**Strategies to improve and maintain the quality of administrative data**

Hélène Bérard, Statistics Canada, helene.berard@canada.ca

Cory Lusk, Statistics Canada, cory.lusk@canada.ca

Josée Cellard, Statistics Canada, josee.cellard@canada.ca

**Abstract**

*Statistics Canada actively obtains administrative data from the public sector, the private sector and various organizations to support the production of official statistics. A number of strategies are in place to facilitate data access and to ensure that the data obtained from the various organizations meet specific quality requirements. The strategies implemented to meet these goals include: performing outreach activities prior to obtaining the data to explain the usefulness of the data, offering technical support, organizing data quality workshops to promote the use of sound methods, and identifying specific requirements in data acquisition agreements to ensure the quality and timeliness of the data. It also includes negotiating access to a test file to assess fitness for use before proceeding with an official acquisition, and establishing a Quality Evaluation Framework to ensure consistency and completeness of assessments across the departments many administrative data files. As well, the strategies involve maintaining post acquisitions communications, and developing mutually beneficial relationships to ensure long term supply, usefulness and quality of the administrative data. One of our challenges is to identify and influence future changes to administrative data sources that may affect statistical use. This paper will discuss specific examples to illustrate the challenges and the lessons learned in establishing these various strategies during the life cycle of the partnership process. It also includes a description of Statistics Canada corporate strategy regarding the statistical use of administrative data, which serves as the foundation for the development of these strategies.*

**Keywords:** administrative data acquisition, partnerships, quality assessment of administrative data

1. **Introduction**

Administrative data, by definition, are collected for administrative purposes by organizations other than the National Statistical Office (NSO). As a consequence, the statistical use of administrative data is inherently more complex than the direct collection of survey data from respondents. For example, the legal or commercial basis on which the administrative data were originally collected may affect the other party’s ability or willingness to disclose the data to the NSO and the ability of the NSO to disclose administrative data to others. In addition, serious privacy and public communications considerations may also arise when making secondary use of data originally collected for another purpose. An early summary of the issues for using administrative data for statistical purposes can be found in Brackstone (1987).

The availability of an increasing number of data sources offers unique opportunities that have been recognized by many NSOs in spite of these challenges. For example, administrative data is used increasingly to replace, complement or support more efficient survey collection. Increasing the use of administrative data provides new statistical and analytical outputs that are relevant to information requirements that are key priorities for Statistics Canada, as explicitly stated in consecutive Corporate Business Plans (see Statistics Canada, 2016 for the 2016/2017 to 2018/2019 Statistics Canada Corporate Business Plan).

Statistics Canada’s long-term strategy regarding its administrative data program is presented in section 2. The strategic partnerships developed to facilitate access to administrative data that are fit for use are presented in section 3.

1. **Statistics Canada Administrative data program**

The Administrative Data Secretariat (ADS) was created in 2012 to develop and coordinate an administrative data program to foster a more efficient statistical use of administrative data. Key outcomes targeted by this program include: accessing and using efficiently administrative data, reducing response burden, lowering costs, improving quality and maintaining public trust.

Five broad areas requiring improvements (or pillars) were identified to help achieve these outcomes: 1) more supportive governance frameworks related to administrative data, 2) improved partnerships with data providers, 3) more effective two-way communications with data providers and stakeholders, 4) improved knowledge and skills together with more comprehensive and coherent corporate tools for NSO employees that are planning to obtain or use administrative data, and 5) enhanced research to efficiently use more new data sources.

Many initiatives have been carried out to support the achievement of these outcomes. An overview of the various initiatives and strategies are described in Table 1. More details related to the governance framework are described in section 2.1.

Table 1. Statistics Canada administrative data program pillars to support a more efficient use of administrative data for statistical purposes while maintaining public trust.

|  |  |
| --- | --- |
| **Pillars** | **Initiatives and strategies** |
| **Governance** | **Organizational changes** with the establishment of centers of expertise* Developed **centralised services** to support the acquisitions in 2014.
* Created the **Data Stewardship Secretariat** in 2018 to better support the development, coordination and implementation of sound management practices for the life cycle of the data source.

New **policy instruments** to support efficient management* Developed the [Statistics Canada Policy on the Use of Administrative Data Obtained under the Statistics Act](https://www.statcan.gc.ca/eng/about/policy/admin_data), directive and guidelines describing roles and responsibilities related to the acquisition and management of administrative data.
 |
| **Partnerships** | **Data provider centric strategies** to support strategic partnerships* Established the role of a **liaison** responsible for negotiating on behalf of the NSO with the data provider to support efficient communication with the data provider and within the NSO. The liaison becomes the **NSO data custodian** of the administrative data obtained and inherits specific roles and responsibilities for its management.
* Promoting a **win-win approach** when negotiating data acquisition agreements.
 |
| **Communications** | **Inward facing** (within the NSO) **communication** facilitated through the development of new tools and processes* Broadcast announcing new potential data sources that serves to inform but also to document corporate needs to be considered when negotiating the administrative data acquisition.
* Working groups discussing challenges associated with specific data sources including exchanging information on the data quality evaluation processes.

**Outward facing** **communication** facilitated through the development of new tools and processes* Website page explaining what is administrative data as well as [why and how it is used for statistical purposes.](https://www.statcan.gc.ca/eng/about/about?MM=as#a3)
* [Frequently asked questions and answers](http://www.statcan.gc.ca/eng/about/admin_data_faq) for external and internal audience regarding the statistical use of administrative data.
* [Brochure](http://www.statcan.gc.ca/eng/about/admin_data/brochure) with key information regarding the statistical use of administrative data
* Communication material developed when contacting data providers that includes information such as the authority under which data is obtained, for what purpose, and how it will be protected.
 |
| **Knowledge, skills & tools** | **Enhanced processes and tools*** Handbook to guide employees planning to obtain or use administrative data. Generic templates for data acquisition agreements
* Administrative data inventory to inform the NSO employees about data sources already obtained.
* Quality evaluation tools for administrative data
* Courses on the new policy instruments
 |
| **Research****& development** | **Developed, enhanced methods to facilitate the efficient use of administrative data*** Continue to experiment with techniques to improve efficiency such as methods to facilitate the automation of specific statistical processes (for example preprocessing and processing).
* Continue to enhance the Statistics Canada statistical register base system to facilitate the efficient ingestion and use of multiple data sources
 |

*2.1 Governance framework*

The legal landscape that must be considered when developing a data acquisition agreement includes relevant Canadian laws and regulatory framework of the various jurisdictions of the ten provinces and three territories, as well as those that relate to specific data providers from the public and private sectors.

Statistics Canada has the legal authority to obtain and use administrative data for statistical purposes and research under Section 13 of the [*Statistics Act*](http://laws-lois.justice.gc.ca/eng/acts/S-19/FullText.html)*.* The coverage of Section 13 includes federal, provincial, territorial and municipal governments, as well as private businesses and other types of organizations.

Statistics Canada operates using a well-established governance framework that relies on many policy instruments such as policies, directives and guidelines, covering the acquisition, management and the statistical use of administrative data along with a broad range of other topics such as information management, privacy, security, etc. The policies and the directives provide respectively formal direction and formal instructions while the guidelines provide guidance and advice.

In particular, the Statistics Canada Policy on the Use of Administrative Data Obtained under the Statistics Act includes a set of ten principles that are conducive to the development of enriched partnerships with different data providers. Some of the key principles include: consideration for the impact on privacy, establishment and maintenance of relationships that are on-going and mutually beneficial and, transparent use of administrative data.

1. **Successful partnership strategies**

Successful partnership strategies draw on many of the initiatives and strategies developed under the five pillars described in section 2. For example, having a robust governance structure with well-defined roles and responsibilities such as those that fall under the liaison (NSO data custodian) facilitate efficient communication with the data provider as well as sound data stewardship. The creation of centralised function to support the acquisition of administrative data contributed largely to improving the acquisition strategy and avoiding duplication within the NSO. Furthermore, well established communication material facilitate the exchange of information and address concerns that the data providers may have.

The sections below present the strategies that were implemented as well as the lessons learned when developing successful partnerships with data providers from both the public and private sectors. They are presented under three broad categories that reflect the life cycle of the partnership: preparatory work, launching the acquisition process, and establishing and maintaining the partnerships.

* 1. *Preparatory work*

A number of steps done prior to launching an acquisition process can be instrumental in initiating and maintaining successful partnerships. They are described below.

* + 1. *Data discovery*

Given the growing availability of data sources, proactively exploring the web can lead to interesting discoveries that complement targeted searches for data on specific topics and help identify potential data providers. As well, a review of the available metadata for a specific data source, depending on its completeness, may be sufficient to decide to exclude a data source from further investigation, if basic requirements are not met.

A number of current Statistics Canada initiatives include conducting environmental scans of various types of open data. Given their nature, open data files are often created by aggregating information produced at more granular levels. Hence, further inquiries with the open data provider can be considered to meet NSO specific needs for more granular data.

* + 1. *Understanding the data provider business model*

Understanding the data product and documenting the data provider business model facilitates efficient communication and helps to orient discussions that address potential concerns but also identify mutually beneficial outcomes. Aspects to consider in this review include the following:

* Data products
	+ Review the information readily available to help determine the potential of the data source and its limitations given the intended statistical purpose.
	+ In the context of a win-win approach, explore how specific products produced by the NSO’s could be of interest to the data provider.
* Regulatory framework
	+ Identify specific laws and rules that apply to this organization and in particular those that could restrict the use of the administrative data for statistical purposes.
	+ Determine if the data provider collects its own data or if it was obtained from another organization. For example, commercially available data originally produced by an organization may be obtained through a third party. Specific rules set by the original data producer may apply to the data obtained through a third party.
	+ Consider intellectual property rights associated with the data product and the legislative framework that governs it.
* Hierarchical structure
	+ Identify the individual that should receive the initial communication material that is being sent by the NSO. This crucial step is not always an easy task. A review of publicly available information supported by direct contact with the organization may be necessary to confirm the contact information.
	1. *Launching the acquisition process*

The efficient launch of the acquisition process greatly depends on well-planned communications within the organization, the data provider, other stakeholders and the public. Over the past few years, we implemented a number of new processes based on lessons learned. Our approach is to continuously improve these processes and respond adequately to emerging needs.

* + 1. *Internal communication*

At Statistics Canada, the data acquisition process is supported by a robust governance framework that allows for organizational buy in. Tools such as acquisition process flows and checklists are communicated to subject matter experts to guide them through the acquisition’s various stages. A number of formal communication processes have been implemented as described below.

* + - 1. *Corporate consultations*

Typically, data acquisitions start with a broadcast that serves to announce new potential data sources and identify interested users within the NSO. These are followed by corporate consultations to gather and document the requirements from the different data users, which often represent various statistical programs with specific needs. This avoids costly duplication of efforts and ensures that all corporate needs are documented.

* + - 1. *Working groups*

Corporate consultations are supplemented by theme based working groups. The goal of these working groups is to promote discussion and co-operation between internal stakeholders and address any issues related to the statistical use of a specific data source. It also serves to coordinate the quality assessment of the data obtained and identify any issues that should be communicated to the data provider. The specific corporate requirements are reviewed and prioritized. Other operational considerations such as how the data will be received, stored, pre-processed, processed, documented and managed are addressed.

Consultations with working groups occur over the duration of the negotiation process given that engagement with the data provider may lead towards the discovery of other data sources or specific limitations that restrict the statistical use of the data.

* + 1. *Communication with the data provider*

The initial contact with a data provider is key. Introductory packages are sent out to data providers that include an opening letter summarizing the uses and impacts that the data will have on Canadians, outlining the NSO authority under which the data is requested and its legal responsibility to protect the confidentiality and security of the data obtained. Other documents provide additional details about the information required and its uses for statistical purposes along with a data acquisition template that outlines items such as legal aspects, terms and conditions of use, information management protocols and security elements.

Shortly after, a conference call is scheduled with the data provider to address any questions or concerns relative to the introductory package. Typical concerns are often centered on elements related to the NSO’s legal authority to obtain the data, as well as information management and Information Technology (IT) aspects related to security and confidentiality.

Typically, a first conference call is followed by other conference calls and face to face meetings that explore more specific concerns. It is important to determine the key individuals that will be involved in these discussions and the negotiations of the acquisition agreement early in the process. As the negotiations evolve, different individuals representing the data provider and the NSOs will participate in the discussions based on expertise (legal requirements, knowledge of the product, technical aspects related to specific requirements such as data transmission, encryption, security requirements, etc.) and appropriate level of authority.

It is found to be beneficial to engage the NSO subject matter experts in the discussion with the data providers early in the process to explain how the NSO plans to use the data for statistical purposes. Often exchanges between the data experts representing the data provider and the NSO lead to innovative win-win approaches. IT experts must also be involved early in the negotiations as technical challenges related to transmission, storage and processing of the data may cause specific challenges for the NSO or the data provider.

A preferred option is to negotiate obtaining a test file to determine the fitness for use of the data that includes a nondisclosure agreement. The agreement states that the data will be used to perform a feasibility study and will not be used for production. This approach also serves to establish the partnership in a progressive manner.

Non-confidential reports regarding the quality assessment of the data (test files, or data obtained for the production of official statistics) that are shared with the data provider can lead to mutual benefits as they can serve to improve the quality of the information produced by the data provider.

* + 1. *External communication*

Supporting a more efficient use of administrative data for statistical purposes while maintaining public trust requires transparency in the NSO’s use of the different data sources that are obtained.

Information is directly available on the Statistic Canada website that gives a definition of administrative data, how it is used, the benefits of using administrative data, and how the confidentiality and security of the data is protected (see table 1 for links to specific products).

* 1. *Establishing and maintaining a partnership*

At Statistics Canada long lasting successful partnerships have been established with a large number of public sector organizations. These partnerships are fostered through efficient communication carried through a number of forums and regular events (for more details see Statistics Canada, 2017, chapter 1.4) and are beneficial for all stakeholders. They serve to identify data gaps and to get advance warning on policy changes that may affect data produced by different stakeholders.

Successful win-win approach with other departments include: 1) the return of processed micro data files as permitted following confidentiality requirements, 2) providing quality reports, and 3) hosting ad hoc Data Quality Workshops with data providers to ensure the quality of administrative data.

More recently Statistics Canada has been establishing a number of new partnerships with private data providers. Strengthening the relationship with these new data providers may require the implementation of different communication strategies to ensure efficient on-going communication. Outreach activities such as participating at conferences organized by the data providers has proven to be a useful exercise.

Examples of mutually beneficial initiatives that contribute to a successful partnership with private sector data providers include: 1) returning more complete metadata prepared by the NSO to the data provider, 2) providing non-confidential quality reports on the data source obtained, and 3) providing non-confidential tabulation for statistics of interest to the data provider.

1. **Moving forward**

Many initiatives were brought forward recently that contributed to streamlining the acquisition process and supporting successful engagement with existing and new data providers.

Some of the challenges ahead include our ability to discover and assess quickly the fitness for use of new data sources and to work collaboratively with data providers to ensure a win-win approach that enables access to better quality data.

We need to continue to develop communication material to better inform organizations and the public about the NSO many initiatives aimed at using administrative data more efficiently.

**5. References**

Brackstone, G. J. (1987), [Issues in the use of Statistical Records for Statistical Purposes](http://www.statcan.gc.ca/pub/12-001-x/1987001/article/14467-eng.pdf), Survey, Methodology, Vol 13, No. 1,pp 29-43, Statistics Canada, Available at <http://www.statcan.gc.ca/pub/12-001-x/1987001/article/14467-eng.pdf> (Accessed: May, 2018)

Statistics Canada, (2016), Compendium of Management Practices for Statistical Organizations from Statistics Canada's International Statistical Fellowship Program, Available at: <https://www150.statcan.gc.ca/n1/pub/11-634-x/2016001/section1/chap4-eng.htm> (Accessed: April 27, 2018)

Statistics Act (R.S., 1985, c. S-19, s. 2; 2017, c. 31, s. 1.), Available at <http://laws-lois.justice.gc.ca/eng/acts/S-19/FullText.html> (Accessed May 6, 2018)

Statistics Canada (2015), Statistics Canada Policy on the Use of Administrative Data Obtained under the Statistics Act. Available at <https://www.statcan.gc.ca/eng/about/policy/admin_data> (Accessed May 6, 2018).

Statistics Canada Corporate Business plan (2016), Available at <https://www.statcan.gc.ca/eng/about/bp> (Accessed April 13, 2018)