A Framework for the Evaluation of Administrative Data

Mylène Lavigne and Christian Nadeau[1]

Abstract

The decline in response rates observed by several national statistical institutes, their desire to limit response burden and the significant budget pressures they face support greater use of administrative data to produce statistical information. The administrative data sources they must consider have to be evaluated according to several aspects to determine their fitness for use. Statistics Canada recently developed a process to evaluate administrative data sources for use as inputs to the statistical information production process. This evaluation is conducted in two phases. The initial phase requires access only to the metadata associated with the administrative data considered, whereas the second phase uses a version of data that can be evaluated. This article outlines the evaluation process and tool.

Key words: Quality of administrative data; Quality Assurance Framework

1. Introduction

The decline in response rates observed by Statistics Canada in recent years, the desire to reduce response burden and the growing budget pressures it faces are forcing the agency to consider effective alternatives to its current statistical information collection and production practices. To address such challenges, the agency aims to make greater use of administrative data to replace, complete and expand on its statistical programs. A number of factors promote this approach, including the richness and diversity of available administrative data sources, their accessibility and the technical capabilities now available that enable their use.

In general, it is advantageous for national statistical institutes to use administrative data (Brackstone 1987). However, it is crucial that decisions about the acquisition and use of a specific administrative data set consider their nature and quality, the type of intended use, how these data would affect the quality of the resulting statistical products, as well as confidentiality, privacy, and legal and financial considerations. In the context of acquiring and using administrative data files for statistical purposes, it is advisable to conduct a systematic evaluation based on these criteria to inform the decision-making process. Daas et al. (2009), Berning et al. (2013) and the Australian Bureau of Statistics (2009) have already suggested such evaluation approaches.

Some activities conducted by various statistical programs already make it possible to evaluate the quality of administrative data at Statistics Canada. However, it is now appropriate to develop an evaluation framework for administrative data to support all statistical programs in their efforts to acquire new files. This framework is the focus of this article. The objectives of the evaluation process and the underlying principles of the functions of administrative data acquisition and use for statistical purposes are outlined in section 2. The evaluation process and tool are discussed in sections 3 and 4 respectively, after which there is a brief conclusion.

---

[1]Mylène Lavigne, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, KIA 0T6 (mylene.lavigne@statcan.gc.ca); Christian Nadeau, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (christian.nadeau@statcan.gc.ca).

2. Principles and objectives

If an administrative data file is being considered for statistical use, there must first be a data acquisition agreement between Statistics Canada and the data provider[2]. It is assumed that the purchaser[3] is already aware of the existence of the administrative data file in question, that he or she is familiar with some of its aspects and has roughly determined some of the intended uses. Before finalizing any data acquisition agreement, the purchaser must conduct a more comprehensive evaluation during the first two phases of the acquisition and use process illustrated in Figure 2-1.

Figure 2-1
Administrative data acquisition and use model

Once the file has been identified, a four-phase administrative data acquisition and use process begins. The four phases are exploration, initial acquisition, acquisition and use. During the exploration phase, the purchaser contacts the data provider to find out about the file's content, concepts and characteristics as well as the context in which it was created, not necessarily by gaining access to the file itself, but through communications with the provider and a review of the documentation and metadata. If uses for the data have not been established precisely beforehand, they must be during this phase, although they could change later. The purchaser must evaluate the information gathered during this phase to determine whether to pursue the process to enter into a data acquisition agreement for the intended uses.

During the initial acquisition phase, the purchaser may negotiate an agreement with the provider to receive a version of the data that will help him or her essentially—but not exclusively—evaluate their accuracy and coherence. The evaluation helps to determine whether a data acquisition agreement should be established.

During the acquisition phase, the purchaser negotiates an acquisition agreement to get a version of the suitable file for the intended uses. The evaluations in the previous phases may have revealed changes to be made to the file (format, content, etc.); the file must then be validated before proceeding to the use phase.

The last phase, the use phase, includes the implementation of the file's intended uses for the production of statistical information. It also includes the design, development and testing steps of the production processes involving the file, the validation and evaluation of the products that used the administrative data, and the subsequent evaluation of the administrative data inputs. In repeated statistical processes, and as shown in Figure 2-1, the acquisition and use phases could be repeated with subsequent versions of the administrative data without necessarily having to go through the exploration and initial acquisition phases.

Although evaluations are required at all steps of the administrative data acquisition and use process, the evaluation framework outlined here only covers the evaluations conducted during the exploration and initial acquisition phases. The primary objective of the evaluation process outlined in Section 3 is to evaluate the fitness for use of the administrative data files as inputs to pre-defined statistical processes, not to evaluate the statistical products derived from the use of administrative data.

Evaluating administrative data files in terms of this objective depends to a large extent on their intended uses. Lavallée (2007) describes seven types of uses: creating and maintaining sampling frames, building sample designs, partial substitution for direct data collection, edit and imputation, indirect estimation (and weighting), direct

---

[2] The "provider" is defined as the organization providing the administrative data.
[3] The "purchaser" is defined as the individual at Statistics Canada responsible for negotiating the data acquisition agreement with the provider.

tabulation and survey evaluation (data validation). The type of use could have a direct impact on the choice of the various evaluation elements and their relative significance. For example, the "population coverage" element may have greater significance if the intended use of the file is to develop a frame.

The six dimensions of quality defined in Statistics Canada's Quality Assurance Framework (Statistics Canada 2002) were used to develop the administrative data evaluation framework. The evaluation elements outlined in Section 3 were identified with these six dimensions in mind. Table 2-1 shows an interpretation of these dimensions in the context of administrative data use.

Table 2-1
Quality dimensions applied to administrative data

The **relevance** of administrative data reflects the degree to which they meet the real needs of the agency with respect to the intended use. To be deemed relevant, administrative data must, for example, fulfill the reasons for their acquisition (reducing costs, reducing response burden, etc.) or contain the information sought about the target statistical units for the target reference period.
The **accuracy** of administrative data is the degree to which the information correctly describes the phenomenon it was designed to measure. It also relates to the errors that may occur during their production, such as reporting, non-reporting, coverage, interpretation, capture and processing errors.
The **timeliness** of administrative data refers to the time 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 to the statistical agency. There is often a trade-off between timeliness and accuracy. A draft version of an administrative data file may, for example, contain a few shortcomings in terms of coverage of all the units or checking of specific variables. The timeliness of the information will influence its relevance.
The **accessibility** of the administrative data refers to the ease with which the agency can obtain them in their entirety. It includes the ease with which the existence of the information can be ascertained, whether they are presented in an appropriate format, and whether the medium through which the information can be accessed is suitable. In some cases, the cost of the information can also be considered an aspect of accessibility.
The **interpretability** of administrative data reflects the availability of the information and the metadata necessary to interpret and use them appropriately. It generally includes information on the variables, classifications and underlying concepts used, the methodology of data collection and processing, and the indicators of the accuracy of the administrative data.
The **coherence** of administrative data reflects the degree to which they can be matched with other data from the statistical agency within a broad analytical framework over time. The use of standard concepts, classifications and target populations promotes coherence.

Given the wide range of possible uses, it was essential to develop a flexible tool that could be adapted to different needs. The suggested process is based on the analysis of a set of elements whose significance varies by the context of acquisition, rather than on the application of strict decision-making rules based on empirical results. In this context, the purchaser is responsible for establishing the relative significance of the various elements on which he or she will base his or her decision and the sequence in which the elements will be evaluated.

Finally, the process also involves documenting and storing information on the quality of the administrative data source. Information about the evaluation is an important source of metadata that will be valuable for other potential users or any future uses that may be considered. This information may also be used as a source of metadata for the statistical products that use the administrative data.

3. Evaluation process to determine fitness for use

Considering the objectives and principles outlined in the previous section, the proposed evaluation process consists of gathering the required information and analyzing it to evaluate the fitness for use of the administrative data file as an input for an intended use. This evaluation process is divided into four steps: description of the intended acquisition initiative, evaluation during the exploration phase, evaluation during the initial acquisition phase, and the overall analysis and decision.

Table 3-1 summarizes the evaluation process by outlining the elements evaluated by theme. In the **acquisition initiative**, the purchaser documents the acquisition initiative, identifies and describes the administrative data file sought, and specifies the objectives and needs of the acquisition. This first part of the evaluation process is critical since the administrative data file should be evaluated based on the objectives and intended uses described in this phase. However, some elements sought may become clearer during the evaluations of the exploration and initial acquisition phases.

Table 3-1
Evaluation process

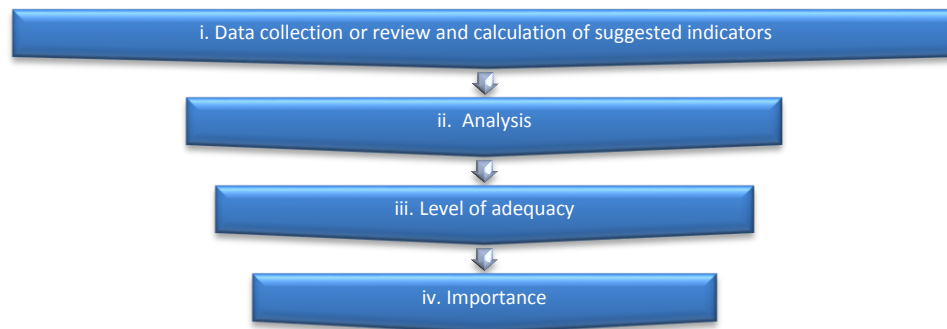|	Theme	Element
Description of the intended acquisition initiative	Identification of the acquisition initiative	Identification of the acquisition initiative and the person in charge
	General information on the administrative data file	Name, description, provider information, information on other organizations involved
	Objectives and needs related to the acquisition initiative	Types of intended use, motivations, objectives, steps to achieve the objectives and needs regarding the target population, reference period, schedule and content
Evaluation during the exploration phase	Institutional environment	Terms and conditions of access and reliability
	Characteristics of the administrative data file	Population (D1, D6), reference period (D1, D6), dates of receipt (D1, D3), content (D1, D6), collection process (D1, D2), file format (D4), variable format (D4), processing (D1, D2), metadata (D5)
	Impact of using the administrative data file	Costs, human resources and response burden
Evaluation during the initial acquisition phase	Evaluation by error type	Coverage, non-response, declaration, measurement and processing errors (D1, D2)
	Data validation	Validation of the various statistics produced from the administrative data file (D2)
Overall analysis and decision	Summary of information collected	Note: Six dimensions of quality: D1 = relevance, D2 = accuracy, D3 = timeliness, D4 = accessibility, D5 = interpretability, and D6 = coherence.

The intent of the **exploration phase** is to evaluate certain characteristics of the administrative data source that do not require access to the data. The evaluation is based on information that can be obtained from the provider through meetings, discussions, information sharing, documentation sharing or other means. It focuses on the relevance, timeliness, accessibility and coherence of the data. The evaluation during the exploration phase looks at the institutional environment[4], the characteristics of the administrative data file as they relate to the intended use and the impact of using the administrative data on some aspects of the statistical process.

During the **initial acquisition phase**, the evaluation focuses on the accuracy and coherence of the data. The process evaluates certain types of errors by calculating various accuracy indicators using one version of the data source and validating it by comparing it to other known sources. A preliminary version of the administrative data file (sample or others) can be used for this evaluation, which also provides an update of certain results from the evaluation during the exploration phase, if required. It complements the first evaluation phase and helps to evaluate the relevance of acquiring the administrative data source.

---

[4] The evaluation of the institutional environment takes into account organizational or institutional factors that could affect the provider's ability to provide data that meet the quality requirements. This includes an evaluation of the provider's quality standards and the procedures required to maintain quality.

For these two phases of the evaluation process, elements are evaluated independently of each other based on the intended use. Thus, it is recommended to evaluate each element of the exploration and initial acquisition phases in four steps. Figure 3-1 shows the four steps of the evaluation process during the exploration and initial acquisition phases.

Figure 3-1
Steps of the evaluation for each element based on the intended use during the exploration and initial acquisition phases

For each phase, the first step is to collect and document the information sought on the element being evaluated. The second step consists of an analysis of the fitness for the intended use of the administrative data file, considering only the information element. This analysis should take into account the information collected during the first step and the description of the acquisition initiative. The third step is to evaluate the level of fitness for use of the administrative data file (high, moderate or low) for this element, which is inferred from the analysis in the second step. Finally, the fourth step is to assign a relative importance (mandatory, important, less important, no importance) to the information element in the context of the intended use. This relative importance will be taken into account during the overall analysis and acquisition decision. However, in some cases, the decision to not enter into an acquisition agreement may be made after the exploration phase, should important shortcomings have been noted.

The overall analysis and acquisition decision are the last parts of the evaluation process. The purpose of the proposed compilation and analysis process is to easily and quickly identify the gist of the data collected and documented through the evaluations during the exploration and initial acquisition phases in order to facilitate a decision. It reflects the administrative data file's suitability for the needs and the importance assigned to each element collected. The intent of this process is more to systematically guide decision-makers through the information collected than to provide absolute decision-making rules.

The evaluation process presented assumes that the administrative data file is not accessible at the beginning of the acquisition process. In this case, the evaluations during the exploration and initial acquisition phases are conducted sequentially. Should the file be accessible at the beginning of the acquisition process, the two phases can be conducted at the same time without any concerns. It is important to note that the evaluation process provides some flexibility with respect to the sequence of evaluation of the elements. This sequence can be based on the importance of the elements evaluated and the level of effort required. We can seek to minimize the effort required to reject sources initially perceived as unsuitable by first evaluating the most important elements requiring the least effort. A complete evaluation may be preferred for sources initially deemed suitable. Some elements may not even be evaluated for some uses.

4. Evaluation tool

The tool used to implement the evaluation process consists of a guide (Statistics Canada 2015a), a questionnaire (Statistics Canada 2015b), an evaluation grid and a recommendation sheet. The guide describes the evaluation process and provides a definition of the errors, concepts, information sought and the levels of suitability and importance of each element. It also describes certain evaluation methods and indicators to help the purchaser evaluate the administrative data file at the initial acquisition phase. Furthermore, instructions are given for the questionnaire, the evaluation grid and the recommendation sheet.

The questionnaire is the tool of choice to collect and document information on the first three parts of the evaluation of an administrative data file. In addition to fulfilling the primary objective of evaluating the suitability of the administrative data file, it also fulfills the objective of documenting and storing information on the data quality. The questionnaire comprises eight sections. Sections A, B and C collect the information needed for the description of the intended acquisition initiative. Sections D through H serve to evaluate different elements of the exploration and initial acquisition phases and to collect the information elements required to evaluate the fitness for intended use of the administrative data file. The questionnaire follows the same format for each evaluation element in the exploration and initial acquisition phases, i.e., the four steps shown in Figure 3-1 of the previous section. For example, for the "Population" element, the questionnaire would look like the one shown in Figure 4-1.

Figure 4-1
Example from questionnaire

Quality dimensions: **Relevance** and **coherence**			
i. Briefly describe the **population** of the administrative data file.			
ii. Determine whether the **population** of the file described in (i) fulfills the needs reported in part C.5. If it does not fulfill them exactly, describe the extent of the gaps and their impact on the intended use. Identify any potential means to mitigate this impact or the lack of such means.			
iii. Based on the analysis done in (ii) concerning the **population**, evaluate the level of fitness for intended use of the administrative data file.			

Level of fitness for use	High	Moderate	Low
	O	*O*	*O*

iv. What **importance** must be given to the "**population**" element for the intended use?			

Importance	Mandatory	Important	Less important	No importance
	O	*O*	*O*	*O*

The evaluation grid and recommendation sheet use spreadsheets to compile and analyze the information collected during the evaluations in the exploration and initial acquisition phases. If the evaluations are conducted sequentially, this step will be completed twice, once after the exploration phase and again after the initial acquisition phase. However, if both are conducted at the same time, this second step would be completed only once, after filling out the questionnaire. An example of the evaluation grid is shown in Figure 4-2. The content of the evaluation grid can be sorted in increasing order of importance or filtered to clearly show the level of fitness in relation to the needs based on the importance given to the elements. The recommendation sheet is used to document and justify the decision to proceed with the acquisition process based on an analysis of the evaluation grid and the relevant data documented by the evaluations during the exploration and initial acquisition phases of the questionnaire. An example of the recommendation sheet is shown in Figure 4-3.

Figure 4-2
Sample evaluation grid

Phase	Element	Importance	Fitness for use
Exploration	D.1 Terms and conditions of access	1-Mandatory	1-High
	D.2 Reliability	3-Less important	2-Moderate
	E.1 Population	2-Important	3-Low
	E.2 Reference period	3-Less important	3-Low
	E.3 Potential dates of receipt	2-Important	1- High
	E.4 Content	2-Important	2-Moderate
	E.5 Collection process	4-No importance	Not evaluated
	E.6 AD file format	3-Less important	3-Low
	E.7 Variable format	3-Less important	2-Moderate
	E.8 Processing	2-Important	1-High
	E.9 Metadata	4-No importance	2-Moderate
	F.1 Costs	2-Important	3-Negative
	F.2 Human resources	3-Less important	2-Moderate
	F.3 Response burden	2-Important	1-Positive
Initial acquisition	G.1 Coverage errors	1-Mandatory	1-High
	G.2 Non-response errors	2-Important	2-Moderate
	G.3 Declaration, measurement and processing errors	3-Less important	2-Moderate
	H.1 Validation based on comparison of statistical data	2-Important	1-High

Figure 4-3
Sample recommendation

Question 1: Recommendation Statement	Recommendation to pursue acquisition	
	Yes	*O*
	No	*O*
Question 2: Justifications based on the analysis of results in the evaluation grid and on information on the mandatory and important elements		

5. Conclusion

Various methods are used to assess the quality of administrative data. However, it is now relevant to develop an evaluation framework specifically for administrative data to support the acquisition efforts of all the statistical programs. The main purpose of the administrative data evaluation framework being developed at Statistics Canada is to evaluate the suitability of administrative data files as inputs to previously identified statistical processes.

A sequential evaluation process and an evaluation tool are proposed to assess the potential fitness for use of an administrative data file. The proposed approach fosters using dedicated analytical resources wisely by identifying the key elements to evaluate. The proposed evaluations are first completed using metadata during the exploration of a new administrative data file, before even gaining access to the file, and then during the acquisition of a version of the data that can be evaluated more in depth. Several elements are evaluated to determine the fitness for intended use of the file. The procedure for each element involves a number of steps—information collection, analysis, and evaluation of the level of fitness and the importance of each element—and is accomplished through a questionnaire and a supporting guide. The process then guides the purchaser through a simple analysis using various dimensions and highlighting the most important characteristics involved in the acquisition decision. While the evaluation elements are listed sequentially in the questionnaire, they may be used in any order, thereby giving the tool a high degree of flexibility to fulfill the requirements of various data acquisition contexts.

The proposed process and tool will be changed and improved based on the feedback received during their use. For example, a potential improvement would involve the tool's transition to an electronic format. In addition, the framework could be expanded to include the evaluations during the acquisition, use and repeated use phases.

Acknowledgement

Bibliography

Australian Bureau of Statistics (2009), "ABS Data Quality Framework", Australia: Australian Bureau of Statistics.

Berning, M., Iwig, W., Marck, P. et Prell, M. (2013), "Quality Assessment Tool for Administrative data", Working Paper, Federal Committee on Statistical Methodology, United State.

Brackstone, G. (1987), "Issues in the Use of Administrative Record for Statistical Purposes", Methodology Survey, vol. 13, n° 1, p. 35-51.

Daas, P., Ossen, S., Vis-Visschers, R. and Arends-Tóth, J. (2009), "Checklist for the Quality Evaluation of Administrative Data Sources", Discussion paper 09042, The Hague/Heerlen, Netherlands: Statistics Netherlands.

Lavallée, P. (2007), "Administrative data usage in the framework of social statistics: Current and future picture", Unpublished report, Ottawa, Canada: Statistics Canada.

Statistics Canada. (2002), "Statistics Canada's Quality Assurance Framework", Catalogue no. 12-585-XIE, Ottawa, Canada: Statistics Canada.

Statistics Canada. (2015a), "Evaluation Guide for Administrative Data: Exploration and Initial Acquisition Phases", Unpublished Paper, Ottawa, Canada: Statistics Canada.

Statistics Canada. (2015b), "Evaluation Questionnaire for the Administrative Data Acquisition Process", Unpublished Paper, Ottawa, Canada: Statistics Canada.