of enrollment in the national identification system, providing them access to government services and protecting them from potential exploitation. Following recommendation by a number of international organizations, this approach is gradually being adopted by countries.[11]

8. Conclusion

Birth registration offers significant advantages for both individuals and governments. As a result, birth registration via government-run civil registration systems exists in almost every country in the world. At the same time, birth registration completeness remains low in many countries and improvement has been slow.[2] There is promise of future improvement however, as growing awareness of the unique advantages of a well-functioning civil registration system is leading donors to offer registration assistance to countries.

As external organizations begin to consider assistance on birth registration they can provide to low-and medium-resource countries, they would do well to apply the lessons learned in assistance to date. Approaches that have been found to be important include the need to involve country partners in a joint assessment of the system, identification of system problems, and agreement on solutions. System improvements can draw on the recommendations mentioned above, among others, always with an eye toward improvements that the country can sustain with its own resources. Another approach that has had positive results is creating *reasons* for families and individuals to register births, either through benefits to enjoy or via enforceable requirements for registration. Finally, integrating birth registration and the national identity systems will help to avoid diverting resources from birth registration.

[10] World Bank Group. Identification for Development: Africa business plan. World Bank, Washington DC, 2017.

[11] Botswana Department of Civil and National Registration. *Integration of civil registration and vital statistics and Identity management systems: Botswana success story.* Department of Civil and National Registration, Ministry of Labour and Home Affairs, Gaborone, 2015.

The multiple faces of trust in statistics and composite indicators: A case for healthy mistrust

Markku Lehtonen
Universitat Pompeu Fabra, Barcelona; Ecole des Hautes Etudes en Sciences Sociales, Paris; University of Sussex, Brighton, UK

Abstract
Declining public trust in official statistics and institutions responsible for statistics and indicator production is frequently highlighted as a key obstacle to reasoned debate on policy options and governance choices. The potentially harmful impacts of Big Data and an alleged "post-truth" era have further accentuated such concerns. To remain trusted and credible, statistical institutions must safeguard their authority as sources of independent and scientifically sound indicators, while at the same time being prepared to innovate and explore new methodological options. However, this paper argues that, in addition to this trust-building work, indicator designers need to embrace mistrust and distrust as essential for the generation of relevant and influential composite indicators. While important, regaining trust should not be seen as the overarching objective let alone a 'silver bullet'. This paper seeks to unpack the notion of trust and makes the case for mistrust and distrust as potential resources rather than mere threats to the credibility and authority of official statistics. For further empirical work, it proposes a conceptual framework consisting of three dimensions of trust and a distinction between mistrust and distrust, and illustrates the framework by concrete examples from indicator work. The conclusions suggest ways for statistical institutions to exploit the potential virtues of mistrust and adjusting their strategies to maintain trust via a more nuanced understanding of the multiple dimensions of trust, mistrust and distrust.

Keywords
Indicators; trust; mistrust; distrust; statistical offices; post-truth

1. Introduction

The alleged decline of trust in statistics and in statistical authorities has in recent years generated increasing concern, often framed as part of broader debates over the dangers of Big Data and post-truth politics. In one of the most prominent among such accounts, Davies (2018) argues that "the basic honesty of mainstream politicians, journalists and senior officials is no longer taken for granted", and pinpoints statistics as a key target of attack by the "populist right": *"...with statisticians and economists chief among the various "experts" that were ostensibly rejected by voters in 2016. Not only are statistics* *viewed by many as untrustworthy, there appears to be something almost insulting or arrogant about them. Reducing social and economic issues to numerical aggregates and averages seems to violate some people's sense of political decency"* (Davies 2017). Ultimately, Davies (2017) argues, the declining trust in quantitative expertise would be unlikely to improve democracy, as the critics would like to claim, but to *"an unleashing of tabloid editors and demagogues to provide their own 'truth' of what is going on across society."*

The virtues of trust are numerous and incontestable. Research has shown its value for a wide range of economic and social processes: interpersonal relations and economic exchange (Dasgupta 1988), financial investments (Kalkbrenner & Roosen 2016, 62), the legitimacy of political power (Tait 2011), societal and economic development and growth (Gallucio 2018), innovation, education, rule of law, good governance, reduction of corruption and violence, subjective well-being (Zak & Knack 2001; Laurent 2009, 14; Volland 2017), environmental performance and the propensity to adopt strict environmental policies (e.g. Owen & Videras 2008; Tjernström & Tietenberg 2008; Carattini et al. 2015), and successful cohabitation of peoples in multinational democracies (Karmis & Rocher 2018).

However, the dim views of Davies concerning the loss of trust merit nuancing. Firstly, Desrosières (2015) traces the origins of the crisis of trust back to the "feedback" of indicators and quantitative information on their users and producers, and the tensions this generates for the identity of statisticians. The very identity of statisticians rests on their role as custodians of objective, incontestable data in defence of democracy and the underprivileged, yet such quantified forms of information have performative effects (e.g. Merry 2016; Ràfols 2019), that is, they "change the world through their very existence, their circulation and their rhetorical usage in science, politics or journalism" (Desrosières 2015, 334). Through institutionalisation – a largely invisible process whereby indicators become objectified, as the quantification and encoding procedures become routine (ibid.) - indicators become a *lingua franca* that allows smooth communication amongst actors (ibid.). Making these processes explicit may be highly desirable for the sake of democracy and openness, yet it also risks weakening the argumentative effectiveness of indicators: the underlying conventions and assumptions would be opened up to contestation, and would probably require lengthy discussions and demonstrations (Desrosières 2015, 341).

Secondly, a distinction should be made between the main clients of statistical offices, the 'users' of indicators and statistics, who usually share the mental frameworks and objectives of statisticians. These users usually hold strong trust in official statistics, and have experience-based reasons for doing so. The argument put forward by Davies concerns primarily the vast majority of 'non-users' – citizens that have a very vague idea of the credibility of indicators and statistics. For statistical offices, a major objective would therefore be to turn these potential users into real users, who would gain first-hand experience and hence presumably regain their trust in indicators.[1]

Thirdly, while trust has certainly declined in some countries and with regard to some aspects of political institutions, empirical analysis reveals a more complex and differentiated picture. When it comes to trust in statistics, the UK – alongside the US the main concern of Davies – appears in Europe as an outlier due to the particularly low trust that Britons express in statistics. In 2010, only slightly over 30% of the UK citizens trusted in statistics, as compared to 70% in Denmark and Sweden, and well below even such low-trust countries as Greece (56%) and Portugal (52%) (Eurobarometer 2010). In Finland, the statistics office remains as one of the most trusted institutions in the country, with lonely very slight decline observed in the past few years (Melkas 2014). Even in France, dubbed sometimes as a "country of distrust" (Algan & Cahuc 2007; Algan et al. 2012; Agacinski 2018), citizens express mistrust towards 'official statistics', but when inquired about their trust in 'public statistics', give far more positive answers (Desrosières 2015; Rosanvallon 2017). Jasanoff and Simmet (2017, 752) contest pessimistic picture even for the UK. A poll from 2016 found that trust in experts had actually increased between 2014 and 2016, as 85% of the surveyed Britons wanted politicians to consult professionals and experts when making difficult decisions.

A final nuance comes from the observation made by Ràfols (2019, 9) about two simultaneous trends in science, technology and innovation (STI) indicators: these indicators are increasingly seen and used as a means to foster STI governance under a New Public Management ethos, while at the same time there has been an "erosion of the uncritical belief in the benefits of STI". The term "uncritical" is the key: unlike Davies, Ràfols does not see the loss of trust as an exclusively phenomenon, but rather as a healthy move away from naivety. In the following, I will follow this thread, arguing that rather than lamenting the loss of citizens' trust in indicators and experts, we should perhaps celebrate this loss of innocence, and certainly seek to harness the perhaps more mature and healthy mistrust to constructive purposes.

Representing – by definition – precisely the kind of reductionism of complex societal issues into "numerical aggregates" evoked by Davies, composite indicators appear as a prime candidate for mistrust. I will therefore address one of the questions listed in the guidelines for the authors of this ISI session: *"how to enhance stakeholders' confidence in indicators? Which*

[1] An argument made by Pilar Martin/Guzman, Universidad Autonoma de Madrid, in her presentation at the CESS meeting in Bamberg, 19 October 2018.

conditions (legal, political, institutional) need to be fulfilled in order to enhance such confidence?" However, rather than focusing on means of 'enhancing confidence', I will seek to provide a more nuanced picture of the multiple roles of trust in indicator work, and highlight the potential downsides of "overtrust" and virtues of mistrust and distrust. The operational question then becomes: what role, if any, might constructive mistrust and distrust play in strengthening rather than undermining the production and use of composite indicators, in the context of alleged loss of trust and post-truth politics? I suggest that the concepts of trust, mistrust, and distrust briefly introduced here could be helpful in attempts to better understand the role of multiple roles of statistical offices and composite indicators. Further empirical research will be needed to corroborate (or invalidate) the applicability and usefulness of the framework.

2. The multiple dimensions of trust, mistrust and distrust

Trust[2] can be defined generally as a stance whereby an individual accepts 'believing without knowing', thus placing herself voluntarily in a position of vulnerability towards 'the other', be it another individual or an institution[3] (Earle & Siegrist 2006). Trust represents a 'leap of faith' (e.g. Davies 2018), because there is always a risk that the 'trustee' proves untrustworthy, yet as a voluntary choice, trust does not have to imply the feeling of loss power and control (Espluga et al. 2009).

Three mutually interacting dimensions of trust can be distinguished. First, **social trust** is interpersonal. It can entail **generalised** trust in other, unknown, members of society (Rothstein & Stolle 2008) or **particularised (specific)** trust in people we already know, with whom we interact regularly, for example in our own social or demographic group (Bäck & Christensen, 2016, 180). For indicators, a key issue then concerns particularised trust in statisticians as individuals. Arguably, this interpersonal type of trust has little to do with current problems of loss of trust in indicators and statistics – few citizens have regular encounters with statisticians, or have particular trust/mistrust relations with statisticians.

Institutional trust denotes the public trust in institutions such as statistical authorities, the government, government regulation, or NGOs. Institutional trust can entail *specific support*, in other words, individuals' judgement of what the institution *does* (its performance), or *diffuse support*,

[2] For the sake of simplicity, I use the term trust to encompass both its traditional meaning as a normative judgement concerning an individual or entity, and confidence, that is, a belief based on earlier experience that certain events will occur as predicted (Earle & Siegrist 2006; Luhmann 2006; Kinsella 2016).

[3] Following Hodgson (2006, 18), institutions are here defined broadly, as "systems of established and embedded social rules that structure social interactions". Organisations, in turn, are a specific type of institution.

that is, what the institution *is* – what it represents for the individual in question (Kestilä-Kekkonen & Söderlund 2016, 141). For example, one can trust specifically the present government coalition or more generally the national government. To evoke trust, an institution needs to be perceived as competent, sincere, transparent, reliable in keeping its promises, and demonstrate ability to deal with mistrust and to avoid mismanagement or entanglement in political scandals (Holmberg & Weibull 2017, 39; Laurian 2009, 383-384; Tuler & Kasperson 2010). The independence of statistical offices is therefore vital and seen as an essential part of the very identity of statistical institutions. Earlier research has shown that in situations of longstanding institutional mistrust, attempts at trust-building, e.g. via participation and openness, can initially undermine trust (Gouldson et al. 2007; Laurian 2009). The above statements from Davies and Desrosières express concern for loss of trust in statistical offices and statistics as an institution.

The distinction between specific and diffuse support to an institution helps us to understand the foundations of institutional trust, yet if indeed specific support stems from evaluation of the performance of an institution, the question arises: on which grounds do citizens judge the performance of a statistics institution? Is scientific robustness enough or does trustworthiness also build upon features such as the sensitivity of the institution to the plurality of contexts, "both in terms of the relevant social spaces and values" (Rafols 2019)? Statistical institutions are faced by a dilemma between independence and scientific robustness on one hand, and relevance on the other. To be 'argumentatively effective' (Desrosières 2015), indicators need to be seen as scientifically reliable, but when this implies foregoing considerations relating to the needs and values of the potential users, indicators risk being seen by the public and stakeholders as increasingly obsolete, irrelevant, and even misleading. Ignoring the feedback and the performative effects of indicators risks hiding underlying value commitments, uncertainties, and the regimes of power' underpinning the indicators behind seemingly neutral discussions concerning the appropriate methodological choices (Merry 2016). Jasanoff & Simmet (2017, 754) draw attention to the "wider realities in which contested public facts are embedded", and argue that any "truth needs to be performed and accepted" (ibid., 756). These insights help to put into perspective the "truths" conveyed by composite indicators and the alleged downsides from the declining trust in indicators and statistics. According to this line of thinking, what is at stake is the ability and moral authority of the state to produce 'serviceable truths', that is, "robust statements about the condition of the world, with adequate buy-in from both science and society" (Jasanoff & Simmet 2017, 756).

The **ideological** dimension of trust is relatively seldom addressed in social science research, yet it is perhaps the most informative from the point of view of discussions on post-truth. This type of trust relates to higher-level institutions, such as democracy, the state, market, and planning, and their legitimate roles in society (Tait 2011, 158). As a more abstract form of trust, it is difficult to capture via quantitative surveys. It concerns schemes of interpretation of reality, relating to means-ends relationships and strategies (Söderbaum 1999; 2013, p. 223), i.e. to "wider abstract systems and ideas", such as economic growth models, the legitimate role of government in intervening in the economy (Tait 2011, 160), technological optimism, centralised or decentralised solutions (Söderbaum 1999, 163), or the legitimacy of the purposes of data collection for production of statistics and indicators. Unlike social and institutional trust, ideological trust does not draw on previous evidence or knowledge, but "on an individual's or institution's place within wider social discursive structures" (Tait 2011, 160).

To the extent that non-users mistrust indicators, this mistrust is largely ideologically founded, precisely because of their lack of first-hand experience. This is by means not to say that 'users' and creators of indicators would be free of ideological postulates and assumptions. The very birth of statistics as a state endeavour, and their further development in the early 19th century in the hands of 'social reformers' (Bascand 2009), involved judgements concerning the legitimate roles of the state, experts, and the civil society. Moreover, the cultural and institutional context is vital: countries and democratic cultures vary not only with respect to their traditions concerning the respective roles of the state, the private sector and the civil society, but also when it comes to the roles that trust and mistrust have played in their development. For example, liberal democracy is arguably founded in mistrust amongst citizens towards the state (e.g. Warren 1999, 310; Lenard 2008), whereas in the Nordic countries, democracy has developed upon a trustful social contract between the state and the citizens (e.g. Montin 2015).

Ideological trust in the area of indicators finds one of its most prominent examples in what has been variously called "indicator culture" (Merry 2016), 'trust in numbers' (Porter 1995), or audit culture (Power 1999; Shore & Wright 2015), including phenomena such as the 'fallacy of misplaced concreteness'[4] and false precision (Hicks et al. 2015). More generally, these cultures and beliefs are founded in "*trust in technical rationality, in the legibility of the social world through measurement and statistics, and in the capacity of numbers to render different social worlds commensurable"* (Merry 2016). Trust in numbers – in science as a form of obtaining information, and the corresponding scepticism about politics – is fundamentally ideological in nature. It is crucial to examine, however, the interaction between the dimensions of trust: the

[4] The fallacy of misplaced concreteness is commonly described as mistaking a theoretical construct for a physical or 'concrete' reality.

degree of trust/mistrust in numbers is shaped by our perceptions of the trustworthiness of institutions responsible for upholding such a culture, as well as by our everyday encounters with and perceptions concerning individuals working within those institutions. Likewise, the greater our buy-in regarding an indicator culture, the more prone we are to trust an individual statistician or the statistical office he/she represents, and – by implication – the products of that institution. Indicator culture is deeply institutionalised, for example in the strong particularised social trust amongst the indictor user and producer communities.

Past and present indicator work also involves choices and judgements between trust in state planning and market competition, respectively. While indicators can be mistrusted for the excessive power they grant to technocratic control, they are today increasingly blamed for fostering the intrusion of market and private-sector logics into the public sphere. Ràfols (2019) talks about the "political economy of quantitative evaluation increasingly shaped by commercial information infrastructures", and about an "ongoing managerial push for standardization via 'platforms' run by commercial oligopolies".

3. Downsides of trust

In his analysis of STI indicators, Ràfols (2019) implicitly addresses the potential downsides of trust and virtues of mistrust. He identifies three false ideological assumptions underpinning STI policy and indicator work. Not only are STI policy and indicators grounded in the assumption that "STI eventually lead to positive outcomes" – hence, the more STI knowledge created, the better, but it is also founded on 1) institutional trust in the benevolence of the state bureaucracy in supporting STI, and hence in its aim of fostering general well-being, and 2) social (interpersonal) trust in the "scientists and experts responsible for developing and applying STI", as purveyors of "the public good rather than particular interests" (Ràfols 2019, 9).

Ràfols acknowledges that the rather naïve view about STI for the public good may have been a reasonable representation in some limited areas in the past, such as *"when indicators were used to challenge the scientific establishment, particularly in low-trust countries with a tradition of nepotism".* Yet, such a highly optimistic view has given way to warranted scepticism. If one takes seriously the argument of Merry (2016) about the neglect of *"the social aspect of indicators… in the face of trust in numbers, cultural assumptions about the objectivity of numbers, and the value of technical rationality",* then mistrust of this kind of an indicator culture should be a healthy and welcome phenomenon, which might help to bring about a different, more reflexive and deliberative indicator culture.

The downsides of trust are indeed well covered in social science literature.

Goel et al. (2005, 203) mentions three harmful consequences from “overtrust”: leniency in judging the trustee, delay in perceiving exploitation, and increased risk-taking. This type of overtrust or “unwarranted trust” (Warren 1999) typically corresponds to a situation in which the trustee is judged positively, but on the basis of poor or inexistent knowledge (e.g. Balme 2003). This can take the form of a blind “trust in numbers”, yet the question is more complex than merely one of lacking knowledge. Especially composite indicators tend to hide essential choices and value judgements behind ostensibly objective and value-free data, thus concealing the complex calculations employed to arrive at the final numbers. In the words of Barré (2019, 3), STI indicators have, in some cases turned from tools informing and enlightening decision-making to *“ignorance producing devices”* that – by virtue of the lock-ins and irreversibilities produced by their institutionalisation and ‘taken-for-grantedness’ – do not express what they pretend to, and may even resist new knowledge gained over the years on the topic in question.

Also social trust can be detrimental to the common good. ‘Bonding social capital’ – which characterises particularised social trust – can feed exclusion, homogeneous social networks, specific norms of reciprocity (Santaoja et al. 2016; Putnam, 2000, 19-21), groupthink, the exclusion of different yet competent others (Kujala et al. 2016, 702), and creation of sharp boundaries between ‘insiders’ and ‘outsiders’ (van Deth & Zmerli, 2010). The tight social ties within the “indicator industry” (Hezri & Hasan 2004) – or the often highly homogeneous community of indicator users and producers (e.g. Sébastien & Bauler 2013) – constitute an example. The greater the internal cohesion within an indicator community, the less likely it is to explore perspectives from outside the community, and foster the emergence of a more reflexive indicator culture (cf. Bhuta et al. 2018) in the spirit of opening up (Stirling 2008; Ràfols et al. 2012). Excessive trust can undermine the attempts towards inclusive deliberation (e.g. Jasanoff & Stimmer 2017) also because trustful citizens may lack the motivation to participate, preferring instead to delegate power to trusted experts and institutions (Parkins & Mitchell 2005, 536).

4. Sources of trust and mistrust

Perceptions of competence and sincerity are often described as they key characteristics that shape social and institutional trust. The competence of statistical authorities is seldom at stake, and hardly a reason for current-day concerns about post-truth and loss of trust in expertise and authority. By contrast, the loss of the naïve belief in sincerity is what underpins a lot of these concerns. Again, what has come under attack is less the sincerity of individual statisticians and indicator creators than the purposes for which indicators are being produced (e.g. increasing control as part of New Public Management, private-sector profit-seeking), and the worldviews and democratic choices incorporated in indicators. In addition to competence and sincerity, the perceived relevance of indicators is a crucial shaper of trust. Hence, in their attempt at maximising the presumed objectivity and scientific quality of indicators, in the spirit of 'indicator culture', statistical offices risk undermining the relevance of, and hence trust in, indicators. The question can also be formulated in the following manner: do we trust the statistical offices and other indicator developers to be sincere in their commitment to producing information that is useful and relevant for our daily practice?

Trust can build on either previous experience or on normative predispositions and broader worldviews. For regular and occasional indicator users, experience plays a significant – and often positive – role, whereas non-users must rely on their normative predispositions in judging indicators and their producers. However, as argued in the previous section, the concept of ideological trust reminds of the pervasive nature of such normative factors: also regular indicator users rely upon their ideological perceptions concerning the respective roles of 'meta-level' institutions in society, and worldviews such as those embedded in indicator culture.

Table 1 summarises the three dimensions of trust and the various sources of trust as employed in this article.

Table 1. Summary of the key concepts relating to trust and mistrust.

<table>
<tr><th>Type of trust/mistrust</th><th>Social</th><th>Institutional</th><th>Ideological</th></tr>
<tr><td>Description</td><td>Generalised
Particularised</td><td>Diffuse support
Specific support</td><td>Legitimacy of and support to meta-level institutions</td></tr>
<tr><td rowspan="3">Sources of trust</td><td colspan="2">Competence
Sincerity</td><td>Worldviews, visions</td></tr>
<tr><td colspan="3">Normative predisposition in relation to an institution or an individual (trust)</td></tr>
<tr><td colspan="2">Predictability, based on previous experience (confidence)</td><td></td></tr>
</table>

5. Mistrust and distrust as assets – towards a more reflexive indicator culture?

On the most fundamental level, as 'healthy suspicion', mistrust towards the powers that be, constitutes a foundation for the vitality of a democratic system – a form of "civic vigilance" (Laurian 2009), responsibility, and countervailing power that helps citizens to hold political, economic and cultural elites to account (Warren 1999, 310; Laurent 2009, 27; Allard et al. 2016, 14). Organisations and procedures of regulation (e.g. auditing, evaluation, ranking, and benchmarking) represent an institutionalised form of mistrust and vigilance (Lenard 2008; van Deth & Zmerli 2010, 2665). The current mistrust of statistics and indicators should not be seen as an exclusively negative phenomenon. Mistrust can, under certain conditions, constitute an asset in efforts to bring about a more reflexive indicator culture. But what are those conditions? When and how can mistrust be turned into a constructive force? I can only provide a few pointers towards possible means of building on mistrust, to be tested and confirmed via further empirical research.

A distinction between mistrust and distrust may be a first conceptual step forward. While mistrust reflects a wait-and-see attitude and associated action, the core of distrust is the desire to take distance. In other words, a mistrustful individual retains the hope and expectation that the trustee might, after all, prove trustworthy, whereas distrust implies the loss of such hope (Kuryo 2011). Lenard (2008, 319) describes mistrust as an unstable situation, in which people "make no decisions in advance about whether to trust others", but "consider a range of questions before making a decision", concerning the immediate risks and benefits involved, possible 'safety nets' in case of trust being betrayed, and possibilities of gaining more information to back up one's decisions in the near future.

For indicators, constructive mistrust might entail vigilance, from civil society actors or other stakeholders, towards the producers of official indicators. Calls for openness and transparency of data, for greater participation by the various potential user communities, and demands for greater relevance of indicators could constitute useful civic vigilance. By keeping a critical eye on producers of official statistics and indicators, and by helping to show their strengths and weaknesses, such vigilance can actually help to improve both the relevance of indicators and the reputation of statistical offices. Trust and mistrust would thereby operate in tandem, with vigilant mistrust helping to strengthen the necessary trust in official statistics. The example provided by Ràfols of the potential usefulness of STI indicators, when "used to challenge the scientific establishment (Martin and Irvine 1983), particularly in low-trust countries with a tradition of nepotism (Ràfols et al. 2016)."

Distrust, by contrast, is potentially more problematic, as it can feed on suspicion close to paranoia, and manifest itself in the spreading of doubt, for instance via social media. These are the kind of phenomena that Davies warns against, the world in which facts, statistics and indicators do not matter, alternative facts, etc. Lenard (2008, 316) describes distrust as a danger to democracy, "an attitude that reflects suspicion or cynicism about the actions of others; people are deemed untrustworthy in part because they have over time provided (what is taken as) evidence that they cannot be trusted". The stability, evoked by Lenard (2008) as a key feature of distrust, reflects the reciprocal, asymmetric and self-reinforcing characteristics of trust and mistrust

(Slovic 1993). Hence, Davies (2018) notes: *"Whereas it is impossible to conclusively prove that a politician is morally innocent or that a news report is undistorted, it is far easier to demonstrate the opposite. Scandals, leaks, whistleblowing and revelations of fraud all serve to confirm our worst suspicions."*

However, unlike Lenard (2008, 318), who sees distrust as "by and large inimical to democracy", but "mistrust is in some ways good for democracy", I argue that even distrust can have its constructive side. The notion of 'alternative facts' indeed points towards the fact that the desire to take distance via strategies of independence, autonomy and autarky can be beneficial for society, in fuelling autonomous civil society activity (e.g. Lehtonen & de Carlo 2019). Instead of false data and fake news used to provide "alternative facts", constructive distrust could manifest itself in the production of indicators that represent framings radically distinct to those of the dominant official indicators and statistics. Instead of contesting and challenging the existing indicators, this type of distrustful indicators would provide an independent alternative, building on either alternative data sources or official statistics. As opposed to mistrustful vigilance, constructive distrust would represent an alternative, these indicators not discussing and challenging directly the dominant indicators, but would provide truly alternative, hitherto neglected, framings. Examples from this type of distrust can be found in the development of radically alternative accounting methods and frameworks, including indicators such as ecological footprint and various well-being indicators. Obviously the distinction between mistrustful vigilance and distrustful autonomy is not sharp and clear-cut: also radically alternative indicator frameworks can indirectly shape mainstream indicator work.

6. Discussion and Conclusion: Post-truth, Big Data, and cultural differences?

In this paper, I have argued that indicator work would do well in drawing on lessons from social science literature on the respective roles, virtues and benefits of trust, mistrust and distrust. In particular, both mistrust and distrust can have virtues that could be intentionally exploited in indicator work, not least in the current context of 'post-truth', Big Data, and multiplication of data sources. Statistical institutions in their attempts to safeguard the public interest against the risk of the emergence of private-sector data oligopolies, yet the traditional model of authority of these institutions is no longer operational: indicator work needs innovate, engagement with data users and new data communication and visualisation technologies, taking seriously the importance of contextualisation, if indicators are to engender trust.

In practice, this could entail measures such as:

- State support for 'civic vigilance' and constructive mistrust and distrust. This could take the form of technical and financial support to civil society organisations for their efforts to hold official statistics production and producers accountable, but also support for organisations producing radically alternative types of indicator framings, relying on official or alternative data sources. Experience from radioactive waste management policy can provide examples of this kind of support.
- Opening up the processes of indicator design, use, and refinement to participation by a broad range of actual and potential user and stakeholder groups, with a view to maximising the range of normative and cognitive perspectives involved (cf. Ràfols et al. 2012; Ràfols 2019).
- Actively seek critical discussion concerning the underlying framings and methodological choices underpinning indicators, in order to highlight the associated uncertainties and the importance of framings on the shape and impact of indicators. This could help resist "the law of inherent conservatism in official statistics" and foster "statistical entrepreneurship", by enhancing the double role of statistical offices as producers of reliable, authoritative and incontestable facts, and as reflexive innovators enlightening the society of the complexities and uncertainties behind statistics, and exploring their practical significance (Van Tuinen 2009, 441). An option might be the kind of inclusive deliberation that associates truth claims (indicators) with the issues of public value and purpose, which Jasanoff and Simmet see as the prerequisite of acceptable truths. As such, they could an antidote against the problems placed – in their view erroneously – under the banner of "post-truth".
- Conduct analysis amongst actual and potential users and the wider public in order to identify and minimise risks of dysfunctional distrust, engendered in particular by feelings of disappointment and betrayal (cf. Lenard 2008).

References

1. Agacinski, D. 2018. Expertise et démocratie: faire avec la défiance. France Stratégie. Décembre. www.strategie.gouv.fr
2. Algan, Y. & Cahuc, P. 2007. La société de défiance. Editions ENS rue d'Ulm.
3. Algan, Y., Cahuc, P. & Zylberberg, A. 2012. **La fabrique de la défiance... Et comment s'en sortir**. Paris: Albin Michel.

4. Allard, O., Carey, M. and Renault, R. 2016. De l'art de se méfier [The art of mistrust]. Tracés 31(2), 7-20.
5. Amnå, E., Czarniawska, B., & Marcusson, L. 2013. Tillitens gränser: Granskningskommissionens slutbetänkande, Göteborg, Granskningskommissionen i Göteborgs Stad.
6. Bäck, M. and Christensen H. S. (2016). When trust matters – a multilevel analysis of the effect of generalized trust on political participation in 25 European democracies. Journal of Civil Society 12(2), 178-197.
7. Balme, R., Marie, J.-L. & Rozenberg, O. 2003. Les motifs de la confiance (et de la défiance) politique : intérêt, connaissance et conviction dans les formes du raisonnement politique. Revue internationale de politique comparée 2003/3 Vol. 10 | pages 433 à 461.
8. Barré, R. 2019. Les indicateurs sont morts, vive les indicateurs! Towards a political economy of S&T indicators: A critical overview of the past 35 years. Research Evaluation, 28(1), 2–6.
9. Bascand, G. 2009. The role and challenges for official statisticians in fostering progress. The 3rd OECD World Forum on "Statistics, Knowledge and Policy": Charting Progress, Building Visions, Improving Life. Busan, Korea - 27-30 October 2009.
10. Bhuta, N., Malito, D.V. & Umbach, G. 2018. Introduction: Of Numbers and Narratives—Indicators in Global Governance and the Rise of a Reflexive Indicator Culture. In Bhuta et al. (eds.), The Palgrave Handbook of Indicators in Global Governance. Pp. 1-29. https://doi.org/10.1007/978-3-319-62707-6
11. Dasgupta, P. 1988. Trust as a Commodity. In Gambetta, D. (ed.), Trust: Making and Breaking Cooperative Relations. Basil Blackwell: Oxford, UK & Cambridge, MA. Pp. 49-72.
12. Davies, W. 2017. How statistics lost their power – and why we should fear what comes next. The Guardian, 19 January.
13. Davies, W. 2018. Why we stopped trusting elites. The Guardian, 29 November. https://www.theguardian.com/news/2018/nov/29/why-we-stopped-trusting-elites-the-new-populism
14. Desrosières, A. 2015. Retroaction - how indicators feed back onto quantified actors. In: Rottenburg, R., Merry, S. E., Park, S.-J. & Mugler, J. (eds.) The World of Indicators: The Making of Governmental Knowledge through Quantification. Cambridge: Cambridge University Press.
15. Earle, T.E., & Siegrist, M. 2006. Morality information, performance information, and the distinction between trust and confidence. Journal of Applied Social Psychology 36(2), 383-416.
16. Espluga, J., Prades, A., Gamero, N. & Solà, R. 2009. El papel de la "confianza" en los conflictos socioambientales (The Role of "Trust" in Socio-Environmental Conflicts). Política y Sociedad 46(1-2): 255-273.

17. Eurobarometer. 2010. Special Eurobarometer 323: Europeans' knowledge of economic indicators. European Commission. Directorate-General for Communication.
18. Galluccio, C. 2018. Trust in the Market: Institutions versus Social Capital. Open Journal of Political Science 8, 95-107. https://doi.org/10.4236/ojps.2018.82008
19. Goel, S., Bell, G. G. and Pierce, J. L. (2005). The Perils of Pollyanna: Development of the Over-Trust Construct. Journal of Business Ethics 58(1-3), 203-218.
20. Gouldson, A, Lidskog, R., & Wester-Herber, M. (2007). The battle for hearts and minds? Evolutions in corporate approaches to environmental risk communication. Environment and Planning C: Government and Policy, 25(1), 56-72.
21. Hezri, A.A. and M.N. Hasan (2004), 'Management framework for sustain-able development indicators in the State of Selangor, Malaysia', Ecological Indicators, 4, 287–304.
22. Hicks, D., Wouters, P., Waltman, L. de Rijcke, S. & Ràfols, I. 2015. The Leiden Manifesto for research metrics. Nature 520, 429–431(23 April 2015). doi:10.1038/520429a
23. Hodgson, G.M. 2006. What are institutions? Journal of Economic Issues XL(1): 1-25.
24. Holmberg, S., & Weibull, L. (2017). Långsiktiga förändringar i svenskt institutionsförtroende [Long-term changes in Swedish trust in institutions]. In U. Andersson, J. Ohlsson, H. Oscarsson, & M. Oskarson (eds.), Larmar och gör sig till. SOM-undersökningen 2016. SOM-rapport nr 70. Pp. 39-58.
25. Jasanoff, S. & Simmet, H. R. 2017. No funeral bells: Public reason in a 'post-truth' age. Social Studies of Science Vol. 47(5) 751–770.
26. Kalkbrenner, B.J. & Roosen, J. 2016. Citizens' willingness to participate in local renewable energy projects: The role of community and trust in Germany. Energy Research and Social Science 13 (March): 60-70.
27. Karmis, D. & Rocher, F. 2018. Trust, Distrust, and Mistrust in Multinational Democracies: Comparative Perspectives. Mcgill-Queen's University Press.
28. Kestilä-Kekkonen, E. & Söderlund, P. 2016. Political Trust, Individual-level Characteristics and Institutional Performance: Evidence from Finland, 2004–13. Scandinavian Political Studies 39(2): 138-160.
29. Kinsella, W. J. 2016. A question of confidence: Nuclear waste and public trust in the United States after Fukushima. In: Hindmarsh, R. & Priestly, R. (eds.) The Fukushima effect: A new geopolitical terrain. London: Routledge.

30. Kujala, J., Lehtimäki, H. & Pučėtaitė, R. 2016. Trust and Distrust Constructing Unity and Fragmentation of Organisational Culture. Journal of Business Ethics 139(4): 701-716.
31. Kuryo. (2011). La confiance : approche historique et sociologique. [Trust: an historical and sociological approach]. Le labo de la confiance Kuryo. 3 novembre 2011. Available at: http://kuryo.typepad.com/lelabodelaconfiance/2011/11/la-confiance-un-%C3%A9tat-des-savoirs.html (accessed 14th August 2017).
32. Laurent, É. 2009. Peut-on se fier à la confiance? [Can we trust confidence?] Revue de l'OFCE 2009/1, no 108, pp. 5-30.
33. Laurian, L. (2009). Trust in planning: Theoretical and practical considerations for participatory and deliberative planning. Planning Theory & Practice 10(3), 369-391.
34. Lehtonen, M. & de Carlo, L. 2019. Community energy and the virtues of mistrust and distrust: lessons from Brighton and Hove energy cooperatives. Ecological Economics.
35. Lenard, P.T. 2008. Trust Your Compatriots, but Count Your Change: The Roles of Trust, Mistrust and Distrust in Democracy. POLITICAL STUDIES: 2008 VOL 56, 312–332.
36. Luhmann, N. 2006. La confiance : un mécanisme de réduction de la complexité sociale. [Trust: a mechanism of reducing social complexity] Economica, Paris.
37. Martin, B. R., and Irvine, J. (1983) 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', Research Policy, 12/2: 61–90.
38. Melkas, J. 2014. Luottamus murenee. Tieto & Trendit, Tilastokeskus. https://www.stat.fi/tietotrendit/artikkelit/2014/luottamus-murenee/
39. Merry, S.E. 2016. The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking. Chicago and London: The University of Chicago Press.
40. Montin, S. 2015. Från tilltrobaserad till misstrobaserad styrning: Relationen mellan stat och kommun och styrning av äldreomsorg [From trust-based to mistrust-based governance: relationship between the state and the municipality and governance of elderly care]. Nordisk Administrativt Tidsskrift, 92(1), 58-75.
41. Owen, A.L. & Videras, J. 2008. Trust, cooperation, and implementation of sustainability programs: The case of Local Agenda 21. Ecological Economics 68(1-2): 259-272.
42. Parkins, J. R. and Mitchell, R. E. (2005). Public Participation as Public Debate: A Deliberative Turn in Natural Resource Management. Society & Natural Resources 18(6), 529-540.

43. Power, M. 1999. The Audit Society: Rituals of Verification. Oxford: Oxford University Press.
44. Ràfols, I. 2019. S&T indicators in the wild: Contextualization and participation for responsible metrics. Research Evaluation 28(1), 2019, 7–22.
45. Rafols, I., Ciarli, T., van Zwanenberg, P. & Stirling, A. 2012. Towards Indicators for `Opening Up' Science and Technology Policy. Presentation at the conference. Internet, Politics, Policy 2012: Big Data, Big Challenges? Oxford Internet Institute, University of Oxford, 20-21 September 2012.
46. Ràfols, I., Molas-Gallart, J., Chavarro, D.A. & Robinson-Garcia, N. 2016. On the Dominance of Quantitative Evaluation in 'Peripheral' Countries: Auditing Research with Technologies of Distance (May 28, 2016). Available at SSRN: https://ssrn.com/abstract=2818335 or http://dx.doi.org/10.2139/ssrn.2818335
47. Rosanvallon, P. 2017. Seminar presentation at the Institut des Etudes Avancées, Paris,24 March.
48. Rothstein, B. & Stolle, D. (2008). The State and Social Capital: An Institutional Theory of Generalized Trust. Comparative Politics, 40(4), 441-459.
49. Santaoja, M., Laine, M. and Leino, H. (2016). "Joku palikka siitä puuttuu" – luottamuksen rakentuminen täydennysrakentamisen suunnittelussa. ["There's a piece missing somewhere" – construction of trust in planning in-fill construction] Yhdyskuntasuunnittelu 54(1). Available at: http://www.yss.fi/journal/joku-palikka-siita-puuttuu-luottamuksen-rakentuminen-taydennysrakentamisen-suunnittelussa/ (accessed 5th June 2018).
50. Sébastien, L. and T. Bauler (2013), 'Use and influence of composite indicators for sustainable development at the EU- level', Ecological Indicators, 35, 3–12.
51. Shore, C. Wright, S. 2015. Governing by numbers: audit culture, rankings and the new world order. Social Anthropology, 23, 1: 22-28
52. Söderbaum, P. 1999. Values, ideology and politics in ecological economics. Ecological Economics 28 () 161–170.
53. Stirling, A. 2008. 'Opening Up' and 'Closing Down': Power, Participation, and Pluralism in the Social Appraisal of Technology. Science, Technology & Human Values, 33, 262-294.
54. Tait, M. 2011. Trust and the public interest in the micropolitics of planning practice. Journal of planning education and research 31(2): 157-171.

55. Tjernström, E. & Tietenberg, T. 2008. Do differences in attitudes explain differences in national climate change policies? Ecological Economics 65(2): 315-324.
56. Tuler, S.P. et Kasperson, R.E. 2010. Social distrust: implications and recommendation for spent nuclear fuel and high level radioactive waste management. A Technical report prepared for the Blue Ribbon Commission on America's Nuclear Future. 29 January, 2010
57. van Deth, J. W and Zmerli, S. 2010. Introduction: Civicness, Equality, and Democracy – A "Dark Side" of Social Capital? American Behavioral Scientist 53(5): 631–639.
58. van Tuinen, H. K. (2009), "Innovative Statistics to Improve Our Notion of Reality," Journal of Official Statistics, 25, 431–465.
59. Volland, B. 2017. The role of risk and trust attitudes in explaining residential energy demand: Evidence from the United Kingdom. Ecological Economics 132(February): 14–30.
60. Warren, M. E. (ed.). (1999). Democracy and Trust. Cambridge University Press, Cambridge.
61. Zak, P.J. & Knack, S. 2001. Trust and growth. Economic Journal 111(470): 295–321.

Of numbers and narratives[1]: Evidence for policy-making in the 21st century

Gaby Umbach
European University Institute, Florence/Italy

Abstract

The paper discusses key aspects of evidence-based policy-making (EBPM), of governance and knowledge effects of statistics and of narratives in politics. It focusses on EBPM stimuli on the creation of knowledge and on its normative effects on defining and identifying (new) evidence. The paper also interlinks the analysis of statistics and politics. It deals with essential elements of EBPM and draws conclusions on its challenges for statistics and data providers. With this focus, the paper speaks to questions of the intentional and unintentional use of indicators and to the question to what extent indicators fulfil their intended purpose (intended by developers, politicians, advocacy groups, etc.). It, moreover, deals with risks of indicators and their interaction with policy (especially aspects of EBPM and the path of indicator information to and from policy-makers). Finally, it addresses issues of informational governance and the conditions (legal, political, institutional) for the enhancement of stakeholders' confidence in statistics, data, indicators and factual evidence. The paper is based on the author's joint research with Debora Valentina Malito and Nehal Bhuta for their co-edited *Palgrave Handbook of Indicators in Global Governance*, her analyses within *GlobalStat* and most recent work for the *Enlightenment 2.0* project and expert group of the European Commission's Joint Research Centre.

Keywords

Data; evidence-based policy-making; measurement; narratives; conceptual frameworks

1. Introduction

Data[2] is a 'vector of acting by knowing' (Bhuta, Malito, and Umbach 2018, 4). As data manifest approximations to social reality, they stand in dialectical relationship with social action and are essential means for politics. In the form

[1] The paper and title are based on the joint research of Bhuta, Nehal, Debora Valentina Malito, and Gaby Umbach. 2018. 'Introduction: Of Numbers and Narratives—Indicators in Global Governance and the Rise of a Reflexive Indicator Culture." In *The Palgrave Handbook of Indicators in Global Governance*, eds. Debora Valentina Malito, Gaby Umbach, and Nehal Bhuta. Cham: Springer International Publishing, 1–29. https://doi.org/10.1007/978-3-319-62707-6_1.

[2] For the purpose of reducing linguistic complexity, the terms 'data' and 'statistics' are used as proxies for the universe of metrological instruments that count as evidence in policy-making and include measurement tools such as indicators, indices, composite indicators, scoreboards, ecc.

of indicators, indices, composite indicators, scoreboards (ecc.), they are key instruments to quantify, qualify and compare complex processes, structures and situations in order to assess the performance of nations and human well-being. As such, they are essential means of collective political action.

The connection between statistics and politics materialises the essential feature of 'governing by knowing' in 21st century evidence-based decision-making. Data and factual evidence 'stand in for' abstract concepts and realities to make them measurable. Hence, quantification and categorisation through indicators and scoreboards define meaning and are neither merely technical nor normatively neutral: they create reality, impact on behaviour, and define and change paradigms, ideas, and relationships between actors. The normativity of measuring defines social reality that is subject to state regulation and policy-making (Bhuta, Malito, and Umbach 2018).

In this understanding, data and statistics are an essential parts of evidence-based policy-making and have fundamental 'knowledge and governance effects' within the political process: systemic conditions alter (participation and network governance develop); 'data cultures' emerge (new avalanche of indicators); new forms of knowledge become relevant and accepted ('post-metrological trends'); and statistics turn into advocacy and policy tools that perform new qualitative functions in politics.

From such multiple impact also stem the instrumental limits of statistics and factual evidence as policy tools. Central questions in this context centre on the 'rhetoric-reality gap' and the 'means-ends dilemma' of the embedment of political interests in measurable scales. Difficulties of common understandings; the life-cycle of paradigms as well as the resulting 'ecology of indicators' and 'data cultures' are to be scrutinised thoroughly as is the criticism on the one-size-fits-all character and the politicisation of data in politics.

The metamorphosis of data evidence into policy tools finally also affects the genuine work of data providers. The use of data in EBPM is subject to the principles of legitimacy, transparency and accountability, but also of preference building and negotiation. Data therefore become subject to public discourse, scrutiny and contestation that target the narratives and power structures that emerge from and result in their production and use. The use of data consequently fosters participatory structures, co-creation and epistemic community building around data communities that perceive their evidence as opportunity structure for the normative framing of policy content. These developments have immediate impact on political institutions and their capacity to select, evaluate and process data, which is directly influenced by difficulties to develop common understandings of complex social phenomena and by the inevitable trend towards oversimplification of such complexity. In this way, statistics become part of the knowledge and evidence production process itself that supports discourse and deliberation in evidence-based processes. As such, they require institutional checks and balances to guarantee neutrality of and open access to data to increase transparency about who defines what at which stage of the evidence-based policy-making process. They demand interlinkages between and responsiveness of statistics creation and the policy context, which data should inform. As a result, the use of data in EBPM becomes open to multiple, strategic and value-based considerations that support interests of central knowledge and data producers, brokers and users.

2. Methodology

Conceptual framing; results of case study analysis.

3. Essential results and key take-aways

- Data use in politics materialises essential features of 'governing by knowing'.
- Statistics have become key instruments of collective political action.
- Evidence morphs into policy and advocacy tools.
- Quantification is normative and political.
- Data need to adhere to principles of legitimacy, transparency, accountability.
- Evidence in politics is subject to preference building and negotiation.
- Statistics have 'knowledge and governance effects'.
- Evidence becomes subject to public discourse, scrutiny and contestation.
- Data use fosters participatory structures, co-creation and epistemic community building.
- Narratives and alternative forms of evidence become relevant.
- Statistics become part of the knowledge and evidence production process.
- Their production and use require institutional checks to guarantee transparency of use.
- Requirements for data and evidence production change.

Today's key challenges for data production for EBPM are:

- Openness to public scrutiny & contestation *(Normative character & implications: What narratives & power structures emerge from use of statistics)*
- Participatory structures in political systems *(Tendency of epistemic community-building: Statistics as opportunity structure; Framing potential; Participation; Co-creation)*

- Institutional capacity to select, evaluate & process data *(Difficulty of common understandings / Oversimplification of complexity: Statistics are part of knowledge production process; Discourse & deliberation in evidence-based processes)*
- Institutional requirements to guarantee neutrality of / access to data (*Who defines what at which stage; Responsiveness statistics creation – policy context; Multiple and strategic/value-based use to support interests & aims)*
- Instrumental limits *('Rhetoric-Reality Gap' , 'Means-Ends Dilemma', Difficulty of common understandings / Oversimplification of complexity / Life-cycles of paradigms: 'Ecology of indicators'/'Data cultures'; One-size-fits-all; 'Post-factual' trends; Politicisation)*

4. Discussion and Conclusion

The Intentional and Unintentional Use of Indicators: Data as a Means of Politics

Statistics are key instruments to quantify, qualify and compare; as such they are essential for collective political action (see Bhuta, Malito, and Umbach 2018). While a lot of academic reflection focusses on the micro level of data development and statistical methods, macro level analysis of the overall position and relevance of statistics in policy-making are essential to understand the challenges ahead for data providers and politicians in view of producing and using data as evidence in 21st century policy-making.

History informs us that the connection between knowing and governing links statistics and politics ever since the 17th century's rise of the early modern state. As highlighted by the previous research of Malito, Umbach and Bhuta (2018), quantification became an essential means of early modern politics through governing by numbers based on delocalised, aggregate knowledge about the state and its component parts. When, within the enlightened state of the 19th century, rationality and effectiveness of government became categories of politics, statistical knowledge was in high demand and the proliferation of statistics led to an 'avalanche of printed numbers' (Hacking 1990: 138). With this increased availability of metrics of the state and the need to justify state intervention, measurable realities and measuring became instrumental for collective political action. The inseparable link between statistics and state power was born.

Indicators and Their Interaction with Policy: Data as Factual Fifth Power of Evidence-Based Politics

Grounding decision-making in factual evidence 'helps people make well informed decisions about policies, programmes and projects by putting the best available evidence at the heart of policy development and implementation' (Davies 2008). Such evidence-based policy-making 'represents both an important set of professional practices and aspirations; and also a political rhetoric seeking to legitimate forms of decision-making' (Head 2010). The more recent concept of 'evidence-informed policy-making' (EIPM) seeks to be more flexible (Martinuzzi and Scholl 2016) and cautious about the impact of evidence on policy-making (emphasising the input side of the political process and indicating at the fact that evidence has been injected into it). The earlier concept of EBPM embraces a more throughput- or output-oriented perspective (assuming that the evidence injected into the policy process influences policy-making, both procedurally and in view of the policy content adopted). With these nuances, EIPM (Chalmers 2005) pays tribute to the fact 'that policy making is an inherently political process ..., involving ideology, vested interests, institutional norms and path dependencies...' (Bannister & O'Sullivan 2014; see Head 2015) as an alternative basis for decision-making.

EBPM entails structural and procedural shifts in policy delivery and outcomes within knowledge conversion processes (Castellani et al. 2016). It embraces evidence and expertise as building blocks of policy design and highlights the efficacy of particular types of evidence. EBPM treats evidence and (statistical) knowledge as a (political) resource and highlights the relevance of knowledge management and communication. The (data) science-policy interface and the relationship between evidence, knowledge and power are essential aspects of EBPM that also impact on the role of statistics on (political actors' strategies in) policy-making.

Evidence informing EBPM can include scientific and research evidence, statistical and survey data, results of impact assessments and policy evaluation or 'contextual knowledge from previous experiences' (Castellani et al. 2016). As one of its most robust types, and going beyond their purely metrological purpose of quantification and measurement, statistics have become one of the most important forms of evidence in such EBPM in the 21st century. Used to provide the factual basis for evidence-informed policy development, statistics serve multiple purposes: they measure and compare; increase insight and knowledge; inform monitoring of progress; support evaluation and assessment; and constitute independent sources of information that open government decision-making to wider scrutiny. Within EBPM, statistics hence take over different functions. They can

- support strategic planning in multilevel political structures;
- define common goals for progress and development;
- enhance multi-dimensional performance assessment;
- increase transparency of decision-making processes and policy instruments;
- inspire innovation;

- and serve various other instrumental, conceptual, tactical, symbolic and political purposes.

Moreover, measuring itself turned into an essential 'way of doing politics' (Malito, Bhuta, and Umbach 2018, 511), justifying and rationalising political decisions and making progress visible and comparable. Statistics are, hence, an essential part of governing with and through knowledge to bring 'objective consistency' (Desrosières 1998, 32) of social phenomena into politics. As such, statistics are important means of identifying policy requirements and of holding decision-makers accountable for their policy choices. In this way, the use of data evidence increases transparency of policy choices and can therefore be regarded as the factual fifth power in evidence-based policy-making in the 21st century.

Issues of Governing through Numbers and Challenges for Data Providers: Limitations of Data as Factual Evidence

Increasing a post-metrological trend in EBPM, new forms of evidence and (robust) knowledge have yet entered in competition with official statistics. Such new forms include evidence not based on narrow scientific definitions or other knowledge claims and propositions about the factual world. Increasing the plurality of knowledge sources are experimental studies (such as randomised controlled trials); observational qualitative and quantitative studies to identify causal mechanisms; functional models; modelling; normative deduction from principle norms; citizens and practice-informed knowledge or indigenous knowledge and experience. While such new evidence sources need to show a convincing degree of validity and reliability, levels of uncertainty and room for interpretation remain, influencing the authority/contestation of evidence and resulting in a (multidimensional) knowledge hierarchy for EBPM. A classification (i.e. abstract, practical, subjective) and quality assessment (i.e. narrative, storytelling, data, anecdotal) of knowledge as well as 'systematic linkages between context and evidence type" of knowledge necessarily follow (Bannister & O'Sullivan 2014; see also Epstein et al. 2014).

A divide over measurement flaws and conceptual approaches accompanies this multiplication of evidence sources that also unveils challenges of previous one-size-fits-all measures that are not entirely actionable or do not provide policy-relevant information. Resulting from the proliferation of evidence and knowledge sources, official (aggregate) statistics are complemented by more disaggregate, micro-level, local experience-based measures to support the development of more targeted policy interventions. At the expense of comparability, the latter are more targeted to trigger action and steer reforms and decision-making on the ground.

Another challenge for data as evidence in policy-making lies within their alleged objectivity: by definition, statistics define normalcy and deviancy and hence codify behaviour. Measuring is therefore not just counting, but defining and creating 'objective' norms (Bhuta, Malito, and Umbach 2018). The decision on what to measure is thus framing reality. In the same vein, deciding not to measure social phenomena creates reality through exclusion and prioritisation. Through active codification and categorisation, (social) phenomena become reality. In an increasingly pluralistic world with contradicting world views, statistics hence become inherently political.

Data used in EBPM also have governance and knowledge effects that need to be accomodated. The 'making of knowledge *about* governance is a medium *of* governance'(Bhuta, Malito, and Umbach 2018, 13) and 'knowing about governance ... is at the same time also governance by knowing' (Voß and Freeman 2016, 5). Quantification is an inherently political process that defines multiple relationships of power and dominance given that data and indicators are instruments of governance in themselves and affect governance. They substitute governance processes by influencing decision-making through forms of soft power. They also re-configurate political relations and create political priorities by influencing assessment and judgments. While their power is highly contextual, they shape political action in new modes of governance, such as bench-marking processes; peer review and monitoring exercises; expert exchange; performance-based management, self-evaluation, audit cultures; and forecasting or horizon scanning. They influence formal and informal regulatory practices through normative frames and paradigms, technical standards, as well as 'shared ontologies, rationalities, models and technical standards of governing often develop momentum as an independent force of collective ordering' (Voß and Freeman 2016, 5). Moreover, by 'activism through numbers', they function as advocacy tools through naming-blaming-shaming exercises in order to promote policy change (Malito, Umbach, and Bhuta 2018).

From EBPM also arise new challenges for the impact and role of statistics in politics as systemic preconditions change and 'truisms' vanish (see Cairney 2016). EBPM results in the dissolution of a fixed policy cycle for the injection of evidence. Instead of few access points for evidence injection into an assumed closed policy cycle, policy-making is to be understood as a continuous framing of policy narratives for which evidence is the basis. Such narratives further develop throughout the EBPM process and evidence is therefore required at every stage of the policy-making process. Policy-making is consequently to be deconstructed into its component parts – actors, institutions (rules and norms), networks, belief systems (core beliefs, ideas, paradigms), policy conditions and events – rather than perceived as a closed cycle. Knowledge production and injection processes need to accompany policy-making throughout the entire process as narrative policy frameworks are not developed in linear stages. A solid relationship and trust base between knowledge producers and users forms a precondition to influence actors in a process rather than stages in a cycle.

Rendering the science-policy interface for the injection of data and evidence into politics even more complicated is the 'paradox of scientific authority' that accompanies the rise of populism and politicisation of evidence that led to contestation and instability of expertise (Newman and Clarke 2018; Fleming and Rhodes 2018); failure of evidence to influence policy-making; mis- or disinformation, post-truth and post-fact phenomena. Research on their foundations, inter alia rooting in strategies of 'deceptive misinterpretations" (Frankfurt 2005), as well as on emotions, values and interests in politics experiences a boom to which not only academics, but also journalists increasingly contribute. As a result of this 'emotional turn', data providers are challenged by perception-based and emotional politics and biases that are strongly influenced by scepticism and change averse attitudes; (dysfunctional) political communication in the age of social media and fake news as well as potential strategies to neutralise negative impacts of false information (d'Ancona 2017; Davis 2017; Gluckman and Wilsdon 2016; Levitin 2017; Rosling, et al. 2018). Cognitive biases[3] and bounded rationality interfere with the power of data evidence in politics and influence the selection of (best possible) evidence for the political process. 'Rational' (goal-oriented selection of evidence) as well as 'irrational' (stricto sensu belief- and/or emotion-based prioritization of information) behaviour of political actors are hence fundamentally influencing the kind of evidence selected for policy-making (Cairney 2016). Emotions and political behaviour can hence be assessed as gatekeepers when it comes to the selection of evidence for policy-making. Additionally, 'the gap between the needs of policy makers and the ways researchers present evidence' (van der Heide, et al. 2016) is identified as a central barrier for the injection of evidence into policy-making that enhances 'the non-systematic or almost accidental feature of the processes leading to its inclusion" (Lin 2005). In this context, the translation and 'processing of this information and expert knowledge is [viewed as] problematic and highly variable across organizations' (Head 2015) and communities.

Moreover, questions of truth and mis-/disinformation in politics; the role and relevance of asymmetries of knowledge; forms of knowledge and knowledge transfer; measurement of evidence use in politics and the difference between EBPM and 'ideological or faith-based policy-making' (Head 2010) become increasingly relevant also for statistical bodies. As both policy problems (e.g. climate change, fracking, cyber security) and required solutions are complex and multidimensional, the data evidence informing the

[3] Such as framing effects, representativeness heuristic, availability heuristic/processing fluency, prospect theory, cognitive dissonance, need for coherence, status quo biases/sunk costs fallacy/optimism biases, or groupthink.

development of policy solutions also requires a broader, if not holistic, and in any case multidimensional approach to cover the facets of the political challenges at hand.

This increase of complexity goes hand-in-hand with an increasing contestation of expertise and experts leading to the paradoxical situation that 'expert advice is being sought with growing urgency across a proliferating array of policy and public questions. At the same time, and often on the same issues, the legitimacy of evidence and expertise has rarely been so fiercely contested' (Gluckman and Wilsdon 2016). The gap between demand for and (contestation of) supply of data evidence and expertise hence widens to an alarming extent for the data science and policy-making communities. Increased research activities on trust in politics and data science communication are hence necessary.

References

1. Bannister, J., & O'Sullivan, A. (2014). Evidence and the antisocial behaviour policy cycle. Evidence & Policy, 10(1), 77–92.
2. Bhuta, N., Malito, D. V., & Umbach, G. (2014). Representing, Reducing or Removing Complexity: Indicators of Sustainability and Fiscal Sustainability (Robert Schuman Centre for Advanced Studies Working Paper, 2014/78). Florence: European University Institute.
3. Bowker, G. C., & Susan Leigh Star, S. (1999). Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press.
4. Bradley, C. (2015). International Organizations and the Production of Indicators: The Case of Freedom House. In S. E. Merry, K. E. Davis & B. Kingsbury (Eds.). The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law (pp. 27-74). Cambridge: Cambridge University Press.
5. Broome, A., & Quirk, J. (2015a). The Politics of Numbers: The Normative Agendas of Global Benchmarking. Review of International Studies, 41 (5), 813-818.
6. Cairney, Paul (2016). The Politics of Evidence-Based Policy Making, Palgrave.
7. Cairney, Paul / Oliver, Kathryn (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? Health Research Policy and Systems (HARPS).
8. Castellani, T., Valente, A., Cori, L., & Bianchi, F. (2016). Detecting the use of evidence in a meta-policy. Evidence & Policy, 12(1), 91–107.
9. Chalmers, Iain (2005): If evidence-informed policy works in practice, does it matter if it doesn't work in theory? Evidence & Policy: A Journal of Research, Debate and Practice 1(2), pp. 227–242.

10. Cooley, A. (2015). The Emerging Politics of International Rankings and Ratings: A Framework for Analysis. In A. Cooley & J. Snyder (Eds.). Ranking the World: Grading States as a Tool of Global Governance (pp. 1-38). Cambridge: Cambridge University Press.
11. Davies, Philip (2008). Making Policy Evidence-Based: The UK Experience, Presentation at the World Bank Middle East and North Africa Region - Regional Impact Evaluation Workshop, Cairo, Egypt, 13-17 January 2008.
12. d'Ancona, Matthew (2017): Post-Truth: The new war on truth and how to fight back, Pengiun Books.
13. Davis, K. E., Kingsbury, B., & Merry, S. E. (2012). Indicators as a Technology of Global Governance. Law and Society Review, 46 (1), 71-104.
14. Desrosières, A., (1998). The Politics of Large Numbers: A History of Statistical Reasoning. Cambridge, MA: Harvard University Press.
15. Desrosières, A. (2001). How Real Is Statistics? Four Possible Attitudes. Social Research, 68 (2), 339-355.
16. Desrosières, A. (2016). Retroaction: How Indicators Feed Back onto Quantified Actors. In R. Rottenburg, S. E. Merry, S. J. Park & J. Mugler (Eds.). The World of Indicators: The Making of Governmental Knowledge through Quantification (pp. 329-353). Cambridge: CUP.
17. Epstein, D., Farina, C., & Heidt, J. (2014). The value of words: narrative as evidence in policy making. Evidence & Policy, 10(2), 243–258.
18. Espeland, W. (2015). Narrating Numbers. In R. Rottenburg, S. E. Merry, S. J. Park & J. Mugler (Eds.). The World of Indicators: The Making of Governmental Knowledge through Quantification (pp. 56-75). Cambridge: Cambridge University Press.
19. Espeland W., & Stevens, M. L. (2008). A Sociology of Quantification. European Journal of Sociology, 49 (3), 401-436.
20. Fioramonti, L. (2014). How Numbers Rule the World: The Use and Abuse of Statistics in Global Politics. London/New York: Zed Books Ltd.
21. Fleming, J., & Rhodes, R. (2018). Can experience be evidence? Craft knowledge and evidence-based policing. Policy & Politics, 46(1), pp. 3–26.
22. Frankfurt, H. (2005): Bullshit, Princeton University Press.
23. Gluckman, P., & Wilsdon, J. (2016). From paradox to principles: where next for scientific advice to governments? [Comments and Opinion]. https://doi.org/10.1057/palcomms.2016.77.
24. Hacking, I. (1990). The Taming of Chance. Cambridge: Cambridge University Press 1990.
25. Hacking, I. (1982). Biopower and the Avalanche of Printed Numbers. Humanities in Society, 5, 279-295.

26. Hacking, I. (2002). Historical Ontology. Cambridge, MA: Harvard University Press.
27. Head, Brian W. (2015). Toward More "Evidence-Informed" Policy Making? Public Administration Review, 76(3), pp. 472–484.
28. Head, Brian W. (2010). Reconsidering evidence-based policy: Key issues and challenges, Policy and Society, 29:2, pp. 77-94.
29. Hurrell, A. (2007). On Global Order: Power, Values and the Constitution of International Society. New York: Oxford University Press.
30. Kelley, J. G. (2017). Scorecard Diplomacy: Grading States to Influence their Reputation and Behaviour. Cambridge: Cambridge University Press.
31. Kelley, J. G., & Simmons, B. A. (2014a). The Power of Performance Indicators: Rankings, Ratings and Reactivity in International Relations (American Political Science Association Annual Meeting Paper). Washington, DC: American Political Science Association. Retrieved from https://ssrn.com/abstract=2451319.
32. Kelley, J. G., & Simmons, B. A. (2014b). Politics by Number: Indicators as Social Pressure in International Relations. American Journal of Political Science, 59(1), 55-70.
33. Levitin, Daniel (2017): Weaponized Lies: how to think critically in the post-truth era, Pengiun Books.
34. Lin, V. (2005). Improving the research and policy partnership: An agenda for research transfer and governance, in: Lin, V, Gibson, B (eds), Evidence-based health policy: problems and possibilities, New York: Oxford University Press, pp. 285-297.
35. Merry, S. E., Davis, K. E., & Kingsbury, B. (Eds.). (2015). The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law. Cambridge: Cambridge University Press.
36. Merry, S. E. & Wood, S. (2015). Quantification and the Paradox of Measurement: Translating Children's Rights in Tanzania. Current Anthropology. 56 (2), 205-217.
37. Merry, S. E. (2016). The Seductions of Quantification: Measuring Human Rights, Gender Violence and Sex Trafficking. Chicago: Chicago University Press.
38. Newman, J.& Clarke, J. (2018). The instabilities of expertise: remaking knowledge, power and politics in unsettled times, Innovation: The European Journal of Social Science Research, 31(1), pp. 40-54.
39. Poovey, M. (1998). A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society. Chicago: Chicago University Press.
40. Porter, T. (1995). Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton: Princeton University Press.

41. Porter, T. (2015). The Flight of the Indicator. In R. Rottenburg, S. E. Merry, S. J. Park & J. Mugler (Eds.). The World of Indicators: The Making of Governmental Knowledge through Quantification (pp. 34-55). Cambridge: Cambridge University Press.
42. Power, M. (1997). The Audit Society: Rituals of Verification. Oxford: Oxford University Press.
43. Malito, D. V., Umbach, G. and Bhuta, N. (2018). *The Palgrave Handbook of Indicators in Global Governance*. Palgrave Macmillan.
44. Martinuzzi, A., & Scholl, G. (2016). Advancing evidence-informed sustainable development policies. Evidence & Policy, 12(3), 311–319.
45. Rosling, H., Rosling, O. & Rosling Rönnlund, A. (2018). Factfulness: Ten Reasons We're Wrong About The World - And Why Things Are Better Than You Think. London: Sceptre.
46. Rottenburg, R., Merry, S. E., Park, S. J., & Mugler, J. (Eds.). The World of Indicators: The Making of Governmental Knowledge through Quantification. Cambridge: Cambridge University Press.
47. Strathern, M. (Ed.). (2000). Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy. London: Routledge.
48. van der Heide, I., van der Noordt, M., Proper, K. I., Schoemaker, C., van den Berg, M., & Hamberg-van Reenen, H. H. (2016). Implementation of a tool to enhance evidence-informed decision making in public health: identifying barriers and facilitating factors. Evidence & Policy, 12(2), 183–197.
49. Voß, J.-P. and Freeman, R. (2016). Knowing governance: the epistemic construction of political order. Springer, 2016.

Tracking multinationals' activities in the shipping industry: A new statistical framework for the estimation of sea transport services in the Greek balance of payments

Athanasios Petralias, PhD & Marios Papaspyrou, PhD
Department of Statistics, Bank of Greece

Abstract

The Greek Shipping Estimation Model aims to provide a coherent statistical framework to estimate Balance of Payments items related to shipping activity, based on administrative sources and commercial databases.

Given the multinational nature of the shipping sector and the complex group structures, the estimation of the shipping activity is one of the most challenging tasks in terms of official statistics. This is of particular importance to Greece, since its merchant fleet has a strategic role in the transportation of the trade and energy of many regions in the world. The statistical framework presented in the current study, can well be applied by other countries for Balance of Payments purposes, as well as by researchers and analysts seeking to estimate revenues and expenses related to shipping activity.

In the Greek Shipping Estimation model a granular level approach is used, with vessel by vessel characterization. A three-stage approach is taken. Firstly, the cluster is defined, the main counterparts and the types of Balance of Payments (BOP) transactions that take place and can be estimated. Secondly, the population is defined, including legal owner companies, operator/ship-manager companies and the vessels to be included in the registry. Finally, all BOP transactions for a particular vessel are estimated on a monthly basis.

The maritime cluster structure is discussed, along with guidelines on how to define the resident population on basis of the economic ownership principle. A detailed statistical framework to estimate all shipping related transactions is presented, including vessels' revenues, bunkers costs, port expenses, manning costs, administrative costs and other BOP items. For this purpose a series of international databases is used including among others Lloyd's List intelligence, IHS Maritime and Trade, Clarksons Shipping Intelligence Network and Drewry.

Keywords

Multinational enterprises; shipping; sea transport services; balance of payments

1. Introduction

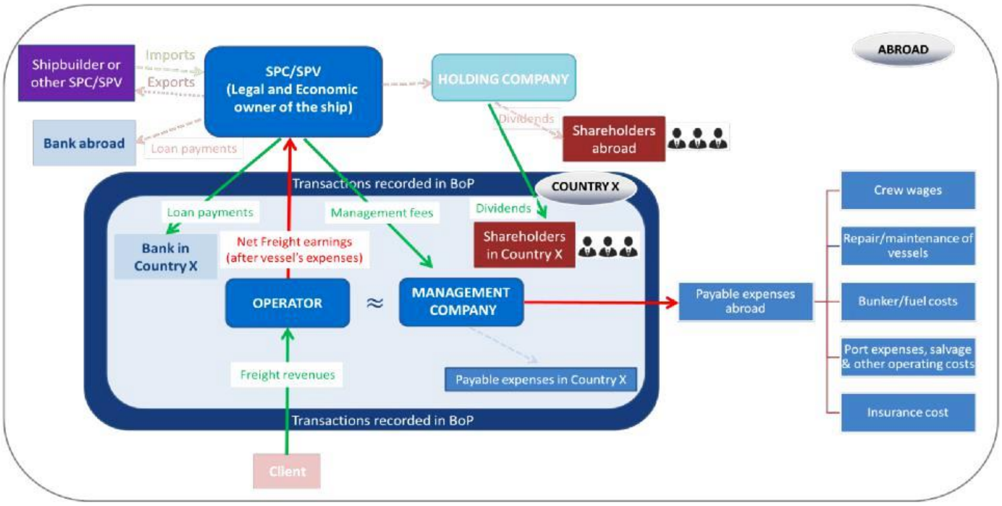
The Greek Shipping Estimation Model aims to provide a coherent statistical framework to estimate Balance of Payments items related to shipping activity, based on administrative sources and commercial databases. Given the multi-territorial nature of the shipping sector and the complex group structures, the estimation of the shipping activity is one of the most challenging tasks in terms of official statistics. This is of particular importance to Greece, since its merchant fleet has a strategic role in the transportation of the trade and energy of many regions in the world and of the EU and shipping activities turnover contributes significantly to Greece's Current Account Balance and the GDP. The statistical framework presented in the current study, can well be applied by other countries for Balance of Payments purposes, as well as by researchers and analysts seeking to estimate revenues and expenses related to shipping activity.

In the Greek Shipping Estimation model a granular level approach is used, with vessel by vessel characterization. A three-stage approach is taken. Firstly, the cluster is defined, the main counterparts and the types of Balance of Payments (BOP) transactions that take place and can be estimated. Secondly, the population is defined, including legal owner companies, operator/ship-manager companies and the vessels to be taken into account. Finally, all BOP transactions for a particular vessel are estimated on a monthly basis.

2. Methodology

The first step is to define the structure of the shipping cluster. The structure presented in Figure 1 refers to the case of Standard Ship Management Agreement. This has been identified as the most common scenario for the Greek shipping cluster, associated with shipping companies engaged to international carriage of goods by sea, but applies in general to cases of standard ship management agreements. The structure of the cluster has been compiled by combining information from: (a) detailed banking transactions data recorded in BOP, (b) contacts with associated banks, shipping experts and companies and (c) commercial databases.

Figure 1: Standard Ship Management Agreement and BOP transactions



Typically, there is a ship management company (not depicted in the graph) which is legally registered abroad, with a branch/office whose main activities take place in country X. The activity is attributed to the branch by aligning with the indication of country of domicile as it appears in the commercial databases. The local branch of the ship management company performs acts of operation of the vessel (there are cases where the operator is different to the ship management company, but these are relatively few). Thus, in Figure 1 the local branch of the ship management company is denoted as the operator and the manager of the vessel.

Also there is a fully legitimate ship owning company (SPC/SPV), which is the legal owner of the vessel, most often registered abroad. Note that the ship management company may manage numerous vessels, each one belonging to a different legal owner (each SPC/SPV typically owns a single vessel for risk-exposure and liability purposes). Note that in case the legal owner is incorporated under the laws of country X, the above diagram and the respective transactions are modified by moving the box of SPC/SPV inside country X.

Typical examples of Standard Ship Management Agreements are the BIMCO “SHIPMAN 98” and “SHIPMAN 2009” agreements. According to the standard ship management agreement terms, the freight revenues are received by the ship management company/operator, on behalf of the ship owner. The ship management company uses these freight revenues to pay the operating expenses of the vessel (i.e. crew costs, insurance costs, bunker costs, port expenses and other operational and non¬operational costs). The rest of the freight earnings are directed to the legal owner (SPC/SPV), under the form of imports of sea transport services. The ship owning company (SPC/SPV) pays management fees to the ship management company, purchases or sells the vessel and also receives the loan drawdowns and pays capital installments and interest. Also, the ship owning company pays dividends to the shareholders. In case a mother-holding company exists, it is typically the mother-holding company who distributes the dividends to the shareholders, after collecting the earnings from all the SPCs/SPVs of the group.

Economic owner of the vessel (i.e. the good, which is associated with imports/exports and the capital costs) is the legal owner (SPC/SPV). The ship management company acts for and on behalf of the ship owning company (SPC/SPV). However the ship management company provides the commercial operation of the vessel, i.e. the transport service and is responsible for the commercial decisions concerning the employment of the ship, and provides also management services.

The second step is to define the population of vessels to be included in the cluster. In order to define the population and estimate the relevant items of the Balance of Payments we have used an exhaustive list of databases, domestic and international, and additional valuable information from relevant maritime legislation, governmental sources and shipping experts both from the academia and the industry. The main data-sources used include IHS Maritime and Trade and Lloyd's List intelligence databases, providing monthly data on a vessel basis for various types of ownership, vessel characteristics, new deliveries/deaths, crew and port movement; Ministry of Mercantile Marine, Aegean and Island policy data with a list of the management companies established in the country; Clarksons Shipping Intelligence Network with detailed monthly time charter rate data by type of vessel and size; Drewry with data regarding operating expenses by vessel type and build year; World Bunker Prices and pricing policy of largest ports worldwide which is available in their websites. Note that whenever a variable appears in more than one database, cross-validations are performed. The above are the main data-sources used for the needs of the compilation of Balance of Payments items. Various other databases were also consulted for developing the Greek Shipping Estimation Model, including the Greek Shipping Directory, a traditional domestic database for Greek shipping; Bloomberg; Greek Maritime Laws; BIMCO types of agreements; UNCTAD; ITF types of agreements; Petrofin; data collected by companies involved in the shipping sector and other.

For each vessel there exist one or many linked companies, including the legal (registered) owner, the shipmanager, the commercial operator, the beneficial owner and potentially the technical manager and/or the third party operator. The relationships among the different players are very complex and identifying their roles is not always easy, having in mind that in the commercial databases are used definitions and terminology not always aligned with the statistical ones.

According to BOP compilation manuals and guidelines and given the Task Force on the recording and compilation of maritime transactions consultation, in the majority of cases, the economic ownership can be identified on the basis of the transport agreement among the legal shipowner (SPC/SPV) and the operator/shipmanager. There have been identified three representative cases of agreements: Time or Voyage Charter (Case A); Standard Bareboat Charter (Case B) and Standard Management Agreement (Case C). Note that as described above, the Standard management Agreement (Case C), is the most common for the Greek case.

In cases of Time or Voyage Charter (Case A) and in Standard Management Agreement (Case C), the legal shipowner (SPC/SPV) is considered as the economic owner of the vessel. The operator/management company acts for and on behalf of the legal shipowner; the legal shipowner receives earnings from the economic activity (in terms of daily fee in Case A and net freight earnings in Case C), whereas the asset (vessel) and the liabilities are attributed to the legal shipowner (SPC/SPV).

In cases of Standard Bareboat Chartering Agreement, the economic ownership can be transferred by the legal shipowner to the operator (chartered) of the vessel, depending on the conditions set in the agreement and whether they fulfill the indicative criteria set in the compilation manuals.

In view of the above the population included in the registry relevant for the Greek Shipping Estimation Model is determined as follows. (A) Cases for which gross revenues and expenditures related to shipping services are included in the Greek BOP: Based on the list of management companies provided by the Ministry of Maritime, identify the companies and their vessels, which are listed in the commercial databases as commercial operators with country of domicile Greece (checking logical coherence between the databases). Exclude from these cases the third party operator is abroad. (B) Cases for which economic ownership lies in Greece and import/exports of vessels (as well as shipping services) are included in BOP: Identify ship owning companies and their vessels, with country of registration of legal owner Greece. Exclude from these cases there is bareboat chartering abroad. Include cases there is bareboat chartering in Greece (these are rare/few cases for the Greek case). The vessel's registry is updated on a monthly basis by incorporating the new information of the commercial databases.

The third step is, given the population of vessels included in the shipping registry, to calculate revenues and expenditures by vessel, coded by BOP item and counterpart country. Types of expenditures that need to be calculated (see Figure 1) include Bunkers/fuel cost, Port Expenses, Manning costs, Vessel Insurance costs, Maintenance and dry docking costs, Stores, spares and lubricant oils costs, Flag state expenses and voluntary tax and Administrative costs. Also transactions among the Commercial Operator and the Legal owner

(SPC/SPV), including Management Fees and net Freight Earnings must be calculated.

The estimation of shipping services revenues (exports of services) for each vessel requires first identifying whether the vessel is hired, then estimating the expected revenues, provided it is hired and allocating the amounts to the respective counterpart countries. Since there is no direct information of whether a vessel is hired, this can be indirectly inferred from its draft which is available in the international databases. A draft above 20-30% of its max depth signals the vessel holds cargo. Also, if there is port movement and the vessel is not laid up or under repair, one can also assume the vessel is hired; note that there are types of vessels (i.e. passenger, tugs) not relevant for draft. Then, for each vessel hired, we calculate expected revenues by vessel type and size (deadweight/TEU) using the respective, monthly Clarksons' time charter rate. Since rates are provided for specific deadweights/teu's, we use linear percentage differences, among two adjacent deadweight/teu categories, to calculate each vessels' revenues, based on the observed exact vessel's characteristics provided by commercial databases. The country of non-resident is set to be the country of the last port visited, which is provided by commercial databases. Although this might not be always exact, it is the best possible proxy on basis of available information and it is aligned with the statistical guidelines. To calculate Bunkers cost, it is first calculated the fuel consumption of each vessel and then world bunker prices are used to approximate the actual bunker expenses. Then, these are allocated to counterpart countries on basis of the last port visited. Commercial databases provide for each vessel the typical fuel consumption (in tonnes per day) at a given (cruising) speed. However it is common practice for vessels to travel at reduced speeds, for fuel economy purposes. There is a well-known theoretical formulae, that links fuel consumption with the third power of speed (see for example, Ronen, 2011). Given the consumption F0 at speed V0, fuel consumption at speed V1 is equal to:

$$F_1 = F_0 (V_1/V_0)^3,$$

which can be directly calculated on basis of the variables that are available by the commercial databases, as described above. Having calculated the actual fuel consumption (in tonnes per day) by vessel, we rely on world bunker indices to calculate the monthly fuel costs.

To calculate port expenses, first we find on basis of commercial databases, the ports and canals wordwide most often visited by vessels included in the registry. Then, we find the pricing policy of these ports, which is publicly available (most of the times in their webpage). Main expenses include Berthing (a fee when entering the port) and Dockage (a fee while staying in the dock).

These fees typically depend on Gross/Net Tonnage or Length and duration of stay. Duration of stay can be calculated either on basis of the commercial databases (on basis of last arrival day and last sailed date by vessel), or use average time in port, provided by the port.

To calculate Manning costs we attribute to each vessel, the expected manning cost, depending on their type, deadweight and age, on basis of Drewry data that have been cross-checked with International Transport Workers' federation "TCC" collective agreements. Commercial databases provide for the majority of vessels total number of crew by nationality, as well as number of officers and ratings. On basis of these data we calculate the distribution of number of ratings and officers by ethnicity and vessel type. In the final step we allocate the manning cost of each vessel to the respective ethnicity, weighted by the ratio of the rates received by officers with respect to rest of crew.

Drewry reports average vessels' Insurance costs, Repair & Maintenance costs, Dry docking costs, Stores, spares and lubricant oils costs, Flag state expenses and Administrative costs by vessel type, deadweight and age. Thus, the respective costs are attributed to the respective vessels in the registry. Note that with respect to repairs, commercial databases indicate whether a vessel is "In Casualty or Repairing" or "Converting/Rebuilding" and thus we assign the respective Dry Docking costs to the port dry docking and repair takes place.

Turning to the transactions between the Commercial Operator and the Legal owner, management fees by type, size and age of vessel are obtained from Drewry. Net freight earnings are paid by the Commercial Operator to the legal owner (see Figure 1). We calculate net freight earnings, by subtracting from the total freight revenues all the vessel's expenses that are paid by the commercial operator both inside the country and outside the country, i.e. manning, insurance, stores, spares and oils, maintenance and drydocking costs, flag and voluntary tax, administrative expenses (except management fees that are considered as income to the operator), bunker costs and port expenses:

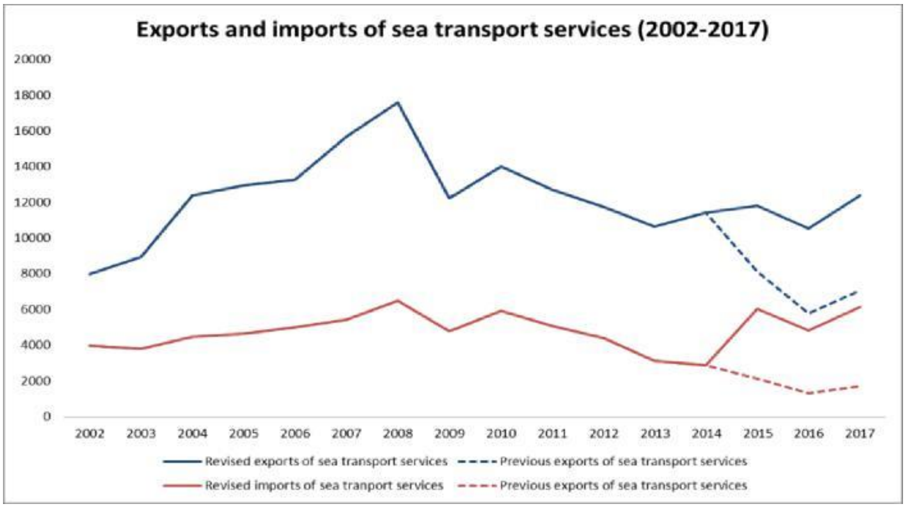
$$\text{Net freight earnings} = \text{Freight revenues} - \text{Vessel's voyage and operating expenses.}$$

Then, on basis of the commercial databases we find the country of registration of the legal owner (SPC/SPV) that receives the net freight earnings from the commercial operator. Net freight earnings enter the Balance of Payments as imports of sea transport services.

3. Result

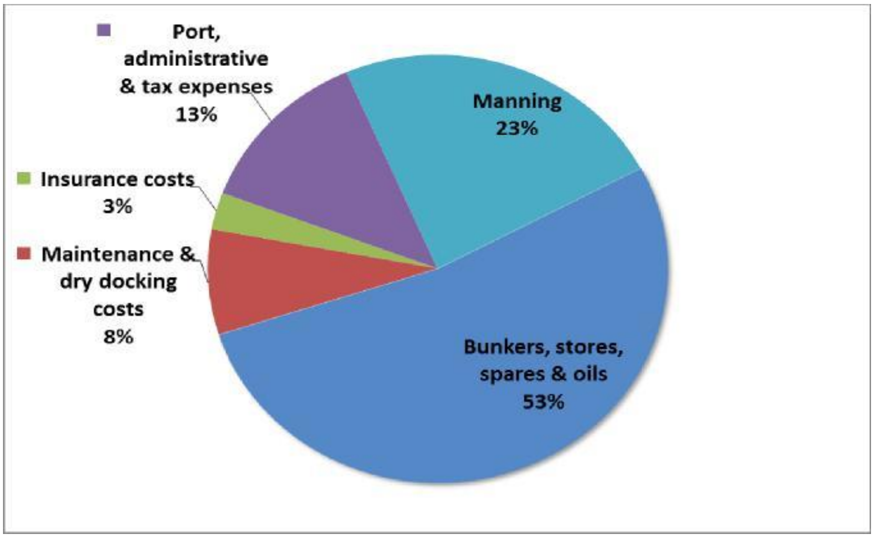
The implementation of the Greek Shipping estimation model allowed to re-establish the continuance of the time series of sea transport services imports and exports, that has been observed in Greek BOP data after the introduction of Capital Controls (Law 18/7/2015) in Greece; see Figure 2 (Bank of Greece, Press Release 20/11/2018).

Figure 2: Evolution of exports and imports of sea transport services in the Greek BOP, before and after the introduction of the Greek Shipping Estimation Model (2002-2017, in million €)



On yearly basis, in 2018 exports of sea transport services (receipts) were estimated up to 14.2 billion € and imports of sea transport services (payments) up to 7.3 billion €. The average utilization rate (i.e. the percentage of days in a month the vessels in the population are hired) was estimated (on basis movement data) up to 95.4%.

Figure 3: Distribution of shipping operating and voyage costs paid abroad (2018Q4)



With regards to the distribution of vessel operating and voyage costs (Figure 3), in the 4th quarter of 2018, 53% of the total vessel expenses paid abroad were attributed to bunkers, stores, spares and oils, 23% to crew wages, 13% to port, administrative and tax expenses, 8% to maintenance & dry docking costs and 3% to insurance costs.

4. Discussion and Conclusion

We have developed a comprehensive framework for calculating all Balance of Payments items related to the shipping activity. The Greek Shipping Estimation Model is, to our knowledge, the first holistic methodological attempt in the EU, to estimate on basis of statistical modelling all Balance of Payments (BOP) items related to shipping activity, using commercial databases in conjunction with administrative data.

It is clear that in order to perform such a task, there were significant challenges to overcome, since a new methodological approach had to be developed, specific for each BOP item. Also in practical terms, combining information from many different and large databases, required database subscriptions, advanced statistical programming and big data management. Indicatively, detailed estimates by vessel, have been produced for all vessels' revenues and expenses (by BOP item) and by counterpart country. Moreover, a detailed transaction tool has been developed, to transform these estimates to approximately 200,000 transactions every month to enter the Balance of Payments.

The model needs to be continuously updated to meet the changing shipping environment. Besides the fact that even the commercial databases are changing over the time their fields (i.e. new indices are published and others stop reporting; new types of vessels appear), the mobility of the shipping business sector is continuously growing. The increased competition from Eastern countries, in conjunction with the new environmental framework for the shipping sector and the developments in the world trade, cause many shipping companies (especially SPC/SPVs) to close an others to open even on a daily basis, purchases, sales and conversions/rebuilds of vessels to increase (including new eco-friendly types), new routes to open and high volatility in freight rates and operating expenses.

The Greek Shipping Estimation Model can be readily applied by other member states, as well as countries outside EU, to estimate some or all of the items in Balance of Payments, related to shipping activity. We believe the current methodological framework can set new standards in official statistics, in terms of both using statistical modelling to incorporate information from commercial databases, as well as on how to handle off-shore companies when there is no information on their income and expenses. Finally, it is hoped that this work can form a basis for future research in shipping, in both the academia and the industry.

References

1. Bank of Greece (2018). Press Release 20/11/2018 - Balance of payments: Revision of sea transport statistics. Available in: https://www.bankofgreece.gr/Pages/en/Bank/News/PressReleases/DispItem.aspx?Item_ID=6 244&List_ID=1af869f3-57fb-4de6-b9ae-bdfd83c66c95&Filter_by=DT
2. Eurostat (2013). European System of Accounts ESA 2010. Luxembourg: Publications Office of the European Union. Available in: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-02-13-269
3. Eurostat (2016). User guide on European statistics on international trade in goods, 2016 edition. Luxembourg: Publications Office of the European Union. Available in: https://ec.europa.eu/eurostat/documents/3859598/7679615/KS-GQ-16-009-EN-N.pdf/073b853a-a4f4-4c55-aaba-162671544c78
4. IMF (2009). 6th edition of the IMF Balance of Payment and International Investment Position Manual. Washington, D.C.: International Monetary Fund. Available in: https://www.imf.org/external/pubs/ft/bop/2007/pdf/bpm6.pdf
5. IMF (2017). Balance of Payments and International Investment Position (BPM6) Compilation Guide. Washington, D.C.: International Monetary Fund. Available in: https://www.imf.org/external/pubs/ft/bop/2007/bop6comp.htm

6. Ronen D. (2011). The effect of oil price on containership speed and fleet size. Journal of the Operational Research Society, 62, 211-216.

Recent improvements of the *Tankan* business survey: Globalization of business and statistical revisions

Yuki Uchida
Bank of Japan, Tokyo, Japan

Abstract

The progressing globalization of business has been observed in enterprises across many countries in recent years. Japanese enterprises have also developed business operations worldwide over the past decade, moving their business strategies from a non-consolidated basis to a consolidated one. With Japanese enterprises developing business operations worldwide, the linkages between domestic business activities and those overseas have become stronger. These linkages are not fully captured in non-consolidated figures, such as domestic investments and exports of enterprises in Japan. The Tankan (Short-Term Economic Survey of Enterprises in Japan) is a statistical survey which aims to accurately assess domestic economic conditions; the Bank of Japan collects answers from enterprises in Japan on a non-consolidated basis, and overseas activities, such as overseas sales and investments, are not fully captured. With the overseas business operations of Japanese enterprises having a larger impact on corporate behaviour in Japan, the Bank faces the challenge of more accurately capturing changes in domestic business trends in a timely manner.

Against this background, in late 2016 the Bank proposed major revisions of the *Tankan*: (1) the introduction of a new survey category on a consolidated basis which captures overseas business activities, such as global sales and investments, by approximately 2020; (2) the addition of R&D investment to its survey items in the March 2017 *Tankan* survey; and (3) the enhancement of the survey on exchange rates.

In addition, the Bank is also considering increasing the frequency of sample extraction through stratification based on population data in the near future. This will further allow the Tankan to accurately capture changes in population enterprises and their business models, including changes through the entry of foreign enterprises into the domestic market.

Keywords

Tankan; Business survey; Globalization

1. Introduction

The *Tankan* (Short-Term Economic Survey of Enterprises in Japan) is a nationwide business survey conducted by the Bank of Japan, which aims to provide an accurate picture of business trends of enterprises in Japan, thereby contributing to the appropriate implementation of monetary policy. [1]

The *Tankan* is a geographical-based survey, using a sample survey framework (outlined in the Appendix) where sample enterprises are selected from a survey population whose target population is private enterprises (excluding financial institutions) in Japan with capital of 20 million yen or more. Of the survey population[2] comprising approximately 220,000 enterprises, approximately 10,000 sample enterprises were chosen in accordance with statistical accuracy. The *Tankan* asks various questions to measure business activities, such as those regarding business conditions, financial statements, and inflation, all of which are surveyed on a non-consolidated basis, collecting answers from enterprises located in Japan.

The above framework meets the Tankan's purpose as it measures, on a quarterly basis, short term domestic economic conditions through the perceived business conditions of the surveyed enterprises and their annual projections via financial statements including exports.

However, it is difficult to capture the factors behind changes in domestic business activities which are occurring due to expanding overseas business activities and which are not fully covered in the existing survey items of the Tankan. With the sales and profits of Japanese multinational enterprises being driven by their overseas business activities, investment decisions are made more and more on a consolidated rather than non-consolidated basis, making it difficult to capture the background behind changes in corporate behaviour and economic conditions in Japan. This poses challenges for the Bank in capturing accurately and in a more timely manner, the changing administrative decisions of firms and business trends in Japan. Section 2 shows recent developments in the overseas business activities of enterprises in Japan and describes how the Bank is responding to the challenges by revising survey items and coverage. Section 3 touches on further improvements the Bank is considering to more timely reflect changes in enterprise business trends and section 4 gives some concluding remarks.

2. Recent developments of overseas business activities and revisions of the Tankan

The sales of overseas affiliates of Japanese enterprises have increased over the past decade (Chart 1). Consequently, these affiliates have increased their

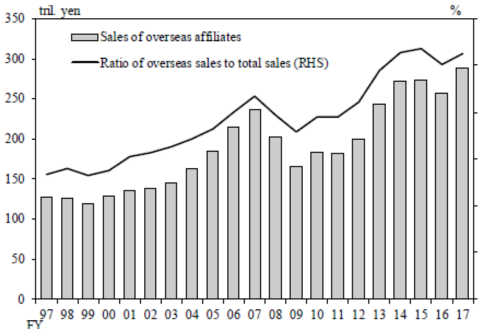
---

[1] For details, see "Explanation of the Tankan (Short-Term Economic Survey of Enterprises in Japan)" (February 2019).

[2] The survey population of the Tankan is drawn from the Economic Census conducted by the Ministry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry, whose enumeration unit is defined as any establishment having one location and being operated under a single management in Japan.

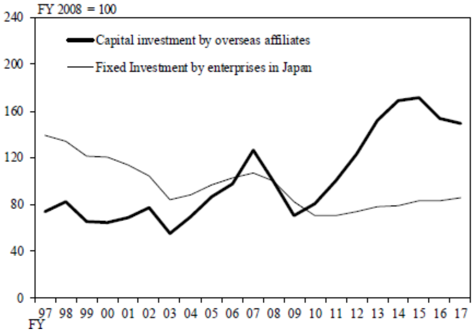
overseas capital investment in response to growing local demand (Chart 2). These charts indicate that Japanese multinational enterprises have been expanding overseas business activities, and that the overseas business is becoming increasingly important for their businesses.

Chart 1 : Sales of overseas affiliates



Note: Ratio of overseas sales to total sales = Sales of overseas affiliates / (Sales of overseas affiliates + Sales of corporations in Japan).
Sources: Ministry of Economy, Trade and Industry, "Survey on Overseas Business Activities"; Ministry of Finance, "Financial Statements Statistics of Corporations by Industry, Annually"

Chart 2 : Capital investment by overseas affiliates



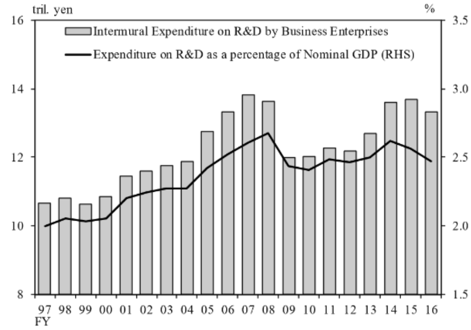
Notes: 1. Capital investment by overseas affiliates is based on simple aggregation of reported companies.
2. Fixed investment by enterprises in Japan is based on estimates of the population total.
Sources: Ministry of Economy, Trade and Industry, "Survey on Overseas Business Activities"; Bank of Japan.

In this situation, where overseas business activities are having a larger impact on the domestic business of Japanese enterprises, the consolidated perspective plays a more important role in building business strategies. Depending on their strategy, Japanese enterprises may increase domestic investments to leverage the production capacity of their mother factories in Japan to meet demand from overseas customers with cash earned through the stronger sales of subsidiaries. Alternatively, they may increase investments by foreign affiliates in response to local demand, minimizing foreign currency risk, rather than domestic investments to re-establish mother factories in Japan. Increasing intermural expenditures on R&D (Chart 3) to enhance their global competitiveness with a view to expanding overseas operations exemplifies the stronger linkage between overseas and domestic business activities.

The current framework of the *Tankan* aims to capture the domestic business activities of enterprises in Japan. With regard to the linkage between overseas and domestic business activities, the *Tankan* only covers the relationship between exports and domestic business activities, such as sales, profit and fixed investments on a non-consolidated basis (Chart 4) in its annual projections of the survey items.

Therefore, the recent developments in operations of Japanese multinational enterprises worldwide and their linkage to domestic business activities are not fully captured within the current *Tankan*.

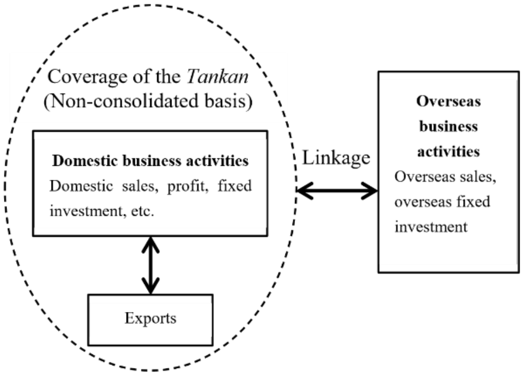
Chart 3 : R&D investment



Note: Business enterprise R&D expenditure is used for R&D expenditures as a percentage of nominal GDP

Sources: Ministry of Internal Affairs and Communications and the Ministry of Economy, "Survey of Research and Development"; Cabinet Office

Chart 4 : The coverage of the *Tankan*[3]



Against this background, in late 2016 the Bank proposed to revise survey items, to more accurately capture the changes in economic conditions reflecting developments in overseas business activities. The main revisions to the Tankan are as follows: (1) the introduction of a new survey category on a consolidated basis which captures overseas business activity, such as global sales and investments; (2) the addition of R&D investment; and (3) the enhancement of the exchange rate survey.

[3] Survey items of the Tankan consist of the following four categories with 26 items surveyed on the basis of nonconsolidated accounting: (1) Judgment survey (business conditions; overseas supply and demand conditions for products; domestic supply and demand conditions for products and services; inventory level of finished goods and merchandise; wholesalers' inventory level; production capacity; employment conditions; financial position; lending attitude of financial institutions; conditions for CP issuance; change in interest rate on loans; change in output prices; change in input prices); (2) Annual projections (sales; exports; exchange rates for exports; operating profits; current profits; net income; fixed investment; land purchasing expenses; software investment; R&D investments); (3) Inflation outlook of enterprises (outlook for output prices; outlook for general prices); (4) Number of new graduates hired.

Chart 5 : Summary of the revisions in survey items[4]

	Before Revision		After Revision		
Annual Projections	Sales	(half year)	Sales	(half year)	
	Exports	(half year)	Exports	(half year)	
	Exchange Rates for Exports (Yen/U.S. dollar)	(half year)	Exchange Rates (Yen/U.S. dollar)	(half year)	Enhanced
			Exchange Rates (Yen/Euro)	(half year)	Enhanced
	Operating Profits	(half year)	N/A		Abolished
	Current Profits	(half year)	Current Profits	(half year)	
	Net Income	(half year)	Net Income	(half year)	
	Fixed Investment	(half year)	Fixed Investment	(fiscal year)	Changed to fiscal-year figures
	Land Purchasing Expenses	(half year)	Land Purchasing Expenses	(fiscal year)	Changed to fiscal-year figures
	Software Investment	(half year)	Software Investment	(fiscal year)	Changed to fiscal-year figures
	N/A		R&D Investment	(fiscal year)	Changed to fiscal-year figures
Overseas Business Activities (consolidated figures)	N/A		Sales	(fiscal year)	Added
			Overseas Sales	(fiscal year)	Added for specific enterprises
			Current Profits	(fiscal year)	Added for specific enterprises
			Fixed Investment	(fiscal year)	Added for specific enterprises
			Overseas Fixed Investment	(fiscal year)	Added for specific enterprises

Notes: 1. Shaded areas indicate revised items.
2. For R&D investment — which has been added from the March 2017 survey — half-year figures has been surveyed at first, as with current investment-related items, and the change to fiscal-year figures will be applied along with these items in due course.

Around 2020, the Bank will introduce a new survey category called "Overseas Business Activities." Specifically, the Bank will conduct a quarterly survey on actual results and forecasts on a consolidated basis for the following five items — sales, overseas sales, current profits, fixed investment, and overseas fixed investment — publishing aggregated amounts for each figure as estimates of the population total.

[4] Based on our survey policy to improve the convenience for users as well as to reduce the burden on survey respondents to the greatest extent possible, the Bank will stop surveying operating profits and half-year figures of investment-related items along with the above revisions in around 2020.

Taking into account ease of response so as to reduce the burden on respondents to the greatest extent possible, the *Tankan* basically collects accounting based figures. However, by doing so, there are technical limitations in its coverage with regards to the treatment of non-controlling interests in the consolidation process. Consolidated sales and investments do not reflect the overseas business activities of associates accounted for using the equity method. If a respondent owns less than 50% of shares in a non-controlling interest of an overseas subsidiary's equity, sales and investment amounts of the subsidiary are not included in the respondent's overseas sales and investment amounts in principle. Regardless of these limitations, the introduction of a new survey category allows the Bank to more accurately capture the impact of overseas activities on business activities of enterprises in Japan.

"R&D investment[5] " has been added as a new item in the survey items of annual projections of the Tankan from the March 2017 Tankan survey. This reflects the fact that Japanese enterprises, facing intensifying global competition, have started increasing R&D investment in order to enhance their competitiveness in the medium to long term, with their domestic offices taking on an increasingly significant role in R&D. Moreover, the addition of R&D investment to the Tankan's survey items makes publishing figures of software and fixed investment excluding land purchasing expenses more in line with the national accounts.

The survey of exchange rates will be enhanced in its coverage of sample enterprises and survey currencies from around 2020. The *Tankan* currently conducts a survey of average U.S. dollar-yen exchange rates for actual results of exports and expected rates on which forecasts of exports are based. The survey thus only collects exchange rates expected by exporting enterprises. However, with increasing outward direct investment, direct investment income, along with increased imports, it has become increasingly important for the business activities of Japanese enterprises, including mainly nonmanufacturing non-exporting enterprises. Collecting the survey with exchange rates for all enterprises becomes increasingly important to accurately reflect expanding overseas business activities. Currencies other than the U.S. dollar, particularly the euro, play a greater role over direct investment income, for instance. In light of these developments, the Bank will take a survey on euroyen as well as U.S. dollar-yen exchange rates expected by all sample enterprises excluding financial institutions and thus exchange rates for overall businesses (not only exports but also imports, outward direct

[5] Total of R&D expenses included in general and administrative expenses as well as in manufacturing expenses for the period.

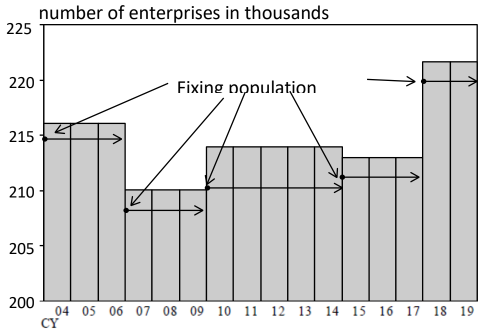
investment, and direct investment income) will be answered by an expanded number of enterprises, including non-exporting enterprises.

These revisions allow the Bank to more accurately capture corporate behaviour and business environment surrounding enterprises in Japan, enabling users of the *Tankan* to analyze the impact of expanding overseas business activities on changes in domestic business activities in a more accurate and timely manner.[6]

3. Further improvements

To reflect changes in the business trends of enterprises in a timely manner, the Bank considers further improvements to its survey framework in addition to revisions to survey items. Currently, in order to conduct the survey as frequently as practicable, the Bank uses a sample survey framework for the Tankan, extracting sample enterprises from the population. This is done by fixing a population at a certain point of time given the technical limitation that population data is only updated every 4 years (Chart 6), while selected sample enterprises based on the population are basically fixed for that period.[7]

Chart 6 : Population size



Sources: Ministry of Internal Affairs and Communications and the Ministry of Economy; Bank of Japan

However, the population data changes over time from the time when fixing the population at a specific time, making estimated figures based on selected enterprises of the *Tankan* less representative of the population. In fact, a fairly large portion of the samples change, every time the Bank updates population data and reselects samples about every 4 years.

The Bank is considering in the near future, conducting sample stratification and extracting sample enterprises more frequently than the current 4 yearly basis. This is in line with the Government of Japan's work on statistical revisions to more accurately reflect economic structures in a timely manner.

[6] For every survey, the Tankan collects responses from enterprises during about one month survey period, and releases results on the day after the end of the survey period.

[7] The Bank examines statistical accuracy regularly (once a year, in principle) and if necessary, adds new sample enterprises, to prevent the lowering of statistical accuracy caused by a decrease in the number of sample enterprises due to bankruptcies, mergers, and other factors. The recent examination showed that the statistical accuracy after the December 2018 survey satisfied the criteria, such as standard error ratio of sales within the criteria 3% or less for manufacturing and 5% or less for nonmanufacturing for each enterprise size. For details, see "Regular Examination of the Statistical Accuracy of the Tankan" (February 2019).

One of its recent developments, for instance, is in the annual business register, which provides, every year, information and an up-to-date database of all establishments and enterprises in Japan, the results of which are used to compile accurate economic statistics.[8]

With more frequent selection of samples from a population using stratified sampling, the Tankan will more accurately capture changes in population enterprises and their business models in a more timely manner, including those changes occurring via the entry of foreign enterprises into the domestic market. By reflecting those changes upon selecting sample enterprises, and maintaining high statistical accuracy, the Bank will be able to more accurately represent the business activities of enterprises in Japan.

4. Concluding Remarks

In this paper, challenges to capture changing business trends in a timely manner have been discussed. With enterprises expanding their overseas business activities, their domestic strategies are likely moving from a non-consolidated basis to a consolidated basis. However, the Tankan as it is, conducts surveys on a non-consolidated basis in order to accurately assess domestic economy, its main aim. To capture changes in business trends more accurately and timely, it is increasingly important to include overseas business activities in the survey.

The revisions, introduction of a new survey category regarding overseas business activities, the addition of R&D investment to the survey items, and enhancement of the exchange rate survey, allow the Bank to more accurately capture the business environment of enterprises in Japan as well as those overseas business activities, enabling users of the Tankan to analyze the linkages between changes in domestic business activities and overseas business activities in a timely manner. Along with further considerations for more frequent sample extraction through stratification, the *Tankan* will be able to capture changes in domestic economy more accurately and in a more timely manner.

[8] For further details about the business register, please refer to Ministry of Internal Affairs and Communications and the Ministry of Economy,
https://www.stat.go.jp/english/data/jsdb/index.html.

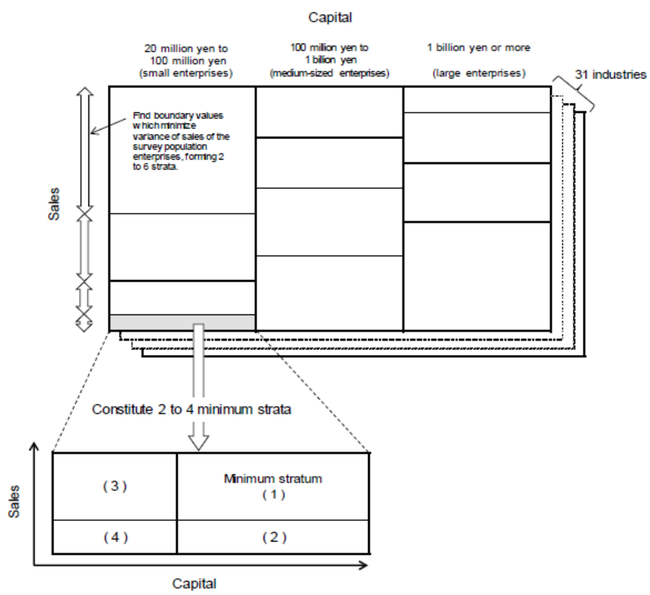
(Appendix)

Sample survey framework of the *Tankan*

The *Tankan* uses a sample survey framework, which surveys part of a population (sample) and estimates the population parameters using the sample data. In a sample survey, a higher number of sample enterprises improves statistical accuracy and also increases the total burden on survey respondents. Stratified sampling has been introduced in the Tankan to satisfy established criteria for statistical accuracy with a controlled number of sample enterprises. Under this sampling method, the survey population is divided into homogeneous groups (strata) using criteria such as industry and capital, then, each stratum is sampled. Detail procedures of the estimation for quantitative survey, which is a part of survey items of the *Tankan*, are as follows.

1. Stratification method of the sample design

The survey population is divided into 93 strata — 31 industries times 3 enterprise sizes (large, medium-sized, and small enterprises). These are then subdivided into 2 to 6 strata in order to minimize the variance of sales of the survey population enterprises with a goal of 300 to 400 strata as a whole. Each stratum is further subdivided into 2 to 4 minimum strata based on capital and sales to test whether there is a statistically significant difference in these strata between the sample and population distributions. The conceptual diagram below shows the stratification method as outlined above.



2. Estimation method for quantitative survey (annual projections, number of new graduates hired) The estimate of the population total is obtained by aggregating the population total in each 93 stratum, which is estimated as follows.

e_i = average amount of responding enterprises in stratum i × N_i

$= \frac{\alpha_i}{n_i} \times N_i$

e_i : Estimate of the population total in stratum i

a_i : Simply aggregated amount of responded figures from sample enterprises in stratum i

n_i : Number of responding sample enterprises in stratum i

N_i : Number of survey population enterprises in stratum i

Complex multinational enterprises: A cooperative approach between Deutsche Bundesbank and Banque de France

Tatiana Mosquera Yon[1], Jens Walter[2]
[1]Banque de France
[2]Deutsche Bundesbank

Abstract
The financial crisis raised concerns about the worldwide interconnectedness of financial institutions and their ability to finance the economy. In the real economy, multinational enterprises generate more and more economic and financial flows due to their international organisation. In order to better understand and explain the contribution of multinational enterprises in their Balance of Payments (BoP), France and Germany undertook a common work on large and complex multinational enterprises operating in the two countries. Based on two work streams: an inter-institutional and an external one (in collaboration with the multinational enterprises), implying exchange of confidential information between institutions, this analysis improved the knowledge on the multinational enterprises involved. It underlined the increasing importance of intra-group real and financial flows that, thanks to this work, will be more coherently recorded in the BoP of the two countries and can be better explained in the future.

Keywords
Multinational enterprises, balance of payments, complex global production arrangements, international fragmentation of production process

1. Sharing of data reported by complex multinational enterprises: a cooperative approach between Deutsche Bundesbank and Banque de France

Links between multinational enterprises and globalisation

Multinational enterprises (MNEs) are the key drivers to the globalisation process in the last decades. International flows are increasing in amount and frequency as MNEs grow. They enhance the interconnectedness of the countries where they operate but also the competition among countries willing to attract them as investors to stimulate economic growth and employment.

The removal of restrictions on the movement of capital, the lowering of trade barriers by the World Trade Organisation GATT and GATS[1] agreements,

[1] GATT: General Agreement on Tariffs and Trade. GATS: General Agreement on Trade in Services.

sinking transport costs and the improvement of information technologies have allowed companies to relocate their production activities to even more remote places around the globe. This all has led to a steady growth in the number of MNEs which in turn intensify the globalisation process. New markets are created, new production chains being established leading to the birth of new leaders. A well-known example is the global production of the I-phone or the development of the digital market which led to the emergence of new actors such as Apple, Amazon, or Google. Their economic development relies on the possibilities offered by globalisation allowing them to grow faster by reaching more customers and to offer more products.

The understanding of the global thinking of MNEs - which are mainly driven by tax minimisation and profit maximisation at a global level - is of utmost importance for politicians today. An adequate statistical measurement of MNEs induced international flows of capital, goods, services and intellectual property is a prerequisite to assess the consequences of national economic and financial policies for employment, income and wealth. Thus, the comprehensiveness of all statistics affected by MNE decisions like the Balance of Payments (BoP), National Accounts (NA) and Business Statistics (BS) are necessary to establish efficient economic, trade or fiscal policies.

Relevant statistics are also important to produce more sophisticated indicators on globalisation, global value chain and international fragmentation of the production process. To give the best information, these indicators need to be produced with high quality data that can only be compiled if the contribution of MNEs is clearly identified.

Location of economic ownership in MNE-Groups

In the International Monetary Fund (IMF)'s sixth edition of balance of payment manual (BPM6), the time of recording of transactions is based on the change of ownership. *"The change of economic ownership is central in determining the time of recording on an accrual basis for transactions in goods, non-produced non-financial assets, and financial assets"* [2].

In the case of MNEs, the BPM6 specifies that "*goods may move between a parent and its branch abroad.* In that case, possibilities exist that either the goods have changed economic ownership or they may have been sent for processing. *The correct statistical treatment is to identify which location assumes the risks and rewards of ownership most strongly (e.g., from factors such as whether the goods are included in the accounts, and which location is responsible for subsequent sale of the goods*)"[3].

[2] Paragraph 3.41 of BPM6, p.55.

[3] Paragraph 3.46 of BPM6, p 56.

For BoP compilers, it is a real challenge to identify which entity assumes risks and rewards in an MNE-Group. It requires having a precise and complete knowledge of how the MNE is organized and operates.

Although the BPM6 gives in several paragraphs (e.g. paragraph 5.3) some guidance for compilers to identify the economic owner inside a group, the explanation of the concept of economic ownership versus legal ownership to their MNEs' correspondents is however not a simple and often time consuming task.

Once this definition is explained, it becomes important to analyse the organisation of MNEs to identify which entity assumes the risk and rewards. There may be several entities or just one, depending on how the MNEs is organised. In the MNEs involved in the analysis, the identification of the economic owner of the produced goods and services was a challenge. Their very complex organisation and the multiple flows between the entities of the group did not allow an unambiguous identification of the economic owner, despite the characteristics given by the BPM6.

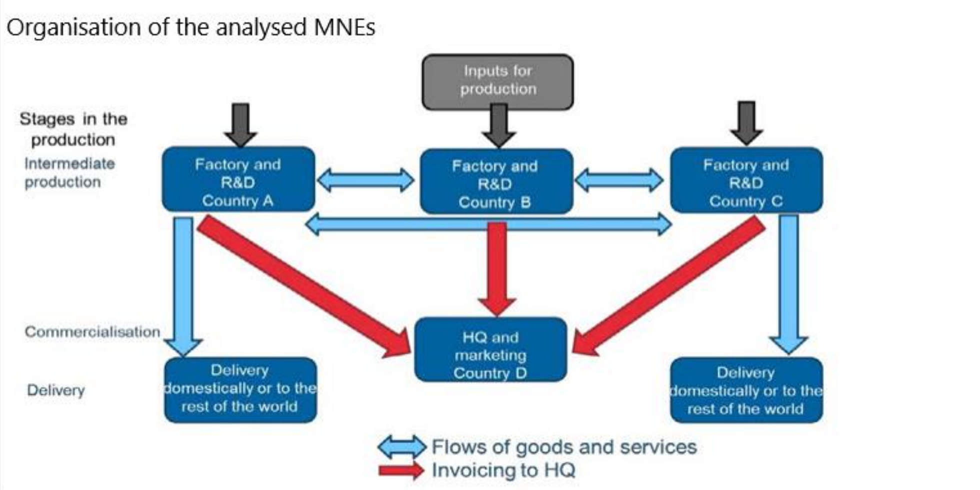
When the economic owner is identified inside the group, it can have impact on BoP compilation since goods may be delivered from Germany, which is recorded in Foreign Trade Statistics (FTS), but sold by the economic owner located in France. Under such circumstances, the export must be reported in France outside the FTS source and to FTS in Germany, using a code which indicates that the goods are not owned by the exporter.

Once these flows are clearly explained by the MNEs, it becomes relevant to share the information with colleagues from the counterpart countries to make sure that the flows are treated in the same way and reported symmetrically in FTS and BoP.

Overview of the MNEs involved in the work stream

The structure of the MNEs involved in the analysis can be schematized as shown below (see illustration 1).

Illustration 1



The affiliates are operating in different countries inside or outside the European Union. At least, one affiliate is established in France and in Germany. The affiliates exchange goods (semi-finished products) and services (use of intellectual property) and invoice each other at transfer prices (that can be agreed in an Advanced Price Agreement with the tax authorities of the countries). Thus, flows can occur between different countries or domestically.

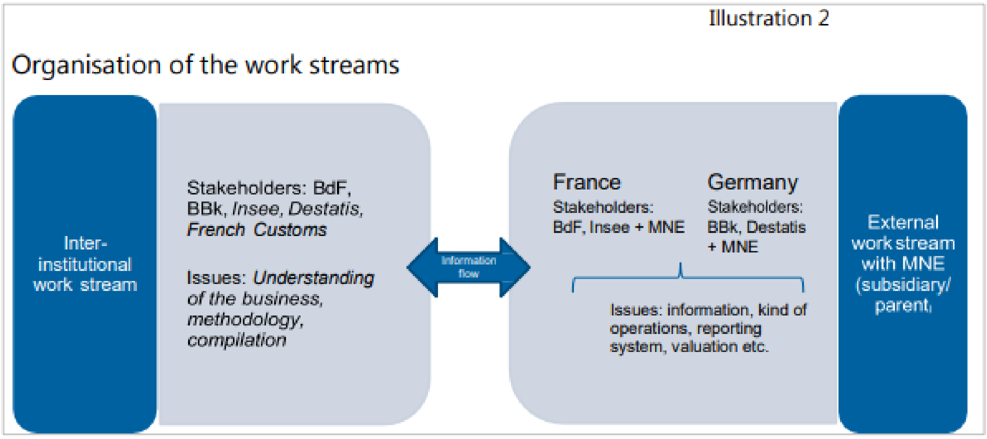
It was recognized by the compilers that the (factoryless) headquarters don't usually manage the whole supply chain. Instead, the affiliates are responsible for certain stages of the chain, usually the production of a semi-finished good or the assembling of the final good. The affiliates invoice their headquarters to cover their operating costs at transfer prices. At the end of the production process, the final goods are owned by the headquarters which is in charge of marketing and customer services.

For the MNE-Group members the correct reporting of these complex arrangements for customs, FTS and BoP is challenging, depending on economic ownership and the movement of the goods. For example, if the headquarters buys semi-finished products from a foreign affiliate and send them for final assembly to another affiliate abroad (without selling them to its affiliate), the headquarter has no obligation to report this to customs/FTS (no movement of goods in its country) but to BoP (as an import, due to the change of ownership) and a subsequent import of manufacturing services (final assembling abroad). If the assembling takes place in the country where the headquarters reside, only a report for customs/FTS is necessary because the movement of the goods coincides with the change of economic ownership.

2. Cooperation between Banque de France and Deutsche Bundesbank

The Banque de France and Deutsche Bundesbank have been cooperating for decades on various fields of balance of payments. From conceptual or methodological issues to resorption of asymmetries, both institutions have been working together to improve their common knowledge by sharing their understanding and experiences. This cooperation is managed at various levels: senior managers or technical experts meet regularly to enhance their collaboration.

The organisation of the cooperation on MNEs mainly relies on two parallel work streams dealing with common topics but analysed under different perspectives. The outcomes of the two work streams are merged at the end of the process in order to have a comprehensive view of MNEs activities and their recording in the relevant statistics, notwithstanding the fact that interactions between the two work streams can happen.



The inter-institutional work stream

The first work stream called "the inter-institutional work stream" deals with conceptual, methodological and compilation topics. It is constituted of experts from different statistical institutions such as the National Statistical Institutes (NSI), National Central Banks (NCB) and the Ministerial Statistical Department of the French customs and indirect taxation authority. Experts from the NSI come from three statistical areas: NA, FTS and Enterprise Statistics.

This inter-institutional work stream meets at two levels: a national level and at an international one. The national level meetings are designed to exchange views on the organisation of the MNEs, to share information on the current reporting, to identify problems and specific reporting practices as well as to clarify the statistical treatment acknowledging the conceptual background. At the beginning of the process, it appeared that each institution had its own comprehension of the MNEs, based on the data it collected. The data reported are usually analysed focusing on the needs of the specific institution/respective statistic and not aimed to design a coherent picture of the MNEs. However, this stovepipe approach sometimes leads to a partial comprehension of the MNEs organisation.

For example, one of the MNEs used the VAT number of one of its subsidiary to report exports of goods from France to Germany to the custom authorities. In the following year, this MNE decided to report its exports under its headquarters' VAT number, with a lower value of the exported goods (as required by the conclusions of an internal audit). For NA, this change was analysed as a major drop of exports of this subsidiary that was not completely balanced inside the group. For BoP compilers, total exports of the group had lowered. Thanks to the cross-checking of data with the custom authorities in one of the "inter-institutional work stream", the information of the change of value and reporting agent was shared, giving sense to the reported data.

The international level of the inter-institutional work stream is aimed at detecting differences in the treatment of cross border flows, to understand the reasons of current asymmetries, clarifying conceptual and methodological issues to get a common view and selecting questions to be addressed to the other work stream. In this work stream, experts can share their understanding of the MNEs organisation developed within the national inter-institutional work stream.

The common understanding of the organisation is a key element of the international inter-institutional work stream to solve asymmetries. Every difference in the assessment of the production process can lead to different customs/FTS codes (e.g. nature of transaction, partner country) reported by the affiliates of an MNE in the countries involved.

Therefore, in order to get a correct reporting, statisticians and reporters must be aware not only of the conceptual differences i.e. physical flows (customs oriented) and the concept of change of economic ownership (BoP oriented) but also have to take into account the whole production chain even if it takes place beyond the borders of their own statistical territory. Throughout the meetings, it was challenging for all the participants to put together the pieces of the puzzle from the external work streams into a picture on which a final decision could be taken, about how these transactions must be recorded in the statistics to provide a consistent dataset.

The external work stream with MNEs

The external work stream with MNEs can only be successful if the MNEs fully agree to cooperate. To reach this level of cooperation, it is really important to communicate and explain what the problems are, how the MNEs will be involved to help solving the problems and what is to be done when solutions are found.

As an in-depth analysis of the MNEs implies to talk about individual data, the principle of confidentiality must be guaranteed in order to allow the MNEs accounting, excise and custom teams to cooperate fully. Regarding confidentiality, it is also important to receive an allowance of the MNEs to share confidential data with the experts of the other statistical institutions of both countries. To reach these two goals it is of utmost importance to create an atmosphere of trust between all stakeholders. Therefore, all steps followed by the other work stream must be clearly presented to the MNEs team, which implies frequent and regular meetings explaining the achieved steps and the coming ones. These meetings took place on a face to face basis, mainly at the beginning and the end of the process. In between, due to practical considerations, these meetings were mostly conference calls.

Once the confidentiality is guaranteed, the conceptual and methodological issues can be debated. The first step is a stocktaking where the MNE team precisely explain how the group operates inside the countries and at the international level. This information is the key element to identify all relevant cross border flows and to characterise them. After the international flows are filtered out, the data reported by the MNEs are scrutinized to evaluate if they reflect the MNEs activities properly.

In some cases, it appeared that the reporting was not relevant, especially when it covers intra-group flows. When the organisation of the MNE is highly centralised, a significant part of the production process is guided by the headquarters, which sometimes also centralise purchases of key components of the final product. In such a case, the headquarters have the economic ownership of the purchased and final goods. Several options of reporting are open to the MNE. The most frequent is the following: Taking the example above i.e. a component is purchased by headquarters located in a country A and is delivered to a factory in a country C from a country B. The factory in country C reports an import from country B, to materialize the inflow of goods to the customs authority of country C. The headquarters in the country A report a financial flow to the country B to the BoP compilers of the country A to materialize the payment of the invoice from country B.

Regarding that reporting, the BoP compilers must be aware that the economic owner of the component is the MNE's headquarters but there is no reporting in the goods item of the country A's BOP. In country C, a final import of goods is reported that the BoP compiler (due to a wrong coding in its FTS) may take into account in the goods item even though there is no change of ownership between the countries C and B. As MNEs generate important flows of this type, this case was analysed by the two work streams to define a homogeneous reporting scheme for headquarters and the factories.

After taking all information about such transactions into account, it was decided that the import of the component in country C should be reported as an import for processing which enables compilers to identify these goods movements and to withdraw them from the BoP of country C because there is no change of ownership. In country A, the headquarters has to report an import for BoP to take into account the change of ownership from country B. If -at the end of the production process - the final product is exported to a third country directly from C, this has to be reported in C as an export after processing (and not as a final sale as it is often done) so that it could be again withdrawn from the BoP of country C. Further, this export has to be reported as a final sale of goods in country A to take into account the change of ownership (again, outside the customs/FTS reporting scheme). To complete the reporting, the factory in country C would also report processing fees charged to headquarters in country A as an export of manufacturing services on physical inputs owned by others; the headquarters in A has to report the corresponding service import.

These in-depth analyses lead finally to coordinated reporting instructions. They must be explained comprehensively to the members of the MNEs to better understand the needs and the interplay of the relevant statistics. The MNEs' accounting, customs and excise teams usually support this work because it is an efficient way for them to get a clearer view of what has to be reported by each subsidiary in the countries where they operate.

The fact that the external work-stream is fed by the inter-institutional work stream gave the experts in the local team much more legitimacy in the discussion with the local MNEs members regarding reporting advices because they can rely on the fact that the explanations given by the local experts are communicated in an identical way to their sister or mother company in the other country.

Once the conclusions were settled by the "inter-institutional" work stream, a final meeting was organised to explain what the new reporting should be, to decide when it could be implemented and to discuss technical aspects (IT development, backward revisions). One MNE, willing to report efficiently, ask French BoP to second experts to work on the changes and help them to identify the accounting elements to include in their reporting. Support all along the process is another key element to make the external work stream successful.

3. Conclusions and Outlook

The increasing relevance of MNE-Groups in a globalised world and the influences of their economic decisions on national economies must be reflected in macroeconomic statistics like the BoP and national accounts in an adequate way. The current concepts of these statistics, focusing on the national territory, are questioned in various ways by users today. The basic question is: do these concepts are still able to reflect economic activities inside the economy and its international relations adequately?

The experience made with the cooperative approach between the Deutsche Bundesbank and the Banque de France turned out that a cross statistical approach combined with a cross-country approach could foster the understanding of MNE activities and enable compilers to measure their activities adequately and consistently without leaving the grounds of the existing concepts.

The insights into a group's operations, in its international production arrangements and internal pricing help to improve the statistical reporting of the MNE-Group members in a common and coherent way in all statistics. Even more, the work in the external work stream improved the understanding about statistical needs and interdependencies between various statistics of the responsible units in the group and it has fostered the internal communication between the group members in different countries.

The work in "two work streams" has eased the communication between the experts (rapid conclusions on conceptual issues and methods) on the one hand and talks with national group members (simple communication without language barriers, openness to admit mistakes) on the other hand.

In addition, what should not be underestimated for the future work between all stakeholders is the confidence in each other combined with the will to improve the meaningfulness and therewith the overall quality of the statistics. However, even with an optimal cooperation between all stakeholders the process is very time consuming. From our experience at least two years are needed from the initial start to a fully "harmonized" reporting in all countries.

Currently Banque de France and Deutsche Bundesbank continue the work to further enhance the knowledge on global production arrangements of MNEs and to fine-tune its cooperation approach. It is foreseen to share any new experiences with the statistical community in the upcoming months.

References

1. Ahmad, N; et al. (2017): Indicators on global value chains: A guide for empirical work, OECD Statistics Working Papers, 2017/08, OECD Publishing, Paris
2. International Monetary Fund (2009): Balance of Payments and International Investment Position Manual (BPM 6)
3. United Nations Economic Commission for Europe (2015): Guide to Measuring Global Productio

Characteristics of MNEs in Canada

Alexandre Fortier-Labonté, Claire Schaffter
Statistics Canada

Abstract
In 2016, less than one percent of enterprises in Canada were MNEs. Though they were a small share of total enterprises, MNEs had a significant impact on the Canadian economy. These enterprises employed one of four Canadians, and owned 67% of total assets in the financial and non-financial sectors.

Keywords
Canada; Multinationals; Industrial Organization; Finance; Employment

1. Introduction

Globalization refers to the economic integration between countries as a result of increasing cross-border trade and capital movements. Multinational enterprises (MNEs) have been drivers of globalization. These enterprises have taken advantage of innovations in logistics and communications technology over the past four decades to diversify their supply chains and expand into new markets. MNEs are also an important source of investment in innovation, technology and skilled labor in Canada. The involvement of MNEs in these aspects of the Canadian economy was discussed in a prior Statistics Canada study.

This analytical study describes the characteristics of MNEs in Canada and abroad, and how their activities in Canada differ from those of Canadian domestic enterprises that do not control businesses outside the country.

The data for the study come from various sources, including Statistics Canada's 2016 Annual Financial and Taxation Statistics (AFTS) program. The AFTS collects and aggregates data from the financial statements of all enterprises in Canada.

2. Methodology

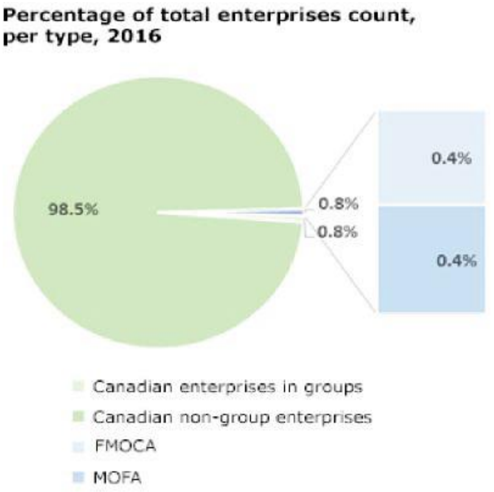
An indicator that distinguishes MNEs from domestic enterprises in the AFTS data is used in this study to compare size, industry involvement and financial performance for both groups. The indicator was developed using two different types of data. The 2016 country of control information from Statistics Canada's Corporations Returns Act program was used to identify foreign majority-owned enterprises operating in Canada. Tax information from 2016 corporate tax and information returns was used to identify Canadian majority-owned enterprises with controlled foreign affiliates. The indicator was developed in collaboration with Statistics Canada's International Accounts and Trade Division.

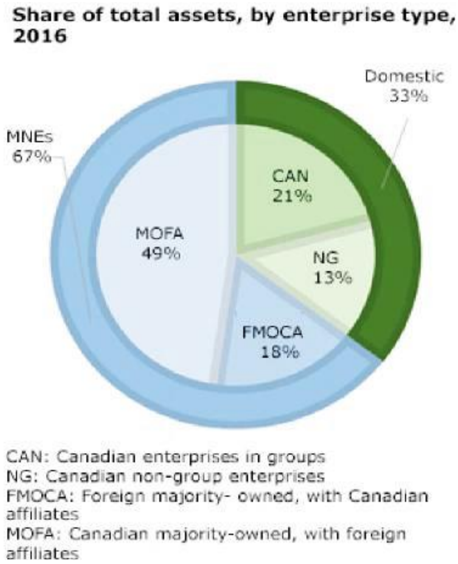
Data collected by the Canada Revenue Agency (CRA) also provide details about subsidiaries of Canadian enterprises abroad.

3. Result

At the end of 2016 (see Chart 1), 99.2% of enterprises in Canada were domestic. Of these, 0.8% were part of an enterprise group, while 98.5% were Canadian non-group enterprises. Only 0.8% of all enterprises operating in Canada were MNEs. Half of MNEs were Canadian majority-owned, with foreign affiliates (MOFAs) and half were foreign majority- owned, with Canadian affiliates (FMOCAs).

Although less than 1% of all enterprises were MNEs, they held 67% of all assets in the Canadian economy. MOFAs owned more assets than FMOCAs, with 49% of the total.

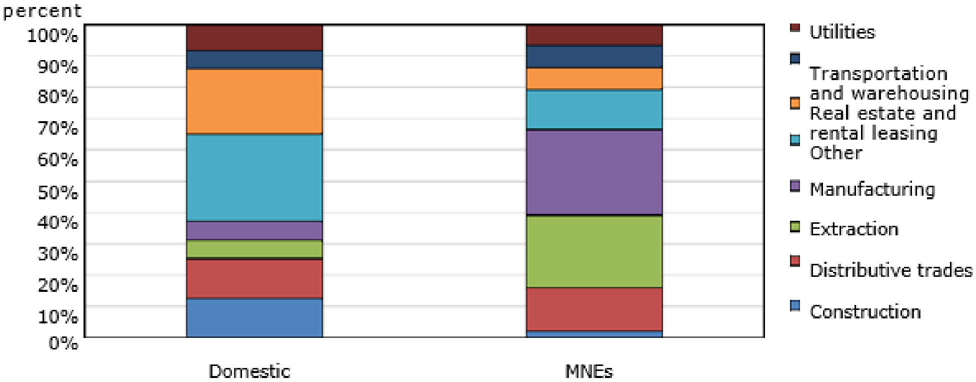




Note: Numbers might not add up to 100% due to rounding.
Source: Statistics Canada, author's calculations based on data from the Annual Financial and Taxation Statistics (AFTS).

Charts 2 and 3 show the distribution of assets for MNEs and domestic enterprises across industries included in the AFTS for 2016. Most MNE's assets in non-financial industries (Chart 2) were held in manufacturing, distributive trades and extraction, while a higher share of domestic enterprise's assets were held in real estate and construction. Overall, domestic enterprises owned 83.6% of all assets in construction and 70.7% in real estate. The local knowledge required to operate in both industries could account for the high share of assets held by domestic enterprises.

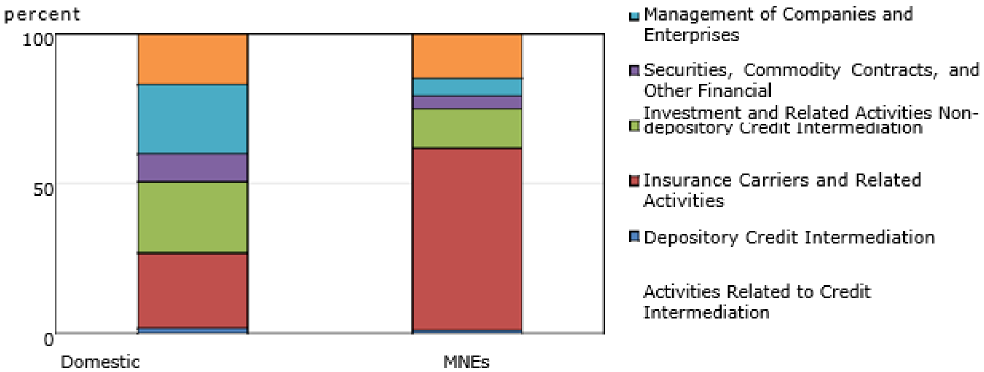
Chart 2
Share of asset holdings by enterprise type — Non-financial industries, 2016



Source: Statistics Canada, author's calculations based on data from the Annual Financial and Taxation Statistics (AFTS).

Chart 3 shows the distribution of assets across financial industries. MNEs held more assets in depository credit intermediation than domestic enterprises. Generally, large banks have operations abroad while local credit unions do not. In general, the financial industries were dominated by MNEs, which held 72.1% of total assets (Table 1). The only financial industry with more assets owned by domestic enterprises than MNEs was Securities, Commodity Contracts and Other Financial Investment and Related Activities, where domestic enterprises owned 57% of total assets.

Chart 3
Share of asset holdings by enterprise type — Financial industries, 2016



Source: Statistics Canada, author's calculations based on data from the Annual Financial and Taxation Statistics (AFTS).

There was also a large discrepancy between the size of MNEs and domestic enterprises in capital intensive industries such as extraction and utilities. In the extraction industry, the median value of assets held by MNEs was \$14.4 million, versus \$0.16 million for domestic enterprises. In the utilities industry, the median value of assets held by MNEs was \$17.34 million, versus

$0.29 million for domestic enterprises. Earlier work (Baldwin and Gellatly) also showed that MNEs were more likely to operate in large-firm industries with significant economies of scale and capital intensity.

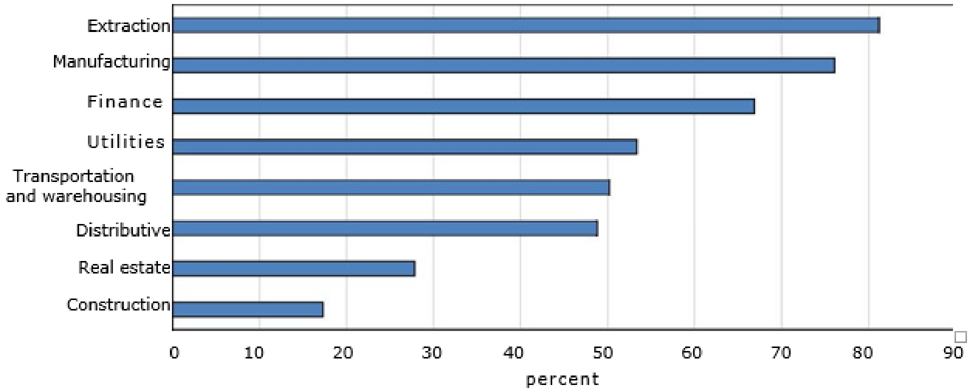
Table 1
Enterprise size in 2016, by type, for selected industries

Industry	MNEs' share of assets	Type	Assets (median)
	percent		$ million
Construction	18.6	MNEs	3.80
		Domestic enterprises	0.13
Distributive trade	60.8	MNEs	6.90
		Domestic enterprises	0.22
Extraction	85.2	MNEs	14.4
		Domestic enterprises	0.16
Finance	72.8	MNEs	3.41
		Domestic enterprises	0.57
Manufacturing	86.5	MNEs	19.52
		Domestic enterprises	0.24
Real estate	33.0	MNEs	3.27
		Domestic enterprises	0.54
Transportation and warehousing	62.3	MNEs	6.43
		Domestic enterprises	0.05
Utilities	53.2	MNEs	17.34
		Domestic enterprises	0.29

Sources: Statistics Canada, author's calculations based on data from the Annual Financial and Taxation Statistics (AFTS).

MNEs had higher operating revenues for the industries where they owned more assets. Chart 4 displays the proportion of operating revenue generated by MNEs in selected industries. MNEs in the extraction and manufacturing industries had the highest share of operating revenue of all industries (80.8% and 75.4%, respectively). MNEs in construction and real estate had the lowest share of operating revenue (17.3% and 27.8%, respectively).

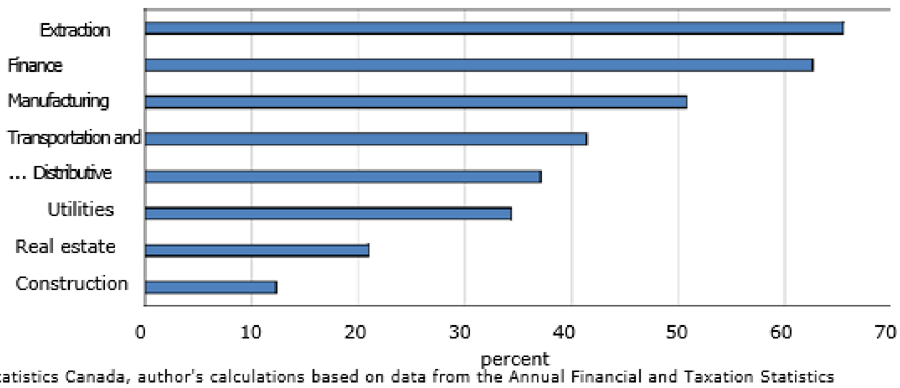
Chart 4
MNEs share of operating revenue for selected industries, 2016



Source: Statistics Canada, author's calculations based on data from the Annual Financial and Taxation Statistics (AFTS).

Overall, one in four employees in the Canadian economy worked for MNEs in 2016. The employment share of MNEs was the largest for the extraction industry (65.5%) and lowest for construction (12.3%) (Chart 5).

Chart 5
MNEs employment share for selected industries, 2016

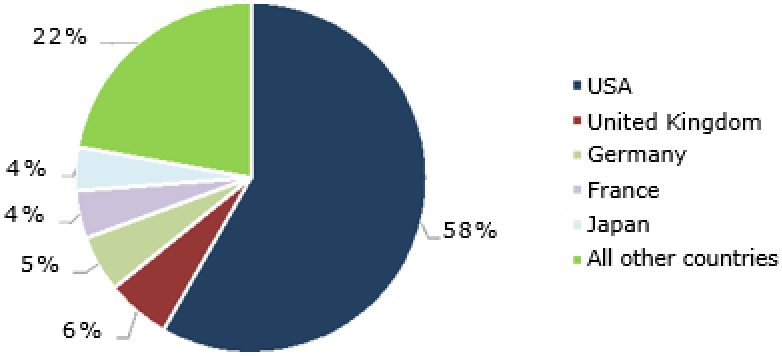


Source: Statistics Canada, author's calculations based on data from the Annual Financial and Taxation Statistics

MNE affiliates abroad

The United States had the most corporate ownership connections with Canada in 2016. The United States had more parent enterprises for FMOCAs than any other country (58.3%), followed by the United Kingdom and Germany (Chart 6).

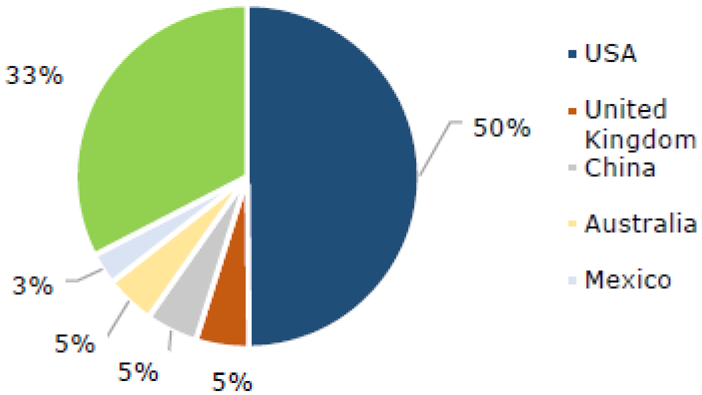
Chart 6
Countries where FMOCAs have the most parents, 2016



Source: Statistics Canada Corporate Returns Act program.

MOFAs reported that most of their subsidiaries (49.9%) were located in the United States. The number of subsidiaries located in the United States was nearly seven times higher than the number of subsidiaries located in the United Kingdom, which had the second highest number of subsidiaries (Chart 7).

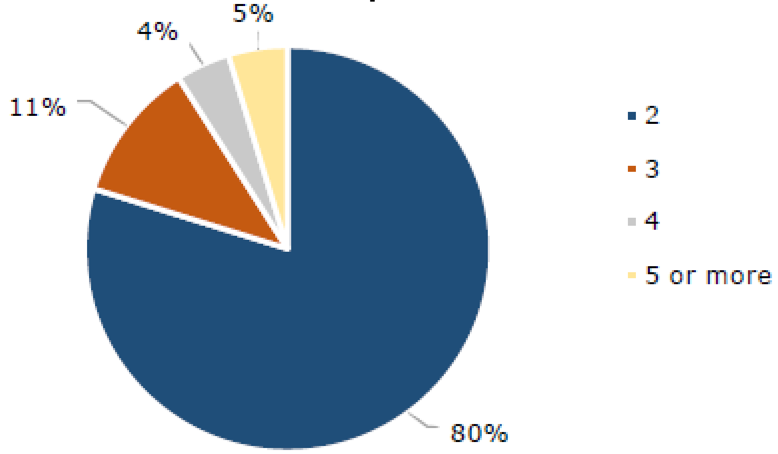
Chart 7
Countries where MOFAs have the most subsidiaries, 2016



Source: Canada Revenue Agency, T1134 information return.

On average, MOFAs directly controlled subsidiaries in only one other country. However, about 10% of MOFAs controlled subsidiaries in more than four countries in 2016 (Chart 8).

Chart 8
Number of countries where MOFAs operated, 2016



Source: Canada Revenue Agency, T1134 information return.

4. Discussion and Conclusion

In 2016, less than one percent of enterprises in Canada were MNEs. Though they were a small share of total enterprises, MNEs had a significant impact on the Canadian economy. These enterprises employed one of four Canadians, and owned 67% of total assets in the financial and non-financial sectors.

The majority of Canadian MNE affiliates are in the United States, where 58.3% of all parents and 49.9% of all subsidiaries are located.

References

1. Baldwin, John R. and Guy Gellatly (2007). Global Links: Multinationals in Canada: An Overview of Research at Statistics Canada. Analytical Studies Branch Research Paper Series. Catalogue no. 11622MIE2007014. Ottawa: Statistics Canada.

Validation of methods and data for SDG indicators

Pietro Gennari[1], Dorian Kalamvrezos Navarro[2]
The Food and Agriculture Organization of the United Nations (FAO)[3]

Abstract
The Sustainable Development Goals (SDG) indicator framework represents a major challenge and a unique opportunity for the advancement of the global statistical system, in terms of both methodological development and governance. Over the last three years, the Inter-Agency and Expert Group on SDG indicators (IAEG-SDG) has gradually developed a number of documents providing criteria and guidelines for regulating the data flows from countries to custodian agencies to inform the global SDG reporting process. The validation of methods and data for SDG indicators, while being apparently two separate procedures, have been closely interlinked in the SDG process. National Statistics Offices (NSOs) when validating country data are also certifying the specific methodology used by the custodian agency for the compilation of the indicator, in particular, the data source used and the adjustments made to harmonize national definitions and classifications. This article highlights some of the main challenges in the practical implementation of the guidelines, identifies areas in need of further guidance from the IAEG-SDG and provides some proposals aimed at improving the global SDG reporting process.

Keywords
Guidelines on data flows; Global statistics governance; International Standards; Non-official data sources; Independence of Official Statistics

1. Introduction

The SDG indicator framework, initially adopted by the UN Statistical Commission in 2016 and subsequently endorsed by UN General Assembly in July 2017, represents a major challenge and, at the same time, a unique opportunity for the advancement of the global statistical system, in terms of both methodological development and governance. Many indicators included in the framework were completely new and did not have an internationally agreed methodology for their compilation (the so called "Tier III indicators").

[1] Pietro Gennari is Chief Statistician at FAO. pietro.gennari@fao.org
[2] Dorian Kalamvrezos Navarro is Programme Advisor at FAO's Office of the Chief Statistician.
[3] The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of FAO 4 UNGA resolution 71/313, Op. Cit.

The intergovernmental body responsible for developing and endorsing new methods is the UN Statistical Commission, supported by the IAEG-SDG. Given that the IAEG-SDG's approval of new methods was a prerequisite for initiating country and global reporting on new SDG indicators, the "reclassification" of Tier III indicators was highlighted as a priority already in the report of the 2016 UN Statistical Commission, and became the biggest focus of the IAEG-SDG's work over the past three years. The modalities for this process of methodological development were not codified at the beginning of the IAEG-SDG process, and have changed significantly over time.

Another key challenge that countries and custodian agencies have faced in global SDG reporting is data validation. Multiple decisions by the UN Statistical Commission have urged custodian agencies to produce SDG indicators based on data produced by national statistical systems. In fulfilling their global monitoring role, however, custodian agencies may need to adjust national data to ensure their compliance with global statistical standards and their international comparability, which is a prerequisite for compiling regional and global estimates. For the same purpose, custodian agencies may also estimate country-specific values of SDG indicators when national official data do not exist, are incomplete, or are of poor quality. In all these cases, custodian agencies are requested "to consult with concerned countries to produce and validate modelled estimates before publication"[4]. Specific guidelines on global data flows between countries and custodian agencies, however, were not available in the initial stages of the process. To address this gap, the IAEG-SDG prepared a series of documents[5] defining "guidelines of how custodian agencies and countries can work together to contribute to the data flows necessary to have harmonized statistics" for monitoring the SDG targets.

While the work done under the overall umbrella of the IAEG-SDG has contributed to clarifying many of the issues involved in the validation of methods and data for SDG indicators and the governance of the international statistical system, numerous issues remain to be solved. This article highlights some of the main challenges in the practical implementation of the guidelines, identifies areas in need of further guidance from the IAEG-SDG and provides some proposals aimed at improving the global SDG reporting process.

2. Challenges for Methodological Validation

The procedure to validate methods was instituted at the third IAEG-SDG meeting in March 2016, when the newly introduced "Tier classification"

---

[4] UNGA resolution 71/313, Op. Cit.

[5] "Guidelines on Data Flows and Global Data Reporting for Sustainable Development Goals", Op. Cit; "Criteria for the Implementation of the Guidelines on Data Flows and Global Data Reporting for Sustainable Development Goals", Op. Cit.; and "Best Practices in Data Flows and Global Data Reporting for the Sustainable Development Goals", Op. Cit..

resulted in the categorization of one third of the SDG indicators in the "Tier III category", meaning that their methodology was still not internationally agreed. Initially though, the exact criteria for validating a new method and thus approving the reclassification of a Tier III indicator, were not known. Only at the fifth IAEG-SDG meeting in Ottawa (March 2017) were "criteria for Tier III indicator reclassification" issued, which were manifestly more stringent than the hitherto unwritten criteria previously used, particularly in their stipulation that pilot tests had to have taken place in a regionally balanced sample of countries[6].

The increased severity of the methodological validation process suggests that many of the indicators reclassified in the initial phase of the process, would possibly not have qualified for reclassification at a later stage. Indeed, a few countries have recently questioned the reclassification of some SDG indicators and have requested that their methodology be once again reviewed. For now, the IAEG-SDG has discreetly distanced itself from such requests, under the UN Statistical Commission's constant drive to "accelerate the development" of Tier III indicators. There is a strong likelihood, however, that requests to review the indicator methodology itself may re-emerge during the 2020 Comprehensive Review process. Greater clarity on how the IAEG-SDG plans to deal with such requests would be beneficial both for countries and custodian agencies.

Greater clarity on the Tier III reclassification criteria themselves would also be beneficial. For instance, the third criterion, regarding "how the methodology has become an international standard", presents a strong paradox: if another international institution has already approved the methodology as an international standard, then what exactly is the role of the IAEG-SDG in this regard? Is it to act as an appellate body and either confirm the decision or reject it? Fundamentally, therefore, the question is whether the IAEG-SDG is in fact the only body that can decide when an SDG-related methodology becomes an international standard. Replying to this question would need to take into account that many UN agencies are governed by intergovernmental bodies that already have a prerogative to approve statistical methods and standards in their area of expertise. It would also need to deal with the fact that many new international definitions and standards have cascading effects beyond the strict confines of the relevant SDG

[6] The criteria for the reclassification of Tier III indicators adopted by the IAEG-SDG in March 2017 are the following: 1) The National Statistical System should be involved in the methodological development of new indicators; 2) New methods should be pilot-tested in a sufficient number of countries with comprehensive regional coverage (at least 5 countries, 1 per region); 3) Information should be provided on how the proposed methodology has become an international standard; 4) Comprehensive metadata should be provided to UNSD according to an agreed template.

indicator. In such cases, it may be necessary to consult a body with an even higher level of country representation, i.e. the UN Statistical Commission.

Custodian agencies have also faced important difficulties in pilot testing new methods/indicators. Although this criterion implies the necessary involvement of countries, in practice, the responsibility for testing new methods/indicators is left entirely to custodian agencies. Pilot testing is a very complex endeavour that requires huge investments in resources and time to verify the feasibility of new methods/indicators, especially for less advanced statistical systems and when new survey tools are needed. In such cases, it has often been difficult for custodian agencies to identify countries willing to participate in pilot tests, as this also implies some additional burden on countries themselves. As a result, custodian agencies have found themselves struggling to fulfil what is a key requirement for obtaining the reclassification of a new indicator. A more proactive role of the Secretariat and the IAEG-SDG Co-Chairs is thus needed in promoting country collaboration in pilot testing at the moment in which the work plan for a Tier III indicator is approved.

Looking beyond the specific list of criteria for the validation of methods, this process has brought with it a whole other set of challenges. The approval of dozens upon dozens of new SDG indicator methodologies has immediately clashed with the crude reality of the existing data collection capabilities of most National Statistical Systems (NSSs). In a situation where most NSSs were already struggling to produce even the most elementary data, there was no easy answer to the question of how countries would actually be producing new SDG indicators. One of the often-cited possibilities was to embrace the "data revolution for sustainable development", which suggested a radical enlargement of the portfolio of potential data sources, particularly by including big data and geospatial information. However, this produced a serious conundrum for the IAEG-SDG. On the one hand, the UN Statistical Commission did not tire in reminding that "the compilation of global indicators should be based to the greatest extent possible on national official statistics provided by countries"[7], on the other hand, expecting already overburdened national official statistics to produce all the necessary data for hundreds of new SDG indicators was clearly not feasible.

To try to solve this conundrum, the 2016 UN Statistical Commission report recommended that "when other sources and methodologies are used, they will be reviewed and agreed by national statistical authorities and presented in a transparent manner". Nevertheless, in practice, a multitude of countries have refused to authorize the use of data produced outside the NSS, even when the approved methodology of the relevant indicators explicitly foresaw this possibility as an interim measure, meanwhile the NSS grapples with how

[7] 47/101 (I), Statistical Commission, Report on the forty-seventh session, 8-11 March 2016.

to produce the indicator with official sources. The main reason cited by those countries who refuse to authorize the publication of country estimates is the non-official nature of the data source. The result is that for those countries, no country value is published, as the NSS itself has not yet produced the indicator. Such an approach effectively presents custodian agencies with a catch 22 situation: they are obligated to request countries' authorization, due to the use of non¬official sources, yet most countries deny the authorization precisely because of the non-official nature of the data source[8].

Another problem with the UN Statistical Commission's recommendation was that, at the time, no agreed mechanism existed for such a review, not even a set of key principles able to give guidance on fundamental questions such as what to do in case countries did not respond to custodian agencies' solicitations or outright rejected the estimate proposed. This was clearly not a satisfactory situation, as it meant that even SDG indicators with approved methodologies could go unreported by the majority of countries. Such an outcome would evidently hamper the ability to monitor progress toward the SDGs and risked undermining the credibility of the entire edifice of the 2030 Agenda's mutual accountability mechanism, grounded on the SDG indicator framework. In the next section, this article will review the main guidelines that were developed to address this gap, and highlight some of their key shortcomings with regard to data validation.

3. Challenges for Data Validation

Acutely aware of the serious implementation challenges posed by the global reporting process, the UN Statistical Commission in 2017 instructed the IAEG-SDG to "develop guidelines on how custodian agencies and countries can work together to contribute to the data flows necessary to have harmonized statistics". The resulting "Guidelines on Data Flows and Global Data Reporting", developed in consultation with the Committee for the Coordination of Statistical Activities (CCSA), are a critical document advancing consensus and collaboration among national statistical systems and custodian agencies. However, they also have several limitations. Firstly, some of the provisions in the Guidelines can be questioned as they may hamper, rather than foster, consensus and collaboration between countries and custodian agencies. Secondly, even where the Guidelines are more specific, many countries are still opposed to their practical implementation, which can put custodian agencies in difficult situations. Thirdly, they are simply "guidelines", mainly outlining general principles, and despite additional supporting

[8] The CCSA has developed recommendations and collected best practices on the use of non-official data for international statistics, see Op. Cit.

documents subsequently produced[9] to clarify implementation modalities, there is still no specific mechanism foreseen for data validation, in case the estimates have not been produced by the national statistical system.

a. *Provisions that may hamper collaboration between countries and custodian agencies*

One of the most controversial provisions in the Guidelines is included in paragraph 24.c), which foresees that "if a country disagrees with the methodology used to produce the internationally comparable indicators or the indicator value itself"… and a mutually agreeable solution cannot be found, then the country data for the indicator should not be published. This provision is problematic on a number of counts, but it also marks a clear instance in which the validation of methods and data can be conflated. Firstly, in principle, individual countries should not be in a position to question the methodology of SDG indicators, which is a prerogative of the IAEG-SDG, through the Tier classification system. Secondly, even if we assume a broad reading of the term "methodology" to also include "selection of a data source", then this raises another question: is the purpose of the validation procedure to provide a country with the opportunity to question the data source per se, or should it rather be to check that the internationally established methodology has been properly applied in calculating the country estimate? In the authors' view, the latter alternative is correct, which is also corroborated by a preceding provision under article 23.b) that suggests that:

"The decision on whether to utilize the data set [provided by an entity outside of the NSS] shall…take into account the professional and scientific independence of the data provider, the use of scientific methods and impartiality, while also keeping in mind the two important principles of country ownership of SDG monitoring and the value of internationally comparable data and statistics".

The third reason that provision 24.c) is problematic is that it foresees the possibility for a country to disagree with the indicator value itself. This effectively means that regardless of whether a methodology is internationally approved, or whether the non-official source is properly vetted, or whether the country estimate has been produced in accordance with the established methodology, a country could still veto the publication of the estimate. As there are apparently no statistical reasons left to claim, countries may refuse the publication of country estimates presumably on the basis of non-statistical reasons. Custodian agencies are further compelled to follow the same procedure "for subsequent years until the country is able to compile the indicator according to international standards and definitions". This means that even if a country has validated an estimate produced by a custodian

[9] Op. cit., 4

agency in one year, the custodian agency should validate new estimates every year even though they may have been produced using the exact same methodology and data source. Once again, one is easily left wondering on what grounds the validation could be denied one year after it has been granted in the past, all else being equal.

The fourth reason that this provision is problematic is the end result that it foresees, i.e. the nonpublication of data in case a mutually acceptable solution has not been found and a country still disagrees with the value. Effectively, this represents a severe limitation on the autonomy and independence of international organizations, which have been assigned the function, by member countries, to collect and disseminate statistics in their relevant domains. Disallowing international organizations to publish their estimates, especially when a country has none of its own, also blindfolds the international community, which in such cases may not have any other instrument to objectively assess a country's situation, or to compare national figures that may have been influenced by nonstatistical considerations. Without independent international statistics, efforts to improve transparency and accountability worldwide are undermined. It is for these reasons that the CCSA strongly recommended that in case a mutually acceptable solution were not found, an international organization could still publish these country estimates under certain conditions (together with the national estimate, when available). This recommendation, though, was eventually rejected by the IAEG-SDG.

This outcome becomes even more controversial as in many countries the independence of the NSO may be vulnerable and its capacity to produce objective estimates of SDG indicators may be challenged, especially when these indicators are politically sensitive. There are known cases where the NSO was not even consulted in the preparation of the national SDG indicator framework. Even where the NSO enjoys a higher degree of autonomy, it still struggles to fulfil the role of "coordinator of the national statistical system". This role may not be formally recognized in national legislation or it may be difficult to implement in practice. Communication between the NSO and other data producers may be an issue and there may not be appropriate mechanisms of coordination and data transmission across data producers. NSOs are often not aware of already existing data flows to custodian agencies, especially for non-statistical indicators. More importantly, NSOs may lack the capacity and experience to certify the quality of datasets produced by other institutions (public or private)[10]. An often-cited reason for not validating estimates is that "we [the NSO] are not in a position to validate these estimates". There is therefore a need to develop systematic data transmission

[10] MacFeely, S. (2018). Op. Cit.

and validation mechanisms at national level and to build the capacity of NSOs to fulfil their role as coordinators of national monitoring in practice[11].

b. *Specific provisions of the Guidelines that are not often followed in practice.*

The Guidelines contain other provisions that are not overtly controversial, but may become challenging when countries and/or custodian agencies attempt to implement them. One such provision is 22.k), which describes the situation where a custodian agency has contacted the designated national focal point but no response has been provided within a reasonable timeframe. In such a case, a non-response "will be taken as agreement with the statistics shared by the custodian agency". Many countries, however, take issue with this provision, which is perceived by them as a way for a custodian agency to bypass the authority of the NSO. Despite the explicit provision therefore, a number of countries have transmitted formal complaints to custodian agencies about this practice, which has led many organizations to self-censor themselves publishing only the estimates of those countries that have explicitly validated them. The consequence is a significant drop in the country coverage for the indicator.

c. *The absence of a specific mechanism for data validation*

The Guidelines may have several inherent shortcomings in their various provisions, but one of the key challenges that countries and custodian agencies have faced in recent years is the absence of a specific mechanism for data validation. This was, and still is, a major constraint, despite the fact that two additional supporting documents on "criteria for implementation" as well as "best practices" were subsequently produced[12] to facilitate implementation. Neither the Guidelines nor any of these supporting documents attempt in any way to suggest particular modalities for how parties should go about the procedure. As a result, different custodian agencies have hitherto had to "go it alone", effectively improvising new procedures and tools. This creates two types of inefficiencies: unnecessary multiplication of work as different custodian agencies try to resolve the same problem; and increased bewilderment and confusion among recipient countries, who are confronted with varying approaches depending on the custodian agency. There is clearly a need to find pragmatic solutions, as data validation is a resource-demanding exercise for both countries and custodian agencies. At the same time, finding a more efficient mechanism for data validation could improve data quality as well as provide an opportunity to find solutions to the decades' old problem

[11] Another symptom of NSOs' struggle with assuming a coordinating role is that less than 60 percent have identified an SDG focal point, "expected to respond to requests from custodian agencies in a timely manner and facilitate the coordination of data transmission within the NSSs for global reporting on the SDGs".

[12] Op. cit., 4

of having different data for the same indicators available in countries and in databases of international organizations.

4. Proposal for a global data validation process

To address this situation, the authors has developed a simple yet effective proposal for a standardized, coordinated approach for data validation among custodian agencies. The proposal is based on two distinct components: a) privileged access to the Global SDG Database that acts as a repository of national estimates awaiting country validation; and b) a web survey for each custodian agency with which countries can provide their feedback in case of disagreement with the methods/estimates, as well as information on the availability of alternative national data and related data sources. This proposal would improve the data validation process in a number of aspects.

Firstly, it would provide a single location where countries could review and validate the country values of all SDG indicators within a set timeframe. Such a repository would eliminate the need for different custodian agencies to come up with their own method of data sharing, and would constitute a common interface that all countries would quickly become familiar with. Countries would be informed that new data are uploaded and would be given login credentials and a password for a privileged access to the unpublished SDG data. They would then be allowed 30-40 days for their review. The schedule of this validation could be synchronized with the February deadline for the transmission of new data for the global SDG database: custodian agencies would thus need to upload their country estimates by an agreed date, so as to provide countries with a reasonable timeframe for reviewing the data and if needed, interact with the custodian agency. The validated estimates could then simply be automatically relocated to the public section of the global database and thus formally released.

Secondly, it would provide a standardized on-line survey tool to facilitate direct communication between countries and custodian agencies. This second component, which incorporates elements of a similar UNICEF proposal submitted to the 9th IAEG-SDG, would allow countries to provide feedback only if and when disagreements on the country values/methods emerges. It would also allow starting a negotiation process whose outcomes would be recorded for future reference.

Such an approach would be in line with the UN Statistical Commission's reiterated call for better coordination, a strengthened Secretariat role, and a more efficient UN Statistical System overall. The proposal will be formally submitted to the CCSA and IAEG-SDG at their next sessions.

References

1. Committee for the Coordination of Statistical Activities (2016): Case Studies: Using Non-official Sources in International Statistics, Special issue paper.
2. Committee for the Coordination of Statistical Activities (2017): Guiding Principles of Data Reporting and Data Sharing for the Global Monitoring of the 2030 Agenda for Sustainable Development - Version 1.
3. MacFeely, S. (2018): The 2030 Agenda. An Unprecedented Statistical Challenge, International Policy Analysis, Friedrich-Ebert-Stiftung.
4. United Nations General Assembly (2017), Resolution 71/313.
5. United Nations Statistical Commission: Report on the forty-seventh session, 8-11 March 2016.
6. United Nations Statistics Commission (2018): Guidelines on Data Flows and Global Data Reporting for Sustainable Development Goals. Report prepared by the IAEG-SDG. Background document, 49th Session, 6–9 March 2018.
7. United Nations Statistics Commission (2019): Criteria for the Implementation of the Guidelines on Data Flows and Global Data Reporting for Sustainable Development Goals. Report prepared by the IAEG-SDG. Annex I, 50th Session, 5–8 March 2019.
8. United Nations Statistics Commission (2019): Best Practices in Data Flows and Global Data Reporting for the SDGs", Report prepared by the IAEG-SDG, Background document, 50th Session, 5–8 March 2019.

The role of combining national official statistics with global monitoring to close the data gaps in the environmental SDGs

Jillian Campbell[1], Jay Neuner[2], Dilek Fraisl[3], Linda See[3], Steffen Fritz[3], Jessica Espey[2], Angela Kim[1]
[1]UN Environment Programme
[2]Sustainable Development Solutions Network TReNDS
[3]International Institute for Applied Systems Analysis (IIASA)

Abstract

The Sustainable Development Goals (SDGs) have elevated the profile of the environmental dimension of development and how we monitor this dimension. However, they also present a challenge to statistical systems in terms of how methodologies and data collection is developed, implemented and reported. For the environmental dimension of the SDGs, there is too little data to formally assess the status of 68% of the environment-related indicators (UN Environment, 2019a).

Many environment-related indicators were not historically part of the purview of national statistical systems and did not have a methodology or data collection system in place prior to the adoption of the SDG indicator framework. In the original classification of the SDG indicators in 2016, 50% of the environment-related SDG indicators did not have a methodology (i.e. Tier III) (UN, n.d.). As of May 2019, this has dropped to 28%; however, socioeconomic indicators in the SDGs are far more developed with only 7% of non-environmental indicators not having a methodology (UN, n.d.).

Methodologies and mechanisms to collect country-level data are being developed by international agencies. However, many countries currently do not have national systems in place for monitoring these environmental indicators and thus there is a risk that much of the environmental dimension of development cannot be captured by using reporting mechanisms which only include national official statistics. For many of these indicators, UN Environment is exploring new data sources, such as data from citizen science and global modelling to fill data gaps. Expanding globally-estimated or modelled data to cover environment-related SDG indicators could build the foundation for a digital ecosystem for the planet, which would provide a basis for developing integrated analysis and insights.

Keywords

Sustainable Development Goals, indicators, environment, tier 3, modelling

1. Introduction

The use of comparable data, which follow harmonized definition and standards, for financial and economic decisions dates to initial systems of writing and trade (Carmona and Ezzamel, 2007). By comparison, the work to measure, account for and value the environment is quite recent with the first attempt of a global environment statistics framework, the Framework for the Development of Environment Statistics published in 1984, the first environmental accounting framework, the System of Environmental Economic Accounting published in 2012, and the initial establishment of the Group on Earth Observation in 2005 (Daguitan et al., 2019, UN, 2014).

The 2030 Agenda for Sustainable Development (2030 Agenda), and its 17 goals, 169 targets and 244 indicators, provides a transformative framework for development which aims to transcend the pursuit of siloed interventions and monitoring (UN, 2015a; UN, 2017). At the core of the 2030 Agenda is an attempt to identify the most pressing issues for global development while keeping within planetary boundaries (Randers et al., 2018): How can natural resources be sustainably managed while at the same time providing food, energy and water for the growing global population? What are the underlying governance and partnership requirements? What is the interaction between human health and the environment? How can we protect biodiversity while still achieving economic growth?

However, in order for the SDG framework to be useful for answering these key questions, data, analysis and science for each of the 244 SDG core development indicators are required. There are still gaps in definitions, methodologies and standards for environmental data, geospatial data and statistics. The data gaps are even more pronounced when trying to understand specific local contexts or the interactions between gender, poverty and vulnerable populations and the environment (Daguitan et al., 2019; UN Environment, 2019b).

2. Methodology

This analysis is based on the 244 SDG indicators which have been adopted for global monitoring by the UN Statistical Commission (UN, 2017b). This indicator list includes some duplication of indicators which are listed under multiple targets and thus there are 232 unique indicators; for the purpose of this analysis, all 244 indicators were included as this provides a method for understanding the availability of data for assessing each of the 169 targets and 17 goals. The SDG Tier Classification was developed by the IAEG-SDGs and endorsed by the UN Statistical Commission and can be summarized as: Tier I: clear, established methodology with 50 percent of data available in every relevant region; Tier II: clear, established methodology, but a lack of data; and

Tier III: no agreed methodology or methodological testing is still underway" (UN, n.d.).

There is no agreed definition of which indicators and targets should be included in the environmental dimension of sustainable development. Different definitions of environment-related have been used in different analysis. The environmental-dimension of development often has focused on specific goals such as climate change (SDG 13), water (SDG 6), energy (SDG 7), consumption and production (SDG 12), land (SDG 15) or oceans (SDG 14) (ISSC, 2015; Lim et al., 2018). The analysis in this paper is based on a classification of SDG indicators which was included as a background document for a committee of the UN Environment Assembly (UN Environment, 2018) and in the 6th Global Environment Outlook (UN Environment, 2019a). This list was then categorized into four indicator types: indicators related to (1) mechanisms, enabling environment or policy; (2) change in behaviour or consumption or production patterns; (3) environmental state and trends; and (4) linkages between people and the environment (access to natural resources, vulnerability to climate change and disasters, environmental mortality, etc.).

To assess data availability, the country, regional and global data in the SDG global database was used. For indicators lacking a global or regional aggregate, an algorithm developed by UN Environment for aggregating data was used (UN Environment, n.d.). The aggregation procedure is only completed if at least 30% of the total population, area, GDP or countries (depending on the weight of the indicator) have a data point for the year being aggregated and 70% must have either a valid data point or a data point within 5 years which can be used for extrapolation or interpolation. For each country, region and at the global level, a determination on sufficient or insufficient data was based on the results of this algorithm or of country-level data as follows: at least one data point for 2010 or later and at least one other data point between 2000 and 2018 for trend analysis.

Table 1. List of environment related SDG indicators by type

Type	List of environment-related SDG indicators
Behavior and SCP	2.4.1, 6.4.1, 6.4.2, 7.1.2, 7.2.1, 7.3.1, 8.4.1, 8.4.2, 9.4.1, 11.3.1, 11.6.1, 12.2.1, 12.2.2, 12.3.1, 12.4.2, 12.5.1, 15.7.1 and 15.c.1
Environmental State or Trend	2.5.2, 6.3.2, 6.6.1, 11.6.2, 14.1.1, 14.3.1, 14.4.1, 15.1.1, 15.3.1, 15.4.2 and 15.5.1
Mechanisms, enabling environment and policy	1.4.2, 1.5.3, 1.5.4, 2.5.1, 4.7.1, 5.a.1, 6.5.1, 6.5.2, 6.a.1, 6.b.1, 7.a.1, 7.b.1, 11.3.2, 11.4.1, 11.b.1, 11.b.2, 11.c.1, 12.1.1, 12.4.1, 12.6.1, 12.7.1, 12.8.1, 12.a.1, 12.b.1, 12.c.1, 13.1.2, 13.1.3, 13.2.1, 13.3.1, 13.3.2, 13.a.1, 13.b.1, 14.2.1, 14.5.1, 14.6.1, 14.a.1, 14.c.1, 15.1.2, 15.2.1, 15.4.1, 15.6.1, 15.8.1, 15.9.1, 15.a.1, 15.b.1, 16.8.1, 17.6.1, 17.7.1, 17.9.1 and 17.14.1
People and the Environment	1.5.1, 1.5.2, 3.9.1, 3.9.2, 3.9.3, 6.1.1, 6.3.1, 8.9.2, 11.2.1, 11.5.1, 11.5.2, 11.7.1, 13.1.1 and 14.7.1

3. Result

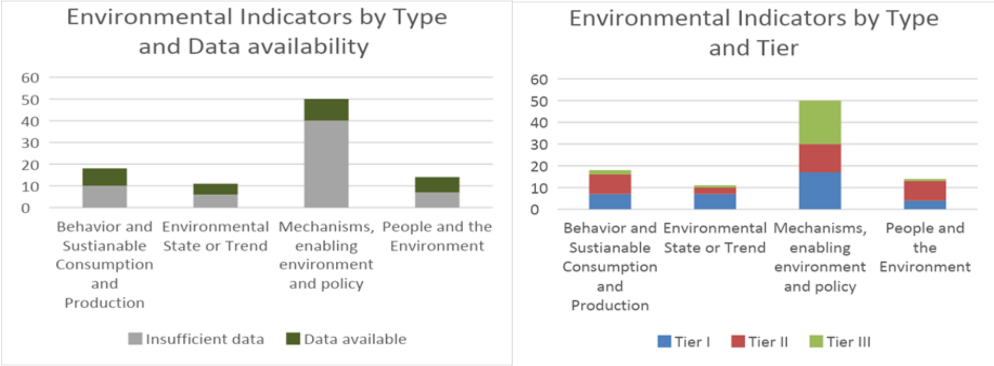
The IAEG-SDGs was established in 2015 to develop the SDG indicator framework (UN, 2015b). In order to do so, the IAEG-SDGs conducted an open consultation involving a wide range of stakeholders from across government, civil society, academia and regional and international organizations (UN, n.d.). From a conceptual perspective, the SDG indicators were developed in order to capture an ideal, ambitious monitoring framework for development, including environmental state, trends and impacts. However, from a practical perspective, the SDG indicator framework took into account an interest in increasing synergies with existing processes (e.g. the Millennium Development Goals, UN Framework Convention on Climate Change, etc.) in order to reduce the reporting burden for countries, increase feasibility and capitalize on prior experiences of national statistical offices (UNSC, 2016). Despite the interest in using existing indicators, when the original tier classification was developed in 2016, 50% of environment-related SDG indicators were classified as Tier III as compared to 28% of the remaining indicators. There has been significant progress to develop methodologies and reporting mechanisms for the SDGs and as of May 2019, 26% of the environment-related SDG indicators are Tier III compared to 7% of the other indicators (Table 2).

Table 2. SDG indicators by Tier, 2016 and 2019

	Percentage of Tier I		**Percentage of Tier II**		**Percentage of Tier III**	
	2016	***2019***	***2016***	***2019***	***2016***	***2019***
Environment-related SDGs	26%	23%	24%	37%	51%	26%
All other SDG indicators	53%	51%	19%	42%	28%	7%

Country level reporting depends on various factors like the capacity of national statistical offices, data availability and political interest. Most indicators rely directly on national capacity and reporting of national data. However, a few use international estimates and models for reporting and gaps filling such as SDG 8.4.2 on material flows and SDG 6.6.1 on water-related ecosystems. Holistic measurement of the environment is complicated by a lack of existing globally-agreed methodologies related to specific SDG targets and the fact that many statistical offices do not have experience in compiling environment statistics or environmental economic accounts (UN, 2015c) thus global modelling provides an opportunity to fill data gaps. Out of the 244 indicators in the SDG framework, less that 5% (11 indicators) are related to environmental state and trends (Figure 1).

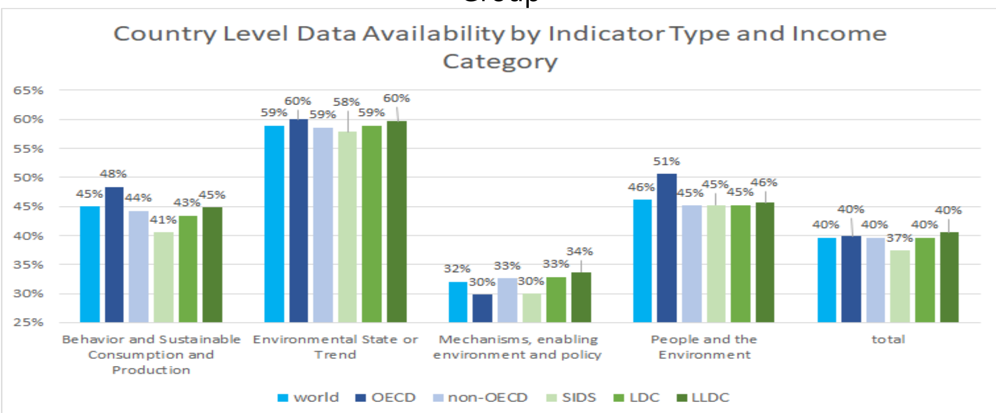
Figure 1: Environmental Indicators (a) Tier and (b) Data Availability



3a. Country level data availability

Analyzing country-level data availability can help identify gaps in national capacity or highlight areas where political interest may be lacking. Overall, there is about a 40% country-level data coverage for the 93 environment-related indicators (including Tier III indicators) (Figure 2). While the biggest differences in country-level data availability are based on the indicator types, there are also some differences in reporting level based on country categories. OECD countries have higher levels of data availability than non-OECD countries for all indicator categories except for “Mechanisms, enabling environment and policy” indicators. On the other hand, SIDS have the lowest levels of country-level data reporting across all indicators. The biggest gap being for indicators in the “Behavior and Sustainable Consumption and Production” category where country-level reporting in SIDS countries (41%) is 4% lower than the global average (45%) and 7% lower than OECD countries (48%) (Figure 2). SIDS small size, remoteness, narrow resource and exposure to global environmental change may be factor to their low levels of country-level reporting.

Figure 2: Country Level Data Availability by Indicator Category and Country Group



3b. Citizen science for monitoring the environmental dimension of the SDGs

New data sources beyond national statistical data sets have yet to be fully exploited for the SDG reporting process. In addition to non-traditional data streams such as Earth Observation and big data analytics, citizen science also has the potential to be used for SDG monitoring. Citizen science is the involvement of citizens in scientific research, from data collection up to higher levels of scientific contribution such as data analysis and hypothesis generation (Bonney et al., 2009). To date, numerous citizen science initiatives have already generated a considerable amount of data in the environmental domain, some of which are being used in national and global biodiversity monitoring frameworks such as the Convention on Biodiversity (CBD). For example, as much as 50% of species occurrence data in the Global Biodiversity Information Facility (GBIF) has been contributed by citizen science projects (Copas et al., 2016).

Data sets sourced from citizen science are also currently contributing to a few of the environmental SDG indicators, but there is still considerable scope for expansion. For example, BirdLife International, which has a huge network of volunteers, compiles the bird taxonomic component of the International Union for Conservation of Nature and Natural Resources (IUCN)'s Red List Index of Threatened Species (covering SDG indicator 15.5.1). For SDG 15.1.2 and 15.4.1 on protected areas, data from BirdLife is included in the development of the Key Biodiversity Areas (KBAs) used as the denominator for the indicator. Hence citizen science is currently contributing to SDG reporting to an important, yet still very limited, degree.

Citizen science may also help to validate or provide sub-national data for indicators that have been at the national level. An example is SDG indicator 14.1.1 on floating marine debris, there are numerous citizen science projects with different purposes, currently, UN Environment, is working with the citizen science community in order to find ways forward, e.g. through establishing acceptable protocols for data collection demonstrated through pilot projects in selected countries or as potential validation for the global modelling approach proposed for this indicator.

4. Discussion and Conclusion

Faced with the SDGs' 169 targets and 244 indicators countries are stymied in both action and monitoring of SDG progress. This is no more true than in regards to environmental indicators which are outpaced among all indicators in globally-agreed methodologies and, regardless of a country's stated priorities, will impact progress on other targets (Nilsson, 2017; Griggs et al., 2017; Weitz et al., 2018).

In order for the SDGs to be an effective tool for galvanizing action, there is a need to urgently uplevel monitoring and build a digital ecosystem for the planet (Jensen and Campbell, 2019). This would provide a foundation to fill data gaps for targets lacking data disaggregated by crucial specificities and disparities (including by gender, income quintile, education, disability and other vulnerable groups) and even lacking data at all. Such an ecosystem would also help build the tools and capacity to conduct integrated analysis that can be used to generate actionable insights (UN Environment, 2019a; Data Revolution Group, 2014). Global data products and global modelling can form a basis for filling data gaps and ensuring that there is some data for all countries. These products also can demonstrate environmental concerns that should be further monitored through high-frequency or high-resolution remote sensing or through in situ collection or other means. Existing global data products like those for SDG target 6.6 (https://www.sdg661.app/) and SDG targets 8.4 and 12.2 (material flows) can provide an example of how such products can and have already been accepted for official SDG monitoring (UN, 2018).

Complementing global products and global modelling is citizen science, which can also help fill data gaps in some countries. Citizen science is already contributing to SDG indicators, e.g. 15.5.1, but further efforts are needed to investigate the potential of citizen science for other environmental SDG indicators, particularly Tier III. Citizen science projects and representatives from the citizen science community need to be brought into the high-level discussions on methodology development and data collection to explore the best way forward. Making progress on one indicator, e.g., 14.1.1. on marine plastic debris, may be provide a blueprint for integrating citizen science in other places.

In economic policy, a single headline indicator – Gross Domestic Policy – is used as a flag for economic progress. There has been some discussion on developing an equivalent indicator, with corresponding sub-indicators, for measuring sustainable development (Ekins, et al. 1999). Building out the portfolio of country-level data on globally-agreed environmental indicators could support consensus on a headline indicator on the environmental dimension of development. However, building the capacity of countries to collect the underlying basic data and better utilizing existing data – including from non-traditional data sources such as satellites, drones and citizen science – are at the crux of being able to monitor the SDGs and develop integrated analyses or indices (Holloway et al., 2018).

The current limitations on understanding environmental indicator progress illustrate a "catch-22 situation" when it comes to achieving the SDGs for the sustainability of the planet and for humanity: We use existing data to identify priorities, but priorities for data collection are identified on the basis of which topics are priorities. It is the role of the international statistical community to disrupt this dynamic, leverage the commitment to monitoring the SDGs to better monitor across all sustainable development issues and to ensure that development occurs within planetary boundaries.

References

1. Bonney R, Cooper CB, Dickinson J, et al (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience* 59:977–984. doi: 10.1525/bio.2009.59.11.9
2. Carmona, S. and M. Ezzamel (2007). Accounting and accountability in ancient civilizations: Masopotamia and ancient Egypt. *Accounting Auditing & Accountability Journal* 20:177-209.
3. Copas, K., Legind, J. and Masinde, S. (2016). Citizen science contributions to EBVs: GBIF.org as a case study. *First International ECSA Conference*, Berlin, Germany.
4. Daguitan, F, C. Mwangi, M. Tan, G. Clark, D. Cooper, J. Donovan, S. Gutierrez, N. Kruglikova, P. Lehohla, J. Seager, W. Sonntag and A. Patel (2019). Future Data and Knowledge Needs in *Global Environment Outlook - GEO-6: Healthy Planet, Healthy People*. UN Environment, Nairobi.
5. Data Revolution Group (2014). A World that Counts: Mobilising the Data Revolution for Sustainable Development. *United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG)*. http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf
6. Ekins, P. and Simon, S. (1999) 'The Sustainability Gap: a Practical Indicator of Sustainability in the Framework of the National Accounts'. *International Journal of Sustainable Development* 2(1):32-58.
7. Griggs, D., M. Nilsson, A. Stevance and D. McCollum (eds) (2017). *A guide to SDG interactions: from science to implementation.* International Council for Science (ICSU), Paris.
8. Holloway, J., K. Mengersen and K. Helmstedt (2018). Spatial and machine learning methods of satellite imagery analysis for Sustainable Development Goals. OECD, Paris.
9. Jensen, D. and J. Campbell (2019). Discussion paper: The Case for a Digital Ecosystem for the Environment: Bringing together data, algorithms and insights for sustainable development. *Science Policy Business Forum*.
10. Lim, M. M. L., P. Søgaard Jørgensen, and C. A. Wyborn (2018). Reframing the sustainable development goals to achieve sustainable development

in the Anthropocene—a systems approach. *Ecology and Society* 23(3):22.

11. Nilsson, M. (2017). Important interactions among the Sustainable Development Goals under review at the High-Level Political Forum 2017. SEI Working Paper 2017-06. *Stockholm Environment Institute*, Stockholm.
12. Randers, J. and J. Rockström, P.E. Stoknes, U. Golüke, D. Collste, S. Cornell (2018). Transformation is feasible: How to achieve the Sustainable Development Goals within Planetary Boundaries. Stockholm Resilience Centre.
13. UN Environment (n.d.). UNEP Live Global database aggregation methods. https://uneplive.unep.org/media/docs/graphs/aggregation_methods.pdf
14. UN Environment (2018). Committee of Permanent Representatives Subcommittee meeting. 20 September 2018. Nairobi. https://www.unenvironment.org/events/subcommittee-meetings/committee-permanent-representatives-subcommittee-meeting-14
15. UN Environment (2019a). Measuring Progress: Towards Achieving the Environmental Dimension of the SDGs.
16. UN Environment (2019b). *Gender and environment statistics: unlocking information for action and measuring the SDGs*. Nairobi.
17. United Nations (n.d.). IAEG-SDG website. Accessed June 2019. https://unstats.un.org/sdgs/iaeg-sdgs
18. United Nations (2014). *System of Environmental-Economic Accounting 2012 Central Framework*. United Nations, New York.
19. United Nations (2015a). A/RES/70/1 Transforming our world: the 2030 Agenda for Sustainable Development. *Seventieth session of the General Assembly on 25 September 2015*.
20. United Nations (2015b). E/2015/24-E/CN.3/2015/40 Report on the forty-sixth session of the Statistical Commission on 3-6 March 2015. United Nations Economic and Social Council, New York.
21. United Nations (2015c). Global Assessment of Environmental-Economic Accounting and Supporting Statistics 2014. *Forty-sixth Session of the Statistical Commission 3-6 March 2015*. https://unstats.un.org/unsd/statcom/doc15/BG-UNCEEA.pdf
22. United Nations (2017). A/RES/71/313 Work of the Statistical Commission pertaining to the 2030 Agenda for Sustainable Development. *Seventy-first session of the General Assembly on 6 July 2017*.
23. United Nations (2018). SDG Indicator metadata, 6.6.1 and 8.4.2, accessed online in June 2019. https://unstats.un.org/sdgs/metadata

24. United Nations Statistical Commission (2016). Technical report by the Bureau of the United Nations Statistical Commission (UNSC) on the process of the development of an indicator framework for the goals and targets of the post-2015 development agenda.
25. Weitz, N., Carlsen, H., Nilsson, M. and Skånberg, K. (2018). Towards systemic and contextual priority setting for implementing the 2030 Agenda. *Sustainability Science* 13(2): 531–48.

Progress in the measurement of German commercial property prices[1]

Thomas A. Knetsch
Deutsche Bundesbank, Directorate General Statistics

Abstract

The compilation of commercial property price indices is a big challenge. In Germany, substantive data gaps prevent the national statistical authority from calculating and publishing official figures. By contrast, policymakers urge for timely, reliable and comprehensive data, thereby making proposals how to aggregate and classify individual price information in order to best serve the intended policy uses. Experimental price indices according to various definitions of commercial real estate are constructed on the basis of two ingredients: the appraisals for transaction prices of houses, apartments, multi-family dwellings, office and retail buildings in 127 German towns and cities provided by bulwiengesa, a real estate consulting company, and corresponding data on floor spaces which make it possible to derive coherent weighting schemes. The overall price trends revealed by the various indices are rather similar while differences in detail can be explained by their specific compositions. Analysts may find the price indices helpful to better understand trends on German commercial real estate markets. Statisticians may acquire from this exercise further knowledge about measurement practices, as official statistics are encouraged to take steps in direction to establishing a thorough reporting on commercial real estate markets.

Keywords

Commercial property price indices; measurement; private data sources; stock weighting; Germany

1. Introduction

The demand for commercial property price statistics has increased in the aftermath of the global financial crisis as both policymakers and academic scholars urge for timely, reliable and comprehensive data to study economic, macroprudential and supervisory issues related to this segment of the real estate market. However, as stated by Diewert and Shimizu (2015), for instance,

---

[1]Address: Wilhelm-Epstein-Str. 14, 60431 Frankfurt am Main, Germany, thomas.knetsch@bundesbank.de The author thanks Julian Barnikol, Alexandra Päckert, Mario Schimmelpfennig and Patrick Schwind for excellent research assistance. The paper also benefitted from discussions with Christine Schlitzer. Of course, all remaining errors are mine. The views expressed in this paper are solely mine and should not be interpreted as reflecting the views of the Deutsche Bundesbank.

the development of commercial property price indices (CPPIs) is "one of the fields in statistics that has perhaps lagged the furthest behind" (p. 131). Some important milestones were recently reached nonetheless. International statistical institutions agreed upon a document informing about source data, conceptual frameworks and methodologies to compile CPPIs (Eurostat, 2017). The collection and dissemination of price indicators available for the G-20 countries were institutionalized within the G¬20 Data Gaps Initiative (DGI), with the Bank of International Settlements (BIS) serving as the data hub and a concrete short-term target being formulated in DGI Recommendation II.18.[2]

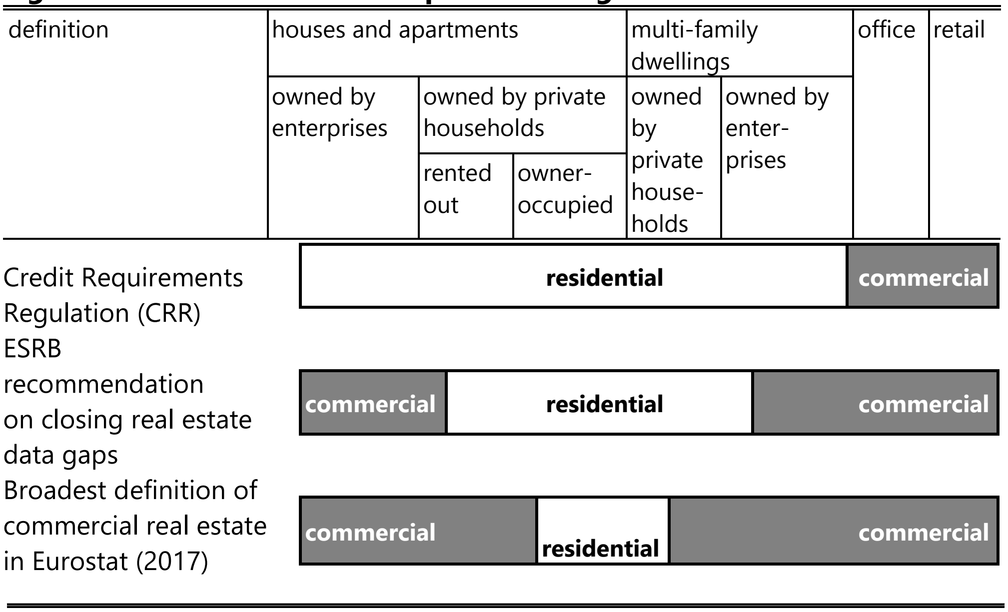
In Germany, CPPIs are currently available only from private data providers while official statistics is reluctant to make vigorous efforts aiming to ensure data provision in the short and medium run. Among the private data providers, vdp and bulwiengesa publish CPPIs which are based on sound measurement and compilation practices to a sufficient degree and reflect price developments with a broad regional coverage. The transaction-based, quarterly vdp indices shall currently be deemed most useful for analysts. The merits of the annual appraisal data from bulwiengesa is not only confined to cross-checks. Rather, as the data is available in the breakdown of 127 German towns and cities as well as for several real estate types such as office, retail, multi-family dwellings, houses and apartments, it is also a valuable source for compiling CPPIs according to several definitions of commercial real estate on an experimental basis.

This paper reports on the attempt of the Deutsche Bundesbank to compile experimental CPPIs for various definitions of commercial property in use on the basis of bulwiengesa data and coherent weighting schemes regarding object types. While office and retail buildings as well as logistics and industrial structures are unanimously considered commercial real estate, there are alternative views on whether rental housing should be classified as residential or commercial real estate. In general, the classification of rental housing may not be decided universally, as the purpose of the analysis matters. Of crucial importance is, on the one hand, the question as to whether real estate is classified according to the user or the owner perspective (e.g. Deutsche Bundesbank, 2013). Amongst the definitions in use, the one laid down in the Capital Requirements Regulation (CRR) implements the user concept and thus considers rental housing as part of residential real estate. If, by contrast, the owner perspective is taken in its pure form, only owner-occupied housing should be regarded as residential real estate, implying that commercial real estate – as it completely includes rental housing – is understood in the

[2] The target is that by 2021 G-20 economies are encouraged to provide nationally available CPPI data to BIS, be it from private sources or sourced from official statistics, the latter obviously being the preferred option.

broadest sense (Eurostat, 2017). On the other hand, for classification, it can further make a difference whether dwellings rented out to tenants are owned by enterprises or private households. In the residential-commercial delineation proposed by the European Systemic Risk Board (ESRB, 2016), the latter, labelled buy-to-let property, is considered residential property because, from a macroprudential point of view, it resembles more owner-occupied housing than other commercial property in terms of credit risks. In Fig. 1, the residential-commercial splits underlying several definitions in use are displayed.

Fig. 1 Residential-commercial split according to several definitions in use



This illustration does not cover the complete stock of real estate properties. For instance, social housing as well as industrial and logistics properties are excluded. The illustration also abstracts from buildings under construction.

While the CRR definition of commercial real estate enables to base CPPI weights on both stock and take-up of space, it is impossible to implement a transaction weighting for the ESRB delineation and the broadest-level concept without making arbitrary choices regarding the assignment of transactions from enterprises to private households and vice versa. In addition, transactions in commercial real estate markets are only incompletely available. By contrast, the real estate stock in Germany is – at least to the extent needed for the CPPI compilation intended here – recorded to a sufficient degree. In particular, the dwellings stock was completely registered by a census in 2011, making it possible to identify and extract rental housing. Data on office floor space in towns and cities is provided by bulwiengesa and, for retail space, there is official information on county level which can be used to perform estimates for towns and cities.

The paper is organised as follows. In the first part of the methodology section, the data and calculation methods applied to extract price trends for office, retail and rental housing in 127 German towns and cities are presented. The second part is devoted to explain the source data used and the choices and assumptions made to derive a coherent weighting scheme for the several definitions of commercial real estate in policy use. In the results section, the price indices for commercial real estate are shown for the total aggregate of 127 towns and cities. Finally, conclusions are drawn and potential further work in this field is sketched.

2. Methodology

The price indices for commercial real estate in 127 German towns and cities are compiled on the basis of two major components. These are, first, the price trends for office buildings, retail space, multi¬family dwellings as well as houses and apartments in each of these towns and cities and, second, the city-level data on floor space which is needed to construct weighting schemes according to the definitions of commercial real estate under consideration.

a. Price data for 127 German towns and cities

The data base RIWIS[3] maintained by bulwiengesa comprises the information needed to calculate prices for office, retail and multi-family dwellings following an income approach. According to this, the market value MV of commercial real estate is inferred using the valuation equation:

$$MV = \frac{NOI}{CAP}$$

where *NOI* is (annual) rental income less of operating costs (including maintenance, administration etc.) and vacancy costs and $CCCCCC$ is the capitalisation rate.

The data is stratified in the sense that the pricing information is available for each of the relevant object types (office, retail, multi-family dwellings) and for the each of the 127 German towns and cities considered. It results from intensive market observations consisting of surveys among realtors, investors, owners and lessees in the context of actual sales and a careful validation by bulwiengesa experts. The information is of an appraisal nature, as the raw data for individual objects are adjusted by expert judgment in order to fit the reference of prime-segment objects. The valuations thus refer to objects of constant quality over time, implying that their percentage changes can be interpreted as price changes.

[3] RIWIS is the acronym for Regionales Immobilienwirtschaftliches Informationssystem meaning regional real estate information system.

Rental housing does not only comprise multi-family dwellings but also extends in part to houses or individual apartments in condominiums. The prices of (detached) houses, terraced houses and apartments in 127 towns and cities are collected by bulwiengesa using a detailed and structured annual questionnaire among local contact persons (e.g. real estate agents, real estate experts at banks, project developers, investors etc.). The outcome of the survey are assessments of transaction prices for typical objects which are characterised in terms of size, location, age, fixtures and fittings, and other price-determining factors. The bulwiengesa experts ensure a careful data validation and form valuations which are representative for houses, terraced houses and apartments in the 127 towns and cities. As characteristics are kept constant over time, the valuations refer to objects of a constant-quality and can thus be interpreted as appraisal prices.

b. Data on the floor space of commercial real estate in 127 German towns and cities

The data on the floor space of commercial real estate is composed by three sources. For rental housing, the 2011 census information is used. Office space is taken from bulwiengesa, and retail space in the 127 towns and cities results from estimates on the basis of official county level data.

The stock of dwellings as existed on 9 May 2011 (i.e. reporting date of the census) is reported in great detail. The tabulated information suits to form the aggregates needed for the derivation of weighting schemes. One crucial breakdown is along the number of housing units in dwellings. This piece of information makes it possible to distinguish between houses (i.e. dwellings with one unit or two) and multi-family dwellings (i.e. three and more units). With the additional information about whether multi-family dwellings are condominiums or owned by single legal entities, it is able to separate out the number of apartments which can be transacted individually. Information about the legal status of the owner help distinguish between dwellings which are either in the hands of private households or private enterprises including housing cooperatives. Dwellings owned by the public sector are not considered in the calculations here due to the supposition that this could be social housing to a large extent. The census also provides information about the primary use of dwellings, yielding a distinction between owner-occupied housing and buy-to-let.[4] A categorisation of housing units in dwellings according to their floor space finally enables to account for the fact that

[4] Additional categories of minor relevance are holiday homes and vacant homes. While the former are split fifty-fifty between owner-occupied and rented out, the latter are completely assigned to rental housing as the owners of vacant homes are assumed to live somewhere else.

average apartment sizes tend to be the smaller the larger the number of units in the dwelling.

For Laspeyres-type CPPIs, a base year must be determined. It seems appropriate to choose 2011 owing to the full recording of the dwellings stock in the census year. This choice is not at odds with corresponding information available for the office and retail segments. For the floor space of office buildings in 127 towns and cities, annual time series are provided by bulwiengesa. Hence, the 2011 figures can just be taken from this source. The floor space of retail structures in 2011 can be estimated using the following official publications: retail space in 2014 broken down in 31 counties, the inhabitants of these counties and of the 127 towns and cities in 2011 and 2014.[5] The estimates are obtained by imposing the assumptions that the retail space per inhabitant in a town or city be equal to the retail space per inhabitant in the county the town or city belongs to and that these per-capita figures do not change over time. An assessment of the quality of these estimates is possible for those towns and cities for which bulwiengesa data on retailers' floor spaces in the period between 2010 and 2012 is available. This cross-check reveals a very good performance.

By summing up all categories of real estate under consideration, the floor space in the 127 towns and cities in 2011 amounts 1.2 billion square metres. Houses and apartments make up almost one-half, multi-family dwellings about one-third and office and retail more or less one-sixth. With comprising only office and retail structures, the smallest concept of commercial real estate (according to CRR) totals 0.2 billion square metres. On the basis of floor space data, office space is given a three times higher weight than retail.[6] The inclusion of residential real estate which is rented out by enterprises doubles the floor space of commercial real estate. Hence, the ESRB definition of commercial real estate includes rental housing with a share of 50 per cent. In the broadest concept of commercial real estate which includes rental housing, regardless of being owned by enterprises or private households, the floor space of commercial real estate is about 0.9 billion square metres, with multi-family dwellings contributing almost one-half, buy-to-let houses and apartments a little more and office and retail a little less than one-quarter.

[5] Four Bundesländer (North Rhine-Westphalia, Bavaria, Baden-Wurttemberg, Hesse) are divided in Regierungsbezirke due to administrative reasons. These 19 administrative units together with the remaining 12 Bundesländer form the county level considered here.

[6] In the vdp CPPI, office is given a weight of 60 per cent using information about the outstanding loans granted by mortgage banks. In the bulwiengesa CPPI, office makes up only a quarter and retail three quarters, representing the shares of the object types in tradable assets. See also Knetsch et al. (2019), p. 6.

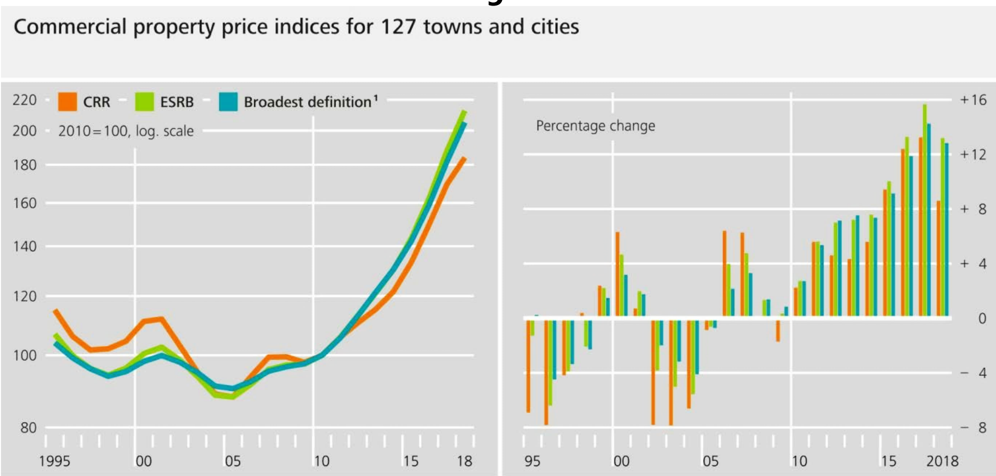
3. Result

CPPIs are compiled on the basis of town-specific price trends and weighting schemes. In particular, the percentage change of the price index according to definition v in period t,t=1,...,T, is given by

$$\Delta \mathrm{p}_t^v = \frac{1}{I \cdot 0} \sum_{i=1}^{I} \sum_{o=1}^{0} \omega_{i0}^{v} \Delta p_{iot}$$

where ω_{io}^{v} is the weight of object type o,o=1,...,O , in the town or city i,i=1,...,I, according definition v and $\Delta \mathrm{p}_{iot}$ is the percentage change in the price for object type i in town or city i in period t. It is worth noting that the same town-specific price changes are considered for rental housing owned by enterprises and private households.

Fig. 2



Sources: Bundesbank calculations based on data provided by bulwiengesa AG. **1** According to Eurostat (2017).
Deutsche Bundesbank S3IN0531.Chart

As depicted in Fig. 2, the aggregate price trends of commercial property in German towns and cities vary somewhat depending on the chosen definition. Averaged over the whole period from 1995 to 2018, the price change is 2.0 per cent per year in the case of the CRR definition. This is lowest because multi-family dwellings which experienced the strongest price increase amongst all object types are completely excluded. The CPPI according to the ESRB definition rose by 3.1 per cent per year on average. This is exactly the same rate as in the case where the broadest definition of commercial real estate is chosen. This similarity can be explained by recalling that the share of multi-family dwellings is uniformly one-half in both definitions and noting that the price trends of the remaining components (which are, in the ESRB, office and retail to the overwhelming part and, in the broadest concept, office and retail as well as houses and apartments in roughly equal parts) shared more or less the same long-run price trends. The CPPIs of the various definitions point to the same direction of change in more of 90 per cent of the years under consideration. The signals regarding the change in momentum are uniform in almost 80 per cent.

4. Discussion and Conclusion

In the paper, experimental price indices for German commercial real estate are presented and compared. The virtue of the compilation here is that the CPPIs, reflecting different definitions of commercial real estate, are coherent in terms of source data for prices and weighting schemes. In particular, the appraisals of transaction prices for office and retail structures as well as rental housing in 127 towns and cities are taken from bulwiengesa and the weighting schemes are derived from data on the floor spaces of these object types in 2011. With respect to the various definitions in use, the main conclusion is that while the CPPIs partly exhibit trends of somewhat different strength, they are very similar in terms of their general time series profiles referring, for instance, to the location of turning points and signals with regard to the direction of change or the change in momentum.

The experimental price indices help users analyse commercial property markets. However, a number of weaknesses have to be mentioned. First, the price data does not cover Germany as a whole but also 127 towns and cities. The reduced regional coverage is perhaps of minor relevance for office and retail because, in these market segments, transactions are concentrated on towns and cities. For rental housing, it is definitely a shortcoming because market activity is also present in rural areas and the trends in house and apartment prices are shown to systematically alter between the more populated and the rural areas of Germany. Second, with office, retail and rental housing, the most important object types are covered. However, at least to some part, the industrial and logistics building stock as well as hospitals and other commercial buildings may also be marketable and, thus, to be included in CPPI measurement. Third, the appraisals collected by bulwiengesa are a second-best solution. Of course, actual transaction prices would be preferred. However, disaggregate transaction-based price data is not available.[7] Fourth, prices are measured at annual periodicity. However, analysts desire price indices at quarterly frequency.

With floor-space data, a stock-based weighting scheme is implemented. With regard to definitions of commercial real estate which require distinguish properties according to use and legal status of the owner, it is only stock weights upon which price indices can be based. The 2011 census provides an optimal source for the floor spaces of rental housing. As the census does not cover non-residential real estate, there is scope for improvement in stock

[7] The vdp CPPIs are compiled on the basis of granular transaction prices. The source data, however, is exclusive to those who construct the price indices.

weighting, too. A transaction-based weighting which is often regarded as the first-best solution is currently not implementable because the data available on the number of non-residential transactions cannot be broken down to a sufficient degree.

By disclosing the current data limits for CPPI compilation, the paper may be of interest for those people who work on establishing a framework for enhancing the source data for official CPPI measurement. Given that macroprudential monitoring demands data of highest quality, this is the ultimate target and should be pursued consequently despite the short and medium-run efforts which are aimed to bridge the information gap with pragmatic intermediate solutions.

References

1. Deutsche Bundesbank (2013). Commercial property prices in Germany: conceptual considerations and statistical status quo, Monthly Report, May 2013, 53-55.
2. Diewert. W. E. & Shimizu C. (2015). A conceptual framework for commercial property price indexes. Journal of Statistical Science and Application 3, 9-10, 131-152.
3. ESRB (2016). Recommendation of the European Systemic Board on closing real estate data gaps (ESRB/2016/14). 31 October 2016.
4. Eurostat (2017). Commercial property price indicators: sources, methods and issues. 2017 edition. Publication Service of the European Union.
5. Knetsch. T. A. Schlitzer C. & Triebskorn E. (2019). The rationale behind a multi-indicator approach to real estate price analysis. Mimeo.

Progress in Japan to implement the DGI-2 on securities statistics and institutional sector accounts

Sayako Konno Mizuno, Haruko Kato
Bank of Japan, Tokyo, Japan

Abstract

With the launch of the second phase of the Data Gaps Initiative (DGI-2) in 2015, the Bank of Japan (BOJ) has initiated various measures in order to compile the financial data sets necessary for implementing recommendations II.7 on securities statistics and II.8 on institutional sector accounts. These measures utilize Japan's Flow of Funds Accounts (FFA) as the underlying data source and also explore complementary data sources.

With respect to recommendation II.7 on securities statistics, this paper introduces our initiatives to compile issuing data with breakdown information on debt securities by sector as DGI-2 intermediate target data. The BOJ plans to investigate new data sources, such as microdata or the disaggregated data of debt securities. This data allows classification of debt securities according to the currency type, interest rate, maturity (original or remaining), and issuer's market and are not available solely from the FFA.

With respect to recommendation II.8 on institutional sector accounts, the BOJ — using new data sources and methodology — looks at the feasibility of compiling the following data: (1) loans and debt securities classified by currency type and maturity (original and remaining) on both the asset and liability sides of each economic sector; (2) breakdown of foreign securities into three transaction items, i.e., foreign equities, foreign debt securities, and foreign investment trusts; and (3) breakdown of foreign direct investment into respective transaction items in alignment with the categories of the DGI-2 template. Of these issues, this paper introduces our initiatives relating to (2) above and the tentative results of foreign debt securities. We are still in the early stages in terms of achieving the recommendation; however, the preliminary results of the time-series data show some trends and features of foreign debt securities held by respective economic sectors. Towards the completion of the DGI-2 by 2021, we will continue to work on the remaining issues for both securities statistics and institutional sector accounts, while taking into consideration factors such as data quality, respondent burden, and data collection processes.

Keywords

Flow of Funds Accounts; DGI-2; Securities statistics; Institutional sector accounts

1. Introduction

Following the financial crisis of 2008, which highlighted the data gaps in economic and financial statistics, the G20 Finance Ministers and Central Bank Governors endorsed the Data Gaps Initiative (DGI) in 2009, listing 20 recommendations for future statistical improvements. Currently, G20 economies and some non-G20 FSB member jurisdictions have been working to advance the implementation of the second phase of the DGI (DGI-2) recommendations, which are due for completion by 2021. Accordingly, the Bank of Japan (BOJ) is implementing initiatives enabling better analysis of economic and financial movements. This paper focuses on our progress on the recommendations II.7 on securities statistics and II.8 on institutional sector accounts. Section 2 outlines our progress in accessing new data sources, methodology and preliminary outputs for securities statistics. Section 3 introduces our progress in compiling the institutional sector accounts with partial outputs. Finally, section 4 concludes.

2. Securities Statistics

2.1 Requirements

Recommendation II.7 on securities statistics requests reporting of issuance, holdings and from-whom-to-whom securities statistics to the BIS.[1,2,3]This section focuses on our development of issuing data corresponding to the intermediate target data for which economies are required to report 2018 data in alignment with their publicly made commitments, by May 2019 at the latest. More specifically, the intermediate target data focus on debt securities issued by residents at the nominal values which are classified by issuing sector, currency (domestic, foreign), maturity (original, remaining), interest rate (fixed, variable), and market (domestic, international).

---

[1] This recommendation requests G-20 economies *"provide on a quarterly frequency, debt securities issuance data to the BIS consistent with the Handbook on Security Statistics (HSS) starting with sector, currency, type of interest rate, original maturity and, if feasible, market of issuance. Reporting of holdings of debt securities and the sectorial from-whom-to-whom data prescribed for SDDS Plus adherent economies would be a longer term objective."*

[2]G-20 DGI templates are available from Principal Global Indicator's website (http://www.principalglobalindicators.org/?sk=E30FAADE-77D0-4F8E-953C-C48DD9D14735&sId=1452784383161).

[3] The Securities Statistics data consist of three reporting templates: (1) reporting template 1 presents debt securities issues statistics by sector, currency, maturity, type of interest rate, and by market of issuance; (2) reporting template 2 presents debt securities holdings statistics by holding sector, residence of issuer, currency, maturity, type of interest rate, and by market of issuance; and (3) reporting template 3 presents from-whom-to-whom debt securities issues and holdings statistics by residence and sector of issuers and by residence and sector of holders. Template 1 has three tables: tabl e 1.1 for stock at nominal value; table 1.2 for stock at market value; and table 1.3 for net transactions at market value.

2.2 Estimation Method

The primary challenge in compiling issuance data in securities statistics is the data constraints to compiling a breakdown of debt securities by currency, maturity, interest rate, and market. The key inputs to compile the DGI-2 debt securities are FFA and its intermediate data, but they have no information on the disaggregated data of debt securities.

The new data sources for debt securities broken down by currency, maturity, interest rate, and market differ according to the type of debt security[4]. There are primarily three types of new data sources: (1) microdata on domestic debt securities excluding government bonds, those with description of issuer name, currency, total amount, type of interest rate, interest rate payment date, date of issue, redemption date, etc. from the Japan Securities Depository Center, Incorporated, (JASDEC[5]); (2) two sets of survey data on (a) government bonds provided by the Ministry of Finance Japan, and (b) monetary claims entrusted with trust banks; and (3) disaggregated data from the BIS international debt securities statistics (IDSS) and private data vendors on the debt securities issued by residents in the international market.

The remaining section explains the estimation methods used to compile (1) the total outstanding amount of debt securities at nominal values and (2) debt securities broken down by currency, maturity, and interest rate. These are estimated separately by type of debt securities as follows: (2)-1 debt securities issued by residents in the domestic market; and (2)-2 debt securities issued by residents in the international market. The methodology is explained below.

(1) Total outstanding amount at nominal values

The total outstanding amount at nominal values issued by respective economic sector are recorded based on intermediate data used in the compilation of the FFA. These intermediate data are basically compiled with the data sources from the financial statements, aggregated data of government bonds, etc. recorded at nominal basis in principle.

(2) Debt securities by currency, maturity, and interest rate

(2)-1 Debt securities issued by residents in the domestic market

With regard to debt securities issued by residents in the domestic market, the breakdown by currency, maturity, and interest rate are

---

[4] The debt securities are described in the international guidelines ("System of National Accounts 2008" and "Handbook on Securities Statistics") as "debt securities are negotiable instruments serving as evidence of a debt. They include bills, bonds, negotiable certificates of deposit, commercial paper, debentures, asset backed securities, and similar instruments normally traded in the financial markets." In accordance with the international guidelines, the BOJ classifies the following FFA transaction items into the debt securities of the DGI-2 template: treasury discount bills; central government securities and FILP Bonds; local government securities; public corporation securities; bank debentures; industrial securities; external securities issued by residents; commercial paper; trust beneficiary rights; and structured-financing instruments.

[5] JASDEC is the central securities depository in Japan.

estimated by combining the total outstanding at the nominal value based on the FFA with microdata which includes individual characteristics of debt securities. For instance, to calculate the debt securities by currency we follow steps 1) to 3) (Table 1).

1) Break down sectorally classified data[6] by currency and then calculate the composition ratio of domestic and foreign currency. The granular data are used solely for calculating the composition ratio since the coverage of microdata is not exactly the same as that of the FFA.

2) Multiply the composition ratio of 1) by the amount outstanding of debt securities at nominal values based on the FFA.

e.g. the outstanding of domestic currency issues in the domestic market are calculated as follows:

Ofd = Of * (Omd / (Omd + Omr))

3) Sum up debt securities issued in domestic and international markets by currency.

e.g. the total outstanding amount of domestic currency issues are calculated as follows:

Tfd = Ofd + Ifd

Table 1: Debt Securities by Currency

Breakdown of debt / Type of securities			Debt securities issed by residents in domestic market (O/o) (2)-1	Debt securities issued by residents in international market (I/i) (2)-2	Total (T) (reporting data for DGI-2)
Micro/Dis-aggregated data (m)	Currency (Tmc)	Domestic (Tmd)	Omd	Imd	
		Foreign (Tmr)	Omr	Imr	
FFA basis (f)	Total (Tf)		Of	If	Tf = Of + If
	Currency (Tfc)	Domestic (Tfd)	Ofd = Of * (Omd/(Omd + Omr))	Ifd = If * (Imd /(Imd + Imr))	Tfd = Ofd + Ifd
		Foreign (Tfr)	Ofr = Of * (Omr/(Omd + Omr))	Ifr = If * (Imr/(Imd + Imr))	Tfr = Ofr + Ifr

1) calculate composition ratio

2) multiply the composition ratio by the total amounts

3) sum up domestic currency issues

note : T;total, m;micro data, d;domestic currency, r;foregin currency, f;FFA, I/i;debt securities issued by residents in international market, O/o;debt securities issued by residents in domestic market, c;currency, m;micro/disaggregated data

(2)-2 Debt securities issued by residents in the international market

The total outstanding amount of debt securities issued by residents in international markets by respective sector are compiled based on the FFA data. Afterwards, the total amount is divided into domestic and foreign currency; short and long maturity; and fixed and variable interest rates with the use of the composition ratio estimated from the BIS IDSS's disaggregated data. In addition, for further maturity breakdown (1 year

[6] The sectorally classified data are compiled by breaking down microdata of domestic debt securities by economic sector in alignment with sectoral classification in the DGI-2 template. The individual data is classified into respective economic sector by examining the respective issuers' name.

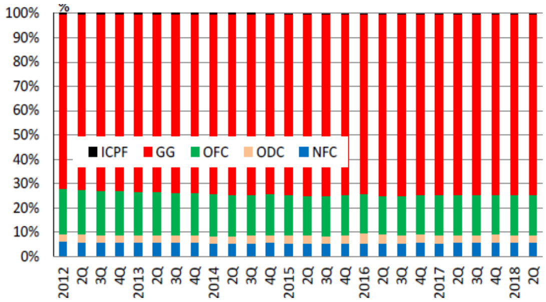
up to 2 years; 2 years up to 5 years; 5 years to 10 years; and more than 10 years), the private vendor data are used as an alternative data source of the BIS IDSS. It is noted that since the BIS IDSS has fewer number of sectoral classifications than those of the DGI-2 template, the same composition ratio is applied to the sub-sectors with that of their parent sector.

2.3 Result

This section presents some features revealed from tentative results.

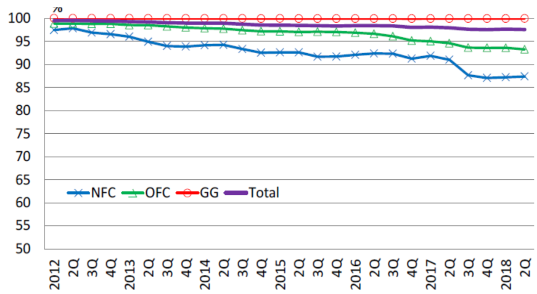
Looking at the overall picture of domestic debt securities issued by residents, the general government issued bonds have the largest proportion, more than 70%[7] of the total amount, followed by the bonds issued by other financial corporations and non-financial corporations (Chart 1). Major issuers in other financial corporations include public captive financial institutions such as Japan Expressway Holding and Debt Repayment Agency, and government financial institutions[8] such as the Japan Housing Finance Agency.

(Chart 1) Proportion of debt securities by issuer



NFC: Non-financial corporations, ODC: Other deposit-taking corporations,
OFC: Other financial corporations, GG: General government,
ICPF: Insurance corporations and pension funds

(Chart 2) Proportion of domestic currency



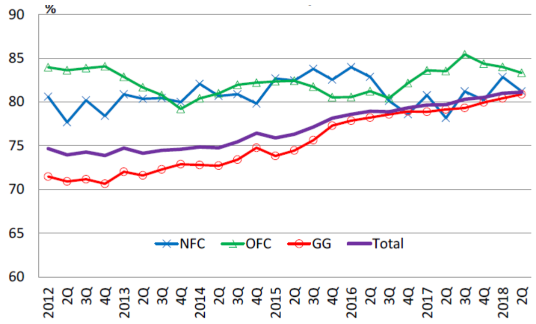
The proportion of domestic currency issues in Chart 2 shows that more than 90% of the total debt securities are denominated in domestic currency.

[7] If government bonds issued by the Special Account of Fiscal Investment and Loan Program Fund (classified in OFCs) were added to the general government issues, the share would be increased up to about 80%.

[8] Public captive financial institutions are defined in the FFA as entities which raise funds only from specific sources such as government and do not plan to raise funds from markets, as well as those organizations which raise funds from markets but only invest in a limited group of destinations or operations.

This feature has not changed significantly since 2012, but OFCs[9], which account for around 20% of the total debt securities, has gradually increased the ratio of foreign currency issues, reflecting the recent increase in TLAC bonds issues by Japanese bank holding companies to support their growing overseas business. The NFCs have also increased foreign currency issues partly due to their strong demands for foreign currency in the midst of recent corporate overseas development and business expansion. Meanwhile, it has to be noted that estimation accuracy for the NFC figures are relatively low as they are calculated as residuals.

(Chart 3) Proportion of issues with long remaining maturity



note: more than one year in the remaining maturity

The maturity structure by economic sector in Chart 3 shows that debt securities with a remaining maturity of more than one year make up more than two-thirds of the total debt securities outstanding. With regard to the government bonds, an increase of the outstanding amount with long remaining maturity reflects the strong demand from institutional investors such as insurance companies and pension funds to balance the maturity of assets and liabilities. In addition, the "Debt Management Report 2018" released by the Ministry of Finance Japan explains that "*in a bid to reduce future interest rate hike risks in the low interest rate environment, the government has lengthened the average maturity of JGB issues in recent years.*"

3. Institutional Sector Accounts

3.1 Requirements

Recommendation II.8 on institutional sector accounts calls for "*the G-20 economies to compile and disseminate, on a quarterly and annual frequency, sectoral accounts flows and balance sheet data, based on the internationally agreed template, including data for the other (non-bank) financial corporations sector, and develop from-whom-to-whom matrices for both transactions and stocks to support balance sheet analysis.*" The template of institutional sector accounts consists of three parts: (1) a general template (principal target for 2021); (2) a template for the collection of data on shadow

[9] OFCs issuing debt securities include: nonbanks; public financial institutions; financial dealers and brokers; financial auxiliaries; and public captive financial institutions.

banking (a more advanced ambition); and (3) a template for the collection of from-whom-to-whom data (a more advanced ambition). This section focuses on our progress on the general template[10]. The BOJ has been working on compiling the sectoral financial balance sheet of the general template of (1) in collaboration with the Cabinet Office.

3.2 What are our challenges?

There are primarily three issues in compiling the financial accounts: (1) breaking down of loans and debt securities by currency type and maturity (original and remaining) on both the asset and liability sides of each economic sector; (2) breaking down foreign securities in the FFA into respective transaction categories, debt securities, equities, and investment trusts in alignment with the DGI-2 template; and (3) breaking down foreign direct investment in the FFA into respective transaction categories, debt securities, equities, etc. The following section introduces the methodology for (2) mentioned above and the tentative results of foreign debt securities whereas both (1) and (3) are left as future issues.

3.3 Methodology

The foreign securities termed as outward investment in securities in the FFA include foreign debt securities, foreign equities, and foreign investment funds by definition, however, these three sub-components are not compiled in the FFA. In the absence of microdata, we take the following steps to measure the three sub-components.

- Investigate data sources for the breakdown of foreign securities. The data availability varies with each sector. Major data sources include financial balance sheets publicly available and sharing information on bank examination and monitoring, etc.

If the source data are available, proceed through either the following (1) or (2).

(1) If the data cover an entire sector, reporting data are compiled by grossing up the source data without any estimation.

(2) If the data have limited coverage, consider if the data can be used as a benchmark to estimate an entire sector.

- If the data are not available, examine an alternative methodology based on relevant assumptions. The alternative solutions include: (a) use the same composition ratio (foreign debt securities, foreign equities, and foreign investment funds) with that of similar industries/entities which have common features in business structures;

[10] G-20 DGI templates are available from Principal Global Indicator' website (http://www.principalglobalindicators.org/?sk=E30FAADE-77D0-4F8E-953C-C48DD9D14735&sId=1452784383161).

(b) conduct a survey of the entities; or (c) calculate figures as residuals and do not estimate the figures directly from the data sources.

3.4 Results

The total foreign debt securities are estimated to be about 369 trillion yen[11] at the end of June 2018. Chart 4 shows that the foreign debt securities held by GG, ICPF, and ODC account for around 80% of the total amount. The holdings by GG including investment by foreign exchange reserves[12] and Government Pension Investment Fund (GPIF)[13] account for around 35% of the total foreign debt securities issues. The time-series data shows ICPFs gradually increase their share over recent years while ODCs decrease amid the recent upward trend in the U.S. interest rate. The BOJ's *Financial System Report* describes the recent movement of insurance companies in the following way, "*a breakdown of portfolios shows that purchases of domestic bonds, which offer low yields, have been restrained amid the prolonged low interest rate environment, while investment in foreign bonds and investment funds, which offer relatively high yields, has increased*."

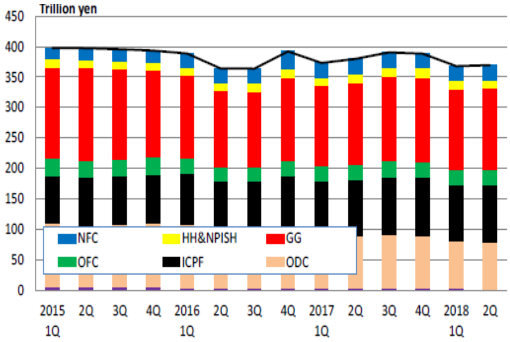
Chart 5 illustrates the ratio of foreign debt securities in foreign securities and shows more than 60% of foreign securities are investments in foreign debt securities, except for the case of OFCs including securities investment trusts sector in the FFA. The time series data shows the share of foreign debt securities, most of which are denominated in U.S. dollar has gradually decreased amid the recent upward trend of the U.S. interest rate.

[11] The spot rate as of the end of June 2018 is 110.64 yen/dollar.

[12] Foreign exchange reserves are foreign financial assets that can be used immediately and under the control of monetary authorities for financing or regulating payments imbalances or indirectly make adjustments for foreign exchange market intervention.

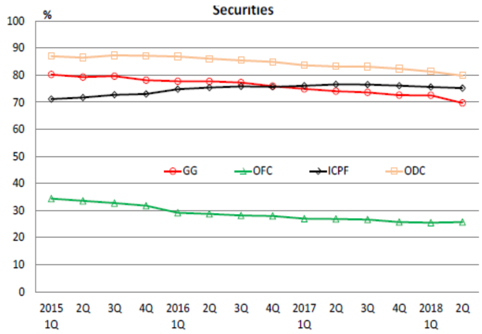
[13] The GPIF "manage and invest the Reserve Funds of the Government Pension Plans entrusted by the Minister of Health, Labour and Welfare, in accordance with the provisions of the Employees' Pension Insurance Act (Law No.115 of 1954) and the National Pension Act (Law No.141 of 1959), and shall contribute to the financial stability of both Plans by paying out profits of investment to the Special Accounts for the Government Pension Plans."

(Chart 4) Holdings of Foreign Debt Securities



CB: Central Bank, NFC: Non-financial corporations, ODC: Other deposit-taking corporations,
OFC: Other financial corporations,
GG: General government, ICPF: Insurance corporations and pension funds, HH&NPISH: Households & NPISH

(Chart 5) Ratio of Foreign Debt Securities in Foreign Securities



3.5 Future issues

We are still in the early stage in terms of achieving the DGI-2 recommendations II.8 on Institutional Sector Accounts. There is still more room to enhance the foreign debt securities data if more granular data are available. In addition, we have other remaining issues as mentioned in 3.2, including the breakdown of loans and debt securities by maturity and currency in both asset and liability sides. We plan to further collect new additional disaggregated/granular data to compile the breakdown of loans, debt securities, and foreign securities, while giving due consideration to the balance between respondent burden and users' convenience for economic and financial analysis.

4. Summary

This paper presented our progress in the implementations of the DGI-2 recommendations II.7 on securities statistics and II.8 on institutional sector accounts. Regarding securities statistics, we describe what type of new data sources and methodology will be applied to identify the breakdown data by currency, maturity, interest rate, and the domestic/international market. The preliminary outcomes on debt securities show some important features of Japan's debt securities market. Regarding institutional sector accounts, we present our initiatives to decompose the foreign securities into foreign debt securities, foreign equities and foreign investment funds with some outstanding features based on the tentative outputs. Toward the final deadline in 2021, we will continue to work on the remaining issues relating to the DGI-2 recommendations.

References

1. Bank of Japan. *Financial System Report*. October 2018.
2. Ministry of Finance Japan. *Debt Management Report*. 2018.
3. Bank of Japan. *Guide to Japan's Flow of Funds Accounts*. March 2018.
4. Bank of Japan. *Compilation Method of Japan's Flow of Funds Account*. March 2018.

A reflection on privacy and data confidentiality in Official Statistics

F. Ricciato, A. Bujnowska, A. Wirthmann, M. Hahn, E. Barredo-Capelot
EUROSTAT – Directorate B: Methodology; Dissemination; Cooperation in the ESS

Abstract

The availability of new digital data sources represents an opportunity for Statistical Offices (SO) to complement traditional statistics and/or deliver novel statistics with improved timeliness and relevance. Nowadays SOs are part of a larger "data ecosystem" where different organizations, including public institutions and private companies, engage in the collection and processing of different kinds of (new) data about citizens, companies, goods etc. In this multi-actors scenario it is often desirable to let one organization extract some output statistics (i.e., aggregate information) from input data that are held by other organization(s) in different administrative domain(s). We refer to this problem as cross-domain statistical processing. To achieve this goal, the most intuitive approach—but not the only one—is to exchange raw input data across administrative domains (organizations). However, this strategy is not always viable when personal input data are involved, due to a combination of regulatory constraints (including lack of explicit legal basis for data sharing), business confidentiality, privacy requirements, or a combination of the above. Furthermore, new data sources often embed a much more pervasive view about individuals than traditional survey and/or administrative data, an aspect that amplifies the potential risks of data concentration. In such cases, performing cross-domain statistical processing requires technologies to elicit only the agreed-upon output information (exactly or approximately) without revealing the input data. This entails addressing two distinct but complementary problems. First, we need to compute the desired output statistics without seeing the raw input data. Second, we need to control the amount of information that might be inferred about individual data subjects in the input dataset from the output. In the field of privacy engineering the notions of "input privacy" and "output privacy" are used to refer respectively to these two problems. We remark that these problems are separable, i.e., they can be addressed with distinct tools and methods that get combined together, overlaid or juxtaposed. In this contribution we review recent advances in both fields and briefly discuss their complementary roles. As for input privacy, we provide a brief introduction to the fundamental principles of Secure Multi-Party Computation (SMPC). As for output privacy, we review recent advances in the field of Statistical Disclosure Control (SDC). Finally, we discuss possible

scenarios for SMPC and SDC integration in the future *"confidentiality engineering"* setup of modern official statistics.

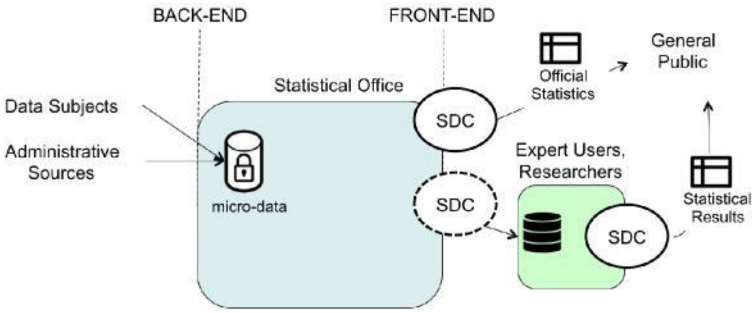
Keywords
Privacy; Confidentiality; Security; Statistical Disclosure Control; Secure Multiparty Computation

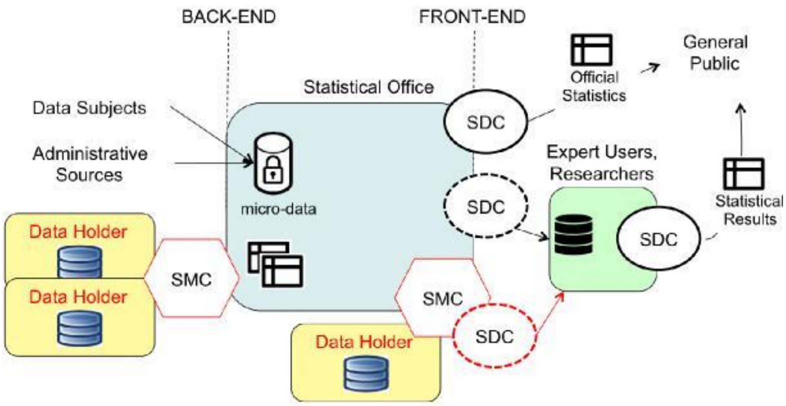
1. Introduction and motivations

The modern society is undergoing a process of massive datafication [1]. The availability of new digital data sources represents an opportunity for Statistical Offices (SO) to complement traditional statistics and/or deliver novel statistical products with improved timeliness and relevance, so as to meet the increasing demands by users. However, such opportunities come with important challenges in almost every aspect – methodological, business models, data governance, regulatory, organizational and others. The new scenario calls for an evolution of the modus operandi adopted by SO also with respect to privacy and data confidentiality. We propose here a discussion framework focused on the prospective combination of advanced (dynamic) Statistical Disclosure Control (SDC) methods with Secure Multi-Party Computation (SMC) techniques.

For decades, the data business has been a natural monopoly centered around SO: no other entity had the technical and legal capability to collect and process large scale data across individuals and organizations. In the traditional operation model, illustrated in Fig. 1, the SO ingests internally all source (micro-)data that were collected either directly from the data subjects, via surveys and censuses, or indirectly through administrative registers. The input source data collected in the back-end are then processed centrally to deliver two types of front-end data in output: (i) official statistics for the general public; and (ii) more detailed data for further processing by expert users and researchers downstream the data flow.

The legal mandate of SO includes two important obligations that can be summarized as 'closed input and open output'. On the input side (back-end) SO must preserve the confidentiality of the (micro-)data in order to protect the privacy of data subjects. On the output side (front-end) SO are committed to publish openly the processed statistics (and in general any output data), so as to ensure that all potential users get the same information and do so at the same time. The motivations and implications of both obligations are intimately connected to the democratic role of official statistics in modern society. However, in terms of real world applications, there is an unavoidable conflict between these two goals, since by definition the output data carry non-zero information about the input data (otherwise they would be useless), i.e., they always reveal something about the input. On the front-end, SO must determine whether what can be inferred from the output about the input can be tolerated or not, i.e., whether it is acceptable or not for the privacy of the individual data subjects. Such determination must be done case-by-case and this is the goal of the so-called Statistical Disclosure Control (SDC) function (ref. Fig. 1). When critical cases are detected, SDC methods seek to strike a reasonable compromise between the two conflicting goals of preserving accuracy and completeness of the output along with confidentiality of the input. In practice, this involves limiting and/or degrading the output data in a controlled manner. Traditionally, this was done statistically by suppression of selected elements in the output table. More recently, following the increasing demand by expert users to go beyond static tables predefined by SO and make their own statistics, be it tables or other forms of output, SDC is evolving towards dynamic models based on data perturbation, as discussed later in the paper. In general, with SDC (both static and dynamic) a trade-off is in place between accuracy and utility of the output on one hand, and confidentiality of the input on the other [2, 3], and SDC methods strive to address this problem.





The new scenario illustrated in Fig. 2 yields several elements of novelty. Instead of dominating the 'data monopoly' as in the past, SO are now one species of a larger 'data ecosystem' where different organizations, including public institutions and private companies, engage in the collection and processing of different kinds of data about citizens, companies, goods etc. For SO this change has implications on both sides. In the back-end, there are new potential data sources to be accessed, in addition to traditional survey/census and administrative data, but the peculiarities of such new sources might require alternative access models other than direct ingestion of raw input data [4]. On the front-end, the expert users downstream the processing flow have now increased possibilities to combine the data obtained by SO with other external data, an aspect that exacerbates the challenge for SDC.

2. Input privacy vs. Output privacy

Hereafter we provide an abstract view about the relations between input data and output data and then present the notions of 'input privacy' and 'output privacy' as introduced in the literature (see e.g. [5,10]). Finally we elaborate about how these categories map to the new scenario described above.

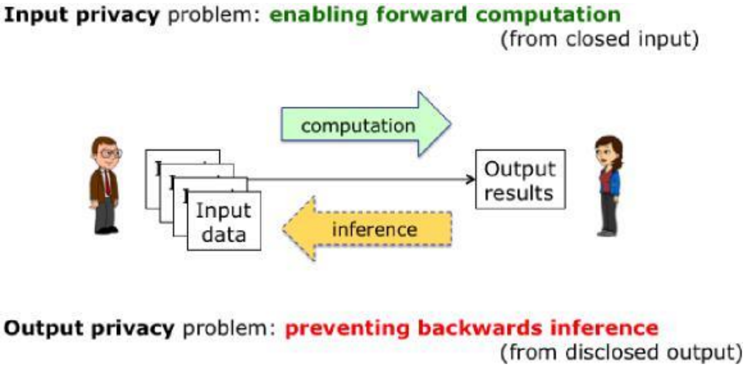


Figure 3 – Input Privacy vs. Output Privacy problems.

We call computation the task of extracting the desired output information (or results) from a set of input data. We call inference the task of extracting some (partial) information about one of the input components based on knowledge of the output (and possibly other external data). It is clear from Figure 3 that computation and inference flow logically in opposite direction. In this discussion the output can take any arbitrary form, including for example a summary indicator, a set of regression coefficients or a whole frequency table, to name some concrete examples. We focus here on the case where the computation function (that may be called algorithm, methodology, procedure, program etc. in different context) is well defined before execution. In other words, our focus is on the stage of statistics production, not on the (logically antecedent) phase of data exploration and methodological development. What is relevant for our discussion is the multi-party scenario where the entity or entities (organizations, institutions, individuals, etc.) holding the input data differ from the entity/entities interested to get the output results. We shall use the terms 'input party' (IP for short) and 'output party' (OP for short) to refer, respectively, to the entities holding the input data and those interested to learn the output result. For example, when processing confidential data held by private business companies for official statistics, such companies take the role of IP, while the OP role is taken by SO. In another citizen statistics scenario, each individual respondent (data subject) can be an independent IP, and again the OP maps to SO.

Given this abstract setting, with IP and OP role, we identify two distinct confidentiality challenges:

- **Output privacy** problem: Given that the computation results will be made available in some way to the OP how to prevent OP from inferring too much about the input data held by IP?
- **Input privacy** problem: Given that the input data are confidential and cannot be disclosed outside their respective IP, how to enable the OP to learn the computation results?

In the particular case where a single IP holds all input data, the input privacy problem admits a very simple solution: the whole computation can be executed internally to the IP, and only the final output is passed to the OP. This has been indeed the case of official statistics for decades, with the statistical office playing the role of IP on the front-end, where the external users (including researchers and the general public) play the role of OP. In this setting, exemplified in Figure 1, the input privacy problem is inherently solved and only the output privacy problem had to be addressed.

Instead, in the new scenario depicted in Figure 2, we foresee the possibility for the statistical office to compute statistics based on confidential data held by other entities (e.g., private companies, other public institutions, or individuals) that we cannot or do not want move into the statistical office domain. In this case, external data holders play the role of IP in the back-end, where the statistical office plays the role of OP. Furthermore, on the front-end, we may want to let our users compute statistics based on the fusion of confidential data held by the statistical office with other external input data. In this case, the input privacy and output privacy problems occur jointly on the front-end.

In the new scenario, we must cope with the input privacy problem in addition to (not in place of) the output privacy problem. Again, if the desired statistics can be computed from the input data held by a single data holder (as IP) in isolation from other data holders, the most natural approach is to let the IP execute the computation and then pass the (final or intermediate) non-confidential data to the statistical office (as OP). Standard technical and non-technical means can be adopted to ensure that the program that is executed by the IP premises does not deviate from what was approved (or developed) by the OP. This is particular relevant on the back-end, where OP maps to the statistical office: we highlight that outsourcing the mere execution of a computation program to the IP does not imply loss of control by the OP over what program is executed.

The input privacy problem is more challenging when the required output results are based on the contribution of several input data sets held by multiple IPs that cannot disclose their data. In some cases, the computation program can be factorized into separate computation instances that are run independently by the multiple IPs, either sequentially or in parallel. However, very often the desired output results do not allow for computation factorization. For example, this is the case when output results must be computed on the intersection records between different IP data sets, or when a regression must be run over variables that are held by different IPs. In these cases, we may resort to Secure Multi-Party Computation (SMC).

3. Solution approaches to Output Privacy problem

The output privacy approach is traditionally addressed by so-called Statistical Disclosure Control (SDC) techniques, possibly in combination with Access Control (AC). SDC aims at restricting what is disclosed, while AC imposes restrictions on to whom it is disclosed. Generally speaking, there is a trade-off between the two: the weakest SDC requires strongest AC, and vice-versa.

AC methods rely on combination of requirements referring to the nature of the potential users, their experience with holding confidential data and legitimate use of the data. Potential users must provide evidence of fulfilling AC requirements which is scrutinized by the data owners. The trustworthy users are confined with more detailed data and better access facilities. SDC methods rely on combination of suppression, perturbation, randomization and aggregation of data.

Historically, SDC was performed manually by dedicated experts, following practices and criteria that were developed through the years in the official statistics community. SO successfully managed the output control as the statistics going out were pre-defined and SO could consistently apply suppressions on primary and secondary confidential cells. The current trend is towards "on-the-fly SDC". Nowadays many users want to calculate their own tailor-made statistics on the basis of the detailed data sources. In response to these needs SO make available dynamic data querying systems that implement modern SDC approaches, addressing in particular the problem of differential disclosure. These new SDC approaches require that the output is always safe, also in combination with any other statistics based on the same source. The random noise protection method developed by the Australian Bureau of Statistics (ABS) is an example of modern SDC approach [11, 12]. The ABS method consists in applying small perturbations (controlled noise) to the data with the predefined probability distribution. A specific pseudo-random mechanism called "cell key method" is adopted to ensure that the injected perturbations are consistent across multiple queries. This approach is robust to differential attacks that, instead, represent the main limitation of pure randomized systems (where noise varies across queries). The cell key method is recommended for protection of European census 2021 round [13]. It is expected that it will ensure consistent protection of the census data in view of making them available via various channels and access systems.

4. Solution approaches to Input Privacy problem

When the input data are held by multiple IPs, and the computation cannot be factorized into independent (sequential or parallel) components, one possible solution approach to the input privacy problem is given by Secure Multi-Party Computation (SMC) methods based on the principle of secret sharing. In a nutshell, with SMC every individual input data element is transformed into a set of so-called secret shares that are passed to a set of (three or more) intermediate 'computing parties' (CP). The CPs form collectively the SMC infrastructure. The secret shares are produced in a way that yields two important properties. First, under certain conditions, defined by the applicable attack model, secret shares do not reveal anything about the input source data to the individual CPs (non-invertibility). Second, they allow to compute exactly the correct output that would be obtained by a direct computation on the clear input (homomorphism). A general introduction to SMC and secret sharing can be found in [6] while examples of practical applications[1] are found in [7, 8].

To preserve confidentiality, each CP must not disclose the received secret shares to other CPs, i.e., CPs must not collude among themselves to break the confidentiality of IP data. SMC can be tuned to be robust against a subset of colluding CPs. In other words, the system preserves input confidentiality as far as at least one CP does not collude with the others. That means, the CPs must be trusted collectively, not individually. Then the problem of ensuring be trusted collectively, not individually. Then the problem of ensuring confidentiality moves up to an institutional level, and translates into the task of identifying a suitable set of CP. An important property of SMC plays in our favour: in practical deployment, the same institution can play multiple roles. For example, one data holder serving as IP can at the same time host one CP instance – obviously he would never collude with other CPs against himself. Also, one entity (e.g., the SO) can play contemporarily the roles of IP, OP and CP.

[1] The relationship between SMC and personal data protection legislation presents some open issues that go beyond the scope of the present contribution, see e.g. the discussion in [15].

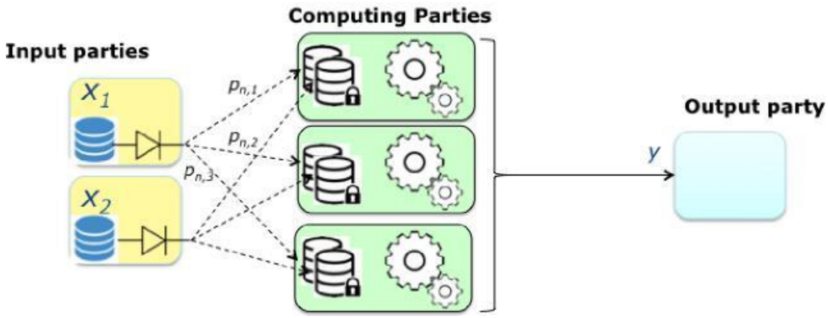


Figure 4 - Principle of operation of SMC (secret sharing)

In theory, any arbitrary function can be computed via SMC, at the cost of increased computation load and communication overhead between the CPs compared to a plain centralized computation. The cost of SMC translates into longer computation time and/or more hardware/bandwidth resources. The cost increase factor might be substantial, but still acceptable for most practical applications.

When the computation cost and/or delay of SMC is too large, we may resort to an alternative solution, hinted hereafter. The key point of both solutions is to let the set of relevant stakeholders (any combination of IP, OP and/or external entities) to exert shared control over the computation process, so as to ensure jointly that no confidential data is disclosed except the agreed-upon final results. Such guaranteed can be delivered, in principle, by a special computation machine that is built (at both hardware and software levels) to execute exclusively code that is cryptographically authenticated by all and only the intended stakeholders, as depicted in Figure 5. Such ideal machine can be built by combining so-called Trusted Execution Environment (TEE) technology with cryptographic solutions for multi-party control (MPC). The TEE technology [14] was developed recently to address the emerging need in cloud computing applications to decouple, also at the hardware level, the physical operation of the computing machine (hosting, powering up, general maintenance) from the control of what is executed over that machine.

The MPC-TEE solution should be distinguished by the simplistic approach of relying on a Trusted Third Party (TTP). The trust models underlying the two settings are completely opposite to each other, as exemplified by the diagrams in Figure 6. The TTP represents an independent entity, outside the control of all stakeholders. On the contrary, MPC-TEE can be thought as a dependent entity that is under direct control of all stakeholders jointly. In other words, full delegation takes place with TTP, while no delegation take place with MPC-TEE.

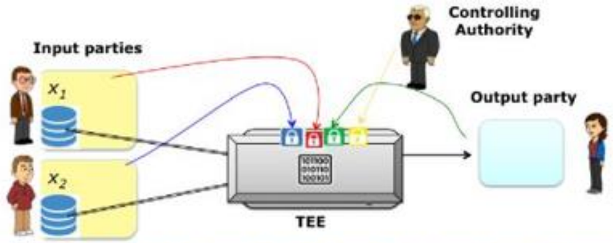


Figure 5 – Concept of Multi-Party Controlled Trusted Execution Environment (MPC-TEE)

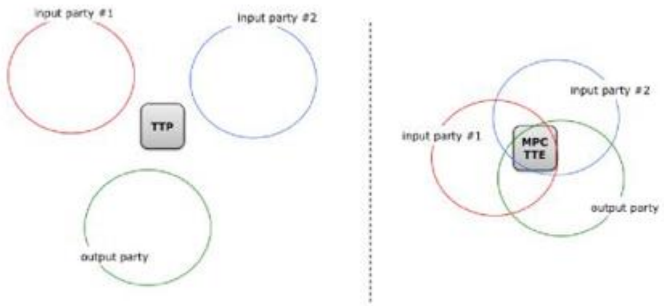


Figure 6 – Logic difference between the TTP model (independent entity, full delegation) and the MPC-TTE model (jointly dependent entity, no delegation). The latter model applies also to SMPC.

5. Conclusion and discussion

Since SDC and SMC are targeting different but complementary problems, it is natural to consider their combination. From the above definitions it should be clear that, in principle, both 'input privacy' and 'output privacy' problems might be encountered at both sides. In other words, in the new scenario we may consider adopting some combination of SMC and SDC in the back-end as well as on the front-end.

As to the back-end, SMC may play an important role when joint processing of multiple data sources from different parties is required but direct ingestion of raw input data by SO is not possible, e.g., due to legal restrictions or business considerations (as in [7]). This includes cases where the source data are held by the private business sector. Considering the special trust endowment of SO, who plays the role of OP in the back-end, it is reasonable to assume that non-disclosure agreements and legal provisions are sufficient to solve the 'output privacy' problem in the back-end, waiving the need to introduce SDC tools on this side.

Conversely, SDC solutions will remain crucial on the front-end. SMC can be used on the front-end to enable joint processing of confidential input data from SO and other data holders (ref. rightmost part of Fig. 1(b)). More in general, a wise combination of SMC and SDC might help to achieve a higher level of overall confidentiality in the new wilder scenario, where increased availability of external data sources amplifies the non-disclosure challenges.

Some initial work in this direction is starting to appear in the field of Official Statistics [9], while commercial implementation of SMC already include some simple safeguards for disclosure control [8].

In conclusion, the new datafied scenario requires SOs to widen their traditional approach to privacy and data confidentiality. Purely regulatory means in the back-end and simple SDC methods in the front-end might not suffice any more. Embracing novel tools such as SMC, in combination with more advanced forms of dynamic SDC, seems to be a promising direction to move forward. More in general, SO need to develop a more systematic and articulated approach towards confidentiality engineering to face the new challenges posed by an increasingly complex data ecosystem.

References

1. K. Cukier and V. Mayer-Schoenberger. The rise of big data. Foreign Affairs, May/June 2013. URL https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data.
2. J. M. Abowd and I. M. Schmutte. An economic analysis of privacy protection and statistical accuracy as social choices. American Economic Review (forthcoming), August 2018. https://arxiv.org/abs/1808.06303.
3. J. Domingo-Ferrer, S. Ricci, and J. Soria-Comas. A methodology to compare anonymization methods regarding their risk-utility trade-off. Int. Conf. on Modeling Decisions for Artificial Intelligence (MDAI 2017), August 2017.
4. F. Ricciato et al. Towards a Reference Architecture for Trusted Smart Statistics. 104th DGINS conference, Bucharest, Romania, 10-11 October 2018. https://tinyurl.com/yco77y62.
5. T. Wang and L. Liu. Output privacy in data mining. ACM Transactions on Database Systems, 36(1), March 2011. DOI: 0.1145/1929934.1929935.
6. R. Cramer, I. Damgard, and J. Buus Nielsen. Secure Multiparty Computation and Secret Sharing. Cambridge University Press, 2015.
7. S. Anspal, M. Kaska, and I. Seppo. Using k-anonymization for registry data: pitfalls and alternatives. IEEE Trans. on Dependable and Secure Computing, 2017. http://dx.doi.org/10.12697/ACUTM.2017.21.05.
8. D. Bogdanov et al. Rmind: a tool for cryptographically secure statistical analysis. IEEE Trans. on Dependable and Secure Computing, 15(3), May/June 2018.
9. K. Shirakawa et al. A proposal of a simple and secure statistical processing system using secret sharing. UNECE Work Session on Statistical Data Confidentiality, Skopje, 20-22 Sept. 2017.

www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/6_secret_sharing.pdf.
10. UN Handbook on Privacy-Preserving Computation Techniques, April 2019. Available online at http://tinyurl.com/y4do5he4.
11. L Leaver, Victoria and Marley, Jennifer K. (2011) A Method for Confidentialising User-Defined Tables: Statistical Properties and a Risk-Utility Analysis; ISI conference, Dublin 2011.
12. G. Thompson et al (2013) Methodology for the Automatic Confidentialisation of Statistical Outputs from Remote Servers at the Australian Bureau of Statistics, UNECE Works Session on Statistical Confidentiality, Ottawa, Canada, 20-23 October 2013.
13. P.P. de Wolf, Perturbative methods for ESS census tables, conference on New Techniques and Technologies for Statistics, Brussels, March 2019.
14. M. Sabt, M. Achemlal, A. Bouabdallah, Trusted Execution Environment: What It is, and What It is Not. 2015 IEEE Trustcom/BigDataSE/ISPA. DOI 10.1109/Trustcom.2015.357.
15. G. Spindler and P. Schmechel, Personal Data and Encryption in the European General Data Protection Regulation, (2016) JIPITEC 163 para 1. Online: https://www.jipitec.eu/issues/jipitec-7-2-2016/4440/spindler_schmechel_gdpr_encryption_jipitec_7_2_2016_163.pdf

Trade repository data - access and sharing

Jose-Maria Serena Garralda
Bank for International Settlements, Basel (Switzerland)

Abstract

Derivatives data collection in trade repositories has grown as a result of the regulatory reforms launched after the Great Financial Crisis. Ensuring that these data are effectively used, not least to guide policy actions, constitutes a priority for policymakers. This paper discusses access to, and sharing of, trade repository data, leveraging on the findings of the survey conducted by the Irving Fisher Committee in 2018 among central banks. The survey shows that most central banks are gaining some access to data from the trade repositories collecting them. Access to granular information in their jurisdictions is relatively good, but access to data held by foreign trade repositories is complex -often requiring authorization, and restricted aggregated data. Data sharing with internal users is relatively widespread. However, the external dissemination of trade repository data is highly restricted, being limited mainly to indicators with a sufficient degree of aggregation. Overall the results suggest that data sharing of trade repository data remains a challenge. Looking ahead coordination between domestic authorities, including between central banks and trade repositories, should be enhanced.

Keywords

Derivatives data; Trade repositories; Data access; Data sharing

1. Introduction

One important consequence for financial statistics of the reforms undertaken after the Great Financial Crisis (GFC) of 2007–09 has been the collection of a very large amount of Trade Repository (TR) data shedding light on the global derivatives market. Yet a key issue for public authorities is to ensure that these data are effectively used, not least to guide policy actions.

Data availability differs across jurisdictions, and it is widespread among the largest ones. Progress has been particularly notable in FSB jurisdictions (as well as in non-FSB countries subject to European regulation) in terms of reporting requirements and richness of the information collected (eg counterparty coverage). Since these jurisdictions host the largest market segments, the coverage of the global derivatives market has significantly improved.[1]

[1] Data availability remains an important issue in several other (smaller) jurisdictions.

However data usage has been prevented by problems to access and share them. Some central banks have been directly tasked with collecting derivatives transactions data. But most of them have to get this information from the TRs set up separately to collect it.

Against this backdrop, this paper summarises key findings of the survey conducted by the Irving Fisher Committee on Central Bank Statistics among central banks on access to, and use of, trade repository data. The survey was answered by 50 central banks.[2] The main conclusions on data access are the following. Central banks have relatively good access to granular information (ie position- or transaction-level data) from the TRs located in their jurisdictions. But access to data held by foreign TRs is more complex, often requiring specific authorisation and being restricted to data with a certain level of aggregation. Concerning data sharing, the survey reveals significant problems to share granular data to external users. Data sharing within central banks, and of aggregated data, is less problematic.

The rest of the paper is structured as follows. Section 2 details the main findings concerning central banks' access to trade repository data. Section 3 outlines the takeaways on data sharing. Section 4 presents the main conclusions.

2. Trade repository data access

The OTC reform agenda sought to ensure an effective and practical access to TR data for public authorities so that they can use this information adequately. Central banks are among the most interested authorities in getting access to this information.

Most central banks have some kind of access to the information collected, although one fifth have no access at all. However, the survey reveals two problems to access data. First, central banks face limits to access highly granular data. In particular 65% of the respondents declared that they can receive "pure" micro information – ie position or transaction-level data (Graph 1).[3] This ratio rises to 77% for disaggregated data,[4] which are less informative

---

[2] The IFC published a report discussing the main findings. See Irving Fisher Committee (2018), *"Central banks and trade repositories derivatives data"*, IFC Report

[3] Micro data allows the identification of entities, unless they are anonymised or masked by deleting or encrypting certain elements of the positions and/or transactions. Micro data are defined as *"data on individual reporting units or specific transactions/instruments, which in most cases allow the identification of individual entities and therefore considered confidential. In addition, publicly available data on individual reporting units are considered non-confidential although they can still be subject to data sharing limitations due to commercial property rights"* (IAG (2017)).

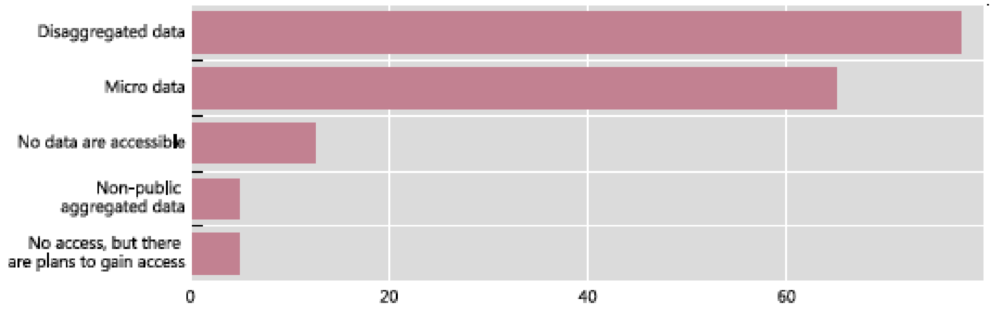
[4] Disaggregated data are below the level of aggregated data, and are more likely than aggregated data to reveal the identity of individual reporting units. Disaggregated data are

but can nevertheless provide useful insights for central banks. In around 15% of these cases, central banks have no access to any non-public data at all (and do not plan to gain access).[5]

Does your central bank have access to non–public derivatives data reported to TRs/TR-like entities (other than the central bank)?

In per cent

Graph 1



The sum of the responses can exceed 100%, as several answers are possible.

Sources: IFC survey on trade repositories, 2018

The second insight is that access to data depends on the location of the trade repository. It should be noted that central bank rarely collects transactions data themselves. Actually, in most jurisdictions central banks are not in charge of this task, which is ensured by separate TRs that can be located within the central bank's jurisdiction or in another jurisdiction. Since TR data are very confidential, access for central banks that are not directly tasked with collecting them has not been straightforward. Overall, access problems are greater when the recognised TRs are located outside a central bank's jurisdiction.

The third insight is that access to highly granular data collected by TRs outside a central bank jurisdiction is particularly problematic. Not only access may be limited, but also arrangements can be complex - central banks often need to access through another authority.

Overall, the survey reveals that central banks face limits to access data collected by trade repositories. This suggests that there is scope to remove barriers to access, and effectively use it.

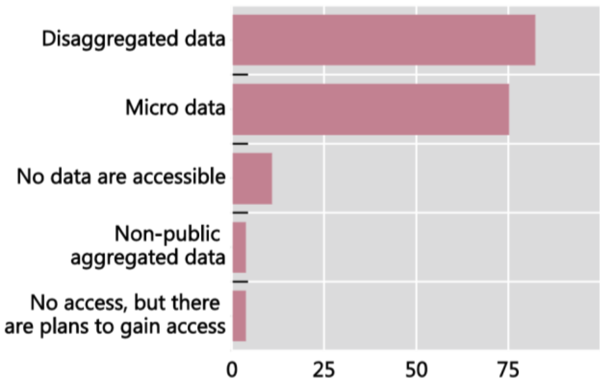
defined as *"data below the level of aggregated data and with a higher likelihood of identifying individual reporting units than in the aggregated data"* (IAG (2017)).

Central bank access to trade repository data

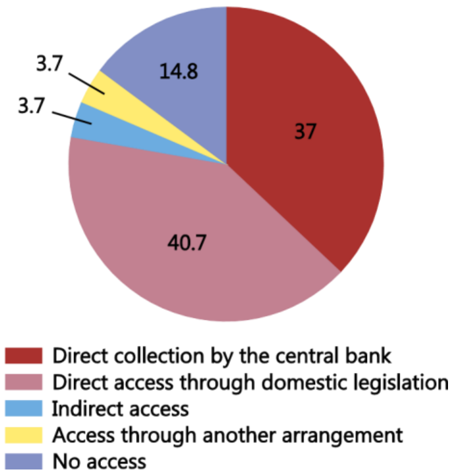
In per cent Graph 2

A. Trade repositories based in the jurisdiction

Does your central bank have access to non–public derivatives data reported to TRs/TR-like entities (other than the central bank)?[1]

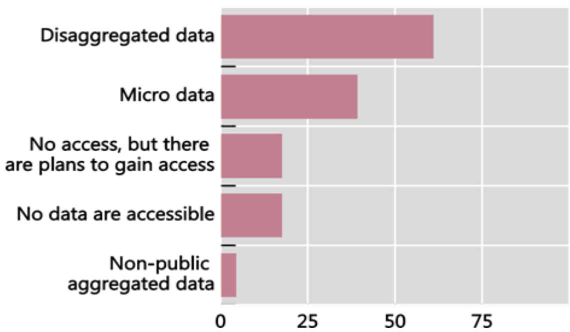


What types of arrangement govern your access to non-public derivatives data collected by TRs/TR-like entities?[2]

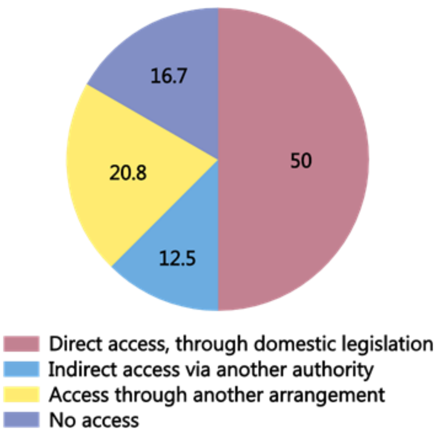


B. Trade repositories based outside the jurisdiction

Does your central bank have access to non–public derivatives data (involving a domestic counterparty), when they are reported to TRs/TR-like entities?[1]



What types of arrangement govern your access to non-public derivatives data (involving a domestic counterparty) collected by TRs/ TR-like entities?[3]



[1] The sum of the responses can exceed 100%, as several answers are possible. [2] Indirect access via another authority (ie by request of the primary authority with oversight of TRs/TR-like entities) [3] Direct access through domestic legislation (since domestic counterparties can report derivatives transactions to TRs/TR-like entities based outside my jurisdiction); Indirect access via another authority (ie by request of the primary authority with oversight of the TRs/TR-like entities)

Source: IFC survey on trade repositories, 2018.

3. Trade repository data sharing

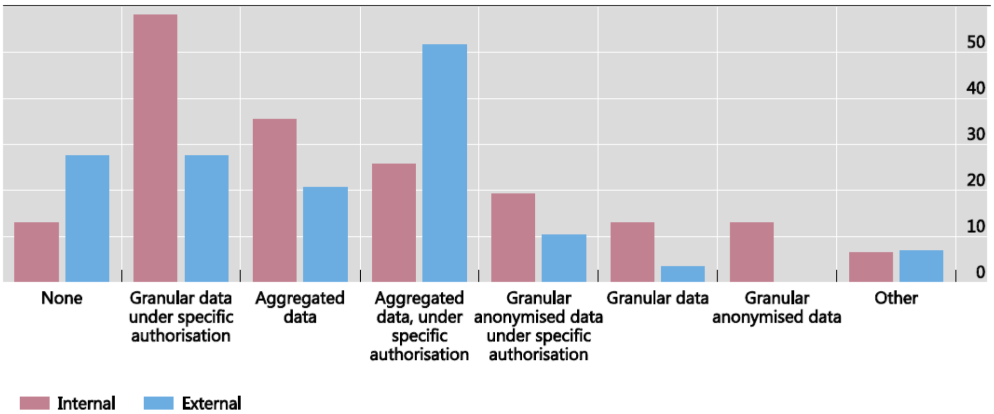
The survey reveals that data sharing is easier within central banks, and at higher levels of aggregation. Central banks' ability to share granular data with external users is very constrained.

Graph 3 summarises central banks' data sharing. In most instances central banks can share some data with both internal and external users. However, sharing of granular data with external users is very restricted. Almost half of the respondents reported that this is not possible at all. Data-sharing with external users is restricted mainly to aggregate data, and often requires specific authorisation (Graph 3). The picture looks brighter for internal users. Data sharing of granular data is allowed in more than 50% of the central banks (although requires specific authorisation, and in some instances data shall be anonymised).

What type of non-public derivatives data can the central bank share with users?[1]

In per cent

Graph 3



The sum of the responses can exceed 100%, as several answers are possible.

[1] Granular data refers to micro or disaggregated data (see Box 3).

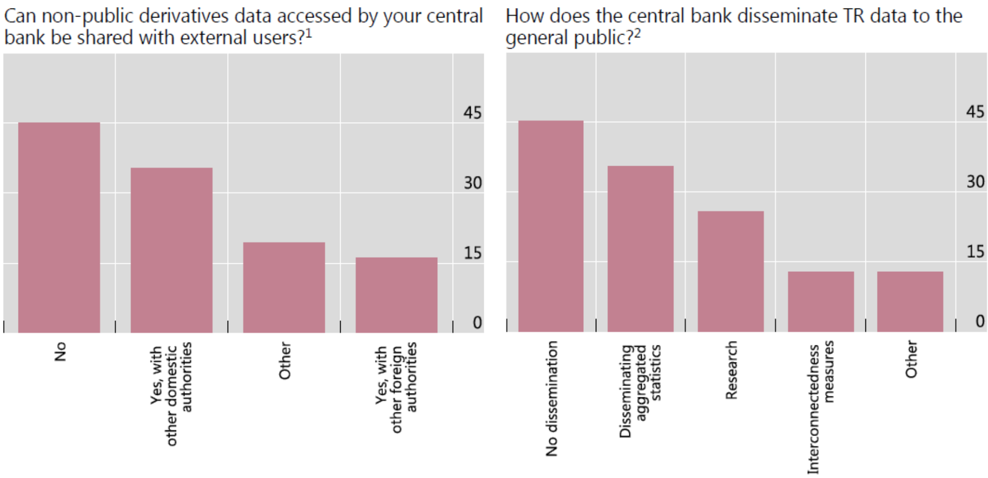
Source: IFC survey on trade repositories, 2018.

Graph 4 provides further details on central banks' ability to disseminate (share) non-public derivatives data with external users. In half of the instances, central banks do not share data at all. Data sharing with other domestic authorities is relatively frequent (30%), but sharing with foreign authorities is rate. Dissemination of (aggregated) TR data to the general public is also rare (Graph 4, right-hand panel). Nearly half of the central banks do not disseminate data at all, and the rest publish aggregated statistics, and analyses. Some central banks conduct careful checks to prevent the identification of individual counterparty data, for instance.

What access and type of non-public derivatives data is shared with external users and how are these data disseminated to the general public?

In per cent

Graph 4



The sum of the responses can exceed 100%, as several answers are possible.

[1] Authorities include supervisory and regulatory authorities. [2] Aggregated statistics include notional amounts, number of transactions etc (see Box 3); research refers to analyses/research conducted using derivatives data.

Source: IFC survey on trade repositories, 2018.

4. Discussion and Conclusion

Trade repositories are collecting a large amount of derivative transactions data. Central banks are facing challenges to access the data, and to share it with other stakeholders. Data access is particularly limited when it is held outside a central banks' jurisdiction. Sharing data to external users is also rare. In both instances, the main constraints are on access and dissemination of highly granular data. More work is needed to unlock data access and sharing, since granular data are required to assess financial stability risks.

References

1. Irving Fisher Committee (2018), "Central banks and trade repositories derivatives data", IFC Report
2. Inter-Agency-Group on Economic and Financial Statistics (IAG) (2015): Update on the Data Gaps Initiative and the Outcome of the Workshop on Data Sharing, March.

Index

Index

ISIWSC2019

Organised by :

DEPARTMENT OF STATISTICS MALAYSIA
MINISTRY OF ECONOMIC AFFAIRS

BANK NEGARA MALAYSIA
CENTRAL BANK OF MALAYSIA

MALAYSIA INSTITUTE OF STATISTICS

Supported by:

MINISTRY OF TOURISM, ARTS & CULTURE MALAYSIA

Malaysia Convention & Exhibition Bureau

Visit Truly Asia Malaysia 2020

ISBN 978-967-2000-61-7

#ISIWSC2019