United Nations

ECE/CES/GE.20/2021/7

Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on National Accounts

Twentieth session
Online, 17, 18, 20, 21, 25 and 26 May 2021
Item 7 of the provisional agenda
Post-COVID-19: Looking forward

Global Environment for the Economic Statistics Ecosystem: A vision for a modernized Canadian System of Macroeconomic Accounts*

Prepared by Statistics Canada[1]

Summary

      The current operating model for the compilation of the Canadian System of Macroeconomic Accounts (CSMA) and its feeding ecosystem of economic statistics was designed several years ago when big data, machine learning, artificial intelligence, cloud computing, and open-source software were still futuristic concepts. Data demands for evidence-based decision making have been expanding in detail and complexity ever since. Furthermore, as demonstrated throughout the COVID-19 pandemic, a high degree of responsiveness to emerging and broader data needs is now expected to be the norm for compilers of statistics.

      The Global Environment for the Economic Statistics Ecosystem (GEESE) is a project which will trigger a complete review, standardization, and optimization of all the information and operating models used in the compilation of the CSMA. The data processing model will be redesigned to incorporate the SDMX standard while making use of new technologies, with an emphasis on open-source software. This new infrastructure will follow the principles of reusability, modularity, traceability, adaptability and transparency. The design of GEESE will be conducive to the ingestion of high frequency data, and the integration of advanced tools such as machine learning and artificial intelligence in regular production streams.

      This paper will present the GEESE project; its vision, components, goals, and approach. It will also highlight the opportunities for international collaboration and co-development. Statistics Canada foresees to build a vibrant community with other organizations for the efficient production of even more coherent official macroeconomic statistics, which will contribute to the improvement of systems of official statistics in all countries.

---

* This document was scheduled for publication after the standard publication date owing to circumstances beyond the submitter's control.

[1] Prepared by Kathryn Young and Michel Chevalier with assistance from Étienne Saint-Pierre, Holly Mullin, and Tony Labillois.

Please recycle

I. Introduction

1.	The technology available today to statisticians and economists is far beyond that which was available when the current processing environment for the Canadian System of Macroeconomic Accounts (CSMA) was developed, many years ago. A modernized environment for the compilation of macroeconomics data at Statistics Canada will set the stage not only to leverage the latest tools and expertise available today, but will allow our organization to remain open and flexible enough to take advantage of new technologies as they become available in the future. Furthermore, a modern environment will provide our analysts with the infrastructure necessary to develop advanced tools, techniques, and modelling approaches themselves. The reality of an ever-evolving economy means our agency must also be ready to integrate any and all new data sources, processing methods, and modelling technologies in order to serve the existing and evolving needs of our data users. This paper will present the new Global Environment for the Economic Statistics Ecosystem (GEESE) at Statistics Canada, and describe how this environment will be developed to offer a modular architecture combining flexibility, scalability, and traceability to achieve the many goals of this ambitious project.

2.	Crucially, this project will benefit from the latest tools, technologies, and resources being shared among the international community. Statistics Canada can learn much from the organizations that have been incorporating modern processes already, and is also willing to help in fill in the gaps where more work and research is required. Overall, by embarking on this modernization overhaul, the GEESE project will allow for Statistics Canada to increase its level of collaboration with international partners, and to closely contribute to the shared goal of finding solutions to ongoing macroeconomic measurement issues.

3.	The subsequent sections of this paper will cover the motivation, vision, and overarching goals for the GEESE project, followed by a summary of the main components and deliverables. It finishes with a description of the technological approach we plan to use to build a fully modern environment for the compilation of macroeconomic statistics.

II. Motivation, vision, and goals

4.	Increased agility and efficiency form core motivations for the establishment of this project. An ideal and sustainable data ecosystem requires enough flexibility to incorporate new technologies and data sources without requiring lengthy timelines for implementation. These concepts, balanced with increased transparency and better data and metadata documentation, will result in a system that works best for those working within it as well as for those who use the data produced by it.
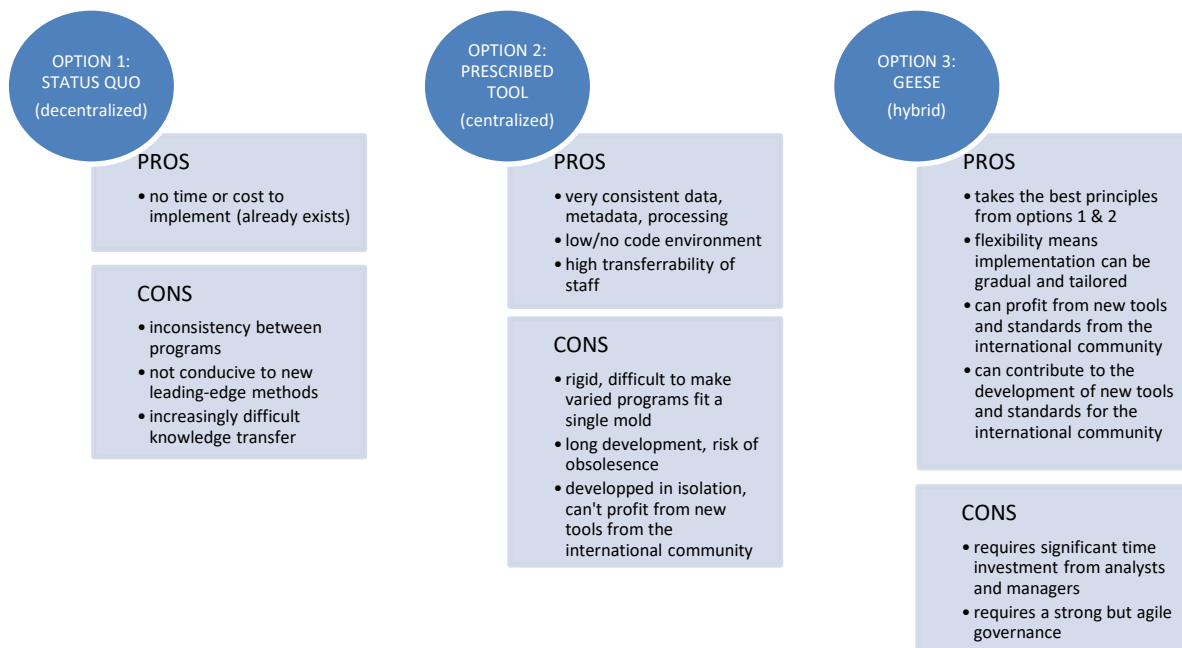
Figure 1
Description of the current CSMA data environment
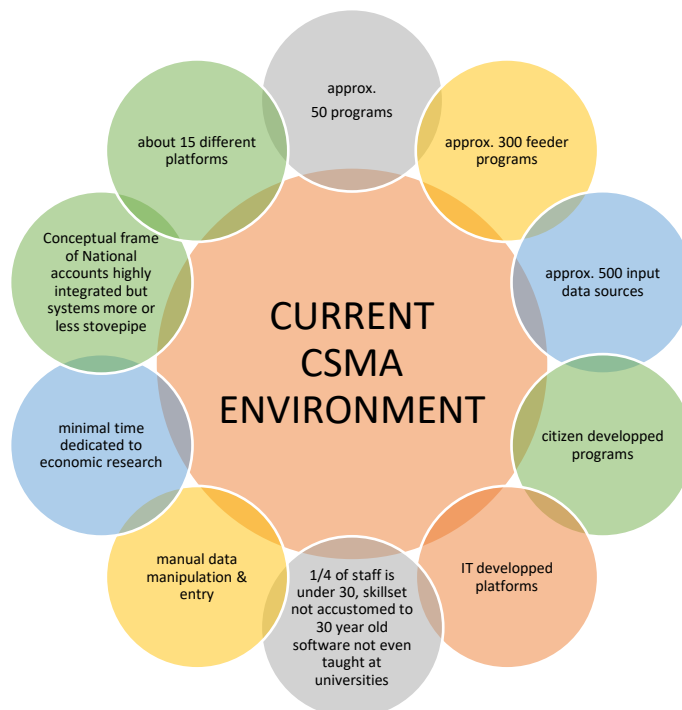
5.   A number of facts outlined in figure 1 above set the scene for a modernization of the CSMA environment. The legacy systems currently in place for the production of the full sequence of accounts in the CSMA have been built over the past number of decades. They use a combination of citizen-developed code and custom-built processing tools comprised of many different technologies and software. All of these systems, whether coded by CSMA analysts or built by IT, were implemented with minimal standardization constraints in order to quickly disseminate existing and new aggregates. While this has allowed our agency to respond to emerging data requirements, we have hit a critical mass of processes that do not speak to one another easily or efficiently. These processes, while currently functional, are not built upon any common coding or information model standard. This means it can be difficult to share processes, data, or metadata between the full suite of economic statistics programs, or track the various adjustments made throughout the production process.

6.   When embarking on the initial planning phase for GEESE, management identified 3 possible options for the way Statistics Canada will produce economic aggregates in the future; 1) status quo system, 2) a newly prescribed in-house tool, or 3) an in-between environment known as GEESE. These options are demonstrated in detail in figure 2 below.

Figure 2
National Accounts production options

OPTION 1:
STATUS QUO
(decentralized)

PROS
- no time or cost to implement (already exists)

CONS
- inconsistency between programs
- not conducive to new leading-edge methods
- increasingly difficult knowledge transfer

OPTION 2:
PRESCRIBED TOOL
(centralized)

PROS
- very consistent data, metadata, processing
- low/no code environment
- high transferrability of staff

CONS
- rigid, difficult to make varied programs fit a single mold
- long development, risk of obsolesence
- developed in isolation, can't profit from new tools from the international community

OPTION 3:
GEESE
(hybrid)

PROS
- takes the best principles from options 1 & 2
- flexibility means implementation can be gradual and tailored
- can profit from new tools and standards from the international community
- can contribute to the development of new tools and standards for the international community

CONS
- requires significant time investment from analysts and managers
- requires a strong but agile governance

7.   A full review of the current production environment led to the conclusion that the status quo (option 1) was not an option – change is necessary for analysts to be able to continue to produce high quality data. The volume of inconsistencies between internal systems is becoming increasingly difficult to maintain, which puts our current and future operations at risk. We believe this to be a common position faced by National Statistical Organizations (NSOs); there is a point where modernization and change must be used to eliminate risks, and we are eager to learn from other organizations how best to proceed through this tipping point.

8.   The decision was also made to not build a proprietary and rigid tool or system (option 2). This option lacks agility and adaptability, likely resulting in a similar position of facing obsolescence a few years down the line, in addition to presenting difficulty in leveraging new tools as they're developed outside of our agency. One of the pillars of this project is flexibility, we want to avoid prescribing a single processing system that will prevent CSMA analysts from using creativity and ingenuity in their methodologies.

9.   The GEESE option (option 3) presents us with an opportunity to identify efficiencies, and to establish a flexible environment grounded in strong metadata-driven information models. This will enable analysts to take advantage of advanced techniques and tools, easily integrating them with existing processes. This environment will also allow CSMA analysts to contribute to the research and development of technologies that Statistics Canada can share with our international partners in return. The GEESE project aims to set the stage for a modular and adaptable ecosystem so our analysts can continue to produce relevant national accounts aggregates with increasing efficiency and flexibility. As the global economy continues to become more complex, national accountants must be equipped to respond to the data needs of policy advisors and other key stakeholders, enabling them to make decisions about economy with the most accurate, timely, coherent, and relevant facts available. These facts mean that analysts must be positioned to make fast and easy adjustments to data and processes in real time, capturing the most up to date information efficiently and effectively. The global strength of international collaboration is certainly the best way to achieve these objectives.

10.   From a human resources perspective, this project will also enable better interchangeability of analysts between the various programs in the CSMA, and will significantly reduce the learning curve that faces all new employees. The GEESE project envisions an environment that will allow our team of analysts to fully take advantage of advanced modelling techniques to help produce high quality macroeconomic aggregates.

Analysts will be able to re-focus their role on the understanding of the contextual emerging trends and factors and removing or significantly reducing some of the frustrating aspects of their current workloads. Overall, we are aiming for a more rewarding work environment that will help us to attract and retain a talented workforce.

11. The new data ecosystem is envisioned to leverage existing open-source innovations by emphasizing software reusability and interoperability. This evergreen global environment will evolve over time with light, efficient, and transparent governance, fostering creativity and innovation while maintaining consistency and cohesion within the economic statistics ecosystem. In addition to the CSMA processes and products, it will encompass the feeding data processes coming from Statistics Canada surveys, administrative data, and high volume/high velocity data flows. This new data environment needs to include structure around a common information model with all components in the CSMA, alignment to solid Information Management principles, and reusable production components which can be assembled and repurposed into custom data and analytics.

12. Current CSMA programs sometimes rely on manual processes for tasks such as coherence analysis and could benefit from increased standardization in order to make use of automation. In terms of planning for the future, the incongruity of the current ecosystem is not conducive to the introduction of AI or machine learning. More importantly, it is difficult to share our tools and processes, or integrate even commonly used tools of today that are developed elsewhere, such as Statistical Data and Metadata eXchange (SDMX) tools or national accounting R packages. The GEESE project aims to address these issues by standardizing the metadata model across the CSMA.

13. Considering the ambitious vision of this project, we are also cognizant that change in and of itself comes with risks. Careful organization and collaboration during the implementation phase must be undertaken by management. Transparency and communication with all stakeholders on an ongoing basis will be the key to the success of this project.

III.   Project components

14. The major phases of the GEESE project are 1) planning, 2) standardization, 3) process redesign, and 4) integration of advanced technologies. Further details are summarized in figure 3 below.

Figure 3
GEESE implementation phases
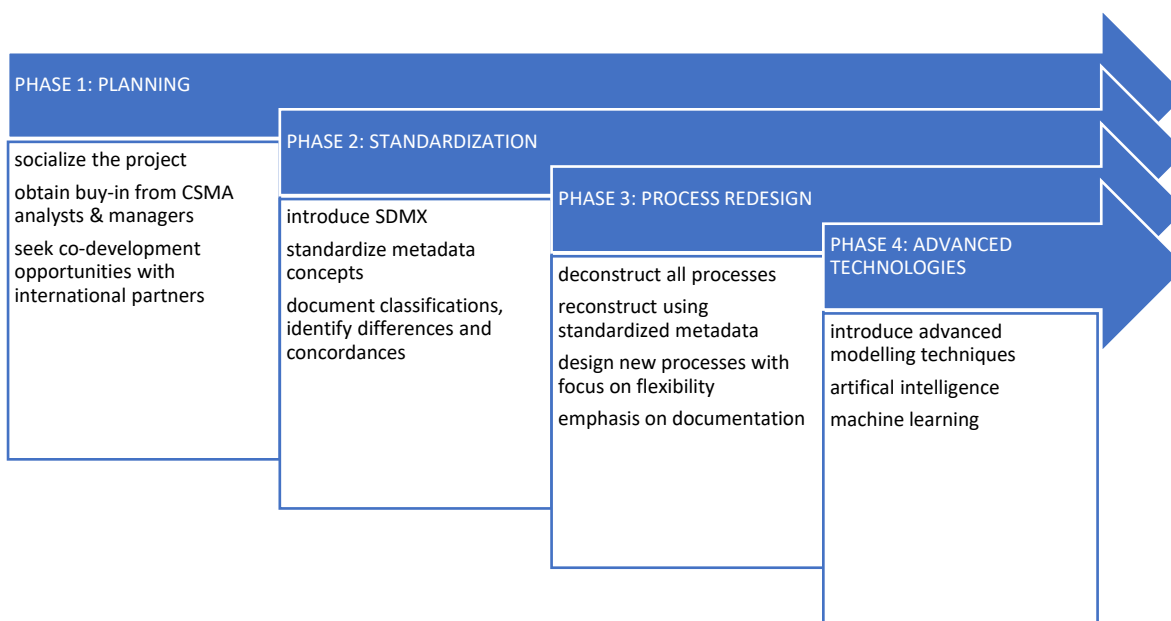
15. To get a little more specific, the short and medium term deliverable components of the GEESE project are: 1) consistent metadata throughout the CSMA and its feeding processes (SDMX compliant wherever possible), 2) a complete A-Z redesign of all CSMA data production processes, and 3) increased automation to reduce production time and allow analysts to shift focus more toward analysis. Once a transparent and harmonized ecosystem has been established, we will ramp up the use of advanced modelling offered by artificial intelligence and machine learning. These types of advanced technologies will open the door to as of yet inaccessible notions such as real-time estimation of a full set of national accounts, automation and integrating big data sets into our regular process flows.

16. The process redesign component of GEESE will begin by reviewing common input datasets (survey data, admin data, etc.) that are used by many programs in the CSMA. The ungoverned citizen-development model used up to now means that many analysts may use their own code to process and aggregate the same input data. We want to establish a 'sandbox' environment where analysts can re-use or adapt code and tools that have already been developed within the agency or by other statistical or international organizations. This will also mean that as new input data are integrated, more programs in the CSMA will be able to leverage new information quickly to inform their estimates. This environment will be complemented with a comprehensive system to track and codify all adjustments made to the data in order to increase the transparency and consistency of the final data aggregates.

17. We have mentioned "modular and flexible processing" a few times in this paper. What this means is that common processes (balancing, seasonal adjustment, benchmarking, imputing, etc.) should be able to be replicated and performed in many areas of the accounts throughout the data production process. Pre-automated modules for common methodologies are envisioned to be used as a default by analysts, with well-documented tools for customization and manual adjustments to be used as complements when necessary.

18. In order to develop easily shareable processing modules throughout the CSMA, there must first be a common metadata structure. The GEESE project will leverage the work done on the international stage which has identified SDMX as the optimal choice of metadata model. We can make use of learning resources to train staff, use modern interoperability approaches such as application programming interface (API) and tools to build coherent databases and track the associated metadata, among other resources. A common starting point will also enable efficient development of advanced models that can use more variables across the CSMA since they will now have more coherent structures and definitions. The current landscape contains a multitude of various data formats, nomenclature, data IDs and systems that make modelling of CSMA data difficult and cumbersome. The status quo also impacts matters such as coherence, agility to adopt evolving international subject matter standards, and the ability to respond quickly to user need by developing tailored products.

19. More consistent metadata also means we can move toward more standard data outputs in terms of format, dimensions, and higher granularity. This means users can more easily compare the outputs from the various accounts within the CSMA. It will also open the door to more data visualization products that can pull from a wider set of macro accounts data at once. Internal to the organization, by adopting SDMX as the data and metadata model, we will be able to streamline the process for international submissions to organizations such as the IMF and the OECD. Throughout this project, we will also review how current output aligns with international standard recommendations. There has been significant work in the past to ensure the conceptual coverage of the CSMA aligns with the latest manuals (System of National Accounts 2008, Balance of Payments and International Investment position Manual, 6th edition, Government Finance Statistics Manual 2014, etc.), however we want to push this alignment even further to cover technical data formats for our submissions to international agencies, and to ensure we are prepared to easily and efficiently incorporate any new recommendations as these manuals undergo periodic updates.

20. As we move through the various phases of the GEESE project, a robust governance structure will establish guard-rails for the citizen-development model. Different components of GEESE will require different levels of governance; for example, the metadata management and standardization governance will require broader oversight than process-flow management which is much more detailed and requires working-level knowledge. We are researching and developing a governance structure that will ideally preserve harmonious

design while fostering an ecosystem with creativity, collaboration, and robustness. Features of our desired governance arrangement include notions of experimental and production sandboxes, certification of components, promotion of mature building blocks, playbooks and business rules, documentation oversight, addressing strategic issues, realigning to evolving Canadian and international contexts, and leveraging opportunities to work with and learn from other organizations.

IV.   Technological approach

21.   In the near term, the technological approach for GEESE begins with using SDMX to document and organize the metadata for all programs in the CSMA. A strong foundation in SDMX principles means we can approach the process re-building stage with a 'technology agnostic' mentality. The idea behind this approach is that, by focusing on coherent metadata-driven systems, we can program the data processing modules using any tool or programming language. We can then easily replicate the same steps in other coding languages so, as an organization, we are not dependent on a single type of software. This leads to an environment that is technologically flexible and can easily integrate newly developed tools so that Canada's system can remain top of the line and won't require top-to-bottom redesigns each time we want to integrate new technologies.

22.   We are also developing small use-cases to demonstrate how current data processing models can be re-programmed using SDMX principles and practices, and also make use of modular processing methodologies. By examining sub-components of different accounts in the CSMA, we aim to demonstrate the viability of this approach while also testing new technologies and tools. A major component of technology testing includes moving process components to a cloud environment.

23.   In the medium term the updated technologies will be managed by the CSMA analysts themselves. These analysts are subject-matter experts in their area of the national accounts. As such, they are the most qualified to evaluate the current systems to determine what works, what doesn't, and what are the most important components for a modernized system. The centralized GEESE team (currently comprised of CSMA analysts and planned for growth in the near term) will provide guidance, planning resources, and education tools to enable staff to rebuild their own systems. The GEESE team will also play a role in documenting the new processes so that lessons learned and best practices can be shared between programs as they integrate to the new environment, building efficiencies as we go.

24.   The building blocks of the GEESE technological approach will be constructed using the principles laid out in figure 4. These principles will be reviewed at each stage of the implementation process by the core GEESE team, IT resources, and the CSMA analysts responsible for the process reconstruction.

25.   The planned approach to technology management will be developed in close collaboration with Statistics Canada IT experts and data architects, and is heavily influenced by the smart data strategies outlined in the paper "Which Strategies for NSOs in the Digital Era? Towards 'Smart Data' Strategies" (Anvar, 2018). Key takeaways include the "need to adapt to the digital era, seizing opportunities and re-define their strategies in a new data ecosystem…agility and responsiveness are called for while remaining committed to independence and quality". In addition to these principles, this paper also makes relevant points regarding organizational structure, highlighting how the ability to modernize requires "tapping into multiple new data sources, adopting new data science techniques, acquiring new platforms, making algorithms available to users – in short, moving from a vertical value chain towards an integrated value chain and step up the readiness to experiment with new sources, methods and partners". The main impediments to this flexibility in our current CSMA ecosystem include data sourcing gaps, data platforms gaps, and data skills gaps, all of which are addressed in the GEESE technological plan.

Figure 4
Pillars of GEESE technological approach
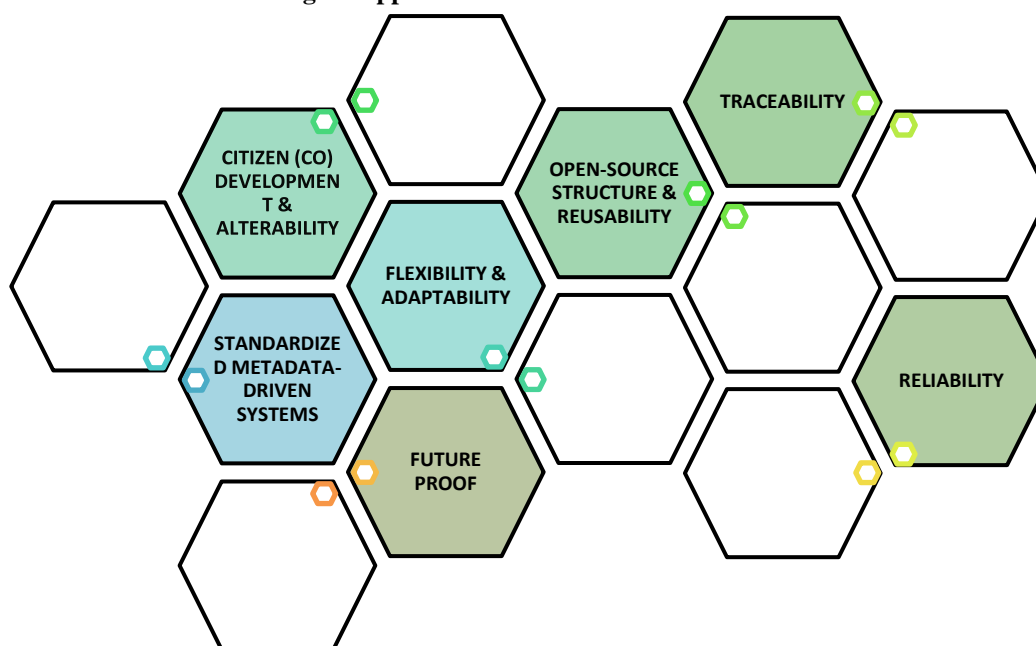
V. Conclusion

26. Statistics Canada is embarking on an ambitious journey to modernize the environment within which we will produce our macroeconomic statistics. In order to maintain our dedication to high quality data outputs, our organization is eager to work with our international partners to ensure we are proceeding with leading-edge knowledge and tools. Our goal for this forum is to learn about what has worked well in data ecosystems for other NSOs, so we don't start our redesign from a blank page. Our contribution in return will be to review current tools and resources, learn where there may be gaps, and target our research and development to filling these gaps.

27. An opportunity is being taken to synergize international efforts in the area of macroeconomic accounts for development of standards-driven open source solutions. Statistics Canada would like to establish a collaborative community and bring together other NSOs with the aim to co-innovate modern macroeconomic measurement approaches. We recognize the international community has achieved great progress in defining models, standards and tools to facilitate the production of official statistics. There has been a strong push for the industrialization of national accounts data production processes, as well as the standardization of the tools used in those processes and this is expected to continue.

28. The result of this cooperation will be an overhaul of the information and processing models used by Statistics Canada. This will further improve the timeliness, coherence, accuracy, transparency, and quality of the CSMA aggregates, while simultaneously ensuring sufficient business agility to continue developing new products and adapt to constantly evolving demands. We are looking forward to embarking on this journey together toward advancing the global effort for the creation of modern macroeconomic measures.

References

Anvar, Eric. (2018), "Which Strategies for NSOs in the Digital Era? Towards 'Smart Data' Strategies", report for 15th meeting of the Committee on Statistics and Statistical Policy (CSSP),http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=SDD/CSSP(2018)7&docLanguage=En

———————