
Good Data and Intelligent Government

Munir A. Sheikh¹

Queens University and Carleton University

Introduction

IAN STEWART, FAMOUS FOR HIS ROLE in the realm of public policy, was a friend of Statistics Canada. He was the very first chair of the National Accounts Advisory Committee for 11 years from 1984 to 1995 and remained a member after he left chair until 2000. This long association with Statistics Canada reflected Ian's conviction that, while an essential ingredient of an intelligent government is intelligent public policy, the latter is critically dependent on good data.

The purpose of this paper is to explore a range of issues related to the development of good data.

1 The paper has benefited considerably from comments by Fred Gorbet, Andrew Sharpe, Ivan Fellegi and Ross Finnie. These are gratefully acknowledged. However, any remaining errors are my sole responsibility.

There are many aspects of decision-making that require good data. Citizens use a variety of information to make decisions on a day to day basis. Policy makers use data to study, develop, implement, monitor and evaluate their policies. Researchers need data to conduct empirical research. Businesses need data to develop their plans and strategies. And, that is also true for other segments of society that need to make evidence-based decisions.

The plan for this paper is as follows. Section 2 provides a description of why we need “good” data. Section 3 discusses a list of attributes of what we mean by good data. Section 4 describes the Canadian data collection system and the role of Statistics Canada in that system. It goes on to evaluate how the Agency measures up to the attributes of good data. The section concludes that the Agency does a reasonably good job in delivering on its mandate. Section 5 examines reasons for the Agency’s good performance. Section 6 describes recent developments surrounding the cancellation of the long form census that has significant implications for the production of good data. With this discussion, and other elements discussed in previous sections, Section 7 pulls together conceptual and practical challenges in gathering good data. Section 8 provides a description of steps that should be considered in ensuring the continuous flow of good data in the future. Finally, Section 9 provides concluding remarks.

Why the Need for Good Data

We describe what data can do for its users by providing one example of data, the consumer price index (CPI).

Data *describe events* as they unfold and thus are the source of information as things change. Every month Statistics Canada releases the CPI which describes the change in consumer prices for the past month. This release may show that the average price measured by the CPI rose by a particular magnitude over the past month. This development over the past month can be compared with time periods beyond the past month to obtain inflation rates over different time periods.

Data can be used to *gain insight* into the phenomenon that occurred. The detailed information contained in the CPI release can pinpoint where prices are changing most. For example, data may show that the main reason the aver-

age price rose last month was because of significant increases in auto insurance premiums. This would allow citizens to understand the reasons for an increase in prices.

Data allow *analysis* of the reasons behind developments that have been observed. Using other relevant data, such as the frequency and seriousness of accidents, it may be feasible to analyze the causes underlying the increase in insurance premiums described above. A better understanding provided by this analysis can be helpful in making improvements in outcomes.

The analysis made possible by data then allows the provision of a *context for decision-making*. The information contained in the analysis may, for example, show that the increase in insurance costs was driven by factors that may not be around permanently, in which case there may be no need for policy action. Or, this information may show otherwise.

Data help in *decision making*. Indeed, it is the most important contribution data make to improve the well-being of citizens. As a continuation of the CPI example, the increase in the inflation rate, along with the details of where the increased pressure may be coming from, gives the Bank of Canada the ability to relate this information to the objectives it is trying to achieve and adjust its policy levers to achieve the objectives.

Data are also used to *monitor progress* in achieving objectives. In the context of the CPI example, the Bank of Canada monitors progress on the inflation front by examining the core rate of inflation, which subtracts the volatile inflation components from the overall rate of inflation, in the context of its inflation targets.

Data are used as well to *build systems*. In the context of the CPI example, it is a key variable in the development of economic models that are used for a variety of purposes.

These systems, built on data, can be used for *forecasting and predicting*. These predictions allow decision-makers to take action in anticipation of adverse events taking hold. For example, CPI models may show, linked with other data, that the inflation rate could fall below the central bank's target range, encouraging the bank to take corrective action in anticipation.

Data are used as well for *evaluation of outcomes*. The evaluation exercise is helpful in determining whether or not objectives have been achieved, and if

yes, satisfactorily or not, and, if not, why not. Such evaluations are a key to making adjustments in decision making.

These uses of data allow a society to achieve a number of fundamental objectives crucial to its civilized functioning, including: democratic accountability, monitoring the performance of society and, in turn, that of government; evidence-based decision-making; research to improve things, including allowing reproducibility of results; and honest information brokerage to settle matters in formula-driven outcomes such as distributing Parliamentary seats, billions of dollars in transfers from the federal to provincial and territorial governments, and labour and business contracts.²

In sum, data provide the foundation for knowing things the way they are and taking steps to making things the way they should be. In this sense, the importance of data in enhancing human well-being cannot be underestimated.

Attributes of Good Data

We begin this discussion with reference to the way Statistics Canada defines its mandate to produce good data. We will then evaluate the extent to which this mandate is achieved.

In its *2010-11 Report on Plans and Priorities* to Parliament, Statistics Canada defines its mandate as follows:

... providing access to a trusted source of information.
Trust can be established only if data are relevant, meaning they meet the most important needs of the users, and if users are confident that the information is of the highest possible quality. Access, relevance and quality can be optimized only if the agency practices maximum organizational efficiency. (Statistics Canada, 2010a)

The four underlined objectives define Statistics Canada's business model, as described in considerable detail in the Agency's first ever *Corporate Business Plan* (Statistics Canada, 2009).

We will argue that these four objectives are the appropriate objectives to achieve as part of a statistical agency's contribution to be a valuable partner in

2 Ivan Fellegi (2011) made these observations in his comments on the first draft of this paper.

intelligent government. The rest of this section explains in some more detail what these objectives mean.

Access

Access refers to citizens' ability to get the information they need from their statistical agency. This access refers to the availability of both non-confidential and confidential data.

On non-confidential data, there are issues of how much data are available, at what cost, how long after the fact, and the ease with which available information can be understood.

On confidential data, there are other issues beyond those described above for non-confidential data. Many users need micro data for research and policy development. However, by their very nature these data cannot be made public. The issue, therefore, is the ability of a statistical agency to achieve both objectives simultaneously, of finding creative ways to make confidential data available to those who have important needs while putting in place safety devices to ensure those users do not make this information public.

Relevance

Relevance refers to the reality that no statistical agency can ever have the ability to gather all data that users may find useful. Choices have to be made. In making these choices, the issue is whether the data produced by a statistical agency are the most important in relation to the priorities and needs of the country.

It is useful to remember that priorities and needs evolve over time. Relevance of data produced by a statistical agency, therefore, automatically declines unless proactive steps are taken to reallocate resources to meet the evolving needs.

Quality

Quality of data refers to how representative the data are of the phenomenon they are supposed to capture. For example, if the data on the unemployment rate, based on a sample, say the rate is 8 per cent, are we confident that taking the labour market as a whole, the rate is really likely to be 8 per cent?

In reality, there is a confidence interval around data estimated from sample surveys. The issue, therefore, is whether the data are good enough for the purposes for which they are produced.

The level of quality needed for different types of data will differ because their uses are different. In view of this reality, a statistical agency may have five types of data gatherings: a voluntary survey; a mandatory survey; a sample census where information is collected from a very large unbiased sample on a mandatory basis; a full census, where information is collected about all citizens on a mandatory basis; and, the use of administrative data, that is information collected for purposes other than gathering statistics, such as to run a government program. An example of administrative data is the tax data that are collected by a tax department but that may also be useful as a source of statistical information.

There are also a number of methods in data collection including survey mailouts, door-to-door visits by enumerators, telephone interviews, and the internet. Again, different methods would be used to meet different needs for data collection.

Going back to representativeness as a hallmark of quality, there are a number of quality indicators that would define this representativeness. These can be of a static nature, meaning they are always present. Or, they can be of a dynamic nature, meaning they change over time. On the static side, these would include:

- *Local Relevance*: This refers to the relevance of a particular survey in gathering data consistent with the needs.
- *Accuracy*: Statistical information is accurate to the degree to which the information correctly describes the phenomenon it was designed to measure. Accuracy is usually characterized in terms of error in statistical estimates and is traditionally separated into bias and variance. Accuracy also includes accurate reporting of data, which is not an insignificant task in view of the fact that in each survey millions of data observations are collected that need to be converted into usable information for dissemination.
- *Timeliness*: This refers to the length of time between the reference point to which the information relates and its availability to users.

-
- *Accessibility*: This refers to the ease with which users can learn of the existence of a particular piece of data, locate it and import it into their own working environment. It includes the suitability of the form or medium through which the information can be accessed and its cost.
 - *Interpretability*: This refers to the ease with which users can interpret and utilize the information appropriately. It includes the availability of “meta-data”, or the supplementary explanatory information on data being provided that answers questions users may have about the data.
 - *Coherence*: The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytical framework. It includes coherence between different variables pertaining to the same point in time, coherence between the same variables for different points in time and international coherence.

On the dynamic aspect of quality, there are two key considerations:

- *Non-response*: This refers to the automatic tendency for response rates to decline over time in view of economic and social evolution and changing technology (e.g. increasing use of cell phones). Maintaining quality would require dealing with this phenomenon.
- *Coverage*: Surveys use samples which need to be representative. The source of information from which samples are drawn is called a “frame”. With the evolution of the society and the economy, and the different mix of data sources (e.g. greater use of administrative data), the frames need to be updated to ensure appropriate coverage.

Organizational Efficiency

Organizational efficiency of a data collection agency is a critical attribute of good data as it allows the achievement of the best possible data within available resources, or high quality data at least cost. The quality of a data collection organization’s governance and management structure should allow it to intelligently handle all issues, existing or emerging, and to have the foresight to anticipate approaching challenges. It will include systems to employ the best available human resources and to keep its capital stock modern and up-to-date. And it should use a well-defined and concrete risk management structure

and an evaluation framework to determine if objectives are being achieved at least cost.

How Does Statistics Canada Measure Up to the Desired Attributes

A comprehensive evaluation of Statistics Canada's performance in the areas described as desirable in Section 3 is not available. Indeed, it would be useful at some stage to estimate the Agency's productivity, which anecdotally seems to have risen considerably in view of the significant increases in data quantity and quality without a commensurate increase in resources.

In the early 1990s, the British magazine, the *Economist* examined the performance of the world's leading statistical agencies and declared Statistics Canada to be the best in the world (The Economist, 1991 and 1993).³ Anecdotal evidence based on views of other statistical agencies around the world would lead one to conclude that Statistics Canada has kept its reputation as a leading-edge organization and can be safely characterized as, at least, one among the best in the world. The recent developments regarding the long form census may have a negative impact but it is still early to tell how much.

In its *Departmental Performance Report* submitted annually to Parliament (Statistics Canada, 2010b), Statistics Canada examines its performance over the past year against a set of quantitative benchmarks and targets. It is useful to reproduce some of that information to obtain a better sense of its performance.

3 Also see Wikipedia article on "Statistics Canada". URL: http://en.wikipedia.org/wiki/Statistics_Canada

Statistics Canada's Performance Highlights

Performance Indicators	Targets/benchmarks	2009/2010 performance summary
Indicators of access		
Number of visits to the Statistics Canada website	Annual increase in visits exceeds 5%	Not met – Visits to the website decreased by 4.2%
Client satisfaction with price to access data	Rating of 4 on a scale of 5	Somewhat met – Client satisfaction rating was 3.6 for economic statistics and 3.3 for social statistics
Indicators of relevance		
Number of data series downloaded from the CANSIM online database	Annual increase in downloaded series exceeds 5%	Exceeded – Downloaded series increased 23.2% for economic statistics, 13.4% for social statistics
Clients' satisfaction with Statistics Canada's ability to meet their needs	Rating of 4 on a scale of 5	Mostly met – Client satisfaction rating was 4.0 for economic statistics and 3.9 for social statistics
Number of media citations	Many	Met all – There were 9,294 media citations
Indicator of quality		
Percentage of statistical outputs that meet set levels of sampling accuracy	95% of major statistical outputs meet set levels of sampling accuracy	Exceeded – 96.0% for economic statistics and 99.7% for social statistics
Indicators of organizational efficiency		
Rating on 21 areas of management, as defined in the federal Management Accountability Framework	Obtain rating of 'strong' or 'acceptable' in most areas of management	Met all – Rated 'strong' or 'acceptable' in all but one area of management
Departmental Staffing Accountability Report (DSAR)	Improve on results from the 2008 DSAR	Exceeded – Overall rating improved to 'strong'
Public Service Employee Survey 2008	Rank among the best in comparison with other federal government departments (response rates and overall results)	Met all – Response rates well above Public Service average; job satisfaction at or above other federal departments
Financial reporting to Policy Committee	Monthly	Met all – All financial reports delivered as planned

Management Accountability Framework: Historical Trend Report Round 7, 2009/2010 Compared with Previously Assessed Round Statistics Canada

Area of management		Asses- sed by TBS	Last	Current	Change
Core					
1	Values-based Leadership and Organizational culture	√	ST	ST	*
6	Quality and Use of Evaluation	√	AC	AC	=
9	Effectiveness of Corporate Risk Management	√	AC	AC	=
10	Excellence in People Management	√	AC	ST	*
17	Effectiveness of Financial Management and Control	√	AC	AC	=
19	Effectiveness of Internal Audit Function	√	OI	AC	↗
Department specific					
12	Effectiveness of Information Management	√	AC	OI	↘
13	Effectiveness of Information Technology	√	AC	ST	↗
Remedial					
7	Quality of Performance Reporting	√	OI	AC	↗
Rotational					
2	Utility of the Corporate Performance Framework	√	AC	AC	=
3	Effectiveness of the Corporate Management Structure	√	AC	AC	=
4	Effectiveness of Organizational Contribution to Government-wide Priorities	√	AC	ST	↗
14	Effectiveness of Asset Management	√	AC	AC	=
5	Quality of Treasury Board Submissions		AC	AC	=
8	Managing Organizational Change		ST	ST	=
15	Investment Planning and Management of Projects		ST	ST	=
16	Effective Procurement		ST	ST	=
19	Effective-management of Security and Business Continuity		ST	ST	=
20	Citizen-focused Service		ST	ST	=
Total number of areas of management assessed by Treasury Board Secretariat in current round		*	↗	=	↘
13		2	4	6	1

ST = Strong; AC = Acceptable; OI = Opportunity for improvement

√ Assessed in current round.

* Unable to compare (e.g., because of methodology changes or unrated areas of management).

International Comparisons of the Periodicity and Timeliness of Data Release, National Accounts, G7 Countries

G7 country	Benchmark	Periodicity			Timeliness		
		SDDS	Result	Rating	SDDS	Result	Rating
Canada	Meet SDDS requirements	Quarterly	Quarterly	Met	1 quarter	60 days	Met
France		Quarterly	Quarterly	Met	1 quarter	50 days	Met
Germany		Quarterly	Quarterly	Met	1 quarter	45 days	Met
Italy		Quarterly	Quarterly	Met	1 quarter	10 weeks	Met
Japan		Quarterly	Quarterly	Met	1 quarter	6 weeks	Met
United Kingdom		Quarterly	Quarterly	Met	1 quarter	1 quarter	Met
United States		Quarterly	Quarterly	Met	1 quarter	31 days	Met

SDDS: Special Data Dissemination Standard.

International Comparisons of the Periodicity and Timeliness of Data Release, Unemployment, G7 Countries

G7 country	Benchmark	Periodicity			Timeliness		
		SDDS	Result	Rating	SDDS	Result	Rating
Canada	Meet SDDS requirements	Quarterly	Monthly	Met	1 quarter	2 weeks	Met
France		Quarterly	Quarterly	Met	1 quarter	Not later than 10 weeks	Met
Germany		Quarterly	Monthly	Met	1 quarter	1 month	Met
Italy		Quarterly	Quarterly	Met	1 quarter	Not later than 1 quarter	Met
Japan		Quarterly	Monthly	Met	1 quarter	1 month	Met
United Kingdom		Quarterly	Monthly	Met	1 quarter	5 weeks	Met
United States		Quarterly	Monthly	Met	1 quarter	3 weeks — not later than 1 month	Met

SDDS: Special Data Dissemination Standard.

We can draw the following conclusions from these results.

- Overall, Statistics Canada's performance is quite good in all areas of its mandate including access, relevance, quality and organizational efficiency.
- This conclusion on the organizational efficiency aspect is confirmed by Treasury Board's independent evaluation.

-
- The Agency does not rest on its laurels. It continues to seek further improvements, as evidenced by improvements in its already strong results in the Management Accountability Framework.
 - Statistics Canada compares quite favourably in international comparisons.

Reasons for Statistics Canada' Success

A number of factors explain Statistics Canada's success.⁴ These are discussed below.

Political Neutrality and Independence

The production of data is a complex business and users have no means to determine its quality on their own. Their trust in data depends on the credibility and reputation of the Agency. Credibility depends on a range of factors described below, but one of the most critical of these factors is the neutrality and independence of the statistical agency. Credibility takes long to establish but can be lost quickly if the Agency lets itself be politically manoeuvred. Whether or not it has the ability to stay independent depends on a number of factors.

The legislative framework is clearly important in this regard. The *Statistics Act*,⁵ however, neither gives the Chief Statistician the final authority to make methodological decisions nor does it provide the Agency independence from Ministerial control.

Despite this legislative weakness, Statistics Canada was able to achieve neutrality and independence in the past because of long-standing conventions on the part of governments not to interfere in its functions. Naturally, there is a real risk that conventions may not be respected. The Chief Statistician's option to resign in the face of political interference is one tool that helps the Agency. Because of the long-established tradition of neutrality and independence, all Statistics Canada staff also guard against governments intervening in an inappropriate manner.

4 See also Fellegi (1996). In his comments on the first draft of this paper Fellegi (2011) groups these factors into two categories: those that provide legitimacy to Statistics Canada data; and, those that provide credibility.

5 URL: <http://www.statcan.gc.ca/about-apercu/act-loi-eng.htm>

Furthermore, there are other tools that help allow the Agency to contribute towards maintaining neutrality and independence. These include the Chief Statistician's prerogative to allocate the Agency's budget appropriations with accountability to Parliament through the Minister, his control on survey questionnaires and survey content design and the existence of a transparent planning process in reallocating resources.

Centralized and Comprehensive Statistical System

The *Statistics Act* provides many strengths in this context. The mandate provided by the *Act* to gather data is virtually unlimited in terms of subject coverage. All requests made by the Agency to gather data are mandatory unless they are explicitly declared to be voluntary. This applies to not only the respondents of surveys but to all governments and other organizations from whom the Agency may seek administrative data.

Privacy and Confidentiality of Data

There is no doubt that the collection of data from citizens is a violation of their privacy. Such a violation is permitted by law, in the *Statistics Act*. This violation of privacy is justified because data and information are the cornerstones of decision-making, both by citizens and their governments. There is obviously a trade-off which is governed by privacy laws and through real life application of the privacy provisions by the Privacy Commissioner.

An important aspect of privacy is the linking of data from various sources, called record linkage. This can be problematic if there are no checks and balances. Statistics Canada has in place a multi-stage review process to ensure data linkage does not occur without careful thought and without the approval of the Chief Statistician.

In contrast to privacy, confidentiality refers to personal information not being available outside the Agency, except in situations where the respondents have given their explicit consent. The *Statistics Act* takes confidentiality extremely seriously. According to the *Act* all employees of the Agency are personally liable for the protection of statistical confidentiality and not even courts have access to individually identifiable information without the informed consent of respondents. The Chief Statistician ensures that confi-

dentiality is upheld and even the Minister has no authority in law to overrule him. All staff in the Agency are extensively trained in understanding their obligations and in ensuring they respect the letter and spirit of the law.

Transparency of Operations

The way Statistics Canada does its business is an open book. Its planning process to establish priorities is public and extremely well-documented. Any and all data collected are made public, except those that violate the confidentiality requirement. There are no special clients: data collected for any client on a cost-recovery basis is also available to all Canadians. The release dates for key data are published in advance so a government would not have the option of manipulating data release dates to its advantage. The Agency's Parliamentary reports are comprehensive and provide all relevant information Canadians want to know.

Stakeholder Input

The Agency cannot produce data that users need most without communication with users. This is done through regular meetings of the advisory committees on various subject matters. The National Statistics Council serves as the advisory body to the Chief Statistician on matters of an overall nature and guides the Agency's operations and advises on its priorities. Contacts also exist with all key data users including federal departments, provincial and territorial governments, municipalities, researchers and the media.

Strong Methodology Group

Quality of data is Statistics Canada's hallmark. Maintaining quality is an ongoing challenge in the face of constant economic and social evolution. Statistics Canada is fortunate to have a world-class methodology group supervised by an Assistant Chief Statistician. They undertake sophisticated research on challenging issues to advise subject matter specialists on how to maintain and enhance quality of their data. They publish an internationally-acclaimed peer-reviewed methodology journal reporting on their findings.

Client-Orientation in Dissemination

Production of data is one thing. Disseminating it in a manner understandable to as many Canadians as possible is another. Statistics Canada takes this aspect of its tasks seriously. It publishes a daily release reporting on new data releases for the day, titled appropriately the *Daily*. It has made considerable efforts recently to improve its web-site.

Management and Governance System

All of the strengths described above are made possible by the management and governance system in place. The *Policy Committee* which is composed of the Chief Statistician and the Assistant Chief Statisticians meets every Wednesday to review major issues that need to be dealt with and makes collective decisions for the good of the Agency.

Sophisticated Resource Allocation Planning Mechanism

Statistics Canada has a well-established, open and transparent annual planning cycle to review its ongoing needs and find resources internally through efficiency-enhancing steps to meet new needs. This process begins in October of each year and is completed by the February of the next year at a marathon meeting of the Policy Committee.

Sophisticated and Comprehensive Program Evaluation System

Statistics Canada's quadrennial and biennial program reviews evaluate each and every one of its programs over a four year period. The Agency follows a unique model for this evaluation based on the following principle: effective program evaluation can be done only by those who actually have the experience in running those programs; however, mechanisms must exist to ensure that these self-evaluations are completely and totally neutral and above-board. To achieve this, those responsible for programs do comprehensive self-evaluations under the oversight of a Quality Secretariat, the centralized program evaluation group and the Policy Committee.

Sophisticated Risk Management System

Under this system each program has an evergreen risk profile which establishes a composite risk in each area. The program then analyzes the means required to bring risk to an acceptable level. The risks and investments required are then evaluated at the annual planning discussions and investments are made to deal with those risks that are beyond an acceptable level.

Strong Analytical and Research Program

Data confrontation, that is testing relevant data against each other, is one way of evaluating data quality and helps in identifying areas that need improvements. This is one of a number of analytical methods that are employed by the Agency in its research and analysis groups to understand the data better for future improvements. The Agency publishes its research, which is accepted as high quality by the research community.

The Census Issue

The message from the above discussion is that Canadian data gathering and dissemination are reasonably good and compare favourably with other countries. The above discussion also describes some challenges that, while important, are natural in a large program.

With that as a context, how does one describe and explain the census controversy of the summer of 2010 when the government decided to cancel the long-form census and replaced it with a voluntary survey, leading to a public outcry and the eventual resignation of the Chief Statistician? These developments reflect substantially on all aspects of Statistics Canada's mandate described above, namely access, relevance, quality and organizational efficiency. In view of the importance of these developments, this section provides a detailed review of the census issue.

The section covers six topics. It begins with a brief history of the census. Second, it describes the process of undertaking the census. Third, it summarizes the nature of the census questions and the uses of the census data. Fourth, having provided a context, it describes elements of the 2011 census drama. Finally, it says a few words on the circumstances leading up to the Chief Statistician's resignation.

Census History

Believe it or not, the first census in the land that we now call Canada was initiated 367 years ago in 1644 by Jean Talon. Given the importance of his statistical contributions to Canada in those early years, Statistics Canada named one of its three buildings in his honor. The census counted the colony's entire population of 3,215 inhabitants and asked questions related to age, sex, marital status and occupation.

There was a long string of 98 colonial and regional censuses during the years leading up to 1871, when the first census was undertaken in Canada following the 1867 *British North America Act*.

The 1871 census contained 211 questions on demography, land holdings, vital statistics, religion, education, administration, the military, justice, agriculture, commerce, industry and finance.

The census with the largest number of questions took place in 1921. There were 565 questions in that census. Questions were asked on population, agriculture, animals and animal products, manufacturing and trading establishments, and certain disabilities.

The 1941 census was the first to introduce sampling within a census by asking questions of every tenth household. This innovation had the objective of collecting high quality data at a much lower cost and significantly lowering the response burden.

Based on the experience of the 1941 census, it was decided to increase the sample size to one in five in the 1951 census.

The current *Statistics Act* was passed in 1971, with Statistics Canada replacing the Dominion Bureau of Statistics. The 1971 census introduced many innovations. It began "self-enumeration". It started a new "short form" questionnaire, with the traditional census now labeled the "long form". It is worth emphasizing that the 1971 census did not begin the long form. Rather, it brought in the new short form. The 1971 census began the process of sending the short form to two in three households and the long form, which included the short form questions, to one in three households.

The 1996 census short form went to 4 in 5 households and reduced the number receiving the long form to 1 in 5 households. As another innovation, the

census could be completed in two official and 49 unofficial languages, 12 of which were aboriginal languages.

The 2006 census took place on May 13, with a total of 13.6 million households receiving a census questionnaire. There were 8 questions on the short form and 53 additional questions on the long form. For the first time ever, internet response was an option, with 18.5 per cent of the population using this mode.

Census Process

The process can be usefully broken into six parts.

The first step is *consultations* with stakeholders and data users. The *2011 Census Content Consultation Guide* was released in July 2007 and the *2011 Census Content Consultation Report* was released in July 2008. Statistics Canada received more than 1200 content-related suggestions from more than 150 organizations and private citizens.

The second step is the *development of questions*. Based on these consultations, and the need for continuity of historical information, changes from one census to the next are generally modest. Sixty percent of the questions asked in the 1971 census still remain. Questions in the 2006 census now dropped are related to unpaid work. Questions added for 2011 include commute time, child care support and its costs, and subsidized housing. Questions are evaluated for privacy objectives, working together with the Privacy Commissioner.

The third step is to determine which questions go into the *short form and which to the long form*. As I mentioned earlier, the distinction between the short and long forms began with the 1971 census.

Statistics Canada does not differentiate among these questions on the basis of their importance, as they are all tied to the needs of a variety of users. For example, the head-count question in the short form may be the most important for governments in the distribution of Parliamentary seats or for federal-provincial transfers. But for a city transportation planning department, the most important piece of information may be about how people travel to work and how much time they spend commuting, a question which is in the long form.

It is, therefore, not the importance of the questions that determines whether they are in the short or the long form. What determines this division is a cost efficiency test: how can we obtain the highest quality data at least cost.

The fourth step in the process is for Statistics Canada to submit the proposed questions to the *Government* and for the Government to review these proposed questions and tell Statistics Canada what the census contents will be.

The fifth step is to *collect* data. As part of this process a census test is conducted about a year before the actual census date to ensure all systems work well. Again, to constantly strive to reduce costs and the response burden, Statistics Canada moved in a substantive way to internet data collection in the 2006 census. The May 2009 census test indicated the potential to double this response rate in 2011 to over 35 per cent.

The sixth step is to *transform* raw data into data useful for users and to disseminate them. Dissemination begins about a year after the census as the large amount of data has to be processed, edited, and checked for accuracy, gaps and consistency. The data are analyzed thoroughly to understand the results before they are released publicly.

The whole process takes about seven years.

Statistics Canada stands out particularly well in any international comparison in collecting census data.

There are four objectives a national statistical office tries to achieve in gathering data: high data quality; low cost; low response burden; and privacy and confidentiality of data collected.

On quality, we are not aware of another statistical organization that produces census data of a higher quality than Statistics Canada.

On cost, we can compare Canada with Australia and the United States, countries that are similar to us in running a survey-based census rather than using administrative data. Statistics Canada's per dwelling census cost for the 2006 census was \$43.77 (in 2009/10 dollars) compared to \$126.18 in the United States and \$49.68 in Australia.

On the response burden, with the innovation of sampling within a census, only a fraction of the population needs to spend time filling the form.

On privacy and confidentiality, there were two complaints from Canadians related to the 2006 census. Statistics Canada has never had a breach of its cen-

sus data because the Agency has in place sophisticated and effective security systems. Indeed, Statistics Canada has been blamed sometimes for being paranoid about privacy and confidentiality.

Census Questions and the Uses of Census Data

From 1971 to 2006, the Census included two forms: the short form and the long form. The short form included questions of a tombstone nature with the main objective being a head count. The long form included the remaining questions that were focused on getting respondents' socio-economic information.

The 2011 census will contain only the short form questions and two questions on language from the previous years' long form. All long form census questions will be part of the new National Household Survey that will be voluntary.

To get a perspective on the uses of past censuses, we can use the 2006 census as an illustrative example.

The short form questions on **demography** have many uses. As examples, this information is essential to enumerate the population, to provide counts of population by federal election district, to redraw the electoral boundaries, and to provide a base for the federal fiscal transfers to provinces and the territories. The language question in the short form allows the government to meet the minority language provisions of the *Charter*.

Labour market and income data are used by many to study economic and social performance of Canadians and are used by governments to develop and improve many social and economic policies. Detailed industry and occupation data for small areas, that are available only from a long form census, are required to assess changing skill needs and labour shortages. The census yields labour market output information for vulnerable groups in the community such as immigrants, visible minorities, aboriginal peoples, and official language minorities.

Transportation data include modes of transportation, place of work and commuting time. This information is used extensively by the federal, provincial and territorial governments, transport associations and the private sector in dealing with critical transportation planning needs.

Education data, in conjunction with other information, are used widely by governments and researchers for labour market analysis and education planning needs. Long form Census education data are fundamental for studies related to groups that draw policy attention such as immigrants, aboriginal peoples and official language minorities.

Language information collected from the long form census is required for federal programs in support of minority language education, for implementing bilingual services, for the promotion of bilingualism nationally, as well as for heritage language programs including aboriginal languages.

Activity Limitations questions provide information to meet the ongoing demands for data on the size and characteristics, including labour force experience, of the population of persons with disabilities.

Housing data provide information on shelter costs of tenants and homeowners, age of dwellings, value of owner-occupied housing, the need for repairs, number of rooms and bedrooms, and the extent of subsidization of housing. These allow governments and researchers to examine issues of affordability and the quality and state of housing. The housing information collected in the census is the only source of data for the condition of housing for aboriginals living on reserves.

Citizenship and immigration data allow governments to develop and enhance immigration and citizenship programs. Combined with other long form data, researchers can gain valuable insights into the performance and integration of immigrants into the Canadian economy and society. The results of such research are critical to governments in managing the Canadian labour market.

Ethno-cultural information allows governments to meet the requirements laid down in the *Charter*, the *Canadian Multiculturalism Act* and the *Employment Equity Act*.

These are just some of the examples of the uses of data gathered under a variety of headings in the short and long form census. Of course, these data can be used jointly to study a range of other questions which geometrically increases their value. Examples include: links between income and education; various indicators of performance for a variety of population groups; and comprehensive report cards on the performance of particular neighborhoods.

Think of a little experiment. Ask yourself what a new neighborhood would have needed as it is established? Just about everything it needed could have used information from the long form census ranging from building roads and sidewalks, to strip malls, to the establishment of schools, to the setting up of health care facilities, to the opening up of a community centre, to the provision of services for minority groups. And the list can go on and on.

The 2011 Census Drama

On June 26, 2010, the Government announced in the *Canada Gazette* that the 2011 census would include only the short form. The form would only have 8 questions.

This was followed by criticism from many users of the data. Close to 370 groups objected to the decision. Critics include, among many others, provincial and municipal governments, academics, think tanks, private sector researchers, NGOs, religious organizations and the media.

The government's response has been that it does not believe that coercion should be used to obtain long form information from Canadians who do not want to provide it voluntarily. The government is against obtaining information on the long form by threatening to send Canadians to jail or to impose fines.

The Minister responsible for Statistics Canada explained at the Parliamentary Committee hearings that his government believed in "balance" in getting the needed data and the citizens' desire to maintain privacy (Standing Committee on Industry, Science and Technology, 2010). This balance naturally produces an outcome where data are obtained voluntarily; hence the voluntary survey that replaced the long form census.

This "balance" approach is a change in course from the existing Statistics Canada practice of achieving both privacy and data quality objectives.

Under existing practice, Statistics Canada consults users extensively on their data needs as we have described earlier.

The questions that Statistics Canada develops are then vetted for privacy objectives, working very closely with the office of the Privacy Commissioner. The two organizations have always worked well together and, to the knowl-

edge of Statistics Canada, it has not been in any situation of significant disagreement with the office of the Privacy Commissioner.

Having assured itself of the privacy objectives, both in letter and in spirit, Statistics Canada finds the most appropriate method of data collection to maximize quality at least cost. This may mean undertaking a voluntary or a mandatory survey depending on client needs. Undoubtedly, the quality of a mandatory census is higher than that of a voluntary survey; just as the power of a 53 foot transporter is greater than that of an SUV which itself is stronger than a small passenger car. However, each is built for its own particular use. One does not use a 53 footer to drive around two passengers in the city and one does not use a small sedan to transport goods to a Walmart on an ongoing basis. Similarly, one cannot satisfy the needs of many data users who rely on a census with the results of a voluntary survey.

Returning to the government's position to seek a new "balance", a number of questions have arisen.

- Did the government analyze carefully the consequences of a loss in data quality as a result of the voluntary survey? Did it consider how this loss in quality would impact the data needs of users? Did it examine the negative consequences of that on policy development, including that at the federal level?
- In undertaking such an analysis, why did the government not consult with data users?
- Did it compare these consequences from the loss of data quality against any privacy gains? In this context, what is the importance of the fact that the Privacy Commissioner had no issues with the long form questionnaire in the context of Canada's strict privacy laws? Furthermore, there have not been many complaints against the census in the past, either to the Privacy Commissioner or to Statistics Canada. To my knowledge, there were two complaints against the census on privacy grounds in the 2006 census and one in the 2001 census. A total of 64 Canadians were referred to the Public Prosecution Service of Canada in a population of over 31 million. Nobody was ever sent to jail for not completing the census questionnaire.
- What was wrong with Statistics Canada's approach to achieving the twin objectives of quality and privacy?

-
- How does the new approach of “balance” improve upon the existing Statistics Canada approach?
 - What are the alternative ways of getting the information that users need to monitor, implement and evaluate their programs?
 - What does this mean for other mandatory surveys of Statistics Canada? The future of the Labour Force Survey? The future of business surveys? For the future of the Agriculture Census?
 - Was the decision on the long form taken with a view that the information so collected is less important than information collected from the mandatory short form, the Agriculture census and other mandatory surveys? Is that assumption valid?

Beyond the specific question of the mandatory or voluntary long form there is a structural issue that has come to the fore as a result of the census debate: the issue here is whether the *Statistics Act* circa 1971 provides enough independence to Statistics Canada and enough authority to the Chief Statistician for them to be known to be at arms length, so the Agency’s products will continue to trustworthy. We believe there is a close relationship between citizens’ trust in data and the independence the statistical agency is perceived to have in collecting these data.

The Minister responsible for Statistics Canada quite correctly observed that the Agency reports to him and is not independent in making its decisions. Data users until then had a perception of Statistics Canada as an arms length type of organization, an impression based obviously on the long tradition of independent decision making by Statistics Canada that was not enshrined in law. An issue in this context is that it is the Minister, not the Chief Statistician, who according to the law retains final authority on technical and methodological issues.

Another issue is whether the determination of census content is a political or a technical question.

And, finally, there is the question of whether the current nature of Statistics Canada, effectively as a government department, is appropriate, or whether a different model should be considered.

These questions are serious and important. They were not dealt with in the past as long traditions made their resolution unnecessary. With the census

debate, it is time to consider these issues and determine whether improvements to the current law can be made.

Chief Statistician's Resignation

As explained in the Chief Statistician's testimony to the Parliamentary Committee, a Deputy Minister has twin roles; one, to give fearless advice to the government; and, second, to implement loyally all of the government's lawful decisions (Standing Committee on Industry, Science and Technology, 2010).

He argued that he had always fulfilled that oath as a public servant and the oath under the *Statistics Act*. He argued he was living up to his obligation to implement the government's decision to replace the mandatory census with a voluntary survey to the best of his ability.

However, during the week of his resignation, following comments the Minister had made on three separate occasions, media stories emerged focused on two points: first, that the quality of the voluntary survey data would be as good as that of the census; and, second, that Statistics Canada, and the Chief Statistician, were firmly behind this decision.

These media stories began to damage the reputation and credibility of Statistics Canada and to cast doubt on its integrity.

The advice Statistics Canada gave to the Minister is protected under the law. Releasing that was not an option.

Having made efforts internally to fix the problem without success, he argued he had no credible option but to resign with an answer to the underlying question in the debate: whether a voluntary survey can be a substitute for the mandatory census.

His response was: "It cannot."⁶

Challenges in Producing Good Data

With this information, we are in position to take stock of the future challenges in the collection and dissemination of good data as part of intelligent

6 See the resignation statement on URL: <http://www.680news.com/radio/680news/article/80353--text-of-munir-sheikh-s-resignation-statement>

government. The challenges include the following in components of Statistics Canada's four part mandate.

Access

On the non-confidential data, the greatest challenge is the price the Agency charges for its regular data. This charge is beyond the cost recovery system the Agency maintains in charging users for the specialized data they ask for. There is no economic rationale for charging a price for a public good beyond its marginal cost, which in the case of Statistics Canada's regularly-produced data, is small to non-existent.

This is a matter for the government of Canada to resolve as the Agency does not have the approximately \$5 million needed to accomplish this objective.

On the confidential data front, the Agency faces a challenge: micro data are essential to researchers to understand critical economic and social issues facing Canada; however, a leak of such data can seriously harm the Agency's data collection business as respondents stop cooperating with the Agency. It is Statistics Canada's obligation to extend the frontiers of this trade-off by finding creative ways to improve accessibility, perhaps by using new technologies, to preserve confidentiality. The Agency is already moving in this direction but it needs to ensure that, rather than letting up on this, it moves full speed ahead.

Relevance

The reality that the relevance of data falls automatically with economic and social evolution, as old data keep being collected and new resources are constrained, has already had real consequences. There are many areas where new and better data are needed but not being produced.

Examples of significant and critical data gaps include, without being exhaustive:

- There is a need for an improved Consumer Price Index that updates weights more frequently, has a wider coverage of commodities and does a better job in dealing with quality changes.
- Canada does not have an ongoing wealth survey. This is surprising since a number of issues the country faces today require information on wealth holdings at the household level. The debate on the adequacy of pensions

is taking place without the knowledge about personal asset levels of Canadians at the time of their retirement. The impact of the financial and economic crisis cannot be fully examined without an understanding of changes to household wealth.

- The Drummond Report, prepared for the Council of Labour Market Ministers, highlighted serious labour market data gaps and urged the government to act expeditiously in view of the weak labour market that exists today following the financial and economic crisis (Drummond, 2009).
- With globalization, trade data have not kept up. There are weaknesses in data for trade volumes, because of inter-industry and intra-industry trade, and trade prices, particularly import prices.
- Service sector data, particularly service prices, need improvement.
- Data on aboriginal populations are particularly weak. This is especially true of populations on reserves where the only source of data for reserves which participate was the census. Even that source is now largely gone with the cancellation of the long form census.

Quality

The Agency has worked valiantly to maintain the quality of its data. However, it is facing a number of challenges on this front as well.

Quality like relevance can be threatened by economic and social evolution and technological change. For example, the increasing use of cell phones makes it harder to contact respondents. Social change is making it more difficult for the Agency to obtain citizen cooperation as the growth of two-earner families and mobility, as examples, makes it harder for the new generation to spare time.

Meeting the challenge requires ongoing investments. However, the option of scaling down existing programs to find these resources is hard as there are clients of all existing data that become very vocal if their cherished program is cut. The alternative is increased funding from the government. The reality is that funding has been cut repeatedly.

The residual result of these realities is the pressure on data quality. There is just so far the Agency can go in finding innovative solutions to this conundrum.

As if this were not enough, the government's decision on the long form census has the potential for a substantial impact in reducing the quality of Statistics Canada data.

Data quality may be affected in a variety of ways.

First, it will affect the long form survey data.

It is a statistical fact that a voluntary survey cannot become a substitute for a mandatory census because of uneven response rates from different population groups and different size geographic areas. Increasing the sample size cannot offset this problem. Hence, many data users including the federal government will lose the data quality they need.

Second, to the extent that the long form census data provide a benchmark for other Statistics Canada surveys, the quality of data from these other surveys would deteriorate.

Third, there is now risk about the quality of the short form data and other data produced by Statistics Canada.

The short form information is even more critical to benchmarking a number of Statistics Canada surveys. Many respondents may be getting confused about what is voluntary and what is mandatory and why is there a distinction between the data needs of the two types. The risk of this confusion may have increased substantially in view of the debate that has unfolded since the announcement of the census decision. One factor Canadians may take into account in deciding whether or not to respond to the mandatory short form or other surveys is the government's argument that Canadians are right in not providing information they think the state does not need to know.

Fourth, what is the future of Statistics Canada's other mandatory surveys, such as the Labour Force Survey, all the business surveys and the Agriculture census? Would they become voluntary in the future and what implications for data quality would follow?

Organization Efficiency

The Agency's philosophy to-date has been to produce the highest quality data at least cost. The cancellation of the long-form census, and its substitution with a voluntary survey, turns this objective upside down: the Agency will now produce low quality survey data with increased expense of \$30 million.

The cancellation of the long form raises more worrisome structural issues regarding the adequacy of the *Statistics Act* in preserving the independence and neutrality of Statistics Canada. If these issues are not resolved they have the potential to damage the trustworthiness of Statistics Canada.

Making Gains in Producing Good Data

Based on the review provided above we are now in a position to suggest improvements. There are some suggestions directed at the government and others directed at Statistics Canada.

We suggest the following to the government.

First, Statistics Canada, and the quality of the data it produces, should not be taken for granted. It should always be on the government's radar screen. No country can be among the league of civilized societies without intelligent policy development. And, intelligent policy development is not possible without good data.

Second, the United Nations Fundamental Principles of Official Statistics⁷ state: "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 dissemination of statistical data."

In this context, the government should ensure that Statistics Canada is at arms length from the government. This indeed was the perception of users until recently, grounded in long tradition, rather than in law. With what has happened on the census, it is now time to revise the *Statistics Act* and achieve this objective in law.⁸

Third, the government should seriously look into Statistics Canada's resources to ensure that it continues to produce data that reflect the highest needs of the country. The government needs to understand that Statistics

7 URL: <http://unstats.un.org/unsd/methods/statorg/FP-English.htm>

8 In a letter sent to the Prime Minister, Cappe *et al.* (2010) recommend inserting this UN principle in the *Statistics Act*. They also suggest that the Chief Statistician and not the Minister, should have authority on all matters of technical and methodological nature. URL: <http://afhimelfarb.wordpress.com/2010/09/10/statistics-canada-letter-to-the-prime-minister/>

Canada is principally a data producer for others to use. It should either find a more efficient way to let the Agency re-allocate resources by terminating data collection programs that have become less relevant over time. The difficulty here is that any decision to stop producing *any* data upsets those who use them, often within the federal government, who make it difficult for the Agency to act. Alternatively, if existing data should continue to be produced, as decided by some users, the government should make funding available for critical data needed for good policy development. Developing pension reform, as an example, affecting a trillion dollar industry without having all the needed data, is short-sighted.

Fourth, data of general purpose use must be free. The government should make funds available to make this happen as quickly as possible.

As for Statistics Canada we suggest the following.

First, it must continue its innovative practices to be a truly efficient organization. It may have some potential to reduce its costs further through re-structuring, which it has already begun. It should complete this process as soon as possible.

Second, it needs to pay more attention to making confidential data available to bona fide researchers *without* sacrificing data confidentiality. It should examine technological solutions, such as remote data access, to achieve the objective. The Agency is already pursuing this course. However, it should give this objective higher priority given the value of these data in achieving Canada's economic and social potential.

Third, until the *Statistics Act* gets changed, it must guard against political intervention and preserve its integrity and reputation.

Fourth, it is understandably becoming quite difficult, but Statistics Canada must continue on the path of reallocating resources from low priority to high priority areas.

Concluding Remarks

We have argued in this paper that Statistics Canada is a world class organization. The quality of data it produces is among the best in the world. It used to enjoy a great reputation internationally until the census issue broke open in the summer of 2010.

The positive outcome was achieved despite a number of challenges including the weakness of its legislation, the difficulty of the reallocation of resources because of interest in all the data it produces, and budget cuts. It has been achieved due to the dedication of the Agency staff who take pride in their Agency and its products.

The census issue has put a pointer on the fact that the good outcomes cannot be taken for granted. This paper has provided a list of issues that should be dealt with, with some focused on the government and others at Statistics Canada.

References

- Cappe, Mel, David Dodge, Ivan Fellegi, and Alex Himelfarb (2010) *Letter to the Prime Minister*; September.
- Drummond, Don (2009) *Working Together to Build a Better Labour Market Information System for Canada*, Advisory Panel on Labour Market Information, Human Resources and Skills Development Canada, July 6.
- Economist (1991, 1993) *Good Statistics Surveys*.
- Fellegi, Ivan (1996) "Characteristics of Effective Statistical System," *International Statistical Review*, Vol. 64, pp. 165-197.
- Fellegi, Ivan (2011) "Comments on good data and intelligent government," prepared for the Ian Stewart festschrift authors' workshop, April 20-21, Ottawa, Ontario.
- Standing Committee on Industry, Science and Technology (2010) *Minutes of Proceedings*, July 27.
- Statistics Canada (2009) *Corporate Business Plan, 2009/2010 to 2011/2012*, Statistics Canada website.
- Statistics Canada (2010a) *Report on Planning and Priorities, 2010-11*, Treasury Board Secretariat.
- Statistics Canada (2010b) *Departmental Performance Report, 2009-2010*, Treasury Board Secretariat.

