High-Level Group for the Modernisation of Official Statistics

The HLG-MOS oversees and manages:

🌟 **Collaboration groups:** Supporting Standards, Capabilities and Communication, Applying Data Science and Modern Methods, Blue Skies and Machine Learning Community

🌟 **Modernisation Projects 2022:** Input Privacy-Preserving Techniques (IPPT) phase 2, Data Governance for Interoperability Framework Project, and Meta-Academy for the Modernization of Official Statistics

🌟 **Specialised topics:** Dissemination and Communication, Data Collection, Statistical Data Editing and Statistical Confidentiality

🌟 **Modernisation Expert group meetings, workshops and seminars**

---

**Generic Statistical Business Process Model (GSBPM)** describes the core business processes undertaken by statistical organisations to produce statistical outputs. It is used by more than 50 organisations worldwide. [Read more here](#)

<table>
<thead>
<tr>
<th>Process</th>
<th>Category</th>
<th>Standard</th>
<th>Type</th>
<th>Data</th>
<th>Activity</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey</td>
<td>Data Collection</td>
<td>ISO 2022</td>
<td>Form</td>
<td>Method</td>
<td>Data</td>
<td>Result</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Generic Statistical Information Model (GSIM)** describes the core pieces of information needed by statistical organisations to produce statistical outputs. [Read more here](#)

---

**Common Statistical Data Architecture (CSDA)** is reference architecture and guidance for the modernisation of their processes and systems. [Read more here](#)

**Strategic Communication Framework** is a guide for Statistical Organisations to the development and implementation of a communication strategy. [Read more here](#)

**Generic Activity Model for Statistical Organisations (GAMSO)** extends and complements the GSBPM by describing overarching activities and processes to support the production of official statistical production. [Read more here](#)
Common Statistical Production Architecture (CSPA) helps statistical organisation create interoperable tools to share within and between statistical organisations. Read more here

Generic Statistical Data Editing Model (GSDEM) is intended as a reference for all official statisticians whose activities include data editing. Read more here

Machine Learning for Official Statistics the 2019-2020 project continued as a collaboration community in 2021. Read more here

Blog on Recent highlights and updates

- **Blog:** New Group on Applying Data Science and Modern Methods created by Wai Kit Si Tou
  04 May, 2022
  High-Level Group for the Modernisation of Official Statistics

- **Blog:** Machine learning for official statistics created by InKyung Choi
  05 Apr, 2022
  High-Level Group for the Modernisation of Official Statistics

- **Blog:** Use of ModernStats Models: Progress and Way Forward created by Wai Kit Si Tou
  20 Oct, 2021
  High-Level Group for the Modernisation of Official Statistics

- **Blog:** New framework sets out a strategic approach to statistical communication created by Taeke Gjaltema
  22 Jul, 2021
  High-Level Group for the Modernisation of Official Statistics

- **Blog:** Upcoming Expert meetings created by Taeke Gjaltema
  08 Jun, 2021
  High-Level Group for the Modernisation of Official Statistics

- **Blog:** Despite the Covid19 pandemic, the involvement in the ModernStats Community increased from 2019 to 2020 update: 2021 involvement created by Taeke Gjaltema
  27 Jan, 2021
  High-Level Group for the Modernisation of Official Statistics
We are glad to announce the establishment of the new group on Applying Data Science and Modern Methods in 2022.

In view of the increasing importance of new data sources and methods for the compilation of official statistics, this new group aims to identify data science initiatives and new methods needed for modernising existing business processes.

Examples of potential activities cover: (i) developing supporting materials to help implement case studies and good practices, (ii) organising workshops and trainings to promote and ensure consistent use of the HLG-MOS supported data science initiatives and modern methods, and (iii) managing the periodic reviews of the new data science initiatives and methods to measure their impacts.

The group has already started conducting a market landscape analysis to take stock of existing work in the field of data science and modern methods. It will help identify specific needs, challenges, and opportunities that support modernising statistical production and services.

Some potential areas of work that the group is considering include (i) use cases on data collection methods, (ii) data platform that facilitates the access to and integration of data from different sources, (iii) responsible AI, (iv) modelling methods for different type of data, and (v) automated dissemination, with user-centric services.

We look forward to welcoming you to participate in the group and other activities.

Please contact Wai Kit Si Tou if you are interested in joining the group.

Machine learning for official statistics

InKyung Choi posted on Apr 05, 2022

Machine learning for official statistics: UNECE help statistical organisations harness the power of machine learning

The new UNECE publication Machine Learning for Official Statistics helps national and international statistical organizations harness the power of machine learning to modernize the production of official statistics.

Machine learning in the context of modernisation for statistical organizations

Statistical organizations produce crucial indicators that portray various aspects of the economy and society that we live in. These include measures such as the gross domestic product, the inflation rate, the population growth, and the unemployment rate on which government and business alike depend when making important decisions. While these may be just a few digit numbers from the end-user side, statistical organizations employ a series of carefully designed and executed processes to distil this key information from the vast amount of raw data.
With increasing challenges arising from new data sources, technological developments and competitions with private companies, statistical organizations have been striving to modernise every part of this production process to provide more relevant and detailed official statistics in a more timely and accessible manner. They utilise the computer-assisted interview and web scraping tool to collect data more efficiently, and build infrastructure for data and IT tools to manage them across the organizations more easily.

Yet, one area that is difficult to modernise is the processes that require “human-like” decision-making, such as reading a textual description to assign a matching classification code or looking at the image to identify what it represents. Traditionally, this has been done either manually or through a complex rule-based system, both of which are costly, time-consuming and hard to manage. This is particularly daunting when statistical organizations try to use big data sources (e.g., price information web-scrapped from online stores) as the cost of resources needed to process such a large amount of data in the traditional manual way is simply too prohibitive.

Machines learning holds great potential for statistical organizations

The recent developments in machine learning technique are pushing the boundary of tasks considered for humans and machines - machines can now draw a painting in the style of an old master and write an article just like humans.

How does this technology work? In one of the most popular approaches called “supervised learning”, machines are first trained on the data that humans labelled, for example, images labelled as “urban” or “rural”. With this data, they figure out patterns associated with labels by adaptively improving their internal logic that maps from the input (image) to output (label). In this way, machines can determine whether an area shown in an image is urban or rural without us providing all possible rules explicitly.

As the machine learning technique can carry out tasks that we used to solely rely on manual works, it holds a great potential to increase the efficiency of statistical organizations, just like the use of machinery powered by steam engines made a huge leap in the productivity in the manufacturer industry few centuries ago. Also, their capability to process various types of data such as text, image and video offers statistical organizations to take advantage of new data sources to produce new statistics that could meet the evolving needs of society.

Challenges in using machine learning for official statistics

Like with any innovation, however, the journey of integrating machine learning in the organization abounds with challenges and setbacks. The technology itself is still relatively new and requires a different skill set that many statistical organizations do not possess; hence it needs to be built inside or acquired from outside.

The real difficulty, however, starts when the machine learning solution needs to move to production, meaning that it is connected to existing processes seamlessly and used for the regular business, beyond an “experiment” stage. Unfortunately, even after successful pilot studies, many machine learning solutions end up being left on the shelf. The difficulty is experienced widely across sectors and domains. It is said that over 80% of machine learning projects never make it to production. Moving machine learning into production requires changes in infrastructure, culture, organizational structure or business processes, none of which is a small task with a lasting effect.

UNECE supports statistical organizations in advancing the use of machine learning
Based on the two international initiatives, the UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS) Machine Learning Project (2019-20) and the United Kingdom Office of National Statistics (ONS) – UNECE Machine Learning Group 2021, the publication Machine Learning for Official Statistics aims to help statistical organizations navigate the difficult journey of advancing the use of this new technology. It presents the practical applications of machine learning in three working areas within statistical organizations and discusses their value-added, challenges and lessons learned. The publication also includes a quality framework that could help guide the choice of methods, demonstrates key steps for moving machine learning from the experimental stage to the production stage, and key messages to facilitate the use of machine learning in the statistical organizations.

The machine learning field is fast evolving with new methods, platforms and approaches coming out every month. To keep up with the pace of change and avoid duplication of efforts, there is a great need for knowledge sharing and collaboration within the official statistics community. UNECE continues its engagement in the international initiative this year, through Machine Learning Group 2022 with the ONS, to support statistical organizations to harness the power of machine learning.
According to the 2021 survey on the use of “ModernStats” models[1], the Generic Statistical Business Process Model (GSBPM) is widely implemented, while the use of Generic Activity Model for Statistical Organisations (GAMSO), Generic Statistical Information Model (GSIM), and Common Statistical Production Architecture (CSPA) is more limited.

To highlight, out of the 45 respondents, two-thirds answered that there is widespread awareness and high level of familiarity with the application of GSBPM. For the organisations that took part in both surveys in 2018[2] and 2021, half of them reported increased use of GSBPM to modernise their statistical production process. Many of the organisations reported to have further developed a national version of GSBPM, with modification of phases and sub-processes in GSBPM and addition of a finer level of details based on the local context.[3] The top three benefits of using GSBPM identified by the respondents are (i) ease of internal communication, (ii) facilitation of quality management, and (iii) ease of comparing processes and identifying inefficiency within the organisation.

Regarding the limited use of GAMSO, GSIM, and CSPA, the major reasons behind are the lack of awareness and related knowledge, while some reported that they have their own model and do not feel the need to adapt to these models. In addition, organisations generally lack human resources and expertise to implement the model. It hinders further application of the model, despite the widely recognized benefits of facilitating the share of methods, service or capability within the organisation as well as easing internal communication.

The ModernStats models discussed above provide a common language and tool to map all activities within and between statistical organisations to a common approach. They enable statistical organisations to collaborate and facilitate exchange of information and the sharing of statistical services, thereby contributing to the advancement of official statistics at the national and international levels for evidence-based policymaking and assessing progress towards achieving the Sustainable Development Goals.

To promote the use of modernisation statistical models, a proactive communication strategy and capacity building programme are essential. UNECE will also continue to work closely with statistical organisations to update the models as well as improve linkage and consistency among the models to support wider integration and implementation.

[1] The online survey was conducted in February and March 2021.


[3] Under the High-level Group for the Modernisation of Official Statistics (HLGMOS), the Supporting Standards Group provides support for the implementation of the “ModernStats models”. A new task team was created in 2021 to compile examples of finer level of GSBPM activities, identify commonalities among them, and prepare a supplementary document to support countries using GSBPM in more detailed level as well as provide input for the next GSBPM revision.
New framework sets out a strategic approach to statistical communication
Taeke Gjaltema posted on Jul 22, 2021
Statistics are all around us. All the more since the COVID-19 pandemic began. Rarely do we see a news story, political debate, press conference or even a social media debate that doesn’t reference statistics. In this environment, the custodians of the figures—the national statistical offices (NSOs) which produce, curate and publish official statistics—have an unparalleled role to play in communicating with their users.

The new **Strategic Communication Framework for Statistical Institutions** serves as a guide to help them navigate the changing demands of this role.

Statistical communication is about more than writing press releases or answering user questions and requests. NSOs need a modern, proactive communication strategy with clearly defined key messages, and must use different channels to reach various target audiences. This may seem obvious, but it is a relatively new concept for many NSOs, which have traditionally focused their efforts and resources on dissemination—publishing their figures in tables, databases and sometimes analytical reports and leaving users to get on with the task of finding what they need, processing it and making sense of it. Dissemination has often been designed principally for expert users who know what they are looking for and how to interpret it.

Increasingly, though, NSOs are embracing the idea that their ‘target audience’ includes all types of users, or even non-users, and that two-way communication with citizens and improving statistical literacy fall within their remit.

With this in mind, the High-Level Group for the Modernization of Official Statistics (HLG-MOS), a group of chief statisticians reporting to UNECE’s Conference of European Statisticians (CES), decided back in 2018 to make strategic communication a key priority. They launched a project to develop a common framework that can serve as a guide for individual NSOs as they rethink their approaches to communicating with data users and with the public. The resulting framework, developed by a group of experts from 11 countries - Australia, Bosnia and Herzegovina, Canada, Croatia, Ireland, Italy, Mexico, Netherlands, Poland, United Kingdom and the United States - plus OECD and Eurostat and endorsed by CES, highlights examples of success and lessons learned from countries in a range of aspects of communication: rebranding in Canada and Poland; issue and risk management in Australia; and crisis communications in the United States of America.

An evolving collection of case studies, published alongside the framework and still growing as countries’ experiences continue to be shared, demonstrates how NSOs have communicated with their users during the COVID-19 crisis—both about the impact of the pandemic on their operations and with tailored statistical communications about the pandemic itself.

The work will not end here. An HLG-MOS Task Team devoted to Capability and Communication is now extending the Framework, developing practical guidelines, examples and tools for managing crises and for brand management.
Statistical data combined with location information can provide critical knowledge through the integration with other data in the data ecosystem to understand multi-faceted issues of the current society. To address the information needs of the users in an increasingly complex and intertwined society, there is a great need for statistical data to be geospatially enabled using consistent and common geographies, in an accessible and usable format.

The production of geospatially enabled statistics should be a routine operation for statistical organisations, not just one-off exercise. The crisis such as global pandemic highlighted that statistical organisations should be prepared to produce them in an efficient and timely manner. To ensure this occurs, geospatial-relevant activities and considerations should be integrated into the regular production processes of statistical organisations, so that the design and production of geospatially enabled statistics can be conducted in a systematic and consistent way.

Using two global frameworks, Generic Statistical Business Process Model (GSBPM) and Global Statistical Geospatial Framework (GSGF), the Geospatial Task Team under the Supporting Standards Group of the High-Level Group on the Modernisation of Official Statistics (HLG-MOS) developed a Geospatial view of GSBPM (GeoGSBPM). It describes geospatial-related activities and considerations needed to produce geospatially enabled statistics using the structure of GSBPM while taking into account GSGF principles so that the resulting statistics have a higher level of standardisation and geospatial flexibility, as well as a greater capacity for data integration.

More details about the GeoGSBPM can be found on this wiki page.
Various Expert meetings or workshops are organized as part of the programme or work of the High-Level Group for the Modernisation of Official Statistics. These statistical process based events bring together experts to discuss innovative, best practices and future work in these areas. We welcome submission of abstracts and you can preliminary register for the events.

- **Expert Meeting on Statistical Data Collection**, 27-30 September 2021, online | [Click for the Information Notice 1](#) | Link to register or submit an abstract.
- **Expert Meeting on Dissemination and Communication of Statistics**, 11-14 October 2021, online (due to the ongoing Covid pandemic, the meeting can unfortunately not be held in Lisbon Portugal) | [Click for the Information Notice 1](#) | Link to register or submit an abstract.
- **Expert meeting on Statistical Data Confidentiality**, 1-3 December 2021, Poznan, Poland (if Covid permits: to be confirmed) | [Click for the Information Notice 1](#) | Link to register or submit an abstract.
Despite the Covid19 pandemic, the involvement in the ModernStats Community increased from 2019 to 2020.

**UPDATE:**

1000+ Participants
250+ Members
100+ Organisations

There were not only more participants in workshops (as these were moved from in-person to online) but also groups and projects saw increase in participants. We thank all of you for joining the ModernStats community and hope to see you all back in 2021!

**UPDATE:**

2020 involvement in HLG-MOS Activities:

6 Workshops
2 Webinars
6 Sprints/
Sessions
4 Expert Meetings/
Workshops & 3 Webinars

1.250+ Participants
300+ Members
50+ Organisations

2019 involvement in HLG-MOS Activities:

7 Workshops
6 Sprints
38 4 Projects
62 members
23 org.

750+ Participants
150+ Members
100+ Organisations

Taeke Gjaltema posted on Jan 27, 2021