eIncome Register in Denmark

– Advantages and Constraints of Using a Unique New Tax

Data Source in Official Statistics

Duer, Karin Holst
Statistics Denmark
Sejrøgade 11
2100 København Ø, Denmark
E-mail: khd@dst.dk

Statistics Denmark uses to a wide extent data that have already been collected in the private as well as the public sector. Consequently, the bulk of social statistics compiled by Statistics Denmark is based on administrative data.

This development is due to several conditions:

- Attention has increasingly been focused in recent years on reducing the burden imposed on businesses when they are obliged to participate in questionnaire-based surveys.
- Good experience with using administrative information for statistical purposes.
- Conducting questionnaire-based surveys is a very resource-demanding task, and subsequently relatively great savings can be gained by using existing data in the best possible way.
- Widespread use of standardised registration units/observation units, e.g. in registers.

Statistics Denmark is continuously making attempts at:

- ❖ improving the quality of the statistics,
- ❖ reducing the response burden and
- ❖ using the resources available to us for compiling statistics in the best possible manner

The use of administrative data contributes in a constructive way to reaching all of the 3 targets, and in consequence of this, Statistics Denmark is continuously making efforts to include external as well as internal data sources, which are instrumental in this process.

Register policy in Denmark

Since 1968 Denmark has operated a register of all persons who are living or who have lived in Denmark, called the Central Population Register (CPR). The CPR is updated daily on the basis of the population registers operated by each municipality in which the citizens live. Information is passed on to the municipal population registers by the different authorities that are of relevance. All registered persons are allocated with a unique civil registration number consisting of 10 digits, which identifies each individual person.

It is common practice for the individual citizen to state the civil registration number (CPR number) when contacting a bank, physician, dentist, library, or engaged in a credit sale, etc. Consequently, it is possible to obtain a highly detailed picture of each individual citizen by combining data from various registers, using the

civil registration number as key.

For the purpose of protecting the individual citizen, legislation relating to registers is very restrictive, which implies that is unlawful to link registers containing personal information. An authority (Data Surveillance Authority) is responsible for ensuring that register legislation is not violated. Citizens and businesses can apply to the Data Surveillance Authority, if they have any complaints against the contents or use of the registers.

Statistics Denmark has been given special statutory status with regard to gaining access to all data. In accordance with the Act on Statistics Denmark, businesses, public authorities and institutions are obliged to submit information as required by Statistics Denmark. In return, Statistics Denmark is under an obligation that the data collected are used for statistical purposes only and that the data are treated confidentially.

The purpose of Skat's (tax authorities) new data source – eIncome

Information about the income of citizens is included in a wide range of contexts, e.g. for calculating income tax and for claiming public transfer payments (unemployment, housing subsidy, social assistance, etc.). Previously, citizens and employers had to report the same information on income several times, e.g. when a citizen applied for both unemployment benefits and housing subsidy, as the benefits are dependent on the citizen's income, but the information was stored in various registers.

In 2005, the Danish government decided to establish a central register containing monthly information on income. The register was to be electronic and designed in such a way that it would not put an unnecessary burden on employers. The general purpose of establishing the new register was to enhance the public administration's efficiency and lay down the basis for simplifications of legal rules. The establishment of the register would thus result in a number of advantages:

- Improved services for the citizens because the case administration in the public welfare system is conducted on the basis of more up-to-date and accurate information.
- Simplifications of legal rules and administrative relaxations, because information is to be reported to the same register and each single item of information is to be reported once only.
- Better possibilities of conducting checks, e.g. with regard to unemployment benefits, because information about each individual citizen is available from the same register and can be accessed by one single entry.
- Electronic tax cards involving a continuous calculation of total wages and salaries and tax withdrawals.
- Compilation of statistics and possibility for the tax authorities to calculate the kilometre allowance (commuting to and from work) as the necessary information is available (information on place of residence of the employee and address of the workplace).

Act on eIncome

The Act on an income register was adopted in 2007, and subsequently data for the new register were to be reported for payments with effect from 1 January 2008. The Act is administered by the Danish tax

authorities.

From the date of when the Act came into effect, all public and private employers as well as public authorities paying out money to citizens at least once every month must report detailed information on the size and the type of payment. Until 2008, the same information were, roughly speaking, to be reported to the tax authorities, but only once annually for the purpose of the yearly tax assessment. The new possibilities with respect to the Act on eIncome in relation to the yearly tax assessments of earlier years to the tax authorities were, in particular:

- Reporting must take place at least once every month, and the period covered by the payment must appear from the data reports.
- In the case of payment of wages and salaries, the number of hours worked for which wages and salaries are paid out must be reported.
- In the case of payment of wages and salaries, the workplace to which the employee is linked must be stated.

eIncome constitutes an electronic solution which enables reporting of data, but it is also possible to report using paper-based questionnaires. However, this possibility is only used by a very few. Most respondents use a solution by which data are generated in the organization's wage and salary system or payment system.

Access to Skat's data on eIncome

Skat's (tax authorities) new register contains sensitive personal data and is subject to current legislation, i.e. access is highly restrictive. As mentioned earlier, Statistics Denmark has in accordance with legislation access to all data in the new register. Others may gain access to data, if this is explicitly authorized by law and access is only gained to necessary, specific information in accordance with the principle "need to know" and not "nice to know".

The authorities are legally obliged, to the greatest possible extent, to collect the necessary information from the register, instead of collecting the information from the employers, the citizens themselves, other authorities, etc. This is to ensure that employers and citizens are not burdened more than is absolutely necessary with respect to requirements of information by the authorities.

Contents of eIncome

eIncome contains monthly reports by Danish employers, authorities and citizens' organizations. The payments give specified details with respect to size and type of payment.

Out of the Danish population of 5.5 mio persons in 2010, 2.8 mio jobs were held by 2.5 employees. Employers must report monthly information on each individual employee, if merely one hour of paid work has been performed. A total of about 4.4. mio data reports from employers is registered in eIncome, i.e. more than 50 mio reports annually. The information is typically generated by means of the employers' wage and salary systems, and the information is, consequently, roughly the same which appears from the employees' pay slips, that is:

- Reference period of the amount with start date and end date
- Number of paid hours of work
- Amount of wages and salaries
- Value of fringe benefits distributed by type (car, board and lodging, gifts, multimedia, etc.)
- Workplace.

On the basis of eIncome it is thus possible to follow each individual employee from job to job over time as well as identifying all paid jobs held by the person in question.

With respect to transfer payments, about 5 mio data reports every month, corresponding to 60 mio reports annually, were registered in 2010. In 2010, the 60 mio reports involved about 2.2. mio persons, who received transfer payments liable to taxation.

A total number of about 10 mio data reports is registered monthly in the eIncome register, which result in vast amounts of data that are to be treated, as the reporting system is, at the same time, very detailed with more than 100 different variables. 85 pct. (2009) of total income is reported to and registered in eIncome.

eIncome DOES NOT contain…

Consequently, eIncome does not cover 15 pct. (2009) of the total income. These 15 pct. relates to:
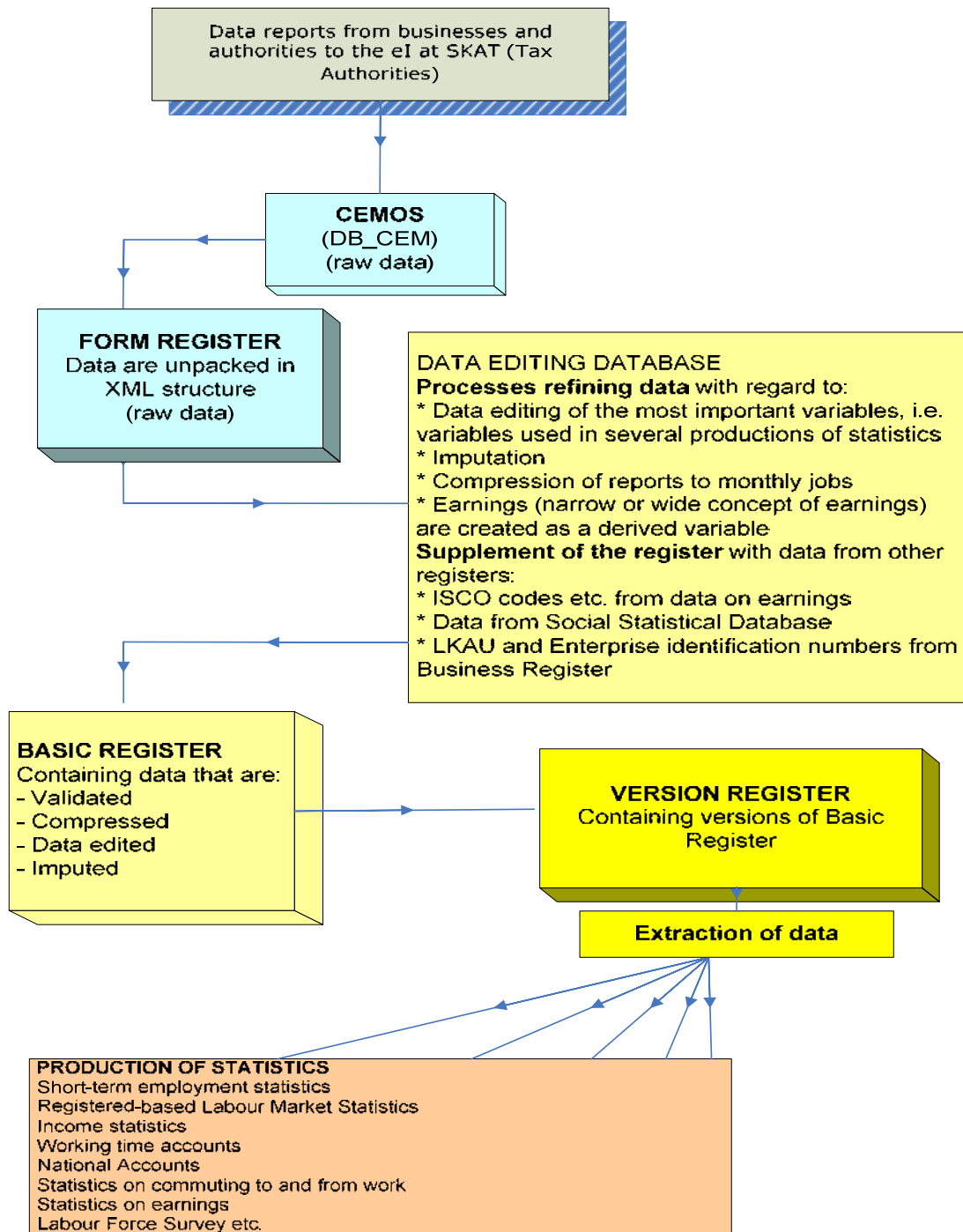
- Income from self-employment and as assisting spouse
- Income/payments from abroad
- Income from shares and capital income
- Pension payments not administered by employers.

Statistics Denmark's eIncome Register System (eIR)

Statistics Denmark receives daily all reports from Skat's eIncome. The data are stored in the form of raw data, which implies that they can, at any time, be accessed again. Subsequently, the data are refined by Statistics Denmark for the purpose of making them suitable for the eIR and thereby production of statistics. This implies, e.g. that the sender and recipient of the transaction have to be transformed into statistical units and accurate reference periods must be linked to the data on a monthly basis, so that it will be possible to compile short-term statistics from this basis.

Another principle in the eIR is to enrich the register with data from other sources at Statistics Denmark to the extent necessary. This information is integrated in the eIR for the purpose of using the information for validation in the eIR or to linking information to the eIR data, if the information is to be used in several subsequent productions of statistics. In this way, it is thus ensured that the same data are used with greater consistency between the resulting statistics.

Consequently, Statistics Denmark has set up an advanced system for receipt, refinement and editing of the data submitted by the Danish tax authorities, which can be seen from the figure below.

Data reports from businesses and authorities to the eI at SKAT (Tax Authorities)

CEMOS
(DB_CEM)
(raw data)

FORM REGISTER
Data are unpacked in XML structure
(raw data)

DATA EDITING DATABASE
Processes refining data with regard to:
* Data editing of the most important variables, i.e. variables used in several productions of statistics
* Imputation
* Compression of reports to monthly jobs
* Earnings (narrow or wide concept of earnings) are created as a derived variable
Supplement of the register with data from other registers:
* ISCO codes etc. from data on earnings
* Data from Social Statistical Database
* LKAU and Enterprise identification numbers from Business Register

BASIC REGISTER
Containing data that are:
– Validated
– Compressed
– Data edited
– Imputed

VERSION REGISTER
Containing versions of Basic Register

Extraction of data

PRODUCTION OF STATISTICS
Short-term employment statistics
Registered-based Labour Market Statistics
Income statistics
Working time accounts
National Accounts
Statistics on commuting to and from work
Statistics on earnings
Labour Force Survey etc.

Schematic diagram of the eIR system

Quality of the information

The Danish tax authorities only conduct editing, to a very limited extent, of the data reported to the eIR in connection with receipt of the data. During the start-up phase of the register, efforts were made to ensure that the great number of new respondents was burdened, as little as possible, in their attempts to submit the legal information to the eIR, and consequently only absolutely necessary checks were to be conducted.

An important precondition of successful electronic solutions is that it is easy and fast to report data. This implies, e.g. that response times must be kept to a minimum. Functionalities, whose purpose is to check the

information reported and return error messages, give rise to considerable delays in the performance, and this is the reason why the Danish tax authorities have only, for the time being, implemented a small number of checks, which are to be taken into account by the respondent.

It has been decided by Statistics Denmark to base its eIncome Register on Skat's raw data and consequently to conduct data editing. The reasons for Statistics Denmark's own data editing is primarily due to the circumstance that Skat has not implemented validation and data editing to a greater extent, and that Statistics Denmark's own quality requirements with respect to the data are subjected to Statistics Denmark's own targets for the data. Furthermore, it is also necessary to conduct data editing as quickly as possible, if they are to form the basis for production of short-term statistics.

Advantages of the eIncome Register

The use of the eIncome Register gives rise to a number of advantages for Statistics Denmark:

1. Use of the same data source as basis for many statistics results in a better quality in contrast to using various data sources for compiling statistics.
2. It is easier to draw comparisons of data across the statistics.
3. The response burden is reduced.
4. Statistics Denmark's workload is reduced when a great variety of data sources are substituted by one source and therefore subject to one data editing.
5. Cooperative possibilities between the Danish tax authorities and various stakeholders with respect to data editing imply that the administrative burden is reduced and the data quality is improved.
6. The eIR can be combined with data from other registers maintained by Statistics Denmark, which open up numerous possibilities of carrying out analytical work.
7. Longitudinally based database opens up a wide range of possibilities of compiling new statistics.

The advantages, which are not listed in a prioritized order, will be briefly discussed below.

re 1. The same data source as basis for many statistics:

Statisticians are familiar with the problems involved in combining information from various sources, which are due to heterogeneous observation sources, different definitions of various concepts, different data-editing methods used with regard to the registered data, etc. Using data from the same source implies that this type of bias is eliminated, which improves the quality of the statistics compiled.

re 2. Easier to draw comparisons of data across the statistics:

The more homogeneous the primary data on which the statistics are compiled the greater the consistency when comparisons are drawn across the statistics. Consequently, Statistics Denmark will make efforts to ensure that the greatest possible part of the editing of eIncome data is conducted *before* the primary data are made available to the producers of the respective statistics. In this way, it is ensured that the primary data remain the same, as far as possible in the production process of the statistics.

re 3. Reduction in the response burden:

The general purpose of eIncome is that employers, etc. who previously were obliged to submit information on income of citizens to different authorities and registers are now obliged to report information only once. This reduces the response burden considerably and thereby gives rise to less irritation.

re 4. SD's administrative workload is reduced:

It has been for many years the policy of Statistics Denmark to make use of already collected data as far as possible. When using already existing data, which have been not been specifically collected for a concrete statistic, there will almost always be a need to adapt the data, to a smaller or lesser degree, to the purpose of the statistic in question. This work, involved in the refinement of the data is frequently very comprehensive, also because many of the statistics are based on various data sources. When data are derived from the same source, it will be possible to standardise a wide range of processes in connection with adaptation of the data to the concrete statistic and the efficiency of the production process is thereby enhanced. For many statistical productions, it is frequently an obvious possibility to conduct data editing only once in order to save resources.

re 5. Several users of the same sources result in improved data editing:

The data are owned by the Danish tax authorities, and consequently they decide how data are reported, how many quality checks are conducted in the reporting process, when the respondents are informed of incorrect information and when reminders are sent out, etc. Like other users of Skat's eIncome data, Statistics Denmark can inform the respondents of any incorrect data reports as well as the Danish tax authorities in the case of systematic errors in the data received from Skat's eIncome. The joint data editing is instrumental in ensuring the data quality.

re 6. Combination of eIR with data from other registers maintained by Statistics Denmark:

The eIR contains information about transactions whereby the authority or employer paying out an amount can be homogeneously identified by means of Statistics Denmark's Business Register and the payee is identified by means of his/her homogenous civil registration number. Both identifications are incorporated in a great number of other registers maintained by Statistics Denmark, and thereby it is possible to combine data from various registers for conducting different analyses of statistics. Some of the major registers that can be used in combining data are the Social Statistical Database and the Building and Dwelling Register:
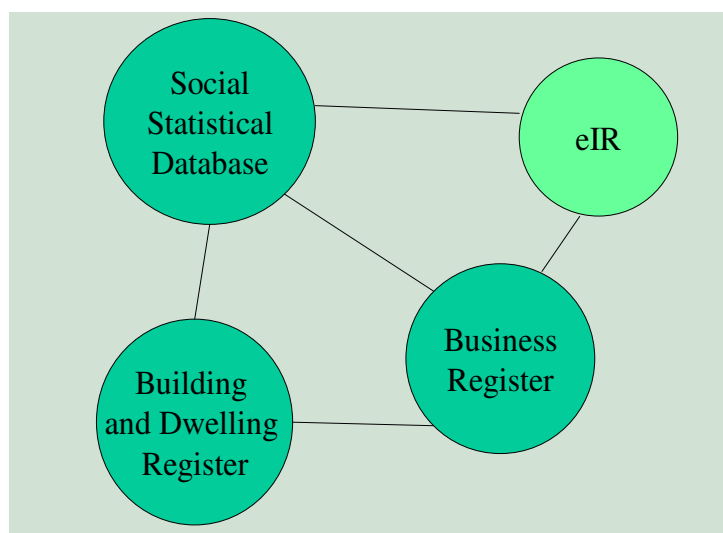
The Building and Dwelling Register was established in 1976 and contains a homogeneous registration of all properties and buildings in Denmark. By means of addresses it is possible to link data from this register to the place of residence of the persons and workplace of the businesses.

re 7. *Longitudinally based database opens up numerous possibilities of compiling new statistics:*

Information on income of citizens must be reported at least once every month and with indication of the precise period covered by the payment. This implies that it is possible to follow a person over time in the labour market as employee as well as outside the labour market, if the person in question receives a transfer payment. Statistics Denmark has initiated work on developing a labour market account, which is a longitudinally based database, where data from the eIR are combined with data also from a longitudinally based database containing persons, who are outside the ordinary labour market as well as data on self-employed persons and assisting spouses' activities.

This opens up a wide range of possibilities to analyse various socioeconomic areas. Some examples are:

- Development in the number of jobs with indication of number of hours worked, which is added to each individual job distributed by geography, sector and industry.
- Development in the employment volume (full-time equivalents) for employees distributed by geography, sector and industry.
- Longitudinal analyses showing (by sector, industry, etc.) where persons, who become unemployed, come from.
- Job mobility between sectors, industries, etc.
- Longitudinal analyses of persons receiving different types of transfer payments (e.g. social assistance) with regard to how long time it takes before they get a job, which type of job, and for how long they remain in the job in question.
- A vast amount of resources is used on job training programme for unemployed persons. eIncome opens up the possibility of analysing the consequences by following unemployed persons participating in training programmes with a view to measuring the usefulness of the efforts made in this respect, e.g. whether the persons participating in training programmes get a job, and if this is the case, which type of job, and for how long the persons remain in the job.

- Longitudinal analyses on the basis of different educational groups, e.g. with regard to how many persons get a job, respectively receiving a transfer payment and how it develops over time.
- Analyses of the activity rate of pensioners, e.g. conducted on the basis of job experience and education of the persons.
- Analyses of the consequences of concrete labour-market policy measures.
- Analyses of attachment to the labour market of students during their study time and subsequently.
- In Denmark, all full-time students are entitled to claim education grants. eIncome opens up the possibility of analysing the consequences of the persons' activity status, following the conclusion of their training and education.

Disadvantages of eIncome

As it appears from the above-mentioned, there are many advantages derived from eIncome. However, there are also some disadvantages:

      A. Data have not been collected for statistical purposes

      B. Skat (Tax Authorities) is the owner of the data, and thereby deciding the design of eIncome

      C. Vast amounts of data give rise to technical problems

      D. When one major single source is used for compiling almost all statistics on employment, Statistics Denmark is highly dependent on a source providing a reliable and secure operation.

Below, the disadvantages are briefly discussed:

re A. Data are not collected for statistical purposes:

It is frequently the case that when statistics are based on already collected data, the data do not quite contain the information as desired, because they are collected for other purposes. As far as eIncome is concerned, there is particularly one thing that gives rise to challenges/problems in connection with the production of statistics, i.e. the choice of registration unit. The payee is identified by means of his/her civil registration number, implying that this does not result in any problems with regard to the production of statistics. In contrast, the business, organization or authority liable to report data are identified by means of an administrative unit, which has been "invented" by the Danish tax authorities and is widely used in connection with tax registration of exports/imports, VAT settlement, etc. The businesses are free to create as many registration units as they need, resulting in a number of problems for Statistics Denmark, when the statistical units are to be identified over time.

re B. Owner of the data/data are decided by Skat:

It is (naturally) the owner of the data who decides the design of platform, the formats and form in which the data are reported, prioritization of developing new functionalities, etc. and it is natural that Skat too has its own focus. One consequence of this is that Skat changes the data format over time because, for example, new rules governing taxation and new objects for taxation are found. It is thus necessary for Statistics Denmark to adapt the new versions, which is a resource-demanding task. Another consequence is Skat's method of data collection, which gives rise to technical challenges for Statistics Denmark. There are, e.g. 3 different ways in which a respondent can choose, when an incorrect data report is to be changed. Another and

major problem for Statistics Denmark is that in 2006 it was decided to use a common data format, when the authorities are exchanging data. This was a sensible decision, but the common data format (XML) that was chosen, and which is suitable for single entries in Skat's register, has proved to be a very inexpedient decision, when vast amounts of data are exchanged. This has given rise to technical problems for Skat with respect to submitting vast amounts of data in XML, and consequently data have until now been submitted in another format. It is the hope of Statistics Denmark that we will continue to avoid receiving data in XML.

re C. Vast amounts of data gives rise to technical challenges:

The vast amounts of data that are involved in eIncome constitute in themselves a major challenge. The eIR now takes up 800 GiB, and although disc space is cost-demanding and requires backup capacity, etc. the most serious consequence is the very long time a computer takes to run a programme. At the moment, it takes about one month to run the monthly data through the eIR system. This is a long time when data are to be used as basis for a subsequent production of short-term statistics. Consequently, very high priority has been given to the task of reducing the time it takes to run a programme and to improve performance.

re C. One single source results in vulnerability:

The more statistics that are based on the same source the greater the vulnerability is with regard to errors and production problems involving entry and refinement of the data in the eIR. If a problem arises in the running of the eIR, and it gives rise to delays, all statistics based on the eIR will be affected . It will not be possible to immediately substitute data with other sources, due to methodological differences and resulting data breaks in the statistics. Similarly, any errors in the data made available from the eIR for the production of statistics will be implemented in all statistics using the data in question. This vulnerability implies that it is extremely important to set aside resources for a carefully prepared system, which is subjected to thorough tests, and where there are appropriate possibilities of support with respect to the technical staff and where the data quality is continuously and thoroughly checked.

Conclusion

It must be concluded that great advantages are gained by Statistics Denmark in using Skat's new source. The very detailed monthly information at the level of individuals and with indication of employees' workplace is a goldmine with respect to the possibilities of compiling statistics. The short-term statistics as well as the structural statistics in the area of employment are to be based on the eIR as source, which opens up the possibility of using more standardised concepts and thereby improving the consistency between the statistics. However, it is also important to take into consideration that very comprehensive administrative efforts are a precondition before the primary data can be used for producing statistics. Furthermore, it is also necessary to spend time on converting the data so that they are manageable for the production of statistics, together with the technical and maintenance work necessary, and the data have to be thoroughly subjected to editing, as the source is of great importance in the subsequent processes of statistical production.

REFERENCES (RÉFERENCES)

Publication: Indkomster 2008, Danmarks Statistik, http://www.dst.dk/pukora/epub/upload/15218/indk.pdf

Treasury Dept.: http://modernisering.dk/da/projektside/andre_projekter/eindkomst/baggrund

Tax Authorities: http://www.skat.dk

RÉSUMÉ (ABSTRACT)

In 2007, the Act on an Income Register was adopted in Denmark. The Act is administered by the Danish tax authorities. The law requires that all public and private employers as well as public authorities report income payments to citizens at least once a month. The reporting includes detailed information on the size of the amount, type of the payment, where the employee is working and the number of working hours covered by the payment. Type of transfer payments is also registered.

Collection and storage of these data in one single register is a storehouse of information for statistical purposes and opens up a wide variety of possibilities. Access to the wide range of sensitive personal data is strongly restricted, but Statistics Denmark has special legal access provided that data is used for statistical purposes only.

Having access to a wide range of income data collected in one source gives a qualitatively great advantage as the statistics produced on the basis of this source become comparable. This is of special interest in this case as the eIncome Register will become the statistical source for short-term as well as long-term statistics.

However, as data are not collected by the Danish tax authorities for statistical reasons, it is necessary to spend resources on refining the data to make them suitable for producing statistics. Moreover, the data are available in very large quantities, including corrections to former reporting of data which provide special challenges.

Although there are considerable constraints of using one huge source of data the advantages by far exceed these. Statistics Denmark has been working for a couple of years on making the data available in a smart way for producing statistics and this work is hardly finished. However, the expectations are that eIncome is an indispensable source of data in the future.