Harvard Data Science Review • Issue 3.4, Fall 2021

The Evolution of Official Statistics in a Changing World

Wesley Yung¹

¹**Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, Canada**

Published on: Oct 28, 2021

DOI: <https://doi.org/10.1162/99608f92.48466abe>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

As part of the 2020 World Statistics Day, a virtual discussion on the future of federal statistics was held to share ideas and points of view. During the discussion a common message was that as National Statistical Offices modernize, they will have to accept changing the way that they do business and will have to embrace the use of alternative data, emerging fields such as data science, and solid data stewardship practices. Holding the discussion in the middle of a pandemic only highlighted the importance that data and the principles of data stewardship play in everyday life to give decision makers the information required to make informed decisions. Several speakers mentioned that National Statistical Offices are at a crossroads, but that with the right choices and embracing change, the future can be very bright. This article presents some ideas on these choices and what changes can be embraced to ensure that this future is indeed bright. Much has changed since World Statistics Day in terms of the pandemic, but the main messages in the discussion still hold. In fact, with all of the data collected during the pandemic, now is the time for National Statistical Offices to show leadership in combining this data with modern methods to produce official statistics that matter and can help avoid similar situations that have arisen during the pandemic.

Keywords: official statistics, fundamental principles of official statistics, modern methods, data science

1. Introduction

In celebration of World Statistics Day on October 20, 2020, a virtual discussion on the future of federal statistics was organized by the American Statistical Association, the Caucus for Women in Statistics, the *Harvard Data Science Review*, the International Statistics Institute, and the International Association for Official Statistics. Many ideas were shared by the panelists and discussants and this article summarizes the views that I expressed, some of which have changed after listening to the other panelists and discussants.

The term official (or federal) statistics is commonly interpreted as those produced by National Statistical Offices (NSOs) or the National Statistical System. Their importance is outlined in the Fundamental Principles of Official Statistics (United Nations, 2014), in particular in the first principle, Relevance, Impartiality and Equal Access (all 10 Fundamental Principles are given in the Appendix for interested readers):

Official statistics provide an indispensable element in the information system of democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation.

Several key points stand out in the above paragraph. First of all, official statistics are an indispensable element of a democratic society. In the words of one of the discussants of the panel discussion on the Future of Official Statistics, “Official statistics are the bedrock of our society” (John Bailer, President, International Statistics

Institute). In the era of evidence-based policy-making, official statistics make up the foundations for policy or decision-making that can impact millions of citizens and/or influence the expenditure of billions of dollars. In order to be of use to policymakers or decision makers, official statistics must obviously be accurate, but they also have to be relevant, timely, and available. The second point that stands out in the above paragraph is that official statistics are not only for policymakers but the public as well. NSOs do not work for the “**government in power**” but they work for the **citizens of the country**” (Sir Ian Diamond, UK National Statistician). Therefore, they must be transparent and free from political interference in order to gain the trust of all data users, be they governments, private sector organizations, or citizens. As mentioned during the discussion, “NSOs must be independent but not isolated” (Sir Ian Diamond, UK National Statistician), underlying the fact that NSOs must produce information to meet the needs of their users, and to do that, they must interact with them in order to know what their information needs are.

Based on this definition, NSOs should be striving to produce *high-quality* and *trustworthy* official statistics for their varied data users. Most NSOs have frameworks that define their dimensions of quality. For instance, Statistics Canada’s quality framework consists of six dimensions: accuracy, relevance, timeliness, accessibility, interpretability, and coherence (Statistics Canada, 2019a).

1. **Accuracy** is important as decision makers need unbiased and precise information in order to make the most appropriate decisions. The level of accuracy will depend on the circumstances as some decisions require a higher level of accuracy than others. Regardless, decision makers should not be expected to make decisions without having an appropriately accurate view of the situation.
2. **Timeliness** is particularly important in times of change. For instance, in the current situation of the COVID-19 pandemic, it is more important to have estimated infection rates in near real time than having it precisely estimated to the n -th degree. If NSOs are not reporting in a timely fashion, then they are only providing a postmortem on events or patterns in society and not allowing decision makers or policymakers to influence society. It is clear that accuracy and timeliness are competing dimensions because 100% accuracy usually takes time, so there is a balancing act that has to be performed, and which way to go would depend on the ultimate use of the data. While not a quality dimension, cost also comes into the picture as high accuracy could be achieved either by extending the timelines or dedicating more resources, both of which usually entail higher costs.
3. **Relevance** can also be closely related to timeliness as evidence of an economic downturn 6 months after the fact is not useful other than for confirming that a downturn occurred. However, by that time there will be no need to confirm it as it will have been evident to all involved. Relevance is not restricted only to reference periods, and NSOs typically consult with key stakeholders and data users to ensure that the statistical information produced is meeting their needs. If it is not, then there is no reason to produce them.
4. **Accessibility** is also paramount as information is only useful if it is in the hands of individuals or organizations that can use it to implement necessary changes or to not implement unnecessary ones. In addition to access to information, NSOs must balance the amount of information available with the

protection of the privacy of individuals or businesses who respond. Common methods of suppressing or perturbing data to protect confidentiality could compromise the utility of the information.

5. **Interpretability** refers to the availability of supplementary information and metadata required to understand and use statistical information appropriately. It is important that NSOs provide the necessary information if their data users are to use the data to their full potential.
6. Finally, **coherence** is important if data are to be compared or combined with other sources. If standard concepts, classifications, and target populations are not used, then combining or integrating data sources could result in very misleading conclusions. It is by considering these six dimensions, combined with the transparency of methods and analyses, that NSOs will produce information that decision makers and citizens can trust and have faith in.

2. The Past and the Present

For the past 50 or so years, NSOs have relied primarily on surveys, censuses, and administrative registers to produce official statistics. Due to the infrastructure required to perform nationwide surveys and censuses, NSOs are one of the few, if not the only, organizations capable of producing statistics on a national scale. This is also true when considering administrative registers where NSOs also have the information technology infrastructure to combine or link different administrative data sources and the legal authority to obtain the required data. As the integration of administrative data gained popularity in the late 1990s and early 2000s, NSOs continued to be at the forefront due to the previously mentioned reasons. Therefore, for the past 50 years or so, NSOs have had essentially a monopoly on producing national statistics as few, if any, private organizations were capable to collect or obtain the necessary data. This monopoly has ensured that data users have been using high-quality statistical information (for the most part) but also that some NSOs may not have evolved as new methods and data sources have become available. Nevertheless, things have changed in the past 20 years that have started to erode this monopoly of NSOs.

The fact that a panel discussion was organized to discuss the future of official statistics indicates that some things have changed. So, what are the things that are changing the way that official statistics are being produced? While NSOs have enjoyed unequal access to data for many years, this has been changing since the early 2000s. Data are now easily available to organizations and even the public through open data providers and providers of data at a cost. In addition, organizations such as Google, Amazon, and cell phone service providers are collecting enormous amounts of data and analyzing them for their own and general research purposes. In fact, the analyses carried out by many companies are at the core of their business model as they try to understand the behaviors of their users or customers to gain an advantage over their competitors. In many countries, the protection of this personal and business information is enshrined in laws under which NSOs operate. However, while similar laws exist that cover private sector organizations, the implementation of and compliance to them are more than likely variable across businesses.

This data availability, combined with the availability of open-source software and free cloud computing services means that everyone with an interest and some training can be an analyst. On top of that, anyone with a digital presence can now share their analyses with the world. NSOs are now competing against these analysts, some of whom may not take the care to produce high-quality information in order to produce very timely results. Unfortunately, the methodologies used in some of these analyses are not freely shared, or if it is found out that the results are incorrect, the analysis is simply removed. A well-known example is the Google Flu Trends indicator, which was hailed as revolutionary until it started to perform poorly, at which time it was quietly removed. Unfortunately for NSOs, they do not have this luxury in cases where an error has been made as the results may have already impacted decisions touching the citizens they serve. On the other hand, fortunately some researchers are combining these new sources of data with rigorous methods to push NSOs to rethink how they do things. For example, the Billion Prices project out of MIT (Cavallo & Rigobon, 2008) uses prices collected from online retailers to conduct research on macro and international economies. Using the prices collected online, the project can calculate an alternative to the Consumer Price Indexes produced by NSOs. Before the availability of online prices, NSOs were the only organizations capable of producing a monthly statistical output to measure inflation. The availability of online pricing data allows the Billion Prices Project to produce daily measures of inflation.

The Billion Prices Project also illustrates another change influencing the production of official statistics. Modern methods, such as web scraping used by the Billion Prices Project, are changing how data are being collected. Instead of sending interviewers to stores to collect prices, they can now be collected online and processed in near real time. While one can argue that online collection does not allow for quality adjustments usually required for fixed basket consumer prices indexes, one could also question whether the concept of a fixed basket is still relevant. Other examples of modern methods include the ability of machine learning algorithms to process images or code products based on text descriptions. Processing of satellite images allows the production of estimates of crop acreage without having to contact a single farmer. The Australian Bureau of Statistics is using convolutional neural nets to analyze photographic images to identify building starts as part of keeping their Address Register up-to-date (Merkas & Goodwin, 2020). Machine learning is also being used to code thousands of products present in scanner data and integrating this information into the Consumer Price Index (CPI) and retail commodity programs. Statistics Canada is using scanner data in their CPI (Statistics Canada, 2021a) and retail commodity programs (Laroche & Tremblay, 2020) to replace some data collection activities, thus reducing collection costs and respondent burden. Most of these modern methods are considered to be part of the data science domain, however I will continue to use the phrase modern methods when describing them.

With advances in information technology, the sharing of digital data has become easier for NSOs. Gone are the days of having to capture paper forms to obtain data in a digital format that could be shared among organizations. With the majority of administrative data now being collected electronically and with secure transfer protocols, the sharing of data among NSOs and data custodians is becoming easier. Also, with the

advancements in record linkage methods, the ingestion of administrative data into the statistical system is easier than it has ever been.

These are but a few examples of changes that are pushing NSOs to modernize and consider new ways to produce official statistics. An exhaustive list would be very difficult to put together as new methods and technologies are being developed each and every day.

3. The Future of Official Statistics

NSOs are feeling the pressure to modernize or face being made irrelevant. So, does that mean that the future of official statistics, and NSOs as well, is bleak? In my opinion, the exact opposite is true and the future of official statistics is very bright. One thing that the COVID-19 pandemic has shown is the importance of having good data and data practices. The first annual report of the Canadian Statistical Advisory Committee (Canadian Statistical Advisory Committee, 2020) notes that major data gaps exist in COVID-19 data in Canada. For instance, “information on COVID-19 confirmed cases and deaths suffered from delays, incomplete and missing data and inconsistent definitions across jurisdictions.” The report goes on to say that “Statistics Canada’s central role as an independent national statistical organization has never been more critical to meeting the need for timely and high-quality statistics in Canada,” thus highlighting the importance of NSOs to continue to play a crucial role in the national statistical system.

However, NSOs cannot afford to sit back and expect to be continuously handed this role. Without adapting to the changing landscape of the digital world, NSOs will quickly lose to competing organizations, or new organizations will take over that role. In my opinion, NSOs need to embrace the opportunities offered by modern methods and alternative data sources in order to remain relevant. If not, then they are destined to fade off into the sunset as many have been saying will happen for several years now.

As NSOs adapt, they cannot forget the principles that have served them well up to now. These include the six elements previously mentioned in this article: accuracy, timeliness, relevance, accessibility, interpretability, and coherence. By adhering to the second fundamental principle of official statistics—Professional Standards, Scientific Principles and Professional Ethics—NSOs can ensure the continued accuracy of their products. As modern methods are integrated into the toolboxes of official statisticians, care must be taken that quality indicators can continue to be produced to inform users of the quality of their outputs. For many NSOs, policies exist that state that any statistical output must be accompanied with some indication of the quality of that output. A commonly used quality indicator used by many NSOs is the variance, or coefficient of variation. However, many machine learning algorithms are designed to minimize the prediction error to ensure accuracy of the predicted value. In general, there is less literature about the variance of machine learning outputs, which is a key indicator for NSOs in communicating quality to their data users. By applying rigorous scientific principles when investigating modern methods, NSOs can ensure that those methods that are adopted will not adversely impact the accuracy, including the variance, of their outputs. A concrete example of an area that

would benefit from this type of research is estimating variance due to imputation. In the survey methodology context, there currently exists a framework to incorporate the variance coming from imputation for missing data into a *total* variance. The estimate of the variance due to imputation is based on the underlying imputation model and is combined with the sampling variance to come up with the total variance (note that there are other frameworks such as multiple imputation, but we will illustrate the concept with this model-based approach). Now if modern methods are used to replace existing imputation methods, how does one incorporate the variance due to imputation? Some modern methods do not have a corresponding model variance as the current imputation methods do, so how does one include the uncertainty due to imputation in the quality indicator? There are many open research questions such as this one and are excellent opportunities for NSOs to collaborate with colleagues in private sector and academia.

One of the major criticisms of official statistics is that they are not available in a timely manner. While many NSOs have made significant strides in terms of timeliness, there is still some room for improvement. As previously noted, timeliness and accuracy are competing dimensions with more emphasis being placed, for the most part, on accuracy. However, the ultimate use of the statistical output should inform where the emphasis should be placed. When the COVID-19 pandemic essentially locked down Canada in March 2020, Statistics Canada quickly put in place a crowdsourcing exercise to obtain information on the impacts of COVID-19 on Canadians (Statistics Canada, 2020a). While nonprobabilistic methods such as crowdsourcing are rarely used by NSOs, Statistics Canada felt that timely information was more important than taking the time to set up a probability-based survey. By being transparent and clearly stating the limitations of the results, Statistics Canada was able to produce useful information but also ensured that the results were not misused. Realizing the importance of representative results, Statistics Canada also put in place a web panel survey (Statistics Canada, 2020b) to produce results that applied to the general population of Canada.

Nonprobability samples could offer NSOs faster and cheaper results but will not allow inference to a population of interest. As well, there is no guarantee that point estimates from these samples are reflective of the true situation. This causes problems for NSOs as many data users concentrate solely on the point estimate, which may be biased. For NSOs, not being able to draw inference to a population of interest is a major challenge to using nonprobability samples, but fortunately, there is a fair bit of research that is being performed on how to combine probability and nonprobability samples to allow inference to a target population (see, for example, Beaumont, 2020, Chen et al., 2020, or Kim et al., in press). Finding ways to combine modern methods with proven techniques is just one way that NSOs can adapt to profit from the best of both worlds. This is an area where the field of statistics can play a large role by having the private sector and academics partnering with NSOs as they integrate modern methods into their everyday practices. This also aligns with the scientific principles point of the second fundamental principles of official statistics.

In order for NSOs to remain **relevant**, they must **embrace alternative data sources** at their disposal. With constantly diminishing response rates and increasing calls to further reduce respondent burden, NSOs need to

think outside of the box in terms of data collection. Alternative data sources such as satellite or photographic images can be exploited by using modern methods to supplement or replace existing survey programs. For example, satellite images combined with crop insurance and agroclimatic data can be used to produce crop yield data (see Statistics Canada, 2018, and Statistics Canada, 2019b, for a more methodological description). By leveraging these alternative data sources, collection costs and respondent burden have been significantly reduced. Other examples of potential alternative data sources include GPS coordinates for tracking freight trucking and scanner data for price indexes and retail sales by commodity. These are just a few examples where NSOs are already harvesting efficiencies by considering alternative data sources. By continuing to be ambitious and thinking outside of the box, NSOs can ensure that they remain relevant.

Another element mentioned earlier is accessibility. Official statistics are useful only if they are in the hands of policymakers or decision makers. Therefore, official statistics must be easily obtained. NSO websites should be such that information on any topic can be easily found, and the data, at an aggregate level, should be easily available. For those less statistically savvy, NSOs should ensure that their statistics are easily understood as well. The story behind the numbers should be told and should be understood by all data users, including the general public, regardless of their statistical knowledge. Great strides have been made in visualizations over the past couple of decades and NSOs should be leveraging these advances to their full potential. Interactive tools should be used when appropriate to allow users to drill down to the level of data they require and then that data should be easily downloaded if desired by the user. By making data easily understood and available, NSOs will be contributing to building statistical capacity in the general population and increasing awareness of the important role that official statistics, and NSOs, have in their lives. Success in raising this awareness could ultimately result in increased participation in surveys or acceptance to use administrative data sources to produce official statistics. Both will only improve the quality of statistical information produced by NSOs. Finally, increasing access to more data and at finer levels of detail will require innovative disclosure control techniques such as controlled perturbation or the generation of synthetic data and is an area of interest for NSOs.

Once data are made accessible, NSOs must continue to provide supplemental information so that data users can use, report, and interpret them correctly. When NSOs control the data generation process, providing this information is straightforward. However, as more alternative data sources are used, NSOs must ensure that data providers freely share this information with them and that the information is accurate. Not doing so could put the quality of their official statistics at risk. This can be achieved by ensuring that alternative data come from reliable sources and that a solid relationship be built between the data provider and the NSO. This relationship will also be beneficial in the situation where changes are brought on by the data provider as it could ensure that the NSO is consulted on the changes and afforded time to evaluate the impact of the changes before they are implemented. This is especially important in situations where the alternative data is used in ongoing statistical programs. In circumstances where the data are obtained from an external data provider, NSOs may do well to

continue to collect some data themselves in order to not put the continued production of statistical information solely in the hands of external data providers.

Turning to coherence, the situation is similar to that of interpretability. By controlling the data generation process, NSOs can use standard definitions and classifications to ensure that data and information can be compared or combined across data sources. With the increased use of alternative data sources, without supplemental information and metadata from data providers, NSOs may not be able to produce coherent outputs. Even worse would be the situation where an NSO is unable to ascertain if multiple data sources are using the same concepts and definitions and decides to integrate the data to produce outputs. The risk of this occurring could increase with the use of commercially available or big data that have been collected for purposes completely different than that of the NSO. NSOs must be vigilant to ensure that any alternative data used is fit-for-use in terms of both understanding the concept measured and the quality of that measurement. These considerations are very similar to those identified in Brackstone (1987, p. 32) who, in the context of administrative data, identified three important factors to consider before using administrative data:

1. The definitions used in the administrative system
2. The intended coverage of the administrative system
3. The quality with which data are reported and processed in the administrative system.

4. Closing Remarks

The future of official statistics can be very bright, but several enablers could facilitate the modernization journey. First of all, as NSOs modernize and use more alternative data sources with modern methods, the informatics technology infrastructure needs to keep pace. It is well known that Big Data, with its four Vs (Volume, Velocity, Veracity, and Variety), requires the ability to store and process large amounts of information. Cloud computing is an option frequently discussed due to the scalability of its storage and computing capacities. However, NSOs must be aware of the cost of such an infrastructure, both the initial and ongoing costs as well as the security challenges of having possibly confidential data in the cloud. If NSOs are to modernize, they may have no other choice but to upgrade their infrastructure, so funding will have to be secured and security measures put in place to allow confidential data on the new infrastructure.

As NSOs move toward using multiple data sources, the importance of data stewardship increases. Good data stewardship will allow NSOs to combine multiple data sources to produce information not available from any of the individual sources and thus benefit from the “power of and.” In order to combine multiple data sets, standards need to be established so that the data are coherent across data sources, quality needs to be maintained to produce valid analyses, and security must be ensured or data custodians will cease to share data. Security in this context does not mean just information technology methods such as controlled access but includes additional steps such as anonymizing data during ingestion and using record linkage methods so that

‘Big Brother’ databases containing large numbers of variables are not required. Proper data stewardship will ensure that these necessities, as well as other important aspects, are put in place.

As with any transformation, human resource capacity will have to be addressed. In the survey-first context, survey methodologists have played a key role in ensuring high-quality statistical outputs by implementing rigorous sample designs, editing rules, imputation practices, and estimation methods. As NSOs modernize, the skill set required will shift toward data scientists, information technology specialists, and modelers. Note that alternative data are not perfect, so processes such as editing, imputation, and outlier detection and treatment will still be required. In addition, sampling techniques are being investigated to reduce the size of Big Data and to ensure high-quality training data for supervised machine learning algorithms. Thus, survey methodologists should not despair as their skills will still be required by modern NSOs. Finally, there is an important role for survey methodologists and statisticians to ensure that modern methods are statistically sound and do not become black boxes. Overall, I believe that data science has a lot of tools to offer to NSOs but that the underlying foundation of the production of high-quality information lies in statistics. Thus, I see that survey methodologists will be returning to their roots and becoming statisticians, as they were when they started their careers. The truly successful ones will be those who do so while learning all of the tools that modern methods have to offer and applying them in a knowledgeable fashion. Individuals with this mix of statistics and data science are in very high demand, so NSOs would do well to leverage their existing pool of survey methodologist/statisticians and to invest in their upskilling.

Clearly, the use of modern methods such as machine learning will play a big role in the modernization of NSOs. Machine learning is already showing that it is a powerful tool in processing data. However, NSOs must be cautious in how this tool is integrated into the production of official statistics. There are still open questions on how to make a fair comparison between traditional and modern machine learning methods as they tend to have different goals (see Efron, 2020). The Machine Learning project of the High Level Group for Modernization of Official Statistics has proposed a Quality Framework for Statistical Algorithms (QF4SA) (United Nations Economic Commission for Europe [UNECE], 2020) to start this discussion. The QF4SA is proposed as a complement to existing quality frameworks and places particular attention on the explainability of machine learning algorithms and how quality indicators can be obtained for outputs from these algorithms. Before NSOs can fully integrate machine learning, these points need to be addressed if they are to continue to have the trust of their data users. Statistics Canada is also considering these questions and has developed a framework for the responsible use of machine learning (Statistics Canada, 2021b).

In closing, the modernization of NSOs offers many opportunities for statisticians and data scientists, and in a perfect world, the two groups will come together to allow NSOs to produce the highest quality statistical outputs using modern tools and data sources. However, as they modernize and strive to arrive at that perfect world, they must not forget the Fundamental Principles of Official Statistics and, in particular, principles 2, 3,

6, and 9. That is, the use of international standards and scientific principles, being fully transparent on data sources and methods and, most importantly, guaranteeing the confidentiality of data.

Acknowledgments

The author would like to thank Wendy Martinez and Xiao-Li Meng, organizers of the World Statistics Day discussion, the panelists and discussants at the World Statistics Day discussion and the anonymous reviewers whose comments greatly improved the paper.

Disclosure Statement

Wesley Yung has no financial or non-financial disclosures to share for this article.

References

- Beaumont, J.-F. (2020). Are probability surveys bound to disappear for the production of official statistics? *Survey Methodology*, 46(1), 1–28.
- Brackstone, G. (1987). Issues in the use of administrative records for statistical purposes. *Survey Methodology*, 13(1), 29–43.
- Cavallo, A., & Rigobon, R. (2008). The Billion Prices Project: An academic initiative to improve macroeconomic measurement. <http://www.thebillionpricesproject.com/>
- Canadian Statistical Advisory Committee. (2020). *Canadian Statistics Advisory Council 2020 annual report—Towards a stronger National Statistical System*. Statistics Canada. <https://www.statcan.gc.ca/eng/about/relevant/CSAC/report/annual2020>
- Chen, Y., Li, P., & Wu, C. (2020) Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, 115(532), 2011–2021, <https://doi.org/10.1080/01621459.2019.1677241>
- Efron, B. (2020). Prediction, estimation, and attribution. *Journal of the American Statistical Association*, 115(530), 636–655, <https://doi.org/10.1080/01621459.2020.1762613>
- Kim, J. K., Park, S., Chen, Y., & Wu, C. (2021). Combining non-probability and probability samples through mass imputation. *Journal of the Royal Statistical Society, Series A*, 184(3), 941–963. <https://doi.org/10.1111/rssa.12696>
- Laroche, R., & Tremblay, P.-O. (2020). Assessing the quality of a coding process generated by a machine learning algorithm. In *JSM Proceedings*, Government Statistics Section (pp. 1112–1120). American Statistical Association.

Merkas, D., & Goodwin, D. (2020). *An ML application to automate an existing manual process through the use of aerial imagery. Numerous areas throughout the ABS will benefit from the development of this ML application.* UNECE.

https://statswiki.unece.org/download/attachments/285216428/ML_WP1_Imagery_Australia.pdf?version=1&modificationDate=1605171622692&api=v2

Statistics Canada. (2018). *Integrated crop yield modelling using remote sensing, agroclimatic data and survey data.* https://www.statcan.gc.ca/eng/statistical-programs/document/5225_D1_T9_V1

Statistics Canada. (2019a). *Quality guidelines* (6th ed.). <https://www150.statcan.gc.ca/n1/pub/12-539-x/12-539-x2019001-eng.htm>

Statistics Canada. (2019b). *An integrated crop yield model using remote sensing, agroclimatic data and crop insurance data.* https://www.statcan.gc.ca/eng/statistical-programs/document/3401_D2_V1

Statistics Canada. (2020a). *Impacts of COVID-19 on Canadians: First results from crowdsourcing.* <https://www150.statcan.gc.ca/n1/daily-quotidien/200423/dq200423a-eng.htm>

Statistics Canada. (2020b). *Canadian Perspectives Survey Series 1: Impacts of COVID-19.* <https://www150.statcan.gc.ca/n1/daily-quotidien/200408/dq200408c-eng.htm>

Statistics Canada. (2021a). *Enhancements and developments in the Consumer Price Index Program.* <https://www150.statcan.gc.ca/n1/pub/62f0014m/62f0014m2021005-eng.htm>

Statistics Canada. (2021b). *Responsible use of machine learning at Statistics Canada.* <https://www.statcan.gc.ca/eng/data-science/network/machine-learning>

United Nations Economic Commission for Europe. (2020). *A quality framework for statistical algorithms.* https://statswiki.unece.org/download/attachments/285216420/QF4SA_2020_Final.pdf?version=1&modificationDate=1607912228387&api=v2

United Nations. (2014). *Fundamental principles of official statistics.* https://unstats.un.org/unsd/dnss/hb/E-fundamental%20principles_A4-WEB.pdf

Appendix

Fundamental Principles of Official Statistics (United Nations, 2014)

Principle 1: Relevance, Impartiality and Equal Access

Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy, and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honor citizens' entitlement to public information.

Principle 2: Professional Standards, Scientific Principles and Professional Ethics

To retain trust in official statistics, the statistical agencies need to decide, according to strictly professional considerations and including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

Principle 3: Accountability and Transparency

To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods, and procedures of the statistics.

Principle 4: Prevention of Misuse

The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

Principle 5: Sources of Official Statistics

Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs, and the burden on respondents.

Principle 6: Confidentiality

Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

Principle 7: Legislation

The laws, regulations, and measures under which the statistical systems operate are to be made public.

Principle 8: National Coordination

Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.

Principle 9: Use of International Standards

The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

Principle 10: International Cooperation

Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

©2021 Wesley Yung. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.