

37th Voorburg Group Meeting

Virtual Meeting hosted by Canada
September 13-22, 2022

Cross-Cutting Paper:

Guidelines for Incorporating Alternative Data Sources in Official Statistics

Central Statistics Office, Ireland - Anthony Dawson

Statistics Denmark - Rohan Draper

Office for National Statistics, UK - Scott Kilbey

Statistics Canada - Martin Beaulieu, Kyle Virgin

Table of Contents

Foreword	3
Definitions & Links	4
Approach	4
GSBPM Considerations	6
Fitness for Use Questionnaire	8
Proposals and Discussions	8
Conclusion	9
Appendix A - Fitness for Use Questionnaire	10
Bibliography	15
Additional Reading and Relevant Links	15

Foreword

This Voorburg Group task force was formed with the mandate of producing guidelines for any National Statistical Institution (NSI) working to incorporate alternative data sources into their statistical outputs. This paper was written with the aim of generating discussion and forming consensus towards the ratification of international guidance. It should be viewed in the context of discussions on data quality and alternative data sources that took place at Voorburg Group 2021. (Cross cutting topic 7: Alternative data - best practice for evaluating fitness for use)

At the time of writing there is little agreement on how to best measure data quality beyond survey sources¹. The most advanced proposal was published by StatsNZ², but this focuses primarily on administrative data sets and starts from the point where data have been acquired.

It is important to note the shift in types of data sources being used by NSIs. Whilst administrative data has been used for official statistics for some time, the 2010s saw the first uses of other alternative data sources³. The first reference to alternative data sources in Voorburg meeting papers was 2015⁴ and their use has only increased since, as exhibited by a recent Voorburg Group survey showing this is relatively widespread. This task force contends that it is important to account for the unique aspects of data source types when incorporating and monitoring their use in Official Statistics.

This paper takes current literature on alternative data sources and data quality and aligns them with the traditional General Statistical Business Process Model (GSBPM) framework - starting from the decisions to move forward with an acquisition and initial research. It will also build upon the alternative data source types identified in a previous Voorburg task force⁵. Included alongside this paper is a practical tool that any NSI can use to implement alternative data sources, suggesting measures that can be taken to assess and mitigate risks to quality and sustainable statistics.

¹ [Transparency in the Reporting of Quality for Integrated Data: A Review of International Standards and Guidelines](#) (Czajka, Stange - Mathematica, 2018)

² [Guide to reporting on administrative data quality](#) (Stats NZ, 2020)

³ [The use of supermarket scanner data in the Dutch CPI](#) (CBS, 2010)

⁴ [Use of alternative data sources in Canadian SPPIs](#) (Garneau - Voorburg, 2015)

⁵ [Voorburg Task Force - Alternative Data Sources](#) (Aizcorbe et al. - Voorburg, 2020)

Definitions & Links

Definitions

- Alternative data source - refers to any non-survey data source
- Subcategories of alternative data sources as identified in Voorburg Task Force on Alternative Data Sources:

Source	Type	Origin	Code	Description	
TRADITIONAL	STRUCTURED	Questionnaire (paper, phone and/or electronic)	QNR	The traditional way of collection price information by asking firms for the information via telephone, paper or electronic questionnaires. This is the default and not an alternative data source.	
		Web Prices (manual)	WPR	The price collection data is derived from websites manually. The data source itself is considered an unstructured data source not designed for statistical purposes.	
EXTERNAL	UNSTRUCTURED	Webscraping (automated)	WSC	The price collection data is derived from websites automatic systems. The data source itself is considered an unstructured data source not designed for statistical purposes.	
		Administrative Data Source	ADM	Data which are derived from the operation of administrative systems by public agencies (e.g. data collected by government agencies for the purposes of registration, transaction, regulation and record keeping). Data is often structured for administrative purposes and is highly transferable for statistical purposes.	
	Corporate Datasets	COR	Survey respondent provided datasets obtained directly from corporate headquarters in lieu of data collectors collecting data in respondent stores or on their websites. Data pertains to the particular company that is providing said data is often structured for organisational purposes and is highly transferable for statistical purposes.		
	Trade Associations	TAD	Industry based surveys that the target industry is producing for themselves.		
	STRUCTURED	Data Vendors (commercially available structured data)	DVS	Data acquired from companies that actively collect and sell data as a business activity. Often such companies provide data on a contractual basis with defined terms and conditions.	
		Consultancies (mandated specific task) (transformed data)	CON	Consulting company and/or specialist company is contracted to collect and/or compile data for a specific purpose (mandated or otherwise). Often such companies are utilised on a contractual basis with defined terms and conditions.	
		Credit card and bank data	CCD	Financial information collected at the moment of a transfer of funds between a card holder's account and a business account. Data is graded based on the level of metadata available about the transaction. This source is considered a structured data source.	
		Other alternative data sources n.e.c.	OTH	Other types of alternative data sources not elsewhere classified. For example, transaction-level data from email receipts (like UBER email receipt data). Other special data delivery from third party data collectors not elsewhere classified.	
	INTERNAL	STRUCTURED	Consumer Price Index	CPI	Data is sourced directly from the Consumer Price Index
			Producer Price Index	PPI	Data is sourced directly from the Producer Price Index
Structural Business Statistics			SBS	Surveys utilised for benchmarking purposes	
National Accounts			NA	Price indices derived from volume and value data (implicit price indices)	

- GSBPM - General Statistical Business Process Model⁶
- NSI: National Statistical Institution - also known as National Statistics Office (NSO)
- Sources of Error (from Stats NZ's Guide to Reporting on Administrative Data drawing on the Total Survey Error⁷ framework)
 - Validity
 - Measurement
 - Processing
 - Frame
 - Selection
 - Missing/Redundancy

Approach

The emergence of new alternative data sources has provided NSIs with excellent opportunities to meet new data requirements, fill existing data gaps or improve the quality of the existing indices. However, alternative data sources can provide challenges for assessing quality that are not met by the existing documentation on standards and procedures for survey design and monitoring. For example, NSIs may be unable to:

⁶ [GSBPM - General Statistical Business Process Manual](#) (version 5.1)

⁷ [Total Survey Error: Past, Present and Future](#) (Groves, Lyberg, 2010)

- Specify or design these data sources, meaning that there is no control over variables
- Collect and process raw data. This can lead to some methods potentially being hidden but included in the received data product. This is an example of a processing error, as defined by the Total Survey Error Framework and expanded on in Stats NZ's Guide to Reporting on Administrative Data.

These challenges may vary depending on the data type. For example, an NSI may have more control over administrative data received from another government agency versus scanner data collected from a retailer. It is therefore important to consider the weaknesses, unknowns and constraints of alternative data sources as well as any possible implementation measures. Structured and considered evaluation of quality will help NSIs to mitigate risks and determine the suitability of a new alternative data source.

It is also important to consider ethics from the start of the acquisition process. For structured datasets sourced from the private sector it is especially important to keep in mind the incentives of each party involved in the discussion of acquisition. This is in order to maintain the impartiality of the NSI and avoid any unforeseen or undesirable impacts, intentional or otherwise.

Non-governmental data sources have different purposes and assumptions due to the compelled nature of administrative sources. For the purposes of this paper, the following unique assumptions are made about administrative data sources as opposed to other alternative sources:

- Data are derived from administrative systems operated by public agencies
- The original purpose is usually registration, transaction, regulation or record keeping
- There is most likely a legal or compelled nature to the original collection
- The data are structured and sometimes highly linkable
- Data are highly transferable for statistical purposes

In order to accurately compare the quality and fitness of use of alternative data sources to existing well established methods it is important to use comparable methods of evaluation whenever possible. It is for this reason the task force has recommended the use of an adapted version of the well established General Statistical Business Process Model (GSBPM) (). The GSBPM is a means to describe statistics production in a general and process oriented way. By structuring evaluation around the GSBPM, new data sources can be evaluated against existing data sources to compare to existing quality standards where a gap is being filled, or compared directly to assess whether a new data source is a quality upgrade on an existing method of data collection.

Certain aspects of the GSBPM are not relevant when assessing alternative data sources for fitness of use and quality. In light of this, the task force has developed an easy reference questionnaire which is tailored towards alternative data sources. Comparing new data sources under this questionnaire will also allow NSIs to “fail fast” when it comes to the implementation of new data sources into the statistical process. It will help with the identification of critical sources of error or issues in data acquisition and implementation that cannot be easily mitigated or overcome.

Outlined below is an overview of the GSBPM and the type of information that should be assessed when determining a new alternative data source's fitness for use.

GSBPM Considerations

The rise in the availability and the use of alternative data sources brings new perspectives at each phase of the GSBPM. Data ethics considerations, while not being new, deserve an increased focus when alternative data are involved. They are involved throughout the statistical process from the Specify Needs phase to the Disseminate and Evaluate phases. The NSI team may require an increased emphasis in skillsets less frequently seen in the statistical process (i.e. contract negotiations, legal/policy experts) to work closely with experts in collection, subject matter and statistical methods.

In the Specify Needs phase, once the data needs and desired outputs are well identified, alternative data play an important role in the Check Data Availability sub-phase. When acquiring new data, the necessity of acquiring these should be clearly demonstrated. The means taken and the level of detail of the data to acquire should be proportional to the need expressed in the preceding sub-phase. Another key element at this phase is to have metadata available to ensure the concepts and definitions are aligned with the needs of the statistical program. In the case where data already possessed by the NSI could meet the need, ethical aspects of using data for a different purpose than the initial acquisition should be taken into consideration.

In the Design phase, detailed examination of the metadata, variables and coverage, ideally compared against a known frame, will determine if the alternative source is fit for the statistical program's purpose⁸. At this stage, open communications with the data provider (when applicable) are important to obtain any clarifications needed. Discussions on data transfer protocols can also occur, at the Design Collection sub-phase. The information gathered will also enable design processing and analysis as well as the production workflow. Other NSIs could also be consulted for particularly new alternative data sources or methods to borrow from best practices.

The Build phase will take advantage of the information to build relevant systems for all stages from the actual data acquisition through dissemination. When available a sample of the data set can be used to test and adapt systems used. Alternative data that will be incorporated into existing statistical programs will presumably have lower effort in this phase than alternative data that will help create new or novel data products.

The Collect, Process, Analyse and Disseminate phases remain similar when using alternative data sources or traditional surveys. These phases execute the steps planned in previous phases. The difference will be in the evaluation of each sub-step and the indicators to be used to assess the quality of the alternative data and the process. One possible exception for dissemination would be the choice to first use a new alternative data source in experimental, rather than official statistics. This approach to dissemination is particularly useful when using a new data source to improve an existing estimate as both estimates can be published simultaneously until the NSI is satisfied that suitable standards for accuracy and quality are met.

Finally, the Evaluation phase will be conducted throughout the GSBPM specific elements assessing the performance of the data source relative to the initial need. For example, how well did it meet the expressed need? Were there difficulties in the statistical process directly linked to the use of an alternative data source? Along the way the NSI needs to consider whether the

⁸ See the Questionnaire below for a list of elements to consider when assessing the fitness for use of an alternative data source.

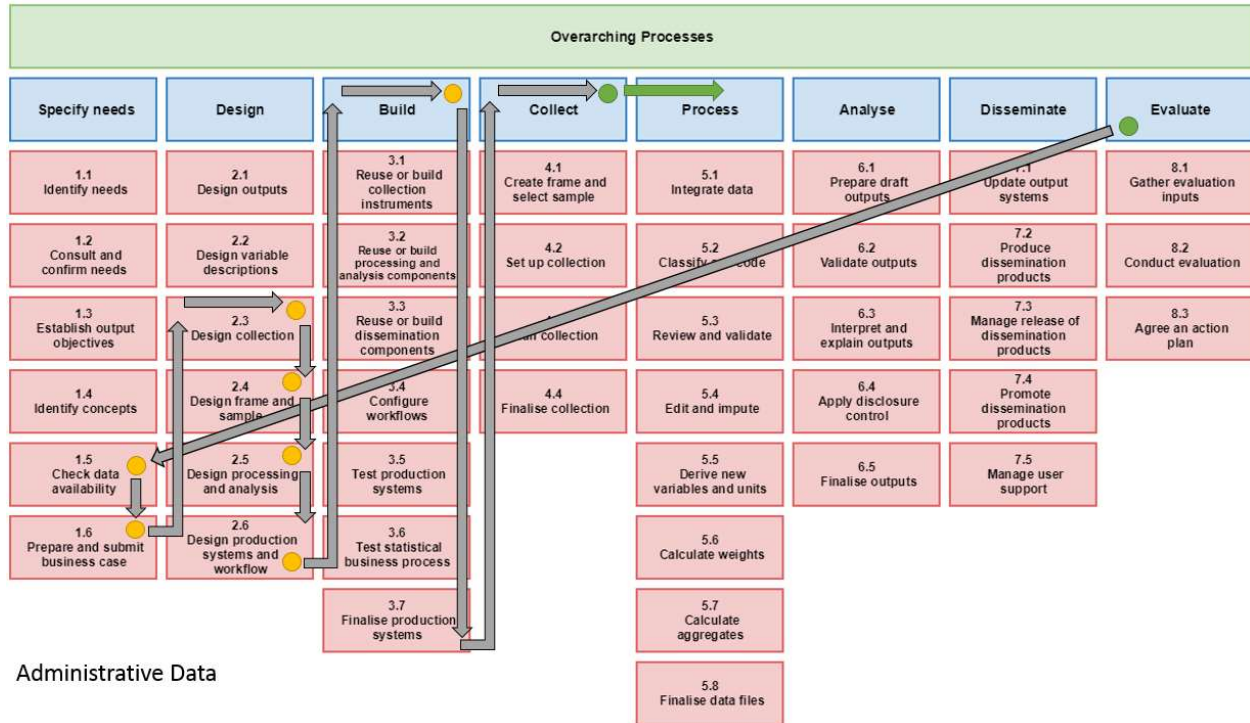
acquisition and use of these data are consistent with the data ethics of the organization. For example, can the NSI consider their estimates to be independent of external influence, either intentional or otherwise. These questions along with continuous communication with the provider will help determine whether the use of the source for this particular need is sustainable through time or not.

When using an alternative source for a statistical process for the first time, it might be wise to perform mini evaluation steps after each phase of the statistical process. This would give the opportunity to readjust some steps and react to any problem arising which could have consequences later on in the process, keeping in mind the objectives of the statistical process.

It will be important to monitor quality indicators as the source sees continued use in official statistics. Similar to using response rates and confidence intervals to indicate the health of a survey data source, NSIs should continue to monitor quality metrics identified in the original incorporation of an alternative data source. This is especially important when data sources are more fluid with regards to variables and population. This task force would recommend setting an acceptable limit which will flag an issue or monitoring processes as well as discussion amongst analysts using the data source.

With the proliferation of alternative sources available, many NSIs already have multiple sources or agreements with providers in their possession. This may lead to situations where the phases of the GSBPM seem to be reordered. For example, in Figure 1, managers of a statistical program may learn of a new source acquired by another program in the same organization and see some potential to use this source in their own program. This would be a situation where the data is already available, and program managers look for a data need to fulfill with this already available source. If a data need can actually be met with this source, then the GSBPM starts again by the Specify Needs phase and the rest of the process can continue. The sub-step Check Data Availability is still required as other sources may answer the need expressed in a better way. However, advantages of reusing a source already available in the organization, such as reduced cost and already existing systems to clean and process the data, should be taken into consideration.

Figure 1. Potential modified sequence of GSBPM



Fitness for Use Questionnaire

The task force recommends that NSIs continue to use existing standards in the evaluation of data quality or fitness for use; however it also found it useful to reframe the discussion around some of those standards when it comes to the use of alternative data sources. A set of questions has been provided in Appendix A based on existing internal Statistics Canada literature and other publicly available standards such as GSBPM. Suggestions for quality metrics to assess and monitor are also included, as suggested in the Stats NZ Guide to Reporting on Administrative Data Quality. These are intended to help the NSI analyst consider where they can assess and control data quality.

Proposals and Discussions

The work presented at the 2022 Voorburg meeting is expected to be part of an ongoing and evolving approach to guidelines on assessing the quality of alternative data sources. This stems from the ever increasing diversity of alternative data themselves as well as the variety of conditions within which each NSI operates (e.g. political, legal, economic/social).

The task force recommends that:

- a. this item be added to the agenda for the 2023 meeting, and
- b. member countries experiment with the proposed questionnaire approach using existing and new alternative data sources in order to provide feedback on its utility

Conclusion

This task force has taken a holistic view from deciding to acquire a data source to monitoring its ongoing use in National Statistics. By mirroring the well established GSBPM for designing a survey, it is possible to build a similar level of trust in alternative data sources to those an NSI directly controls. However, it is important to commit time and resources to documenting and implementing mitigation strategies to build this trust. In other words, this task force urges against taking the path of least resistance.

It is also important to consider the differences in risk when incorporating alternative versus administrative data sources. NSIs could use the tools provided alongside this paper to consider the unique circumstances of a particular data source and identify aspects which should be monitored moving forward, serving as warning signs of quality concerns to be investigated.

Further work is needed to evaluate the suggested tools and framework proposed here. The Voorburg Group could provide a platform to trial this and generate case studies serving as a useful resource to NSIs.

Appendix A - Fitness for Use Questionnaire

This tool is intended to help the NSI assess the quality impacts of implementing an alternative data source, suggesting measures that can be taken to assess and mitigate risks to quality and sustainable statistics. In some cases measurable quality metrics have been guided by Quality indicators for phase 1 errors, as published alongside Stats NZ's Guide to reporting on administrative data quality. These instances will be labeled *SNZ metric X*, where X refers to the metric number in that document.

Questions are categorized along the same lines as the Generic Statistical Business Process Model (GSBPM) and each question shows the predominant quality dimension⁹, if relevant.

R - Relevance	R	A	T	I	C	A
A - Accuracy	R	A	T	I	C	A
T - Timeliness	R	A	T	I	C	A
I - Interpretability	R	A	T	I	C	A
C - Coherence	R	A	T	I	C	A
A - Accessibility	R	A	T	I	C	A

1. Specify Needs

- a. What are the intended goals and future uses of this information? R A T I C A
- i. *A description of user needs and of how users intend to use the data*
 - ii. *Analysis plans that include a description of tables to be released*
 - iii. *A business case establishing the gap between user needs and intended program outcome*
- b. Is there a fee for acquiring this data set? Does the need to acquire the information outweigh the cost? R A T I C A
- i. *A description of user needs and of how users intend to use the data*
 - ii. *Analysis plans that include a description of tables to be released*
 - iii. *A business case establishing the gap between user needs and intended program outcomes*
- c. Given the time and effort required to acquire the data, will the timeliness of the release of information be in line with the needs of the key users and stakeholders? R A T I C A
- i. *Reasonable amount of time between the reference period or date, and the product dissemination date*
- d. Does the data set exhibit the characteristics of an administrative data set? Or an alternative data set?
- i. *Record risks unique to this type of data source (differences between structured/unstructured etc)*

2. Design

- a. Is the geographical coverage of the data adequate for your purposes? R A T I C A
- i. *Report to check adequate coverage of the target population*
 - ii. *SNZ metric 17 (Frame): Undercoverage*
 - iii. *SNZ metric 18 (Frame): Overcoverage*
- b. Does the population covered by the data align with your needs?
- i. *Report to check adequate coverage of the target population*

⁹ [Statistics Canada Quality Guidelines](#), Sixth Edition, December 2019

3. Build

- a. What new components may be required as a result of a new statistical program or any modification to an existing statistical program? (e.g. data acquisition channel, data processing component, machine learning model evaluation, dissemination component). Will these components be designed, built, and tested with the alternative data source?

4. Collect

- a. Does the collection process (performed by the data provider) of the admin data file have any impact on the intended use? If so, are there any means that can be used to mitigate or eliminate this impact?

R A T I C A

- i. Coverage rate of the database used
- ii. SNZ metric 7 (Measurement): Percentage of records from proxies
- iii. Total and partial response rate (see also Stats NZ metric 5: Item non-response & Stats NZ metric 23: Unit non-response)
- iv. Refusal rate
- v. Impact of follow-up strategies
- vi. Impact of collection mode (suggestion: mode effect)
- vii. Capture or coding error rate
- viii. SNZ metric 13 (Processing): Percent of transcription errors

- b. Is the data available at the level of granularity (detail) that is required to fulfill the needs of your statistical program?

R A T I C A

- i. Coverage rate of the databases used
- ii. SNZ metric 25 (Missing/Redundancy): Percentage of units that have to be adjusted to create statistical units

- c. Are accuracy indicators available for the variables that are most important to you? If so, do they fall within a range that is acceptable for your needs? If not, is there any plan to address it?

R A T I C A

- i. Capture or coding error rate
- ii. Detailed description of the instruments and methods for data collection, preparation and analysis. A detailed description of the methodology used will also be available for surveys or data integration
- iii. SNZ metric 13 (Processing): Percentage of transcription errors
- iv. SNZ metric 14 (Processing): Modification rate

- d. Do the variables that are most important to you have enough valid values for your purposes?

R A T I C A

- i. Total and partial response rate (see also Stats NZ metric 5: Item non-response & Stats NZ metric 23: Unit non-response)

- e. Is there sufficient consistency across records in the file to meet your needs?

R A T I C A

- i. Report on births and deaths of objects in the alternative data set
- ii. Report on stability of variables and metadata - refer to data dictionary
- iii. Total and Partial Response Rate - item response, unit response
- iv. Edit failure rate
- v. In some instances, the volatility of the variables in the data source may warrant tracking. Eg. How often do new items appear in scanner data?
- vi. SNZ metric 12 (Measurement): Stability of variables
- vii. SNZ metric 19 (Frame): Authenticity
- viii. SNZ metric 20 (Selection): Adherence to the reporting period
- ix. SNZ metric 21 (Selection): Dynamics of births and deaths
- x. SNZ metric 22 (Selection): Inconsistent objects/units

5. Process

- a. How will missing values in the data be handled? R A T I C A
- i. *Total and Partial Response Rate (see also Stats NZ metric 5: Item non-response & Stats NZ metric 23: Unit non-response)*
 - ii. *Edit failure rate*
 - iii. *Where models have been used, a description of the models' assumptions and an assessment of their likely effects on data quality is available*
- b. Are there any circumstances associated with the chosen reference and collection periods (e.g. recession, pandemic) that might cause issues in the quality or completeness of the data? If yes, how can the issues be addressed? R A T I C A
- i. *Total and Partial Response Rate*
 - ii. *Edit failure rate*
 - iii. *Where models have been used, a description of the models' assumptions and an assessment of their likely effects on data quality is available*
- c. What types of response errors are expected and what is the likelihood of their occurrence (e.g. reporting error, incorrect information)? How will these risks be mitigated? R A T I C A
- i. *Errors attributable to survey eligibility or ineligibility*
 - ii. *A description and justification of the methodology used for each phase of data profiling and preparation is available, with supporting results*
 - iii. *SNZ Metric 3 (Validity): Percentage of inconsistent records*
- d. Is there evidence of bias in the data? (Bias means the systematic deviation from the actual value or the tendency of the measurement process. For example, errors may happen due to the malfunction of instruments that are used for data collection or lack of coverage of some subgroups of the population.) Does the NSI have the ability to maintain the independence of their statistical outputs with respect to the objectives of the data provider or the originally intended use of the data? R A T I C A
- i. *Assessment of data derived from alternative source as compared an existing statistical program for which it will replace or complement*
 - ii. *Ensure that agreement with provider addresses provider responsibilities and potential legal mechanisms*
- e. Are standard concepts and/or classifications being used in the data file? R A T I C A
If not, how will this be addressed?
- i. *A data dictionary and a user guide are available, as needed*
 - ii. *Detailed description of the main statistical concepts, including statistical measures, population, variables, units, domains and reference period*
 - iii. *Accurate references for the concepts, variables and standard classifications used*
 - iv. *Record percentage of items in data set that deviate from target concepts and/or classifications. Note - this may change over time if the data set is unstructured/dynamic*
 - v. *SNZ metric 1 (Validity): Percentage of items that deviate from target concept definition*
 - vi. *SNZ metric 2 (Validity): Percentage of items that deviate from international standards or definitions*

- f. Will the data be linked with other data sources using record linkage techniques? R A T I C A
 - i. Composite quality indicators for combined data sources¹⁰
 - ii. See also Stats NZ's Guide to reporting on administrative data quality, phase 2 on combining datasets to meet statistical purposes
- b. Will established statistical methods be used to create the indirect estimates, direct tabulation or analysis for official release (including quality indicators)? R A T I C A
 - i. *Quality indicators, accuracy measures and/or quality assurance measures are available for the various phases*
- c. Will the product derived from the data be compared with historical data? R A T I C A
If so, are there any limitations that would hinder the comparison?
 - i. *Report outlining any differences from the historical series in concepts/classifications, geographic coverage, population coverage, and survey methodology*
 - ii. *Edit failure rate (historical edits-microdata, historical edits-final estimates)*
- d. Has the mechanism for data transmission been identified, built and tested?
 - i. *Have provider send test file matching format of intended production file and verify transmission as well as file contents*
- e. Have measures been identified for monitoring the quality of data transmitted on an ongoing basis?
 - i. Reporting protocols or systems in place to monitor quality measures identified throughout questionnaire
 - ii. Ongoing monitoring measures could consist of suggested quality metrics from the Stats NZ Guide to reporting on administrative data quality or bespoke measures - for example monitoring of machine learning algorithms as described in the Framework for Responsible Machine Learning Processes at Statistics Canada (Theme: Sound Methods, Guidelines for Quality learning data)¹¹.

6. Analysis

- a. Are there obligations to the data provider or the constituent target population on the dissemination of data derived from the alternative data source? Do specific disclosure control measures need to be put in place? R A T I C A

7. Disseminate

- a. Will the final data products replace existing data products or will they be new to the NSI? R A T I C A
- b. Will the final data products be considered as “official statistics”? Or will they be released as “experimental” statistics? R A T I C A
 - i. Documentation and communication to users about the quality of data being disseminated

¹⁰ [Measuring and Communicating Quality for Programs Using Administrative Data Sources Exclusively](#) (Beaulieu et al., Statistics Canada, 2021)

¹¹ [Framework for Responsible Machine Learning Processes at Statistics Canada](#) (Statistics Canada, July 2020)

8. Evaluate

Each section of this questionnaire provides an opportunity to evaluate the statistical process as well as questions on data ethics. The NSI should review this questionnaire and record their reflections at various intervals (i.e. before acquiring or implementing an alternative data source but also during data development and periodically after implementation) to ensure that expectations are realized and/or re-evaluated.

Bibliography

1. [Transparency in the Reporting of Quality for Integrated Data: A Review of International Standards and Guidelines](#) (Czajka, Stange - Mathematica, 2018)
2. [Guide to reporting on administrative data quality](#) (Stats NZ, 2020)
3. [The use of supermarket scanner data in the Dutch CPI](#) (CBS, 2010)
4. [Use of alternative data sources in Canadian SPPIs](#) (Garneau - Voorburg, 2015)
5. [Voorburg Task Force - Alternative Data Sources](#) (Aizcorbe et al. - Voorburg, 2020)
6. [GSBPM - General Statistical Business Process Manual](#) (version 5.1)
7. [Total Survey Error: Past, Present and Future](#) (Groves, Lyberg, 2010)
8. Placeholder - Published Questionnaire Worksheet (Voorburg Group, 2022)
9. [Statistics Canada Quality Guidelines](#) (Sixth Edition - December 2019)
10. [Measuring and Communicating Quality for Programs Using Administrative Data Sources Exclusively](#) (Beaulieu, Lebrasseur, Gagnon, 2021)
11. [Framework for Responsible Machine Learning Processes at Statistics Canada](#) (Statistics Canada, 2020)

Additional Reading and Relevant Links

- [The Use of Administrative Sources for Economic Statistics](#) - An Overview (Vale, 2006)
- [A framework for the evaluation and use of alternative data in the Consumer Expenditure Surveys](#) (BLS, 2021)