REPORT OF THE WORKSHOP

1.   The ModernStats World Workshop 2020 was held online, from 27 to 30 in October 2020. It was attended by 148 participants from Antigua and Barbuda, Argentina, Belgium, Bulgaria, Canada, Egypt, Finland, France, Georgia, Hungary, Indonesia, Ireland, Italy, Japan, the Kyrgyz Republic, Latvia, Malta, Mexico, Montenegro, Nepal, Netherlands, Norway, Poland, Portugal, Qatar, Serbia, Slovenia, South Africa, Spain, Sweden, Switzerland, Turkey, Ukraine, the United Kingdom and the United States of America, Eurostat, International Labour Organization (ILO), International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), World Bank, the Data Documentation Initiative (DDI), the United Nations Secretariat as well as private sector.

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

3.   The main goal of the workshop was to share experiences on implementing ModernStats models, create a better understanding of the use and benefits of the models individually and collectively and to progress works on development and maintenance of the ModernStats models.

4.   The opening remarks were given by Ms. Marina Signore, Chair of the Supporting Standards Group. The programme of the workshop was organised with the following sessions:

- Generic Statistical Business Process Model (GSBPM) - chaired by Ms. Marina Signore (Istat, Italy) and Mr. Zoltán Vereczkei (Hungary)

- Generic Statistical Information Model (GSIM) - chaired by Mr. Franck Cotton (Insee, France) and Mr. Cory Chobanik (Canada)

- Common Statistical Production Architecture (CSPA) - chaired by Mr. Juan Muñoz (INEGI, Mexico) and Mr. Carlo Vaccari (Istat, Italy)

- Generic Activity Model for Statistical Organization (GAMSO), other standards and innovations - chaired by Ms. Márta Nagy-Rothengass (Eurostat) and Mr. Taeke Gjaltema (UNECE)

5.   In total, 24 presentations were made during 4 sessions. All abstracts, papers and presentations from the workshop are available at the UNECE webpage at: https://www.unece.org/index.php?id=53612

6.   The workshop was concluded with a panel discussion on the role of standards in the expanded data ecosystem.

7.   Throughout the workshop, participants discussed on the use of the models inside and outside the national statistical organisations (NSOs) as well as potential future works for the models. The remarks and proposals include:

- The role of the national statistical offices (NSOs) in coordinating and collaborating with other producers of the official statistics in the national statistical system (NSS) becomes increasingly important. ModernStats models can be introduced to other agencies in NSS to facilitate the communication, increase mutual understanding and ensure the quality of data they produce.
- ModernStats models are flexible and adaptable for new data sources and technologies. They provide a good minimum set of standards to start.
- ModernStats models help NSOs to become more agile. With Covid19, NSOs needed to change how they produce statistics and were asked to produce new information quickly. The models enable NSOs to become more efficient in re-organising production and reacting to new challenges faster.
- Modernisation involves multiple stakeholders, each with own viewpoints and problems. Several standards and models have been created to improve different aspects related to the production of official statistics which results in a complex model landscape. To maximise benefits and avoid overlaps and inconsistencies, it is needed to put them together as an integrated framework.
- Communication about ModernStats models should be more targeted to specific goals (e.g. new data source, new uses) with different communication methods (e.g. videos, storytelling). Benefit of the implementation of the models should be promoted nationally and internationally.
- For countries considering using ModernStats models, Modernisation Maturity Model can provide the recommendations and suggestions regarding how to introduce the model in the organisations and proceed to the next level of implementation.
- Understanding the models and their interlinkage might be easy but implementing the models in the organisations is more complicated (e.g. finding what GSIM object means in the actual process). Support and guidance on this area would be very helpful for countries trying to implement the models.
- Telling stories with GSIM objects can show practically how these objects can be used. It is useful to visualise for people how these seemingly abstract objects move around and interact with each other.
- Once SDMX and DDI are fully released, GSIM mapping to these standards should be updated.
- It will be important to incorporate the developments that are being done in data science and machine learning in the future of CSPA.
- Using open source in a strategic way is important. Themes on open source should be a separate track in the next workshop to discuss on topics such as governance, responsibility and culture.
- Once countries standardise processes, GAMSO could provide help in corporate supporting roles. GAMSO corporate activities such as quality management and metadata management can only effectively be done once processes are harmonised and standardised.

8.   More details about proceeding and discussion can be found in the Annex.

Annex: Summary of proceeding and discussion

1.   Session 1 "Generic Statistical Business Process Model (GSBPM)" was chaired by Ms. Marina Signore (Istat, Italy) and Mr. Zoltán Vereczkei (Hungary). The session consisted of an introductory presentation on GSBPM, presentations on the new development from Supporting Standards Group (Geospatial task team) and country implementation as below:

- Introduction to GSBPM, by Mr. Carlo Vaccari and Ms. Marina Signore (Istat, Italy)
- Overall Review of Statistical Production Processes from the GSBPM perspective: The Case of Kyrgyzstan, by Mr. Omurbek Ibraev (National Statistics Committee, Kyrgyzstan)
- Managing VUCA with VUCA in Statistics Indonesia, by Mr. Joko Parmiyanto (BPS, Indonesia)
- New development from HLG-MOS Modernisation Group - Geospatial task team, by Mr. Juan Muñoz (INEGI, Mexico)
- Ireland - a new approach to monitoring the COVID-19 Outbreak through geostatistical visualisation, by Mr. Kevin McCormack (Central Statistics Office, Ireland)
- Using GSBPM for production of geospatial data, by Mr. Juan Muñoz (INEGI, Mexico)

2.   In addition to clarification questions about the presentations, the points raised during the discussions include:

- Non-linearity of GSBPM is a strength of the model, it is suitable in environments following agile methodologies as users can design processes using any phase and sub-process without "breaking" the model. It is important to move away from the old way of thinking that statistical process is a static list of processes that are followed one after another.
- In the beginning, GSBPM was mainly used for documentation of processes, but the model goes beyond definition and documentation of processes. GSBPM can also be used as a basis for cost allocation for statistical processes and a tool to improve quality through its quality indicators.
- When organisation uses centralised production system, GSBPM is helpful as it can set the boundary where the work of one directorate starts and where that of another directorate starts.
- The role of NSOs in coordinating other agencies in the national statistical system becomes increasingly important. GSBPM can be introduced to these government agencies to facilitate the communication, increase mutual understanding and ensure the quality of data they produce.
- Benefit of the implementation of the models should be promoted nationally and internationally. It would be helpful to have GSBPM implementation cases (presentations, abstracts, etc) from previous UNECE events linked to the UNECE wiki page.
- Once countries standardise processes, GAMSO could provide help in corporate supporting roles. GAMSO corporate activities such as quality management and metadata management can only effectively be done once processes are harmonised and standardised.

3.   Session 2 "Generic Statistical Information Model (GSIM)" was chaired by Mr. Franck Cotton (Insee, France) and Mr. Cory Chobanik (Canada). The session consisted of an introductory presentation on GSIM, presentations on the new development from Supporting Standards Group (Linking GSBPM-GSIM task team and GSIM task team) and country implementation as below:

- Introduction to GSIM, by Ms. Jenny Linnerud (Statistics Norway)
- New development from HLG-MOS Modernisation Group - Linking GSBPM and GSIM task team, by Mr. Flavio Rizzolo (Statistics Canada) -
- New development from HLG-MOS Modernisation Group - GSIM task team, by Ms. Francine Kalonji (Statistics Canada)
- Defining a practical approach to realize a Statistical Datawarehouse platform using GSBPM and GSIM, by Mr. Freddy Maetens (Flanders Statistical Authority)
- GSBPM, GSIM, GAMSO! OMG!, by Ms. Andrea Petres (Statistics Hungary)
- Data architecture for statistical modernization: an integrated approach, by Mr. Flavio Rizzolo (Statistics Canada)

4.   In addition to clarification questions about the presentations, the points raised during the discussions include:

- Understanding GSIM might be easy but implementing and using the model in the organisations are more complicated (e.g. finding what abstract object means in the actual process). Support and guidance on this area would be very helpful for countries trying to implement GSIM.
- Telling stories with GSIM objects can show practically how these objects can be used. It is useful to visualise for people how these seemingly abstract objects move around and interact with each other.
- Modernisation involves multiple stakeholders, each with own viewpoints (e.g. CSDA for data and information, GSBPM for production process) and problems. This created a complex model landscape. Models have some overlaps, but there are also differences which can be used to complement each other. More works are needed to make the models aligned.
- Once SDMX and DDI are fully released, GSIM mapping to these standards should be updated.

5.   Session 3 "Common Statistical Production Architecture (CSPA)" was chaired by Mr. Juan Muñoz (INEGI, Mexico) and Mr. Carlo Vaccari (Istat, Italy). The session consisted of an introductory presentation on CSPA, presentations on the new development from I3S Project and country implementation as below:

- Introduction to CSPA, by Mr. Jean-Marc Museux (Eurostat)
- I3S – architecture guidance, by Mr. Jakob Engdahl (Statistics Sweden)
- Presentation on cloud native Service deployment, by Mr. Trygve Falch (Statistics Norway)
- Communicating about the sharing of statistical services, by Mr. Benoît Rouppert (Insee, France) and Mr. Pedro Cunha (INE, Portugal)
- ModernStats standards supporting the implementation and sharing of statistical services, by Mr. Mauro Bruno (Istat, Italy)
- BREAL in real life: the OJV use case, by Mr. Tomaž Speh (Statistics Slovenia) and Mr. Frédéric Gallois (Insee, France)

6.   In addition to clarification questions about the presentations, the points raised during the discussions include:

- It will be important to incorporate the developments that are being done in data science and machine learning in the future of CSPA.
- For countries starting to use CSPA, Modernisation Maturity Model can provide the recommendations and suggestions regarding how to introduce the models in different development levels.
- Developing services in CSPA principles creates benefits even if the services are not shared because CSPA triggers architecture decisions that could allow developers and architects to do a more sustainable and long-term thinking.
- Using open source in a strategic way is important. Even if the open source services and techniques are not used directly, they can provide more information than simple description which can initiate a new idea and create better services. In areas such as machine learning and big data, open source is becoming default, and this will influence internally to move toward open source in the organisation.
- Providing open source service is one problem but providing its maintenance and quality service is another problem. If there is no support when the service has an issue, the service will not be used whether it is open or not.
- Themes on open source should be a separate track in the next workshop to discuss on topics such as governance, responsibility and culture.
- LIM connects GSIM and CSPA and it should be updated after GSIM minor update and CSPA 2.0 are finalised.

7.   Session 4 "Generic Activity Model for Statistical Organization (GAMSO), other standards and innovations" was chaired by Ms. Márta Nagy-Rothengass (Eurostat) and Mr. Taeke Gjaltema (UNECE). The session consisted of an introductory presentation on GAMSO, presentations on the integrated view of ModernStats models, other standards and new innovations as below:

- Introduction to GAMSO, by Ms. Jenny Linnerud (Statistics Norway)
- Interconnection of ModernStats models, by Marina Signore (Istat, Italy)
- SDMX and GSBPM, by Mr. Juan Muñoz (INEGI, Mexico)
- New Developments with the Data Documentation Initiative (DDI), by Mr. Dan Gillman (Bureau of Labor Statistics, USA)
- BREAL: making Big Data real in statistical processes, by Ms. Monica Scannapieco (Istat, Italy)
- HLG-MOS Machine Learning project, by Mr. Claude Julien (HLG-MOS Machine Learning Project Manager)

8.   The workshop concluded with a panel discussion on the role of standards in the expanded data ecosystem. The panel discussion was moderated by Mr. Taeke Gjaltema (UNECE) and consisted of following panellists:

- Mr. Cory Chobanik (Canada)
- Ms. Monica Scannapieco (Istat, Italy)
- Ms. Marina Signore (Istat, Italy)
- Mr. Edgardo Greising (ILO)

9.   In addition to clarification questions about the presentations, the points raised during the discussions and panel discussion include:

- NSO should collaborate with owners of administrative data as well as other data who may not be the part of NSS. In coordinating and collaborating with other producers of official statistics, ModernStats models and other standards can serve as a very important set of tools for NSOs to ensure data life-cycle support and data quality in the new data ecosystem.
- Compared to GSBPM, GAMSO is less understood and known. It would be good to go from well-defined and world-wide used GSBPM and develop a set of clear overarching activities on development and management (which are in GAMSO sphere) and integrate those back to GSBPM.
- Modernisation Maturity Model is a self-assessment tool that can help organisations to evaluate where they stand in terms of implementation of the models and to understand how to progress to a higher level of the implementation.
- For people who are considering using GSIM which is often perceived abstract, it could be useful to have relatively simple use-case-based guidelines with recommendations on which parts of GSIM (and other models) to use for which practical situations. Linking GSBPM and GSIM task team can provide practical examples of how GSIM objects are used for different contexts.
- There is an ecosystem of coexisting standards and models that have been created to improve different aspects related to the production of official statistics. As these models evolve, they tend to converge and cover gaps. In order to obtain the best benefits and to avoid overlaps and inconsistencies, it is needed to put them together as an integrated framework.
- Mapping ModernStats models to implementation-based standards can put them into a context which is less abstract and show how the models relate to each other.
- Work is needed to cross-link the Awesome List of Statistical Software and CSPA services (e.g. adding interactivity layer on top of the awesome list).
- ModernStats models are flexible and adaptable for new data sources and technologies as the work of geospatial task team has shown.
- ModernStats models provide a good minimum set of standards to start although there needs some more effort to fill gaps and align them with what is going on with respect to new data sources and technologies. They can be used for communicating with other communities.
- ModernStats community has been highly productive, churning out various materials, but it is often occupied with development and jump into creating a solution. Stepping back and checking what has been done could save time and effort.
- Single agency cannot produce all the information that is needed these days, provision of sound information evidence should rely on series of institution that are not traditional statistical producers and data sources that have been originally designed for statistical data sources. With this regard, building and coordinating institutional network are becoming more and more important, this task is often more difficult than obtaining data or preparing technical platforms. Perhaps extending GAMSO for national statistical system could be considered as a future work for Supporting Standards Group.
- Outside the national statistical organisations, the models are not well known, there is a great opportunity to reach out to different communities and propose these models.
- ModernStats models help NSOs to become more efficient and agile. With Covid19, NSOs needed to change how it produce statistics and were asked to produce new information quickly. The models enable NSOs to become more efficient to re-organise production and react to new challenges faster.

- Communication policy should be more targeted to specific goals that we want to (e.g. new data sources, new uses) and develop different communication methods (e.g. videos, storytelling).
- The outputs that have been produced by ModernStats community are valuable to all statistical organisations around the world. The community should make sure that the outputs are visible and accessible so that the knowledge, information and tools are transferred to the developing countries.