

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting of the 2015/2016 Bureau  
Luxembourg, 9-10 February 2016

For discussion and  
recommendations

Item III (d) of the Provisional  
Agenda

**OUTCOMES OF THE ACTIVITIES OF THE HIGH-LEVEL GROUP  
FOR THE MODERNIZATION OF OFFICIAL STATISTICS  
IN 2015 AND PLANS FOR 2016**

**Note by the High-Level Group**

*The note provides a summary of the results of the projects overseen by the High-level Group for the Modernization of Official Statistics during 2015. **The Bureau reviewed the work done and provided advice on future work.***

**I. INTRODUCTION**

1. The High-level Group for the Modernization of Official Statistics (HLG-MOS) was created in 2010 by the Bureau of the Conference of European Statisticians. It comprises the heads of eleven national and international statistical organizations, and has a mandate to reflect on and guide strategic developments in the ways in which official statistics are produced.
2. Each year HLG-MOS organizes a workshop, inviting representatives of various groups and projects related to modernization of official statistics. These workshops help to ensure coordination of activities. They also review progress and determine the key priorities for the following year. In 2014, the HLG workshop decided that **further implementation of the Common Statistical Production Architecture and developing the use of Big Data sources for official statistics were the two highest priorities in 2015**. The HLG launched international collaboration projects to address these priorities, which ran during the calendar year 2015. This paper summarises the main results of these projects, as well as other activities overseen by HLG-MOS.
3. At the request of the United Nations Statistical Commission in 2015, the HLG-MOS will also provide regular progress reports to that body. The 2016 report is included as an annex, and will be discussed under agenda item 3(b) at the 47<sup>th</sup> Session of the Commission in March 2016.

**II. COMMON STATISTICAL PRODUCTION ARCHITECTURE (CSPA)**

4. The Common Statistical Production Architecture (CSPA) was created during 2013 in an international collaboration project under HLG-MOS. It provides a standard framework for developing the components of statistical production, in a way that they can be much more easily shared within and between organizations than has previously been the case. It builds on key international standards developed under the HLG-MOS, such as the Generic Statistical Business Process Model (GSBPM), and the Generic Statistical Information Model (GSIM).

5. The 2015 project focused on providing practical support to countries and international organisations implementing CSPA. The final outputs of this project included:

- CSPA v1.5 – An updated version of the CSPA specification, featuring enhanced guidance for implementers
- The CSPA Logical Information Model v0.9 – Providing a bridge between the conceptual nature of the Generic Statistical Information Model (GSIM) and the practical requirements of CSPA service designers, builders and implementers
- CSPA Global Artefact Catalogue (including the new Investment and Capability catalogues) – Tools to find information about what other organizations can offer, or are planning.
- Updated high level information on CSPA – Brochures, presentations, a video, and other training material.

6. These outputs, together with a wide range of supporting materials, can be accessed via the CSPA wiki - <http://www1.unece.org/stat/platform/display/CSPA>.

7. A new “CSPA Implementation Group”, comprising representatives of 16 statistical organisations, was created in January 2016, to oversee and manage CSPA activities. This group is responsible for the maintenance of CSPA and the CSPA Global Artefact Catalogue, assessment of candidate services, provision of implementation advice and further development of the CSPA Logical Information Model (LIM). Organizations wishing to be involved in this work should inform the UNECE Secretariat ([support.stat@unece.org](mailto:support.stat@unece.org)).

### III. BIG DATA

8. This project brought together over 40 experts from national and international statistical organizations around the world to identify and tackle the main challenges of using Big Data sources for official statistics. It ran from January to December 2015 and focused on three main areas:

- **Experiments** - Task teams explored the potential of several Big Data sources to produce official statistics. This activity involved accessing data from the following sources, processing them and assessing their feasibility as sources for official statistics. Detailed results are reported for each source:
  - Wikipedia page views
  - Trade data from the UN ComTrade Database
  - Social media data from Twitter
  - Data scraped from enterprise websites
- **Sandbox** - The project team developed a sustainable model for the continuation of the "Sandbox" environment as a resource for the official statistics community, including identifying new use cases beyond Big Data.
- **Training and outreach** - Various events took place during the project. Workshops were held in Brussels in March 2015 and Ljubljana in September 2015. Presentations were given to various international fora, including the UN Conference on Big Data, the Eurostat Conference on New Techniques and Technologies in Statistics, and various UNECE expert meetings.

9. All outputs from the Big Data project are available at:  
<http://www1.unece.org/stat/platform/display/bigdata/2015+project>

#### IV. OTHER HLG-MOS ACTIVITIES IN 2015

10. In addition to the projects outlined above, the HLG-MOS oversees the work of four modernization committees and a Task Team on Communicating Modernization. The key outputs of these groups in 2015 included:

(a) Modernization Committee on Organizational Frameworks and Evaluation

- An on-line set of Guidelines for Managers in Statistical Organizations:  
<http://www1.unece.org/stat/platform/display/GFM>
- Generic competency profiles for Big Data teams, and Big Data team leaders:  
<http://www1.unece.org/stat/platform/display/bigdata/Competency+Profiles>

(b) Modernization Committee on Production and Methods

- A set of Generic Statistical Data Editing Models (GSDEMs):  
<http://www1.unece.org/stat/platform/display/kbase/GSDEMs>
- A study on the state of the art in the use of machine learning techniques in statistical organisations:  
[http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2015/Topic3\\_Canada\\_paper.pdf](http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2015/Topic3_Canada_paper.pdf)

(c) Modernization Committee on Products and Sources

- Inventory of Big Data projects in statistical organisations (note: at the request of the organisations involved, most projects are only visible to people with UNECE wiki accounts):  
<http://www1.unece.org/stat/platform/display/BDI/UNECE+Big+Data+Inventory+Home>
- Surveys on outreach, promotion, branding and communication of official statistics, to support the work of the Task Force on the Value of Official Statistics

(d) Modernization Committee on Standards

- A new Generic Activity Model for Statistical Organisations (GAMSO) -  
<http://www1.unece.org/stat/platform/display/GAMSO>
- Draft quality indicators for the phases and sub-processes of the Generic Statistical Business Process Model (GSBPM):  
<http://www1.unece.org/stat/platform/download/attachments/112133421/QI.pdf>

(e) Cross Cutting Task Team on Communicating Modernization

- A new brand and logo for HLG-MOS activities:



- A Modernstats YouTube channel, featuring videos about modernization, presentations and webinars:  
[https://www.youtube.com/channel/UC1GxankkLEdqd\\_Hkj73VQYA](https://www.youtube.com/channel/UC1GxankkLEdqd_Hkj73VQYA)
- Communication packs for key modernization standards and activities:  
<http://www1.unece.org/stat/platform/display/hlgbas/HLG+in+brief>

## V. PROJECTS FOR 2016

11. The two major international collaboration projects identified as key priorities for 2016 are:

- **Data integration** – Exploring how to better make use of a growing range of potential data sources, and integrate them with traditional sources to produce official statistics more efficiently -  
<http://www1.unece.org/stat/platform/display/hlgbas/Data+Integration>
- **Implementing Modernstats Standards** – Developing tools to apply the range of standards and models needed for statistical Modernization, and to support statistical organisations at different stages in the Modernization, including a Modernization Maturity Model and a Modernization Roadmap -  
<http://www1.unece.org/stat/platform/display/hlgbas/Implementing+Modernstats+Standards>

## VI. OTHER ONGOING ACTIVITIES

12. In addition to the two projects, activities are underway in the following areas, overseen by the Modernization Committees:

- Identifying best practices in risk and change management in statistical organizations
- Identifying best practices in capability development and staff engagement
- Overcoming organizational barriers to international collaboration
- The use of mobile devices for collecting and disseminating official statistics
- Sharing information on open data strategies
- Developing a metadata / modernization glossary
- Further developing quality indicators for the phases and sub-processes of the GSBPM, with a focus on non-survey data sources
- Review of the Generic Activity Model for Statistical Organizations (GAMSO)
- Support for the implementation of CSPA

13. Any statistical organisation wishing to get involved in any of the above activities is invited to contact the UNECE Secretariat ([support.stat@unece.org](mailto:support.stat@unece.org)).

## VII. THE SANDBOX AND THE STATISTICAL MODERNIZATION COMMUNITY

14. As the work on modernising official statistics matures, more national and international statistical organizations are looking for ways to align with the various HLG-MOS initiatives. Two opportunities for greater involvement at the organisation level are now available:

- **The Sandbox** – This is a shared computing environment hosted by the Irish Centre for High-End Computing. It provides space to share (non-confidential) data, test new software tools, develop methods, and share training materials, in a collaborative environment. It was originally configured to support collaborative experiments on Big Data sources, but is now available for use for any other activities that would benefit from a shared computing environment. It is operated on a non-profit basis, supported by subscriptions of participating organisations, which jointly form the governance body overseeing future developments. For more information, please see the sandbox prospectus: <http://www1.unece.org/stat/platform/download/attachments/112133045/Sandbox%20Prospectus%20Final.pdf>
  - **The Statistical Modernization Community** – The HLG-MOS has recognised the interest of several organizations in becoming part of a community where the rights and obligations of partners are clearly specified, particularly in regards to jointly developed outputs. This community is a collaborative partnership, open to all interested statistical organisations who publicly endorse a “Statement of Intent”. For more information, please see: <http://www1.unece.org/stat/platform/display/smc>
-

## **ANNEX: REPORT OF THE HLG-MOS TO THE 47<sup>th</sup> SESSION OF THE UNITED NATIONS STATISTICAL COMMISSION**

### **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 United Nations Economic Commission for Europe High Level Group on the Modernisation of Official Statistics**

#### **Note by the Secretary-General**

In accordance with Decision 46/101 (paragraph IV (e)) of the 46<sup>th</sup> Session, the Commission has before it a report of the UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS). The report outlines the outcomes of the activities of the HLG-MOS and its subsidiary groups during 2015. The Commission is invited to take note of this report.

#### **I. Introduction**

1. The UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS) was created in 2010 by the Bureau of the Conference of European Statisticians, the inter-governmental body on statistics of the United Nations Economic Commission for Europe (UNECE)<sup>1</sup>. The HLG-MOS comprises the heads of eleven national and international statistical organizations<sup>2</sup>, 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 HLG-MOS was to develop a strategic vision to provide the necessary coordination and direction to the many international initiatives working on topics related to statistical modernisation. This vision was endorsed by the Conference of European Statisticians in 2011<sup>3</sup>. It was followed by a strategy to implement that vision, which the Conference endorsed on 2012<sup>4</sup>. 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 HLG-MOS has been overseeing a number of implementation activities. These activities are a collaborative effort of interested national and international statistical organisations, with participants from every continent. The activities carried out under the HLG-MOS are strongly demand-driven, in that

---

<sup>1</sup> The CES is comprised of the Heads of statistical offices of 56 member countries of the UNECE, OECD member countries and some other countries outside the region for example Brazil, China, Colombia Mongolia and South Africa.

<sup>2</sup> Australia, Canada, Ireland, Italy, Netherlands, New Zealand, Republic of Korea, Slovenia, Eurostat, OECD and UNECE

<sup>3</sup> <http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+HLG>

<sup>4</sup> <http://www1.unece.org/stat/platform/display/hlgbas/HLG+Strategy>

they respond to the needs and priorities expressed each year by participating statistical organisations, whilst also contributing to the realisation of the vision.

4. This paper summarises the outcomes of the activities of the HLG-MOS and its subsidiary groups in support of the vision, focusing on those achieved during 2015.

## II. Governance of HLG-MOS activities

5. The HLG-MOS 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 modernisation 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 Architecture. Figure 1 shows the governance structure for HLG-MOS activities from 1 January 2016.

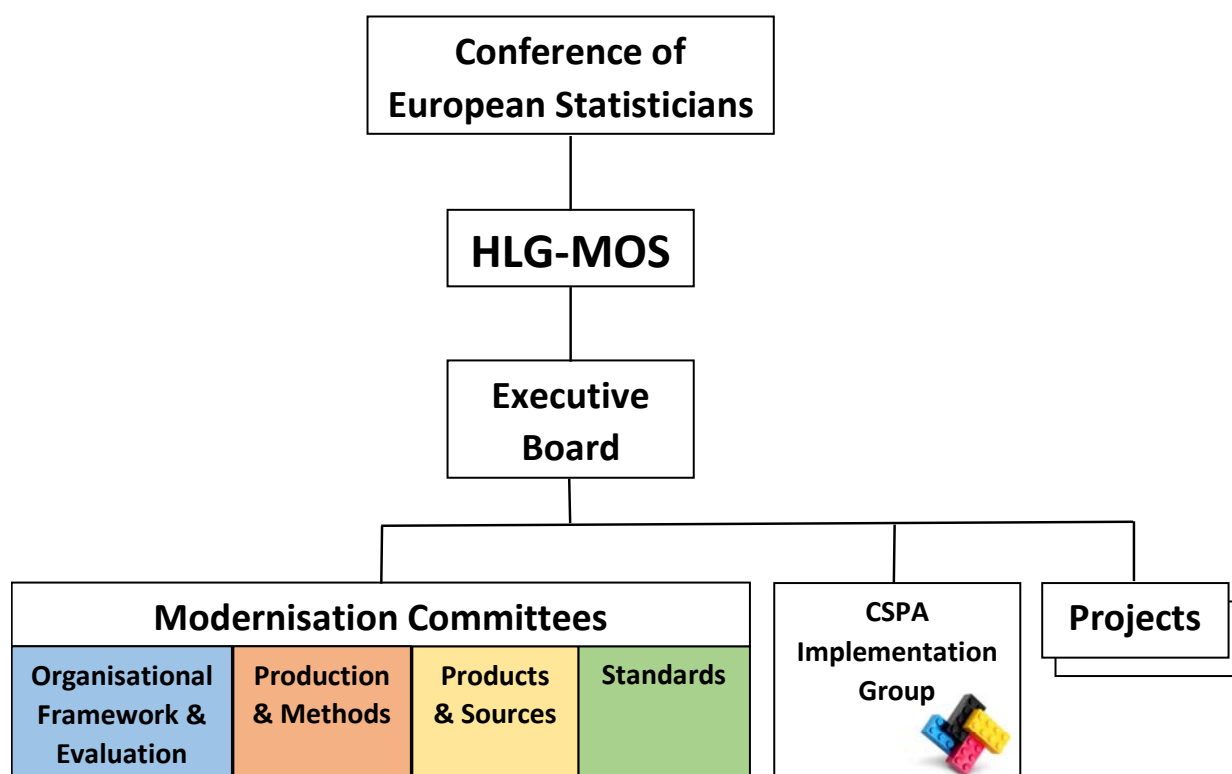


Figure 1: Governance structure for HLG-MOS activities

6. International collaboration projects have a strict twelve-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 HLG-MOS, its subsidiary bodies, and representatives of other expert groups or initiatives related to statistical modernisation.

7. Participation in HLG-MOS activities is open to any interested statistical organisations, and, where relevant, selected organisations outside official statistics (e.g. academics and standards bodies). So far, around 40 national and international statistical organisations, from every continent, have participated to some extent in HLG-MOS activities. All of these

activities are voluntary and un-paid, 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 organisations. HLG-MOS 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 early 2016 the HLG-MOS launched the “Statistical Modernisation Community”. This is a new partnership open to all statistical organisations wishing to jointly benefit from collaborative modernisation activities. Partners in the Statistical Modernisation Community are required to endorse a Statement of Intent, setting out the basic principles for collaboration. For more information, please see the Statistical Modernisation Community wiki.<sup>5</sup> The HLG-MOS has also launched the “Modernstats” brand to give a clear identification to its outputs, with associated accounts on YouTube, Twitter and LinkedIn. Figure 2 shows an example of the branding used.



Figure 2: Example of the branding used on HLG-MOS outputs

### III. Summary of key outcomes before 2015

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

- The Generic Statistical Business Process Model (GSBPM)<sup>6</sup> - The GSBPM describes and defines the set of business processes needed to produce official statistics. It provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components. The GSBPM 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.
- The Generic Statistical Information Model (GSIM)<sup>7</sup> - This model is a companion to the GSBPM. Whilst the GSBPM describes the stages of the statistical production process, the GSIM describes the different objects (e.g. data, metadata, editing rules, and classifications) that flow between those stages.
- The Common Statistical Production Architecture (CSPA)<sup>8</sup> - The CSPA builds on the GSBPM and the GSIM to create an agreed set of common principles and standards designed to

<sup>5</sup> <http://www1.unece.org/stat/platform/display/smc>

<sup>6</sup> <http://www1.unece.org/stat/platform/display/GSBPM>

<sup>7</sup> <http://www1.unece.org/stat/platform/display/gsim>

<sup>8</sup> <http://www1.unece.org/stat/platform/display/CSPA>



promote greater interoperability within and between statistical organizations. It provides the “industry architecture” for official statistics, and provides a blue-print for designing and developing statistical production components in a way that makes it much easier to share them within and between organisations.

- Research on the use of Big Data for official statistics<sup>9</sup> - The HLG-MOS 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 “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.

#### 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 (CSPA). The main outcomes of these projects are described below.

11. **Big Data** – 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. Whilst 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 reported on the UNECE Big Data wiki<sup>10</sup>, 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 it remains available in the future. The sandbox is greatly appreciated by the UN Global Working Group on Big Data, which intends to leverage it to support less developed statistical organisations.
- 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 organisations. Initial results from the sandbox, and work reported by the national statistical organisations 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

<sup>9</sup> <http://www1.unece.org/stat/platform/display/bigdata>

<sup>10</sup> <http://www1.unece.org/stat/platform/display/bigdata>

implications for the processing of data from the 2020 round of population censuses, and other traditional statistical activities that involve relatively large data sets.

12. **Implementing CSPA** – This project set in place a number of mechanisms to support the on-going implementation of CSPA in statistical organisations. It also included the development of several CSPA-compliant services. The detailed results are reported on the UNECE CSPA wiki<sup>11</sup>. Key outcomes include:

- A new governance model for CSPA development and support, including a dedicated committee of experts to oversee the maintenance and enhancement of CSPA and related materials, and to provide technical support to implementers.
- A new method and tool to collect information from statistical organisations about their investment intentions over the next few years, to help identify clusters where collaboration between organisations could help to reduce costs and improve efficiency.
- A multi-layered catalogue giving information about the investment intentions of statistical organisations, existing capabilities and solutions available to be shared (not limited to IT solutions) and CSPA-compliant services, as well as resources to support the development of new capabilities.
- A new version of the CSPA (version 1.5), and enhanced guidance documentation for implementers.

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

- Competency profiles for teams working with Big Data in statistical organisations, and the leaders of such teams<sup>12</sup>
- Guidelines for managers in statistical organisations, including best practices<sup>13</sup>
- Generic Statistical Data Editing Models<sup>14</sup>
- Inventory of Big Data projects in statistical organisations<sup>15</sup>
- Surveys on communicating the value of official statistics
- Generic Activity Model for Statistical Organisations - which extends the GSBPM by adding the non-data activities that take place in a typical statistical organisation.<sup>16</sup>

14. The following events were also organised during 2015 by the HLG-MOS or the modernisation committees. For more detailed information, including papers, presentations and outcomes, please see the meetings section of the UNECE statistics web site<sup>17</sup>:

<sup>11</sup> <http://www1.unece.org/stat/platform/display/CSPA>

<sup>12</sup> <http://www1.unece.org/stat/platform/display/bigdata/Competency+Profiles>

<sup>13</sup> <http://www1.unece.org/stat/platform/display/GFM>

<sup>14</sup> <http://www1.unece.org/stat/platform/display/kbase/GSDEMs>

<sup>15</sup> <http://www1.unece.org/stat/platform/display/BDI/UNECE+Big+Data+Inventory+Home>

<sup>16</sup> <http://www1.unece.org/stat/platform/display/GAMSO>

<sup>17</sup> [http://www.unece.org/stats/stats\\_h.html](http://www.unece.org/stats/stats_h.html)

- Workshop on Big Data, Brussels, 9 March
- Workshop on the Modernisation of Statistical Production, Geneva, 15-17 April
- Workshop on Statistical Communication, Washington DC, 27-29 April
- Workshop on Data Collection, Washington DC, 29 April - 1 May
- Workshop on International Collaboration for Standards-based Modernisation, Geneva, 5-7 May
- Conference of European Statisticians Seminar "Modernisation of statistical production and services, and managing for efficiency", Geneva, 16 June (see "Second Seminar" tab)
- Work Session on Statistical Data Editing, Budapest, 14-16 September
- Work Session on Statistical Data Confidentiality, Helsinki, 5-7 October
- Workshop on the Modernisation of Official Statistics, The Hague, 24-25 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 organisations
  - 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 GSBPM sub-processes
  - The development of a glossary of modernisation and metadata terminology
  - Workshops on topics including risk management, human resources management and training, CSPA implementation, data collection, statistical communication, and standards-based modernisation.
17. All of the above activities are open to any national or international statistical organisation that is interested and willing to contribute. Work is done mainly through virtual task teams, using wikis and web conferencing. For more information, please contact [support.stat@unece.org](mailto:support.stat@unece.org).
18. The Commission is invited to take note of the report.