United Nations E/CN.3/2016/5

Economic and Social Council

Distr.: General
16 December 2015

Original: English

Statistical Commission
Forty-seventh session
8-11 March 2016
Item 3 (b) of the provisional agenda*
Items for discussion and decision: transformative agenda for official statistics

Report of the Economic Commission for Europe High-level Group for the Modernization of Official Statistics

Note by the Secretary-General

In accordance with Economic and Social Council decision 2015/216 and past practices, the Statistical Commission has before it a report of the Economic Commission for Europe High-level Group for the Modernization of Official Statistics. The report provides an outline of the outcomes of the activities of the High-level Group and its subsidiary bodies during 2015. The Commission is invited to take note of the report.

* E/CN.3/2016/1.

15-22301 (E) 060116

Report of the Economic Commission for Europe High-level Group for the Modernization of Official Statistics

I. Introduction

1. The Economic Commission for Europe High-level Group for the Modernization of Official Statistics was created in 2010 by the Bureau of the Conference of European Statisticians, the intergovernmental body on statistics of the Economic Commission for Europe (ECE).[1] The High-level Group comprises the heads of 11 national and international statistical organizations,[2] and has a mandate to reflect on and guide strategic developments in the ways in which official statistics are produced.

2. The first main task of the High-level Group was to develop a strategic vision to provide the necessary coordination and direction to the many international initiatives working on topics related to statistical modernization. This vision was endorsed by the Conference of European Statisticians in 2011.[3] It was followed by a strategy to implement that vision, which the Conference endorsed in 2012.[4] Both the vision and strategy were reviewed in 2014 to reflect developments and ensure continued relevance.

3. Following the endorsement of the vision and strategy, the High-level Group has been overseeing a number of implementation activities. These activities are a collaborative effort of interested national and international statistical organizations, with participants from every continent. The activities carried out under the High-level Group are strongly demand-driven, in that they respond to the needs and priorities expressed each year by participating statistical organizations while also contributing to the realization of the vision.

4. The present report provides a summary of the outcomes of the activities of the High-level Group and its subsidiary bodies in support of the vision, focusing on those achieved during 2015.

II. Governance of activities of the High-level Group

5. The High-level Group reports annually to the Conference of European Statisticians. It is supported by an Executive Board, with members typically at the level of deputy director general. Four modernization committees provide expert-level input on topics such as human resource implications, information technology, methodology, data sources, statistical products, standards and quality. A new group provides support for the implementation of the Common Statistical Production

[1] The Conference of European Statisticians is comprised of the Heads of statistical offices of 56 member countries of the Economic Commission for Europe (ECE), member countries of the Organization for Economic Cooperation and Development (OECD) and some other countries outside the region, for example, Brazil, China, Colombia, Mongolia and South Africa.

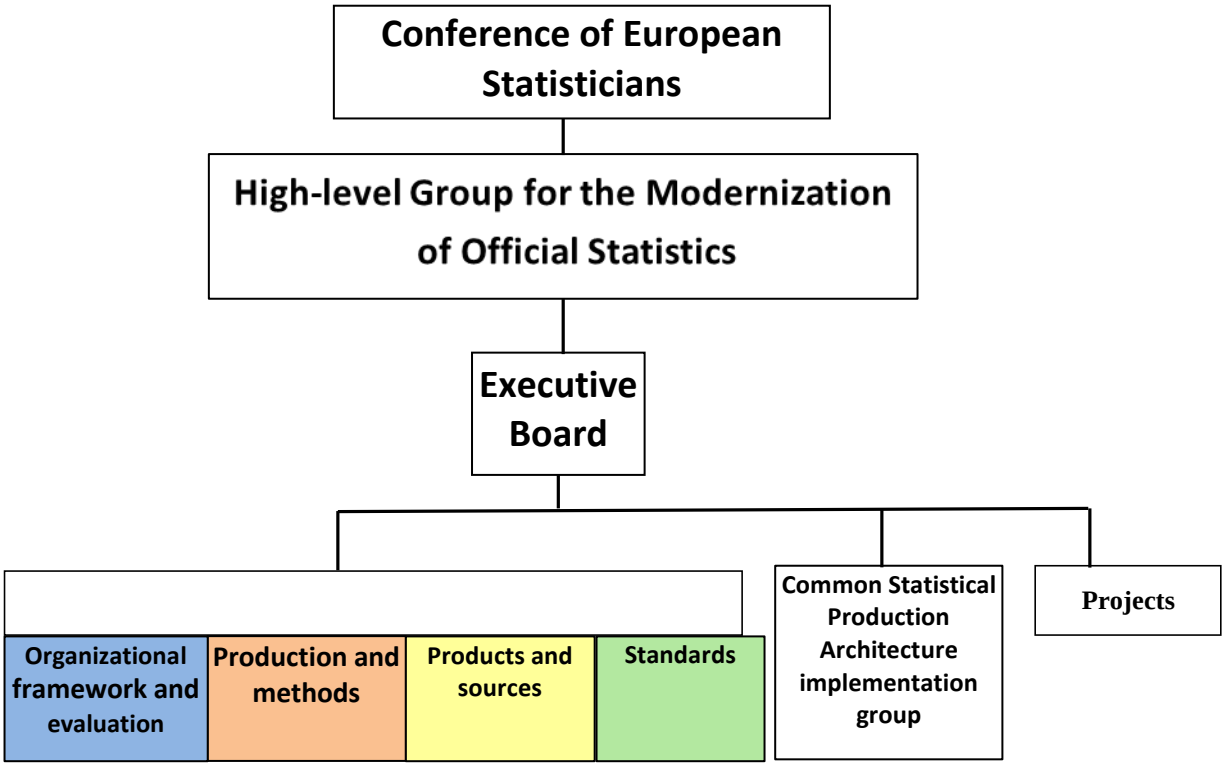
[2] Australia, Canada, Ireland, Italy, Netherlands, New Zealand, Republic of Korea, Slovenia, Eurostat, OECD and ECE.

[3] See www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+HLG.

[4] See www1.unece.org/stat/platform/display/hlgbas/HLG+Strategy.

Architecture. The governance structure for activities of the High-level Group from 1 January 2016 is shown in the figure below.

Governance structure for activities of the High-level Group for the Modernization of Official Statistics



6. International collaboration projects have a strict 12-month time limit, and are chosen in November each year for the following year. This choice is made at a workshop bringing together members of the High-level Group, its subsidiary bodies and representatives of other expert groups or initiatives related to statistical modernization.

7. Participation in the activities of the High-level Group is open to any interested statistical organizations, and, where relevant, selected organizations outside official statistics (e.g., academics and standards bodies). So far, around 40 national and international statistical organizations, from every continent, have participated to some extent in the activities of the High-level Group. All of these activities are voluntary and unpaid, and are sometimes referred to as a "coalition of the willing". This means that they naturally tend to align closely with the priorities of the participating organizations. The activities are supported by donations in cash (to fund central coordination activities) or in kind (typically in terms of staff time, hosting of events, or technical support).

8. In late 2015, the High-level Group launched the "statistical modernization community". This is a new partnership open to all statistical organizations wishing to jointly benefit from collaborative modernization activities. Partners in the community

are required to endorse a statement of intent, setting out the basic principles for collaboration. (see http://www.unece.org/stat/platform/display/smc). The High-level Group has also launched the "Modernstats" brand to give a clear identification to its outputs, with associated accounts on YouTube, Twitter and LinkedIn.

III. Summary of key outcomes before 2015

9. The High-level Group has overseen the development of a number of products, which have been shared with the international statistical community, including the following:

- **Generic Statistical Business Process Model**.[5] The model describes and defines the set of business processes needed to produce official statistics. It provides a standard framework and harmonized terminology to help statistical organizations to modernize their statistical production processes, as well as to share methods and components. The model can also be used as a template for process documentation, for harmonizing statistical computing infrastructures and to provide a framework for process quality assessment and improvement.
- **Generic Statistical Information Model**.[6] This model is a companion to the Business Process Model. While the Business Process Model describes the stages of the statistical production process, the Information Model describes the different objects (e.g., data, metadata, editing rules, and classifications) that flow between those stages.
- **Common Statistical Production Architecture**.[7] The Common Statistical Production Architecture builds on the Business Process Model and the Information Model to create an agreed set of common principles and standards designed to promote greater interoperability within and between statistical organizations. It provides the "industry architecture" for official statistics, and provides a blueprint for designing and developing statistical production components in a way that makes it much easier to share them within and between organizations.
- **Research on the use of big data for official statistics**.[8] The High-level Group has commissioned a number of activities to better understand the importance and impact of "big data" and other new data sources on official statistics. The first output was the paper entitled "What does big data mean for official statistics?", which was presented to the Conference of European Statisticians in 2013. This was followed by a major international collaboration project in 2014, resulting in guidelines on privacy and partnership issues, a big data quality framework, and the documented outcomes of a series of experiments to test big data methods and tools. To support these experiments, the "sandbox", a shared computing environment containing big data sets and software tools was created in partnership with the Irish Central Statistics Office and the Irish Centre for High-End Computing.

[5] See www1.unece.org/stat/platform/display/GSBPM.
[6] See www1.unece.org/stat/platform/display/gsim.
[7] See www1.unece.org/stat/platform/display/CSPA.
[8] See www1.unece.org/stat/platform/display/bigdata.

IV. Outcomes in 2015

10. The two major international collaboration projects during 2015 covered further exploration of the use of big data for official statistics, and the implementation of the Common Statistical Production Architecture. The main outcomes of these projects are described below.

Big data

11. This project consisted of a number of experiments to better understand the usefulness of big data for official statistics. It focused on four data sources, Twitter, Wikipedia, web-scraping of enterprise data and data from the United Nations Comtrade Database. While the last of these is more of a traditional data source, it exhibits many of the characteristics of big data, particularly in regard to volume. The detailed results of the project are available from the big data wiki;[9] however, several key lessons were learned, including that:

- The scope for big data sources to be used on their own to produce new statistics, or replace existing ones, seems rather limited. However, they have considerable potential as complementary sources, used in conjunction with other types of data sources (such as survey data or administrative data).
- The sandbox environment shows proven potential as a shared resource for the international statistical community to facilitate collaborative research and development activities. A new governance model for the sandbox is being put in place to ensure that it remains available in the future. The sandbox is greatly appreciated by the United Nations Global Working Group on Big Data for Official Statistics, which intends to leverage it to support less developed statistical organizations.
- Software tools developed by the information technology industry for processing big data show clear potential for reducing processing times when used with traditional data sets within statistical organizations. Initial results from the sandbox, and work reported by the national statistical organizations of Mexico and Ecuador suggest that in some cases, processes that currently take hours or days can be reduced to minutes. This has potentially significant implications for the processing of data from the 2020 round of population censuses and other traditional statistical activities that involve relatively large data sets.

Implementing the Common Statistical Production Architecture

12. This project set in place a number of mechanisms to support the ongoing implementation of the common architecture in statistical organizations. It also included the development of several services that are in compliance with the common architecture. The detailed results are reported on the ECE Common Statistical Production Architecture wiki.[10] Key outcomes include:

- A new governance model for Common Statistical Production Architecture development and support, including a dedicated committee of experts to

[9] http://www1.unece.org/stat/platform/display/bigdata.
[10] See www1.unece.org/stat/platform/display/CSPA.

oversee the maintenance and enhancement of the common architecture and related materials, and to provide technical support to implementers. A new method and tool to collect information from statistical organizations about their investment intentions over the next few years, to help identify clusters where collaboration between organizations could help to reduce costs and improve efficiency

- A multi-layered catalogue giving information about the investment intentions of statistical organizations, existing capabilities and solutions available to be shared (not limited to information technology solutions) and services that are in compliance with the common architecture, as well as resources to support the development of new capabilities
- A new version of the Common Statistical Production Architecture (version 1.5), and enhanced guidance documentation for implementers

13. In addition to the two major projects described above, the four modernization committees have produced a number of outputs during 2015, including:

- Competency profiles for teams working with big data in statistical organizations, and the leaders of such teams[11]
- Guidelines for managers in statistical organizations, including best practices[12]
- Generic statistical data editing models[13]
- Inventory of big data projects in statistical organizations[14]
- Surveys on communicating the value of official statistics
- Generic Activity Model for Statistical Organizations — which extends the Generic Statistical Business Process Model by adding the non-data activities that take place in a typical statistical organization[15]

14. The following events were also organized during 2015 by the High-level Group or the modernization committees. More detailed information, including papers, presentations and outcomes, is available from the meetings section of the ECE statistics website at www.unece.org/stats/stats_h.html):

- Workshop on big data, Brussels, 9 March
- Workshop on the modernization of statistical production, Geneva, 15 to 17 April
- Workshop on statistical communication, Washington, D.C., 27 to 29 April
- Workshop on data collection, Washington, D.C., 29 April to 1 May
- Workshop on international collaboration for standards-based modernization, Geneva, 5 to 7 May
- Conference of European Statisticians Seminar on the topic “Modernization of statistical production and services, and managing for efficiency”, Geneva, 16 June

[11] See www1.unece.org/stat/platform/display/bigdata/Competency+Profiles.
[12] See www1.unece.org/stat/platform/display/GFM.
[13] See www1.unece.org/stat/platform/display/kbase/GSDEMs.
[14] See www1.unece.org/stat/platform/display/BDI/UNECE+Big+Data+Inventory+Home.
[15] See www1.unece.org/stat/platform/display/GAMSO.

- Work session on statistical data editing, Budapest, 14 to 16 September
- Work session on statistical data confidentiality, Helsinki, 5 to 7 October
- Workshop on the modernization of official statistics, The Hague, 24 and 25 November

V. Work in progress and future plans

15. The two major themes for international collaboration projects chosen for 2016 are:

- Data integration
- Promoting and implementing models and standards

16. Other topics on which work is currently in progress include:

- The development of a set of guidelines on risk and change management in statistical organizations
- Studies on machine learning in official statistics
- Development of a common model of methodology architecture
- The use of mobile devices for data collection and dissemination
- The development of quality indicators for Generic Statistical Business Process Model subprocesses
- The development of a glossary of modernization and metadata terminology
- Workshops on topics that include risk management, human resources management and training, implementation of the Common Statistical Production Architecture, data collection, statistical communication and standards-based modernization

17. The activities listed above are open to any national or international statistical organization that is interested and willing to contribute. Work is done mainly through virtual task teams, using wikis and web conferencing. Additional information is available by contacting support.stat@unece.org.

18. The Commission is invited to take note of the present report.