Adapting Quality Assurance Frameworks to the Fast-Evolving World of Official Statistics

Martin Beaulieu[1], Roxanne Gagnon[1]

[1]*Quality and Data Ethics Secretariat, Statistics Canada*

Abstract

The world of official statistics has evolved rapidly in recent years. The need for more detailed and more timely data was already fuelling this evolution, and this need has become even more acute with the COVID-19 pandemic. The proliferation of data sources and the development of new technological tools offer great opportunities for National Statistical Organizations (NSO). These opportunities, however, bring with them new challenges that are only partially, or not at all, covered by National Quality Assurance Frameworks.

Although the latest edition of Statistics Canada's Quality Assurance Framework (QAF) is relatively recent (2017) and includes some elements related to the acquisition and use of administrative data, the QAF is primarily written with probabilistic surveys in mind. For this reason, the QAF is currently under review, to take into account the issues and responsibilities that come with the use of alternative data sources and techniques such as those requiring artificial intelligence (AI) and machine learning (ML). The assessment the quality of input data, the evaluation and communication of the quality of statistics produced, the ethical issues that come with data acquisition and the responsible use of data, and the data stewardship role of NSOs are just some of the requirements highlighted in the current context. NSOs must adapt their practices related to these issues in order to maintain public trust. Several NSOs, including Statistics Canada, have developed tools to define principles and processes to meet these requirements. The integration of these into the QAF, the highest-level governance tool for quality management, will clearly demonstrate to Statistics Canada employees, data providers and users, and the public that the Agency is committed to meeting these requirements.

This talk will discuss how these issues will be addressed in the new version of the QAF, with the aim of not only of bringing it up to date, but also of ensuring that it remains relevant in the long term.

Keywords: Quality frameworks, Quality assessment, Data ethics, Trust, Relevance

1. Introduction

National Statistical Organizations (NSOs) play a key role in the lives of the citizens of any country. The official statistics they produce are essential for governments to make data driven, evidence-based decisions. The proliferation of data sources and the increased use of modern methods, such as machine learning, combined with the ever-increasing demand for more timely and more granular statistics has changed the context in which NSOs operate. At the same time, NSOs face important challenges such as declining response rates and respondent burden concerns, leading many of them, including Statistics Canada, to innovate and use more and more alternative data[1] to produce official statistics. While NSOs have rightly earned the

---

[1] The term "alternative data" is used here as a generic term to include all non-survey data, *i.e.,* administrative data and data from other sources which include, but are not limited to, web-scraped data, satellite data, privately held data (such as scanner data) etc.

reputation of producing high quality statistics by developing and applying frameworks in combination with scientific methods, this new context in the digital world is forcing them to change the way they produce official statistics.

To ensure the quality of the official statistics they produce, most NSOs and international organizations have developed and applied Quality Assurance Frameworks (QAFs). These frameworks describe the measures taken to manage quality and provide guidance to statistical program areas as they develop and implement quality management strategies to meet their users' needs. Many QAFs exist (see for example, United Nations, Statistics Canada) with significant overlap among them and some features customized for their specific situation. Most of these frameworks were developed in a context where most statistical programs were relying on survey data, for which NSOs had full control of each step of the statistical process. The definition of concepts, the design of the questionnaire and sample, the mode of collection, the post-collection processing, the estimation and the dissemination were all under the NSO's responsibility. With many NSOs prioritizing the use of alternative data to adapt to the new context, they are no longer data producers only, but have become data users as well. Data can be ingested or shared with external partners at any point during the data lifecycle. In this context, additional quality considerations emerge from the fact that all aspects of the statistical process are no longer confined within the organization. QAFs should be adapted to formally reflect these considerations, and thus maintain public trust.

This paper will look at existing QAFs through the lens of the new context in which NSOs find themselves and consider necessary modifications or additions. An overview of current QAFs will be introduced in section 2 and common themes among them will be identified. Section 3 will discuss the current context and highlight quality elements which may require modifications or additions to the existing QAFs. Some thoughts on these key elements and their relationship with current QAF structures will be shared in section 4, followed by a conclusion in section 5.

2.   Current Quality Assurance Frameworks

Several QAFs exist and follow a similar structure. In this section we will introduce the QAFs from the United Nations (2019) and Statistics Canada (2017). It is important to note that the choice of these QAFs should not be interpreted as a rejection of others.

2.1 United Nations National Quality Assurance Framework Manual

The United Nations National Quality Assurance Framework Manual (UN NQAF) consists of 19 principles organized into four 'levels'. The first level, called Managing the Statistical System, covers the entire national statistical system, the coordination of which typically lies with the

NSO, and the management of relationships with data users, data providers and other stakeholders.

The second level of the UN NQAF, Managing the Institutional Environment, includes principles relative to assuring professional independence, impartiality, objectivity, transparency, confidentiality and data security, a commitment to quality and adequacy of resources. These principles are there to ensure that NSOs have the trust of their users and data providers, in particular their respondents, as their data is crucial to produce official statistics.

The third level of the UN NQAF is titled Managing the Statistical Process and its principles aim at assuring methodological soundness, cost effectiveness, appropriate statistical procedures and managing respondent burden. Underlying high quality statistical outputs is the rigour that NSOs must use within their statistical processes. The first three principles under this level underline the importance of using statistical methods that are backed up with sound theoretical frameworks touching all phases of the statistical process.

The fourth and final level of the UN NQAF, Managing Statistical Outputs, consists of assuring relevance, accuracy and reliability, timeliness and punctuality, accessibility and clarity, coherence and comparability and proper metadata management. This level is important as it clearly shows that the quality of statistical outputs is not unidimensional and that NSOs must consider it as a whole. Since quality is relative to the users' needs, there will be times when it is appropriate to relax one or more dimensions to give more importance to others.

2.2   Statistics Canada Quality Assurance Framework

The Statistics Canada QAF was last updated in 2017. This edition was inspired by the UN NQAF, with many common principles. It consists of 12 quality management principles which are, in the same spirit of the UN NQAF principles, divided into three overarching themes: The Corporate Environment, the Statistical Program and Statistical Outputs. These overarching themes are similar to the levels of the UN NQAF except for the lack of an overarching theme related to the statistical system. In addition, there is a slight difference in scope of the second Statistics Canada QAF theme which concerns the statistical program versus the statistical processes of the UN NQAF. In the Statistics Canada QAF, the quality management principles under the Corporate Environment concern quality management, sound implementation of statistical methods and assurance of confidentiality, privacy and security.

The second overarching theme is the Statistical Program. It contains principles around management of input data and relations with data providers, allocation and management of resources and management of relations with data users and stakeholders.

The final overarching theme of the Statistics Canada QAF is Statistical Outputs and includes the following principles, which we commonly refer to as Statistics Canada's quality dimensions: Relevance, Accuracy and reliability, Timeliness and punctuality, Accessibility and clarity, Coherence and comparability, Interpretability and management of metadata. These principles match those presented in the fourth level of the UN NQAF with the addition of the 'Interpretability' in the last principle. However, this does not change the spirit of the principle as access to metadata is essential to statistical information being interpretable.

The principles are not always grouped in the same fashion in the UN NQAF manual and Statistics Canada's QAF, but overall, the principles overlap significantly. Although this review of these two existing QAFs is far from exhaustive, they cover the important principles of quality assurance well. Removing the duplicates, we obtain the 19 guiding principles presented in Table 1. These principles are presented using a combination of the four levels of the UN NQAF and the three themes of the Statistics Canada QAF.

Table 1 :19 Quality Assurance Principles

Managing the Statistical System	Institutional/Corporate Environment	Statistical Process / Programs	Statistical Outputs
Coordinating the national statistical system	Professional independence	Methodological soundness	Relevance
Managing relationships with data users, data providers and other stakeholders	Impartiality and objectivity	Cost effectiveness	Accuracy and reliability
Managing statistical standards	Transparency	Appropriate statistical procedures	Timeliness and punctuality
	Statistical confidentiality and data security	Managing response burden	Accessibility and clarity
	Commitment to quality		Coherence and comparability
	Adequacy of resources		Interpretability and managing metadata

While these principles remain valid and relevant, the next section will discuss key modifications or additions to the existing QAFs that may be required in the current context.

3.   Key quality elements in the new context

As discussed in section 1, many NSOs, including Statistics Canada, have modernized their practices and work in an "alternative-first" context to better meet the users' needs for more timely and more granular statistics. Beyond administrative data, there are more data sources

available and new technologies such as machine learning (ML) and artificial intelligence (AI). In this context, NSOs are no longer only data producers, they have also become data users.

The NSOs' new data user role should be reflected in QAFs to ensure quality throughout the statistical process, despite having less control in the data collected or acquired to feed statistical programs. The Food and Agriculture Organization of the United Nations (FAO) for instance, released a new Statistics and Data Quality Assurance Framework in 2023 which includes a principle that reflects the reality of using multiple data sources to produce official statistics. Principle 8, "Suitable and trustworthy data sources", states the data used for statistical purposes may be drawn from all types of sources and that these sources are chosen "according to strictly professional considerations. The availability of official data as well as the accuracy, the relevance, the independence, the stability and the expected availability over time of the input data are among the most important factors considered…" (FAO, 2023).

While assuring the quality and the fitness for purpose of the data used as input in statistical processes is essential, the NSO data user position can also be reflected via three key elements: (1) data stewardship; (2) trustworthiness and social acceptability; and (3) explainability of ML/AI models used to process these sources of data.

3.1 Data Stewardship

As NSOs look to integrate data from different sources, data stewardship principles, and the corresponding role of NSOs, are becoming more prominent. With NSOs ingesting or sharing data at any point in the data lifecycle, the importance of good data stewardship outside the NSO has increased. For example, availability of metadata and the use of standards are key elements required for data sharing and reuse, to ensure data are not misused and facilitate integration with other sources. NSOs can, and should, play a leadership role in promoting the use of standards by all players in the data ecosystem, within the confines of their legal mandate. The need to continue to evaluate other quality aspects such as coverage and the presence or absence of bias would remain.

NSOs can also play a leadership role in helping non-statistical organizations to better understand what data quality means. These organizations may not use their data for statistical purposes, but still need to assess whether data are fit for their purpose. A good understanding of data quality will contribute greatly to efficient data sharing and interoperability. Statistics Canada played a leadership role as co-leaders of a multi-departmental working group, which developed the Government of Canada Data Quality Framework. The framework establishes a government-wide approach to the definition and assessment of data quality and aims to strengthen government-wide capabilities in data quality management and control with a view

to improve the availability, interoperability, usability and public value of data, and facilitate data sharing and reuse. The framework is generic, giving each department the opportunity to adapt or expand the framework based on their own needs.

3.2 Trustworthiness and Social Acceptability

As the number of potential data sources increases, social acceptability is essential for NSOs to maintain trust from their data users, data providers and the general population to benefit from them. In the survey context, respondents (people or businesses) control the information they provide to the NSO by choosing to respond to the questionnaire. Their consent to respond is well informed as they are provided with clear objectives of the survey, who will use their data and why. When the NSO acquires data via a third party, the subject to which the data pertains may not know the NSO possess this data about them, hence a perception of loss of control of their information. In this context, transparency on the type of data the NSO acquires and for what purpose, is key to maintain social acceptability.

To demonstrate trustworthiness and the necessity to acquire the data with means proportional to the information need, NSOs need clear data ethics principles. It helps NSOs make sure the data has been collected, used and shared responsibly. At Statistics Canada, data ethics are founded on six guiding principles: benefits for Canadians, fairness and do no harm, privacy and security, transparency and accountability, trust and sustainability and data quality.

To be meaningful, they need to be implemented in day-to-day activities, in a consistent fashion. Data ethics principles embedded in a QAF, for instance under the Corporate Environment theme, would be a way to ensure they are applied throughout the agency's statistical programs.

3.3 Model Explainability

Once the data are ingested, modern methods such as ML algorithms are often used for various purposes, such as classification, outlier detection or imputation. Modern algorithms can build very accurate models but other aspects of quality such as explainability, reliability and inferential quality also need to be considered. NSOs are facing the challenge of finding the right balance between accuracy and these other dimensions of quality.

The right balance may depend on the purpose for which the model is built and the need it is trying to meet. For instance, when the model is used to understand the relationship between different factors or variables, simpler but well-understood models may be required. Users may accept less precision for a better understanding of a phenomenon. On the other hand, when

predictions are needed, high precision may be required. There might be cases where this high accuracy objective can only be met by a much more complex, more difficult to explain, model.

4.   Adapting Statistics Canada's Quality Assurance Framework

Statistics Canada has initiated a review of its QAF. The 2017 version provides a solid foundation for quality management with principles grouped under three overarching themes: the corporate environment, the statistical program and statistical outputs. These principles remain valid. The purpose of the review is to expand on these principles to provide guidance on the challenges relative to data quality under the current data ecosystem. It is also an opportunity to distinguish terms that are often used interchangeably: fitness for use[2] (data producer perspective) and fitness for purpose[3] (data user perspective).

The three key elements discussed in section 3, can be linked to principles under the first three QAFs overarching themes (see table 1) or linked to one or more quality dimensions in the statistical outputs theme.

For example, data stewardship can obviously be linked with the principles Managing relationships with data users, data providers and other stakeholders as well as Managing statistical standards. The new part would be looking at data stewardship through the data user's lens. In terms of quality dimensions, data stewardship has links with Coherence and comparability as well as Interpretability and managing metadata.

The data ethics component can be linked to Transparency and Commitment to quality. The link may not be as obvious to quality dimensions, but it can be linked to Accuracy and reliability to ensure the objectives justifying the data acquisition will be achievable. Also, with the necessity and proportionality[4] component, the relevance of each acquisition is challenged.

Finally, model explainability has obvious links with the Methodological soundness principle. The sought-after balance between accurate models and simple or understandable models can be translated as a balance between accuracy and interpretability. One could also argue explainability could be a dimension on its own.

Table 2 summarizes the links between the key elements in the new statistical ecosystem and the structure of current QAFs.

---

[2] In this context, "fitness for use" means data producers ensure the data meets the six quality dimensions and provides users with sufficient information (metadata) for them to assess if it meets their needs.

[3] In this context, "fitness for purpose" means the users has all the information necessary to determine whether a data source can be used for a specific purpose.

[4] See https://www.statcan.gc.ca/en/trust/address

Table 2: Links between potential Quality Assurance Framework (QAF) additions and typical QAF structure

Key Element	QAF Principles	Quality Dimensions
Data Stewardship	Managing relationships with data users, data providers and other stakeholders Managing statistical standards	Coherence and comparability Interpretability and managing metadata
Social Acceptability/ Data Ethics	Transparency Commitment to quality	Accuracy and reliability Relevance
Model Explainability	Methodological soundness	Accuracy and reliability Interpretability and managing metadata

5. Conclusion

Major changes have occurred in the data and statistical world in the last few years. NSOs have implemented changes in the way they work to adapt to this new reality. However, Quality Assurance Frameworks should be updated to reflect the new context and bring to the forefront elements that have gained increased focus in this new context such as data stewardship, social acceptability and data ethics as well as model explainability.

Statistics Canada's upcoming new QAF should be a foundation to assure quality remains a core value not only for the organization, but also for all partners, including data providers. A well-established culture of quality management, data stewardship and data ethics will help maintain and improve the trust of Canadians towards their NSO and official statistics produced.

Acknowledgement

Disclaimer

This paper represents the views of the authors and may not necessarily represent that of Statistics Canada.

References

FAO. 2023. FAO Statistics and Data Quality Assurance Framework. Rome. https://www.fao.org/3/cc6683en/cc6683en.pdf

Statistics Canada. 2017. Statistics Canada's Quality Assurance Framework. Third Edition. Statistics Canada Catalogue no. 12-586-X. Ottawa, Ontario https://www150.statcan.gc.ca/n1/pub/12-586-x/12-586-x2017001-eng.htm

United Nations. 2019. United Nations National Quality Assurance Frameworks Manual for Official Statistics. New York, USA, https://unstats.un.org/unsd/methodology/dataquality/un-nqaf-manual/