REPORT OF THE WORKSHOP

1.   The ModernStats World Workshop 2022 was held in Belgrade, Serbia, from 27 to 29 June 2022. It was attended by 54 participants from Albania, Bosnia and Herzegovina, Canada, France, Hungary, Iceland, Italy, Mexico, Montenegro, Netherlands, Poland, Portugal, Saudi Arabia, Serbia, Slovenia, Spain, Sweden, Türkiye, the United Kingdom of Great Britain and Northern Ireland, the United States of America, Uruguay, Eurostat, International Labour Organization (ILO), International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), the United Nations Interim Administration Mission in Kosovo (UNMIK), the United Nations Economic Commission for Europe (UNECE), CODATA as well as private sector and academia.

2.   The workshop was organised by the Supporting Standards Group of the High-Level Group for the Modernisation of Official Statistics (HLG-MOS).

3.   The aims of the Workshop were to progress work on development and maintenance of the ModernStats models (e.g., GSBPM, GSIM, GAMSO, CSPA, CSDA), and to provide a platform for the users to exchange experience and lessons learned. The programme of the workshop was structured with the following three sessions:

- ModernStats – Implementation: share experiences from users in implementing ModernStats models, challenges and lessons learned;

- ModernStats - Evolution: share experiences from users about how ModernStats models have been used for new data sources, methods or technology, present recent developments related to the evolution of these models, and discuss the future direction of the models;

- ModernStats - Integration: share the experiences of users in using multiple models together, present recent developments on this front, and discuss the way forward.

4.   The opening remarks were given by Mr. Miladin Kovacevic (Director General, Statistical Office of the Republic of Serbia) and Mr. Zoltán Vereczkei (Hungarian Central Statistical Office), the chair of the Supporting Standards Group.

5.   In total, 31 presentations were given in three sessions including 4 soap-box presentations. In addition to regular presentation, several interactive activities were organised (e.g., brainstorming exercise called "six-thinking hats", hands-on activity on Linking GSBPM-GSIM, small group discussion). For more information about the proceeding and remarks, please see Annex.

6.   The discussions during the workshop identified a number of ideas for further work which include:

- Documents and guidelines available online about the models are helpful for conceptual understanding. However, for the implementation (especially for GSIM), real life and concrete examples would be needed;
- The modernisation efforts within the official statistics field requires buy-in from the high level. We need more documentation and materials to communicate the value of the models to the top managers;

- Maturity model and tool kits are still valid, they need to be promoted more and used to help guide organisations in implementing the models;
- Establishing and implementing the governance for the metadata management (e.g., alignment of terminology, codelist) is challenge for the development of metadata-driven system;
- It is important to think how the models should be adapted to new trends (e.g., ML, transition to clouds, automated processes) and new data sources (e.g., registers, sensor data, web data);
- The new developments outside the official statistics field can help statistical organisations improve the process (e.g., MLOps can help make process design and flow more flexible);
- Several organisations are coming up with similar ideas and trying to do similar things (e.g., process modelling) without knowing each other. It would be beneficial to have a community of practices among these organisations to exchange experiences on top of the in-person workshops;
- If standards are interoperable, it does not matter which standards we use. We need to work on mapping for interoperability of these models and standards;
- Challenges for the statistical oragnisations - How to become real innovative organisations? How to move faster from R&D to real implementation? How to navigate the complex data management with more and diverse data that the organisations should handle? How to be flexible while preserving the foundations (e.g., quality) as the producer of official statistics?

7.   All abstracts, papers, presentations and other relevant materials from the workshop are available at the UNECE webpage at: https://unece.org/statistics/events/MWW2022

Annex: Summary of proceeding and discussion

1.   Session 1 "Implementation" was chaired by Mr. Franck Cotton (Insee, France) and Mr. Juan Muñoz (INEGI, Mexico). The session had the following presentations:

- GSBPM as a Backbone of the Internal Documentation System, by Ms. Tina Šijanec (Statistical Office of the Republic of Slovenia);
- Implementation of the Statistical Production Process Model in Statistics Poland, by Ms. Anna Długosz and Mr. Janusz Dygaszewicz (Statistics Poland);
- GSBPM Implementation: Experience from the IMF, by Mr. Marini (IMF);
- New Developments from the HLG-MOS Supporting Standards Group - GeoGSBPM Task Team, by Mr. Juan Muñoz (INEGI, Mexico);
- New Development from HLG-MOS Supporting Standards Group - GSBPM "Task" Task Team, by Mr. Carlo Vaccari (Istat, Italy);
- Industrialization of Official Statistics Production Systems: What Is It and How to Get There?, by Ms. Tuğba Değirmenci (Turkish Statistical Institute);
- Piloting Surveys with GSIM, by Ms. Sandra Déprez and Mr. Christophe Dzikowski (Insee, France);
- Modelling a Methods and Tools Catalogue compliant with Official Statistics standards and frameworks, by Mr. Paolo Francescangeli (Istat, Italy);
- New Developments from the HLG-MOS Supporting Standards Group - GSIM Revision Task Team, by Ms. Francine Kalonji (Statistics Canada);
- INEGI's Information Strategy, by Mr. Manuel Cuéllar-Rio (INEGI, Mexico);
- Transparency and UNECE Metadata Standards, by Mr. Daniel Gillman (Bureau of Labour Statistics, USA);
- INEGI Metadata Management Strategy, by Mr. Manuel Cuéllar-Rio (INEGI, Mexico).

2.   Feedback and comments made during discussion raised the following points:

- The users of process documentation system may not see the value of the system in the beginning of the implementation, but gradually they realise the documentation is actually helpful for their own work. After all, it is the users who need to provide documentations, so it is important that the users are convinced about the benefits, otherwise the quality of documentation might not be good;
- Different stakeholders play different role in the process documentation (e.g., subject matter sections, infrastructure sections), the entire organisation should be engaged for the implementation;
- Governance procedure for the management of metadata (e.g., alignment of terminology, codelist) is one of most difficult issues;
- GSBPM is useful to delineate accountability and responsibility across different units in the organisation, and to prioritise and plan resources between competing projects;
- Geospsatial information has always been used in the statistical organisations, however, GeoGSBPM could help recognise in a more systematic way different stages where the geospatial information is used and find ways to strengthen the current practices;
- Speaking the same language is important, the models provide a common language for international official statistics community;
- Sometimes, it is needed to hide the complexity of the models from the users (e.g., instead of bringing up GSIM classes to subject-matter experts, just focusing on the core concepts that are relevant to them);
- Documents and guidelines available online about the models are helpful for conceptual understanding of the models. However, for the implementation (especially around GSIM), real life examples and concrete examples would be needed;

- GSIM provides a rich set of conceptual classes to model different applications (e.g., business services in methods catalogue, description of metadata-driven collection system). However, there still remains issues with missing classes and ambiguity with the interpretation;
- Statistical organisations need well-designed implementations to obtain the benefits from the standards and models;
- Transparency depends on documentation; the reference models help with transparency and trust;
- Maturity model and tool kits are still valid, they need to be more promoted and used to help guide organisations in implementing the models;
- Ten years ago, GSBPM used to be more theoretical. The model become more and more concrete framework for statistical organisations and other data producing agencies, helping them to make their processes more efficient and effective;
- For the implementation of the models, it is useful to establish champions who can promote and explain the models in the organisation;
- The modernisation efforts within the official statistics field requires buy-in from the high level. We need more documentation and materials to communicate the value of the models to the top managers. Standards are the basis for collaboration, we need to make sure that top management understand this well.

3.   Session 2 "Evolution" was chaired by Mr. Zoltán Vereczkei (Hungarian Central Statistical Office). The session had the following presentations:

- Adaptation of the GSBPM Model to Statistics Production, based on an Integrated System of Statistical Registers, by Mr. Federico Segui (National Statistical Institute of Uruguay);
- Ethics Management, Leadership and Performance, by Mr. Fabrizio Rotundi and Ms. Angela Leonetti (Istat, Italy);
- An Updated Process Model for a Modernised Production Process, by Mr. Jakob Engdahl (Statistics Sweden);
- Use of GSIM for Modelling Machine Learning Process, by Ms. InKyung Choi (UNECE);
- Innovation Meets Standardisation, But Where?, by Mr. Olav ten Bosch and Mr. Matjaz Jug (CBS, Netherlands);
- Integrating Technological Advancements into a Modernised Statistical Production Process, by Mr. Jakob Engdahl (Statistics Sweden).

4.   Feedback and comments made during discussion raised the following points:

- There are difficulties in using the GSBPM for the statistical registers (creation, update and maintenance), we need to look at how to use GSBPM with new data sources (e.g., registers, sensor data, web data) and see if there are any adaptions needed (e.g., more emphasis on building relationship with data providers);
- Ethics cut across the entire business architecture framework, understanding how it comes through strategies to production process is a key to turn ethical principles into action;
- It is important to think how the models should be adapted to new trends such as ML, transition to clouds and automated processes;
- ML is an example where we can learn how to be more flexible with process flows and process design (e.g., from MLOPs). It would be a missed opportunity if we treat ML just like sub-phase where it fits our traditional way of process flow;
- Standardisation provides guidelines to guide everyone towards the goal. If we see standardisation as a means to achieve goal, there would not be competition between innovation and standardisation;

- Innovations need standardisation when we try to scale up. PET and synthetic data are the areas that needs to be scaled up, and need standardisation;
- It is important to have an open space where everyone can bring their ideas and proposals for the models so that the models can continuously evolve;
- Several organisations are looking at the process modelling with the models, could we develop a common solution?

5.   In addition to the presentations, the "Evolution" session had a guided brainstorming exercise called "six thinking hats" to identify challenges, opportunities and issues that statistical organisations have. The challenges identified during this exercise include:

- Official statistics is not changing fast enough while other data producers change fast with more appealing data products;
- How NSOs become real innovative organisation? How to move faster from R&D to real implementation? How to be flexible while preserving the foundations as the producer of official statistics?;
- With more data (volume, diversity), data management becomes more complex. How statistical organisations should maintain the trade-off between data quantity and quality?;
- How to better anticipate the user needs and provide data rapidly?;
- Can there be a universal metadata model for official statistics, not just for NSOs, but also for the whole statistical system?

6.   Session 3 "Integration" was chaired by Mr. Carlo Vaccari (Istat, Italy) and Ms. Christine Laaboudi-Spoiden (Eurostat). The session included the following presentations:

- New Developments from the HLG-MOS Supporting Standards Group - Linking GSBPM and GSIM Task Team, by Mr. Flavio Rizzolo (Statistics Canada);
- A Model for Documenting and Monitoring Quality of Statistical Registers According to GSBPM and GSIM, by Ms. Giorgia Simeoni (Istat, Italy);
- Promoting Modernization: The Integrated Use of ModernStats Models (GSBPM, GSIM, GAMSO) at the Hungarian Central Statistical Office, by Ms. Andrea Petres (Hungarian Central Statistical Office);
- DDI-CDI: A New Type of Metadata Specification for Data Sharing and Integration, by Mr. Arofan Gregory (CODATA);
- The Added Value of SDMX 3.0, by Mr. Edgardo Greising (ILO);
- HLG-MOS Project Data Governance Framework for Interoperability, by Mr. Juan Muñoz (INEGI, Mexico);
- Semantic Web, RDF, Linked Data, by Mr. Franck Cotton (Insee, France);
- Towards Open and Linked Data for Statistical Classifications and Metadata Catalogues, by Ms. Christine Laaboudi-Spoiden (Eurostat);
- New Developments from the HLG-MOS Supporting Standards Group - The Core Ontology for Official Statistics, by Mr. Franck Cotton (Insee, France).

7.   Feedback and comments made during discussion raised the following points:

- Several organisations are coming up with similar ideas and trying to do similar things (e.g., process modelling) without knowing each other. Supporting Standards Group is more about development, it would be beneficial to have a community of practices to exchange experiences on top of the in-person workshops;

- It would be helpful to have new attributes in GSIM to describe data flow and process flow so that countries having the similar issues do not need to invent on their own;
- DDI can help NSOs reach out the non-usual users in scientific community. SDMX and DDI can complement each other, they could be deployed side by side;
- If standards are interoperable, it does not matter which standards we use. We need to work on mapping for interoperability of these models and standards;
- There are several proposals made to adapt the models, but it remains open how to do it (e.g., update explanatory text, addition of sub-process, extended version) and how to manage the changes;
- With the level of maturity growing, there are more countries using the models outside usual community, we need to find a way to connect with those;
- The works linking different models are very relevant, it helps improve understanding the models that are less popular. The recent developments toward more integration seem to be right direction;
- CSPA remains absent from the workshop, how to proceed with it remains as an open question.

8.   Soap box session had the following presentations:

- ONS-UNECE Machine Learning Group, by Mr. Eric Deeben (ONS, UK);
- Rapid Survey System, by Mr. Nebojsa Tolic (SORS, Serbia);
- INEGI's Data Architecture, by Mr. Manuel Cuéllar-Rio (INEGI, Mexico);
- Onyxia – a shared data processing and data science platform, by Mr. Franck Cotton (Insee, France).