

The return of the mandatory census long form : end of the statistical confidence crisis in Canada ?

Jean-Pierre Beaud, CIRST, UQAM

1. Systems of credibility, labels and controversies

In today's world, figures, numbers, graphs, tables and statistics do represent the most powerful weapon for convincing people. I recently attended a conference on diseases and their relationship to environment. The speaker, who presented a radical view on the effects of mercury, electromagnetic fields and pesticides on health, overwhelmed the audience with graphs, tables and figures. The proof was in the figures and the trends they revealed. Some of the figures and facts came from official producers like the World Health Organization; others were the results of some academic researches or medical experiments. Sure, it was not easy for the audience to appreciate the accuracy of such figures, but it gave the thesis an apparently solid background. One knows that statistics are the result of series of choices, interactions, demands, that they represent a compromise between what is desired and what is possible, financially and technically, and above all that it depends on the capacity and on the willingness of people to understand what the question means (if it is the result of the passing of a questionnaire). In health statistics, for sure, some facts do heavily weigh on the figure (you have or do not have a cancer; your age is something quite solid); but your health history is much more evanescent and depends on memory and strategies of representation. Furthermore, in order to understand what those figures and graphs signify, one needs to have not only a certain level of literacy but also a certain level of numeracy. And as stated by our speaker, the effects of environment on QI could be to move the Bell curve drastically on the left. The intelligence, he said, is in danger!

The weakness of reasoning based on non-official figures is that, as the term says, it is not official. It means that you do not have a label that indicates clearly that it is the result of a serious investigation. Thus, the name of the researchers, the methods used, and the journal where the results are published will indicate the value of the figures. But one must have a solid and up to date knowledge of the field (health, international relations, and so on) to be able to appreciate quality and trustfulness. In some fields and for some researches, the value could be as high as gold, but for most of the people in the society, it is quite impossible to decide. So one generally needs the intervention of someone or of some organization to decide if it is good or not, or to rely on a private label (an ISO norm, a ranking) or on professional markers like university titles. Our speaker, a well-known physician and cancer specialist, distilled many marks of his notoriety in his talk so as to solidify the architecture of his reasoning and ultimately win public support. In such a case, thus, agreement with the discourse of a speaker depends on things that are not only linked with the quality of data. But if there is something that destroys an argument or even the whole reasoning of a person, the major part of the system of credibility of the field is preserved.

For national statistical institutes, the pattern is much different and much the same. It is much the same because one never has the ability to evaluate the quality of data, except if you have the time and the capacity to read and evaluate the methodological appendix that accompanies the institute research. The difference with my speaker is that Stats Can, for instance, is a very powerful agency with a history that speaks for it (the best statistical agency in the world in the 1990' according to *The Economist*) and resources (human, financial, e.g.) that impress (at least the agency has enough resources for an efficient communication). It is different because the national

institute produces official statistics. The label is provided by the state and therefore does not need to be presented every time, for every research. One trusts the agency, and so one trusts the numbers provided by it. It is also much different because if the credibility of an inquiry, for instance, depends on the credibility of the whole agency that conducts every day many inquiries and is daily deducted from it, a mistrust generated by critics about an enquiry can have an effect on all the enquiries, and all the work done by the agency.

To sum up, in the first case the system of credibility does not depend on every action generated by a person or a group. It is more decentralized than centralized. Even if all the actors would be university professors and therefore actors of a national scientific field, the divisions, the battles, the controversies are numerous and to a certain point necessary for the well being of such a field. For many sociologists of science, relative heteronomy is a precondition for the development for science, for instance. And competition (for rewards), whose rules are quite complex, leads to the proliferation of controversies (that can be viewed as the expression of disagreement). In the second case, the system of credibility is highly centralized and could be affected by any action that will generate a controversy. Here a controversy is not something that would guarantee high returns for the national statistical agency and for the system of credibility. On the contrary!

But it is necessary to distinguish scientific controversies and public controversies. If the first ones can be easily mastered and regulated (even by organizations like a statistical bureau, because they remain internal), the second ones could severely damage such organizations. In science, the recourse to non-scientific (or considered as non-scientific) arguments (like environment when it is considered as a general variable) is generally indicative of a move toward public debates. For national statistical institutes, a too visible and strong political pressure could generate the same change. That is what happened in Canada in 2010.

2. A context unfavourable to census practices

But, be the inquiry directed privately by a researcher like my speaker or officially by a state agency like Stats Can, it has to deal with a few trends according to questionnaires, respondents, and personal humours. Four main trends that are partially connected can be observed.

1. The decreasing percentage of respondents;
2. The fact that one should call upon the memory as little as possible and require as little as possible a calculation or an intellectual reasoning;
3. The developing susceptibility about personal liberty;
4. The low rate of literacy and numeracy.

When official statistical agencies developed during the 19th century, the mandatory condition seems to have been applied without having been discussed in fact, at least from a general point of view. The history of the mandatory condition certainly needs to be undertaken. If in Argentina, for instance, this condition is presented as essential by the head of the bureau of statistics of the city of Buenos Aires at the end of the 1880's, surprisingly no discussion on that topic can be found in the reports of the first sessions of the *Congrès international de statistique* (see Quetelet, Bruxelles, 1873). And the mention of that principle, for instance, appears in France only in 1951 (and seems to date back to 1938 with "décrets Sauvy"). Prior to that, nothing can be found

(Royer, *Statistique et Société*, 2-2016). In Canada, in the Census Act for the 1871 Canadian census (1870), it is clearly stated that if you refuse to give an answer you could be fined.

How to explain such a mess? I would say that it is possibly because of its logical obviousness that the principle of a mandatory enquiry is not necessarily mentioned or discussed. But it is also true that in the 19th and (beginning of) 20th centuries, the instructors had a strong asset on answers because they completed the questionnaire for the respondent. They were often seen like fiscal or taxation agents and like them apprehended with fear and suspicion. They could even change an answer (cf. instructions for the first Canadian censuses). That procedure was also a good way of taking into account the very low literacy rate of the population at that time.

But this is a global procedure that would appear today as very costly and, certainly, as a violation of personal integrity. Today, for many reasons, financial but not only, one accepts:

1. That people answer freely without any interaction with instructors (except for some small populations as Native people in Canada).
2. That you are right to give an answer that would even seem impossible from a logical point of view. For instance, one can choose any combination of answers to the Canadian question on visible minorities.

The 2010-2011 Canadian controversy appeared in such a context. But one has to add something much important; the fact that today the census is considered more and more as a costly, too slowly and old way of gathering information. It cumulates things that are now considered as problematic. It often calls upon memory, while administrative registers do record facts as close as possible to the moment of their production. It is also often too long, at least the Canadian long form with its 65 questions. One knows that answering to it could seem to some persons as a burden. The experts of the bureau are well aware of that, and methodological expertise is often in fact expertise on how to increase response rate, how to deal with unequal coverage of the whole population and how to improve quality of data according to the more or less goodwill of the potential respondents.

In those conditions, what about the quality of data?

3. The Canadian 2010-2011 controversy

At the end of June 2010, the conservative government announced that the mandatory long form administrated to a sample of 20% of the population is to be replaced by a voluntary questionnaire administrated to 30% of the population for the 2011 census. A public controversy arose during a period that usually is quite calm, the summer. From a statistical point of view, summer 2010 was very hot. The reasons presented by the Conservatives were essentially structured around the idea that a compulsory enquiry is problematic from the perspective of liberty, of private life, as shown by Clement's statement.

Standing Committee on Industry, Science and Technology

Tuesday, July 27, 2010

- [Hon. Tony Clement \(Minister of Industry\):](#)

- Thank you, Mr. Chairman and honourable members. I'm pleased to be here today to talk about the government's decision to replace the mandatory census long form with a voluntary form.
- Our government's reason for replacing the mandatory census with a voluntary national survey on the long form is clear. We do not believe it is appropriate to compel Canadians to divulge extensive private and personal information. We do not believe Canadians should be forced under threat of fines, jail, or both to divulge the answers to questions such as these: How many sick days did you take last year? Were you paid for those? What were your total payments for your primary dwelling last year? Do you have any broken floor tiles in need of repair in your bathroom?
- We recognize that the information gathered in the long-form census is valuable. However, we also recognize that a balance must be drawn when the government is collecting data under the threat of fines, jail, or both.
- Now, although the census goes back to 1871, the long-form census has only been around since 1971. The level of detailed personal information that Canadians are being asked to hand over to the government has grown. The basic eight census questions have remained constant for decades. The additional questions that suddenly appeared in 1971 have been modified, with new ones added each census.
- *[Translation]*
- The short-form census is mandatory and consists of eight mandatory questions on basic demographic information, such as date of birth, gender, marital status and mother tongue. This form will be distributed to all Canadian households.

As it is shown above, in Canada, the questions of the census must receive the approval of the government. It seems to be problematic as regards the independence of the statistical bureau with respect to politics. As Ivan P. Fellegi, who was in charge of Stats Can from 1985 to 2008, used to say: « it is not always easy to keep the statistical bureau at arm's length from the pressures of the political power ». But what was new in 2010, it is that apparently the political intervention was related as much to the form of the questionnaire as to its content. To replace a long and mandatory questionnaire (to eliminate it, if one wants) by an long and voluntary enquiry (the National Household Survey, NHS) could seem to be an intervention on the form and not on the content, and the form, which was considered as the technical side of the census, was for many a question of method and so the affair of the statistical bureau. It had never happen before. A widespread mobilization occurred from across the political spectrum, apart from the Conservatives. Churches, trade unions, business associations, university professors, journalists, the media industry, social groupings, and many citizens, protested, wrote letters asking for some explanation.

For the government, as Tony Clement said, it is to protect the Canadians, their freedom of choice, their privacy that the decision was taken, and not for financial reasons. For the opponents, it was a funny pretext and they gave no credit to that.

So what kind of decision was that? Was it an ideological decision? Was it a political one? Did the Conservatives want to reduce significantly the influence of a bureau that produces massive data for everybody? Or simply did they want to satisfy all those who find that there are always more and more controls, questionnaires to fill, bureaucracy? What can be said?

First, the strong reaction of the Canadian associations was a surprise. The government did not think that such a change would provoke such a mobilization. It must be said that the coalition against the abolition of the long and mandatory form was heterogeneous. Almost everybody suddenly became positivist! No more constructivist, postmodern, radical analyses of statistics. The criticisms of the use of numbers had been for a certain time forgotten. Strange! But it is also true that the survival of the government could not be affected by such a decision. And the next general election was to take place a pretty long time after the 2011 census. Afterwards, Minister Tony Clement, who orchestrated the debate on the Conservative side, admitted, however, that the decision had not been successful and that it would have been preferable not to impose it. "Looking back, yes, I would have done things differently, absolutely [...]. I think we could have had a broader debate on Data collection in our society [and ask] what is the best way to do this to strike a balance between respect of Privacy and reliability of the data. Rather than just making that decision, we could have had a broader debate. [...] I take the blame. I should have asked the question six years ago". November 6, 2015.

It is certainly important, now that the debate is over, as we will see below, to listen carefully to what Clement said and to view it in parallel with the ethical principles of the ISI, for instance, but as Bodin shows, many other statistical organizations share the same ethical principles as for the protection of liberty, the refusal of intrusive methods and questions (United Nations, Eurostat, for instance).

12. Protecting the Interests of Subjects

Statisticians are obligated to protect subjects, individually and collectively, insofar as possible, against potentially harmful effects of participating. This responsibility is not absolved by consent or by the legal requirement to participate. The intrusive potential of some forms of statistical inquiry requires that they be undertaken only with great care, full justification of need, and notification of those involved. These inquiries should be based, as far as practicable, on the subjects' freely given, informed consent. The identities and records of all subjects or respondents should be kept confidential. Appropriate measures should be utilized to prevent data from being released in a form that would allow a subject's or respondent's identity to be disclosed or inferred.

For the Canadian statistical profession, summer 2010 was a very difficult period. From a mathematical or probabilistic point of view, it is clear that a sample constituted of persons who accept to answer (it means without the ones who refuse) is problematic according to representativeness and certainly such a strategy would produce a response rate much lower than the usual strategy of Stats Can with a sample associated with the mandatory principle. And the statisticians of the bureau were well aware of that. The Statistician in Chief at that time, Munir Sheikh, was under pressure. How could he accept such a decision that would have strong bad effects on the quality of data, on the representativeness of certain groups and regions? The comparability through time and space would be problematic. In July 2010, Munir Sheikh eventually resigned and declared in in press release:

"I want to take this opportunity to comment on a technical statistical issue which has become the subject of media discussion ... the question of whether a voluntary survey can become a substitute for a mandatory census. It cannot. Under the circumstances, I have tendered my resignation to the prime minister. He said he had opposed from the outset the abolition of the Long and compulsory questionnaire.

Statistics Act

R.S.C., 1985, c. S-19

False or unlawful information

31 Every person who, without lawful excuse,

- (a) refuses or neglects to answer, or wilfully answers falsely, any question requisite for obtaining any information sought in respect of the objects of this Act or pertinent thereto that has been asked of him by any person employed or deemed to be employed under this Act, or
- (b) refuses or neglects to furnish any information or to fill in to the best of his knowledge and belief any schedule or form that the person has been required to fill in, and to return the same when and as required of him pursuant to this Act, or knowingly gives false or misleading information or practises any other deception thereunder

is, for every refusal or neglect, or false answer or deception, guilty of an offence and liable on summary conviction to a fine not exceeding five hundred dollars or to imprisonment for a term not exceeding three months or to both.

- 1970-71-72, c. 15, s. 29.
- <http://laws-lois.justice.gc.ca/eng/acts/s-19/fulltext.html>

The 2011 Canadian census proceeded as planned with a short form, still compulsory and administrated to the whole population of Canada, and with a National Household Survey administrated to a sample of 30%. The results were better than expected, it means that the response rate was much higher than anticipated (almost 70% instead of 50% and 98.1% for the mandatory short form).

The general election of 2015 was a good surprise for the Liberals and for the opponents of the abolition of the long and mandatory form. One of the first decisions of the Trudeau government was to restore the long and mandatory form! The 2016 Canadian census was even administrated that way (as the 2006 Canadian census). And the results as for participation were number 1!

Stats Can High Marks

Canadians' overwhelming response enables 'best ever' Census in 2016

August 29, 2016

OTTAWA — The willing participation of Canadians has enabled Statistics Canada to achieve the 'best Census ever,' the agency announced today.

"The 2016 Census will provide high-quality information for virtually all communities across Canada," said Chief Statistician Wayne R. Smith. "We thank Canadians for making this the best census since 1666."

The overall collection response rate of 98.4 per cent for the 2016 Census of Population is higher than for both the 2011 and 2006 Census programs.

Canadians' response to the 2016 long form was simply outstanding. In 2016, the collection response rate for the long form was 97.8 per cent, the best ever recorded. This response rate will enable the provision of high-quality information for virtually all communities.

The 2016 Census also stood out in two other regards: self-response and Internet response. Almost 9 in 10 Canadian households completed their long or short form questionnaire without any assistance from Statistics Canada staff. The rate of 88.8 per cent makes this the most efficient among traditional censuses conducted in the world.

Response rates to individual questions are also high and very uniform throughout the questionnaire, further improving data quality.

As for the Internet response, Canadians delivered a gold medal performance with an Internet response rate of 68.3 per cent, surpassing the ambitious initial objective of 65 per cent and setting yet again another world record. High rates of self-response and Internet response contribute to both the efficiency of data collection and exceptional data quality.

Statistics Canada will now turn its focus to providing high quality Census information. An accelerated release schedule is planned to provide Canadians with results in a more timely fashion than for any previous census. First results from the 2016 Census will be published on February 8, 2017, with the release of data on population and dwelling counts.

As well, Statistics Canada continues to plan for future censuses. Consultations on census content for 2021 will take place in 2017. The agency will seek to further refine collection processes, with a view to ensuring the 2021 Census continues to improve inclusiveness, data quality and efficiency.

4. Is this the end of the statistical crisis in Canada?

With the Liberal decision to go back to the old census formula with a compulsory long form administrated to a sample of 25% (this is the only difference in fact with the 2006 census; the sample at that time covered only 20% of the population), it seems as if the crisis would be over. This is true and not so true.

Let examine a few trends that characterize the present situation as for enquiries.

1. The declining response rate

It is well known that the cooperation of people to enquiries is decreasing. It is true for surveys administrated by companies, research centers, private investigators. This is less true with questionnaires administrated through the web, but in that case the problem of the representativeness is crucial.

For a national statistical institute, the question is more about resources allocated to a good coverage of the population than that strictly of the response rate. If you do many reminder calls, you could get a high return, but is this a good allocation of resources, and will the data be automatically of good quality? As stated by Jelke Bethlehem, Fannie Cobben et Barry Schouten, « Many survey organisations focus on the response rate as being the quality indicator for the impact of non-response bias. As a consequence, they implement a variety of measures to reduce non-response or to maintain response at some acceptable level. However, response rates alone are not good indicators of non-response bias. In general, higher response rates do not imply smaller non-response bias » (“Indicators for the representativeness of survey response“, *Survey Methodology*, June 2009, Vol. 35, No. 1, pp. 101-113, Statistics Canada, Catalogue No. 12-001-X, <http://www.statcan.gc.ca/pub/12-001-x/2009001/article/10887-eng.pdf>). In the 2010 Canadian debate, that question of quality of data was mainly discussed through the prism of the response rate. A non-compulsory long form would provide a lower response rate (50% according to Stats Can before the 2011 census; almost 70% after many reminder calls as compared to the 94% for the 2006 long form) and it is true that the coverage is to be unequal and thus problematic as for decision taking. But, what for the real quality of data, it means the match between what people say or want and what the enquiry shows? For the 2010 Conservatives, a high return was not automatically associated with high quality and, worse, could damage that quality. That was not a so bad argument!

Surveys Face Growing Difficulty Reaching, Persuading Potential Respondents

	1997	2000	2003	2006	2009	2012
	%	%	%	%	%	%
Contact rate (percent of households in which an adult was reached)	90	77	79	73	72	62
Cooperation rate (percent of households contacted that yielded an interview)	43	40	34	31	21	14
Response rate (percent of households sampled that yielded an interview)	36	28	25	21	15	9

PEW RESEARCH CENTER 2012 Methodology Study. Rates computed according to American Association for Public Opinion Research (AAPOR) standard definitions for CON2, COOP3 and RR3. Rates are typical for surveys conducted in each year.

→

<http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/>

The national statistical institute, as one knows, possesses three main tools in order not to get into the same bad situation as the survey industry: first it is backed by the State with all what it implies; second it has a solid expertise on how to deal with disaffection for polls and enquiries (<http://www.statcan.gc.ca/pub/12-539-x/2009001/response-reponse-eng.htm>); third it can rely on the mandatory principle.

But if the last one is to be abandoned, the cost of collecting information, as the Canadian experience shows, could be very high.

To sum up, a high response rate is a good indicator of the prestige of the enterprise (of the fear or respect it generates) and of its efforts and thus needs to be envisioned, but the quality of data, which is the ultimate goal, is not reducible to it. Administrative records, if not error-free, could provide a better result in a period of growing distrust according to State demands.

2. The cost of memory and of calculation

The information structured on the basis of memory (facts or conditions of the past) is generally considered as critical. In the short form (Canadian census of 2016), only one question needed a return in the past: question 9, “What is the language that this person first learned at home in childhood and still understands? ». It is a not a very problematic question. In the long form, questions 23, “Where did this person live 5 years ago, that is, on May 10, 2011?”, 24, “Where was each of this person's parents born?”, 46, “How many weeks did this person work in 2015?”, 47, “During most of those weeks, did this person work full time or part time?”, and question F5 on the “period in which the building was completed”, needed a return in the past. Stats Can is well aware of that problem and that is why certainly those questions could be considered as not too problematic.

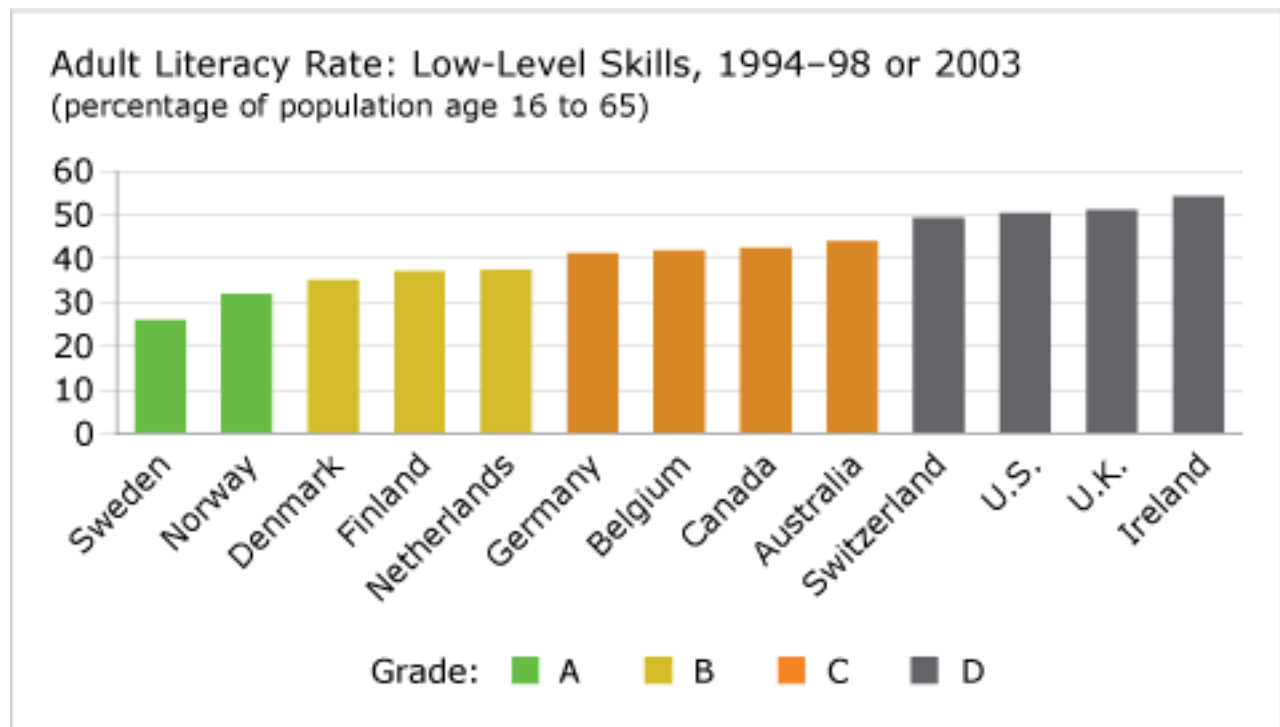
Another problem is related to the fact that some questions also need a kind of calculation. The

mixture of recourse to memory and calculation in a form constituted of 65 questions, some of them related to private matters, can be seen as really problematic.

3. The growing susceptibility about personal liberty

In France, a web page « Je suis libre, je dis non, je sors du rang, pas de recensement » was active during the 2016 French census and will reopen on January 1, 2017 for the 2017 census (in France, the method used is called rolling census). (<https://www.facebook.com/pages/Refus-de-repondre-au-recensement/1534761340083352>, <http://recensement.free.fr/>, <http://recensement.free.fr/ArgumentaireAntiRecensement.pdf>). Even if this is something quite confidential and if it is not sure that such a “movement” exists in Canada, some of the arguments in those pages are quite similar to the ones developed by the Conservatives. It says that in fact a lot of data could be obtained without the exercise of census, that some questions are about very private matters, that ... Such an attitude, I believe, will not decrease in the future, but, on the contrary, be more present. The electronic media can rapidly add some publicity to what are called “resistance acts” by people from the far left as well as people from the far right. What is interesting here is that for the French web campaign “pas de recensement”, I could not determine from where (left or right) came the propagandists!

4. The low rate of literacy and numeracy



Adult Literacy Rate—Low-Level Skills (chart)

The International Adult Literacy Survey (IALS) was a seven-country initiative—including Canada—first conducted in 1994. The idea was to create a comparative adult literacy rate for adults aged 16 to 65. Second and third waves of data were collected for 16 additional countries in 1996 and 1998, leading to a data set for 23 countries. Bermuda, Canada, Italy, Norway, Switzerland, the U.S., and the Mexican State of Nuevo Leon participated in another survey, the Adult Literacy and Life Skills Survey (ALLS) in 2003. The varying participation in the IALS and ALLS surveys means our comparison is based on the most recent test for each country. For most, this is the IALS survey, but for Canada, Norway, Switzerland, and the U.S., the results are from the ALLS survey.

Three categories of literacy were tested:

1. **Prose literacy:** the ability to understand and use information from texts such as editorials, news stories, poems, and fiction.
2. **Document literacy:** the ability to locate and use information from documents such as job applications, payroll forms, transportation schedules, maps, tables, and graphs.
3. **Quantitative literacy:** the ability to perform arithmetic functions such as balancing a cheque book, calculating a tip, or completing an order form.

In Canada...

42% of Canadian adults between the ages of 16 and 65 have low literacy skills.

55% of working age adults in Canada are estimated to have less than adequate health literacy skills. Shockingly, 88% of adults over the age of 65 appear to be in this situation

Impoverished adults often do not have the literacy skills required to get into job training programs. They may need literacy skills upgrading before they can succeed in training programs but only about 5 – 10% of eligible adults enroll in programs

<http://www.literacy.ca/literacy/literacy-sub/>

The problem is, do those people understand well some census questions?, even if Stats Can, well aware of those literacy statistics certainly tries to construct questions so as to be easily understood. A 65 questions form, such as the long form, could be a real burden for those persons.

If the crisis is now over, the context remains the same with all the problems it generates for a national statistical institute.

What about data obtained through census exercise compared to data obtained through administrative records?

What about data obtained through public investigation compared to data constructed through private investigation (Big Data)?

5. The “chant du cygne” of the Canadian census model?

There is a French expression “le chant du cygne” (the swan song) used when one wants to say that before disappearing the old world gets a kind of ultimate new life. Some say that Trump’s election in USA could be the swan song of an electorate that is going to gradually disappear. In Canada, the return of the long form could also be the swan song of the old way of collecting data.

The 2010 Canadian census controversy showed:

1. That the idea to abandon the long and mandatory form was not without good reasons;
2. That the response rate with a non mandatory questionnaire, as shown by the quasi scientific Canadian experiment, is much lower than the response rate with a mandatory one;
3. That there is not necessarily a causal association between obligation to answer and quality of data;
4. That the practice of a statistical bureau is not isolated from what affects the society (distrust, critical attitude toward science, State, elites, medias, politics).

In September 2016, a press release showed that the Canadian bureau of statistics is planning to eliminate the short and mandatory form for the 2026 census and to use instead administrative records in order to proceed to a virtual analysis of the Canadian population!

In this text, I wanted to show what this controversy reveals as for a more general confidence crisis, that pre-existed the controversy about the long form, that affects most statistical offices (and perhaps even some of the statistical principles a long time considered as sacrosanct, of which that of the obligation to answer) and that did not spare Statistics Canada. So if this is certainly the end of the controversy about the mandatory long form, this is perhaps not the end of the statistical confidence crisis in Canada and certainly not the ultimate change in the census practice of the Canadian bureau of statistics!