

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Work Session on Statistical Metadata (METIS)
(Geneva, Switzerland, 10-12 March 2010)

QUALITY INDICATORS AND THE ROLE OF METADATA REPOSITORIES

Contributed Paper

Submitted by Statistics Canada¹,
Session (iii) Metadata and Quality Reporting

I. INTRODUCTION

1. Statistics Canada has recently introduced a number of initiatives to strengthen our quality management and assessment practices. As part of this work, a number of quality indicators have been developed – both at the aggregate level and at the survey program level.
2. At the aggregate level, Statistics Canada has redeveloped its Departmental Performance Report around 26 indicators required by the federal government's Management Accountability Framework. One of these indicators, *quality of outputs*, was developed based on Statistics Canada's Quality Assurance Framework (QAF).
3. At the detailed level, the QAF relates to the overarching process of Quality Management in the General Statistical Business Process Model (GSBPM), where the model was used to review the quality assurance practices of selected survey programs and to identify sub-processes that have a greater risk of errors in quality. Furthermore, quality indicators and good practices were revised as part of the 5th Edition of the Quality Guidelines, which also refers to the GSBPM, and will likely do so more in the future.
4. While the Agency's corporate metadata repository, the Integrated Metadatabase (IMDB), contains a few quality assessment indicators, we are now considering expanding and improving this part of the metadata model, based on the recent initiatives described above, and in order to provide data users these indicators at the survey and statistical program level.
5. This paper outlines the aggregate quality indicators used in Statistics Canada's performance report, the detailed quality indicators in the Agency's Quality Guidelines, and future directions to reporting quality

¹ Prepared by Alice Born, Claude Julien and Laurie Reedman of Statistics Canada.

indicators at the survey level more systematically in the IMDB as part of our overall quality assessment and quality management practices.

II. AGGREGATE QUALITY INDICATORS FOR PERFORMANCE REPORTS

6. The strategic outcome of Statistics Canada is to provide Canadians access to objective, high-quality, non-partisan statistics, statistical products, services and analyses on Canada’s economy and society. As part of this goal, the Agency is expected to provide users measures on data quality – including indicators on access to the data, relevance of the data, and data accuracy.

7. Statistics Canada has redeveloped its Departmental Performance Report around 26 indicators required by the federal government’s Management Accountability Framework. One of these indicators, *quality of outputs*, was developed based on Statistics Canada’s Quality Assurance Framework (QAF) (Annex I). Aggregate quality indicators are now available the Agency’s Departmental Performance Report² for the period ending March 31, 2009.

8. Indicators have been developed for the following dimensions of the QAF – access to data, relevance of the data and data accuracy. Also, indicators measuring organizational efficiency were developed (Table 1). These quality indicators were reported for the economic statistics, social statistics and Census, Demography and Aboriginal domains.

TABLE 1. Aggregate quality indicators

Number of data series downloaded from online database (CANSIM) per year (relevance)

Data series downloaded by external users from CANSIM. The current benchmark is an annual increase of 5%, which is five times the annual population growth rate.

Annual percentage increase in number of data series downloaded from online database (CANSIM) (relevance)

Client satisfaction with Statistics Canada’s ability to meet their needs (5 point scale) (relevance)

Data derived from the client satisfaction survey conducted by programs. The surveys use measures that are consistent across the federal government.

Number of page views of electronic publications (relevance, accessibility)

Annual percentage increase in the number of page views of electronic publications (relevance, accessibility)

Number of print publications sold (relevance, accessibility)

Percentage of major statistical outputs released as planned (timeliness)

Each year Statistics Canada publishes planned release dates for major statistical outputs for the coming year on its website.

Percentage of major statistical outputs whose sampling accuracy is within objectives set (accuracy)

Most surveys are based on statistical sampling. Sampling is an important means of achieving timely and cost-effective results. Sampling accuracy objectives are set on a survey-by-survey basis.

Percentage of reloads of data releases (accuracy)

Outputs that require correction after release. In 2008, Statistics Canada began an ongoing process of monitoring the quality of its data releases to reduce the risk of errors after an official release.

² Available at: <http://www.tbs-sct.gc.ca/est-pre/estime.asp>

International periodicity and timeliness of national accounts and unemployment data release set by Special Data Dissemination Standard (SDDS) of the IMF (timeliness)

Net census undercoverage (accuracy)

The number of people missed less the number of people counted more than once. Following each census, a reverse record check (RRC) is carried out to measure census population undercoverage. RRC estimates the number of people missed, and is combined with the estimate from the census overcoverage study of the number of people enumerated more than once to calculate net undercoverage.

Response rate (accuracy)

Accuracy of disseminated data is directly related to the percentage of respondents asked to provide data who actually do so. Generally, the higher the response rates, the greater accuracy of the survey or census results.

Number of business surveys using tax and administrative data (organizational efficiency)

Percentage of business surveys using tax and administrative data (organizational efficiency)

Contacts avoided because of tax replacement units (organizational efficiency)

Use of tax data and other administrative data sources as an alternative to obtaining data directly from respondents. Reduces number of questionnaires going to respondents and associated response burden.

Average response burden hours per Canadian business establishment (index of response burden hours where 1991=100) (organizational efficiency)

Estimate of response burden hours is calculated annually using the frequency of each survey, the average time to complete the questionnaires and the number of survey respondents. Dividing by the number of Canadian businesses gives the average burden per establishment. The average is indexed with the year 1991 set to 100.

9. The data accuracy indicator, *percentage of major statistical outputs whose sampling accuracy is within set objectives*, is based on comparing planned to actual co-efficients of variation (CVs) for a subset of Statistics Canada's key survey programs³ (e.g., Survey of Employment, Payroll and Hours, Monthly Retail Trade Survey, Monthly Wholesale Trade Survey, Monthly Survey of Manufactures, Quarterly Financial Survey, Labour Force Survey and field crop surveys).

III. DETAILED QUALITY INDICATORS

10. At the detailed level, Statistics Canada's Quality Assurance Framework (QAF)⁴ relates to the overarching process of Quality Management in the General Statistical Business Process Model (GSBPM), where the model was used to review the quality assurance practices of selected survey programs, and to identify sub-processes that have a greater risk of errors in quality. Quality indicators and good practices were revised as part of the 5th Edition of the Quality Guidelines⁵, which also refers to the GSBPM.

11. Quality measures and indicators usually consist of information that is a by-product of the statistical processes. In Statistics Canada's 5th Edition of Quality Guidelines, both quality measures, and when they exist, quality indicators are provided for each phase of the statistical business process. These quality measures and indicators should accompany statistical outputs of the Agency in order to assess the quality of statistical products and as a basis for monitoring performance in terms of quality the processes and products in program areas.

³ Available at: <http://www.statcan.gc.ca/release-diffusion/2009-eng.htm>

⁴ Available at: <http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=12-586-X&CHROPG=1&lang=eng>

⁵ Available at: <http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=12-539-X&CHROPG=1&lang=eng>

12. In Chapter 14 of the guidelines, there are measures related to the whole survey and not associated to anyone phase of the statistical business process. These indicators cannot be measured until the product has been released. These indicators are the most appropriate for reporting in the IMDB but need to be related to the departmental indicators discussed in the previous section. They include measures of timeliness, relevance, interpretability (including completeness of metadata), accuracy (including % of planned estimates meeting CV targets, imputation rates, error rates, etc.), coherence (i.e., differences in results from previous iterations) and accessibility (i.e., number of times a survey product was viewed, formats of products).

13. The Agency's *Policy on Informing Users on Data Quality and Methodology*⁶ requires that each IMDB record contains a statement of key data accuracy issues including measures of coverage error, sampling error, error due to non-response, response error and processing error. Under the category Data Accuracy, survey managers are required to report information on coverage error, co-efficients of variation for key variables, response rates and other relevant data accuracy measures. Additional information on data accuracy, including internal and external quality assessments should be included. Currently this information is not reported consistently across surveys, and it is difficult to compare quality indicators across survey programs.

IV. NEXT STEPS

14. Data quality statements have existed in the IMDB since 1999 under the category Data Accuracy. With the recent development of quality indicators at the aggregate level as measures of the Agency's performance, and the updating of quality indicators in the Agency's Quality Guidelines, internal discussions have started to standardize set of quality indicators reported in the IMDB so that they can be compared across surveys and statistical programs, and be compared with the departmental performance indicators at the domain level (i.e., for economic statistics, social statistics, and census, demography and aboriginal statistics).

15. Statistics Canada has recently adopted the GSBPM as a reference model for describing the survey process and its sub-processes, and the model is used to review the quality assurance practices of selected survey programs and to identify sub-processes that have a greater risk of errors in quality. Referred to as the Quality Reviews, managers responsible for the delivery of a survey or statistical program provide a description of the data quality management tools that are used in the various processes (e.g., questionnaire testing, interviewer monitoring); an assessment of the impact of changes (i.e., as the result of a survey redesign) to these processes on quality, costs, response burden and confidentiality; and a description of how accuracy is managed (i.e., how sampling and non-sampling errors are prevented, identified, corrected and evaluated). It is anticipated that the results of the quality reviews will provide valuable input to identifying quality indicators and quality assurance practices that should be reported in the IMDB and to meet the needs for reporting quality management practices in the Agency.

⁶ Available at: http://www.statcan.gc.ca/about-apercu/policy-politique/info_user-usager-eng.htm

ANNEX I. QUALITY ASSURANCE FRAMEWORK

At Statistics Canada, the dimensions of quality are defined as follows:

The relevance of statistical information reflects the degree to which it meets the real needs of clients. It is concerned with whether the available information sheds light on the issues of most importance to users. Hence, relevance is the most important dimension of quality; one could even consider it among the pillars of a statistical agency. It is largely in the domain of users of the information; it is not something that a statistical agency can establish by itself. Comparatively, the other dimensions of quality are much more within the control of the statistical agency.

The accuracy of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, non-response, response).

The timeliness of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available. It is typically involved in a trade-off against accuracy. The timeliness of information will influence its relevance.

The accessibility of statistical information refers to the ease with which it can be obtained from the Agency. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of accessibility for some users.

The interpretability of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately. This information normally covers the underlying concepts, variables and classifications used, the methodology of data collection and processing, and indications of the accuracy of the statistical information.

The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys. Coherence does not necessarily imply full numerical consistency.