United Nations ECE/CES/2023/9

Economic and Social Council

Distr.: General
11 April 2023

Original: English

Economic Commission for Europe

Conference of European Statisticians
Seventy-first plenary session
Geneva, 22–23 June 2023
Item 2 of the provisional agenda
Work of the High-level Group for the Modernisation of Official Statistics

Report on the work of the High-level Group for the Modernisation of Official Statistics

Prepared by the Secretariat and the Executive Board of the High-level Group for the Modernisation of Official Statistics

Summary

The document provides a summary of the outcomes of the work of the High-Level Group for the Modernisation of Official Statistics (HLG-MOS) in 2022. Addendum 1 to the current document presents the updated terms of reference for the continuation of the work. Addendum 2 presents the work planned under HLG-MOS for 2023.

At the February 2023 meeting, the CES Bureau reviewed and approved the outcomes of work under HLG-MOS in 2022, the updated terms of reference, and the plans for 2023. The Conference of European Statisticians is invited to approve the report and the updated terms of reference of the High-Level Group and take note of the further work planned in the area of statistical modernization.

I. Introduction

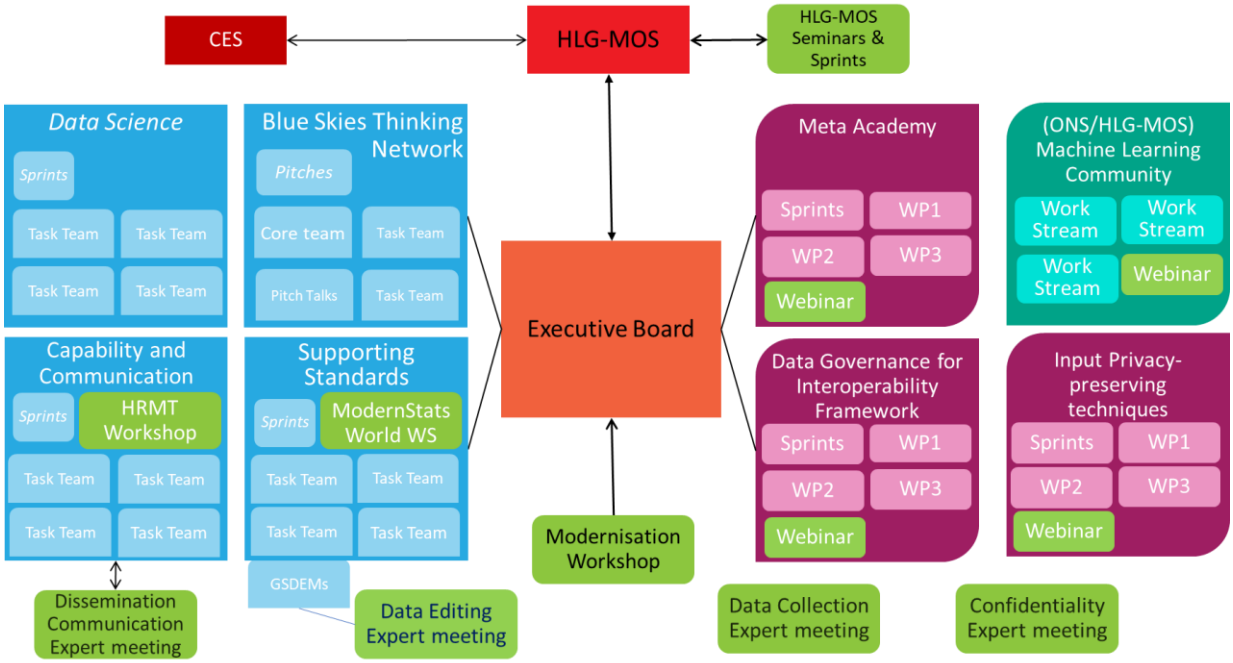
1. This report summarizes the achievements under the Work Programme of the High-Level Group for the Modernisation of Official Statistics (HLG-MOS) in 2022.

2. HLG-MOS was established by the Conference of European Statisticians to advance the modernisation of official statistics. It is led by Chief Statisticians of thirteen organisations who set the vision, mission, and priority topics and approve the work programme. The HLG-MOS Executive Board (EB) was set up to actively monitor progress and provide strategic direction and to adjust activities where necessary. Further information is available from the public HLG-MOS website.

3. In 2022, the HLG-MOS work programme consisted of three projects, four Expert Groups (each with additional task teams), a Community of Practice and several expert meetings and workshops organized by steering committees. A schematic overview of the groups, network, projects and workshops active under HLG-MOS in 2022 is presented in the figure below.

4. In total, Modernisation Groups, Task Teams, Steering Committees and Projects had over 650 members from over 100 different organisations. This number was significantly higher than in 2021, mainly due to the inclusion of over 350 participants from the Machine Learning Community. Expert meetings and workshops were attended by over 750 colleagues, which is lower than in 2021 as several meetings were held in person again.

Figure 1
HLG-MOS Structure in 2022



A. High-level Group for the Modernisation of Official Statistics

5. HLG-MOS regularly updates its priority areas in line with its vision and mission. The COVID-19 pandemic accelerated the changing roles and way of working in NSOs. HLG-MOS held a mini sprint in March 2022 that focused on how HLG-MOS can align its work with efforts to sustain the accelerated innovation observed during the COVID-19 pandemic. It was concluded that the focus should be more on higher-end statistical services (for example, linked to data stewardship) than 'traditional' statistical production. Understanding fully what the stakeholders need from HLG-MOS is important (user centricity). It was also stressed that the engagement impact of HLG-MOS activities as a collaboration platform cannot be underestimated. During the COVID-19 pandemic this had led to various bi- and

multilateral activities that allowed to share innovative practices and solutions. Some further key points were raised during the sprint for HLG-MOS to work on:

(a) Further articulating the HLG-MOS modernisation frame for the future;

(b) Measuring and promoting the HLG-MOS successes;

(c) Supporting continuous improvement of statistical business activities (process innovation).

6. The COVID-19 pandemic made clear that it is necessary to add new capabilities and to change office culture to respond well to the needs of users. HLG-MOS should continue supporting culture change, the move from data analytics to data engineering, and attracting and keeping new staff. HLG-MOS has the momentum to keep defining directions in terms of services and creative solutions.

7. There was strong support and appreciation of the foundational work HLG-MOS has done to innovate and modernise official statistics. HLG-MOS should keep supporting implementation of the outputs that were developed (standards, project outcomes etc.). The key role of HLG-MOS in facilitating transformation was emphasized. The Executive Board has a pivotal role in this. The Executive Board is trusted to keep on moving into new directions in line with the HLG-MOS vision and mission to sustain innovation. It was emphasised that the caveat for a continued success is that CES members provide resources for the work. The Executive Board is also committed to seek further collaborations beyond NSOs, for instance, with other international bodies, with academia and the private sector in order to advance the HLG-MOS modernization agenda.

B. Executive Board

8. In line with the outcome of the HLG-MOS sprint described above, the Executive Board (EB) will revamp and update the modernisation framework. It will seek to:

(a) Redefine continuously our level of ambition and where to focus efforts;

(b) Prioritize statistical processes and emerging services with highest potential;

(c) Shift from producers of statistics towards providers of services;

(d) Methods/standards/technologies are at the heart of developments, but also centred around people (staff, users, stakeholders, data providers);

(e) Increase sharing/collaboration to accelerate progress and delivery of results;

(f) Further optimize our expert meetings;

(g) Identify experts beyond NSOs in modernization fields of interest.

9. The EB oversaw the work continuously and met every month, including with project managers and the chairs of the groups. After a few years under the impact of the COVID 19 pandemic and staff shortages at UNECE, a full work programme was in effect. In fact, one additional project was included. The cycle of expert meetings and workshops had been changed to 24 months, with the option for online events in-between. The Applying ModernStats Methods Group started in 2022.

10. The monthly virtual meetings alternate between modernisation updates and meetings with a strategic focus. At any time, chairs and project managers can request assistance or interventions from the EB. The EB members are assigned as ‘champions’ of specific activities to follow the work in more detail. This helps to ensure that the focus of the expert meetings, workshops and the output of the projects and groups is aligned with the mission, vision, and priority topics of the HLG-MOS. Together with the UNECE secretariat, the group is responsible for organising the HLG-MOS Modernisation Workshop. Monthly progress updates of the groups and projects are available here: Modernisation Updates.

II. Projects

11. For 2022, three projects were selected: the Input Privacy-Preservation Techniques (extension), Meta-Academy for the Modernization of Official Statistics, and the Data Governance for Interoperability Framework project.

A. Input Privacy-Preservation project

12. The Input Privacy-Preservation project was extended to 2022. Statistics Netherlands made available Dennis Ramondt as a part-time project manager. In total, 23 colleagues from seven organisations participated in the project. Project members had online plenary meetings once a month and more regular meetings of the specific work packages and subgroups. Despite the fact that a substantive specialist to co-lead the project was not found and that in-person sprints were not possible, the project progressed well.

13. The project was structured around generalized use cases. After refining the scope and first documenting and investigating various use cases, generalized use cases were defined. This led to three tracks of work:

(a) Private set intersection with analytics;

(b) Private machine learning;

(c) A public consultative survey.

14. The private set intersection is a secure multiparty computation cryptographic technique that allows two or more parties holding sets to compare encrypted versions of these sets in order to compute the intersection. In this scenario, neither party reveals anything to the counterpart except for the elements in the intersection. This allows for queries (aggregations), and analysis can be done on the combined data set without revealing the micro data or other than agreed information of the counterpart.

15. The private machine learning (ML) pilot aimed to build a simulated environment to validate the concept of multi-party privacy preserving Machine Learning (PPML) for both training and interference. The scope was to investigate best practices and open-source tools for distributed and collaborative ML training among multiple organisations in a low-trust environment whilst mutually benefitting from the outcomes (the final model) or allowing safe 3rd party access. A simulated multi organisational set-up with several NSOs gathering data from individuals (sensors) to predict their activities (time use and well-being surveys) was tested successfully.

16. The informal and open technical consultation focused on identifying challenges in practical implementation of a trustworthy Multi-Party Secure Private Computing-as-a-service infrastructure for official statistics (end-of November 2022 was the deadline for replying to the consultation). It targeted experts and stakeholders such as privacy and security experts, potential users, digital activists and civil society representatives and researchers and developers. The project team will summarize the main outcome of the consultation in a public report that will be made available on the project page.

17. Results were shared at a well-received seminar and preparations were made for a public consultative survey on multi party private computing-as-a-service. Presentations are available from the meeting site: https://statswiki.unece.org/x/FIECFg. The project was completed by the end of 2022 and a final report will be produced.

B. Meta-Academy for the Modernization of Official Statistics

18. The Meta-Academy for the Modernization of Official Statistics project was led by Kate Burnett-Isaacs from Statistics Canada and Eric Anvar from OECD. In the project, 29 colleagues from thirteen institutions collaborated. The overall project group met on a regular basis via online video calls. These meetings included several presentations of methodological and practical work undertaken by the members. A seminar related to the Meta Academy and Carpentries was organized adjacent to the HLG-MOS workshop, which provided an

introduction to Git and version control for NSOs, as well as identifying approaches to incorporate generic training within our respective NSOs and obstacles to getting their employees to undertake training.

19. The project explored the opportunities for statistical offices to develop a common framework for *co*-creation of high-quality training, and to reuse content at an international level. It was concluded that currently significant barriers exist:

(a) There is no internationally shared understanding of what the training needs are, and no shared methodology or pedagogic approach to create learning content;

(b) There is no forum or community for 'academy managers' or 'trainers' to exchange practices and learn from each other, to share priorities or to decide to co-invest in, or mutualise, training efforts for a particular topic, including by constantly evolving with the industry and emerging needs.

20. Exploring national and international training initiatives provided insights in what a common framework should entail for statistical offices to join forces in upskilling for modernisation in NSOs. Several potential avenues emerged in this exploratory project of which the critical ones where:

(a) Pedagogy: Self-learning approaches and synchronous learning experiences with a coach (onsite or online) are complementary, and self-learning (including certifications) can be an interesting foundation to make the most of synchronous engagement. However, especially for more advanced skills, human engagement through collective time and a coach-trainee relationship are essential;

(b) Content: Some initiatives take a broad approach at covering all regular official statistics activities (including statistical standards); others are focused on data science or data intensive scientific activities – the latter including initiatives from the market or outside official statistics;

(c) Audience: Some initiatives target a national audience (national statistical system only, or broader civil service) while others aim to grow a broader international network (e.g., with a data for development agenda). However, multilingualism remains a difficulty for many initiatives;

(d) Platform: All initiatives crystallise around a digital platform, through which training content and services are created and exchanged, and where a brand is promoted to unite a community of trainers and trainees. Crucial to success is that training content is not limited to presentations but includes several artefacts (including data samples, algorithms and notebooks, storyboards and training intentions, certification questionnaires etc.).

21. The most promising model identified is that of "The Carpentries" which focuses on synchronous learning within the arena of skills associated with data science. A ModernStats Carpentry could enable NSOs to develop a national ecosystem of trainers / trainees using an existing platform (to minimise IT costs). Within the Carpentries model, all of the artefacts mentioned above are shared as open assets under Creative Commons license in GitHub, reusable at will in other contexts. They could also be managed under the ModernStats branding and governance. Therefore a project proposal was prepared to initiate and develop a ModernStats Carpentry.

C. Data Governance for Interoperability Framework project

22. The Data Governance for Interoperability Framework project is lead by Juan Muñoz from INEGI, Mexico. The twenty participants from thirteen organisations met on a regular basis. The team is composed of experts with different backgrounds providing diverse points of view. Significant effort was spent on scoping the project and defining the main aspects that the framework must cover. Various core concepts for data interoperability in the context of statistical offices, have been defined. After further scoping and discussing the purpose and identifying what output would have most added value, the structure of the statistical data governance framework to achieve data interoperability was set.

23. The framework consists of a system of rules and theories that will help to align efforts, unify concepts, and organize the core elements that are needed to establish and manage an interoperable platform of data, metadata and systems. This will serve as the structure for a document describing the reference framework that contains the main elements to achieve data interoperability. The framework will drive the reader through a straightforward process of building a data governance environment to achieve interoperability in their organization. The document aims to explain why we need a framework for interoperability, the goals it needs to achieve and how to achieve them.

24. The team members have started creating content to populate the different sections that were agreed. However, due to the complexity and limited resources available in the project, the group developed a proposal to extend the project to 2023.

D. Synthetic Data project (follow-on activities)

25. While the synthetic data project finished in 2021, there were some significant follow-on activities in 2022 which continued to be led by Statistics Canada. This included a hackathon in January 2022 to test the guidance developed by the synthetic data project. It was followed by the development of these guidelines into a published report which was finalised by December 2022. This publication, entitled *Synthetic Data for Official Statistics: A Starter Guide*, will be released in January 2023.

III. Modernisation Groups

26. In 2022, there were four Modernisation Groups under the HLG-MOS structure: the Supporting Standards Group, the Capability and Communication Group, and the Blue Skies Thinking Network and starting in 2022, the Applying Data Science and Modern Methods Group. Under each group, various Task Teams or Streams are set up to work on the activities that were identified for the 2022 work programme. Each Group and the Task Teams met regularly, in general at least once per month. The groups coordinate their activities between themselves, but also with the HLG-MOS projects and other international activities in similar areas (for example at Eurostat, OECD, and the United Nations Statistical Division). The chairs of the Groups report to and discuss issues with the Executive Board on monthly basis.

A. Applying Data Science and Modern Methods Group

27. The Applying Data Science and Modern Methods Group was launched in early 2022. Reflecting the increasing importance of new data sources and methods for the compilation of official statistics, the group aims to go beyond conceptual frameworks for data science and modern methods, and identify concrete opportunities to further modernize NSO’s business processes.

28. The group consists of 21 colleagues from 10 organizations. To define the scope of the work, the group has conducted a market landscape analysis to take stock of existing works in the field of data science and modern methods. Eight potential topics were identified under data collection and integration, data editing, data confidentiality, data dissemination, etc.

29. Based on the expertise and interest of the members, three topics were prioritized, and task teams were set up. The group met once a month and the different task teams met on monthly bases or if needed, more frequently.

(a) **Task team 1 - Understanding and Selecting Models**: There is a large and rapidly developing collection of modern methods and different machine learning models, but this information is scattered in silos with little overarching guidance. This task team thus aims to consolidate knowledge from NSOs into guidance that enables users to have a comprehensive understanding on how different methods and models fit different types of problems, data and purposes, thereby facilitating the selection of the most suitable solutions.

(b) **Task team 2 - Accelerating the Implementation of ML-based Solutions in Data Editing**: Machine Learning has a good potential for efficiency gains in complementing

or replacing traditional methods, as well as for improving quality in ways that may be difficult to achieve with traditional methods. The purpose of this task team is to develop a generic guidance on key organizational issues when implementing ML-based solutions in data editing to accelerate the implementation process.

(c) **Task team 3 - International Framework on Responsible Artificial Intelligence (AI) for Official Statistics**: With the wide-scale utilisation and adoption of AI and ML, core principles such as ethics, privacy, fairness, and legality are more important today than ever before. The aim of this task team is to set up a common standard for responsible design, development, and deployment of AI-based solutions in NSOs, ensuring these new methods are acceptable from an ethical and human perspective.

30. The three topics were shared at the ONS Data Science Campus International Collaboration event in Newport, UK, from 12 to 14 July 2022, with strong support and interest from the participants.

B. Blue Skies Thinking Network

31. The Blue Skies Thinking Network (BSTN) is the ideas factory of the statistical modernisation community. The network provides a research and innovation platform where members can share ideas and look for partners to explore new innovations to the production process that can benefit statistical organisations. The objective is to generate and evaluate proposals for HLG-MOS activities and where needed, to do short time-boxed follow-up studies. The network is an umbrella structure for a core group and short-term task teams to investigate new ideas and opportunities through short evaluation projects supported by the HLG-MOS Executive Board. Any idea that is in line with the HLG-MOS strategies can be considered. To allow for new innovations, space is also given to out-of-the-box thinking.

32. The network was led by the innovation manager Barteld Braaksma from Statistics Netherlands and supported by UNECE. The core expert group consists of around fifteen members that meet at least once a month on average.

33. The BSTN reviewed various pitch presentations during the year. These are a type of mini sprint, in which members and outsiders can present briefly an idea or a project in the area of modernisation. The core group and invited experts gave suggestions for how they could be refined or elaborated. This led to an activity proposal on Knowledge Sharing for Open-Source Transformation in NSOs, plus four project proposals for the 2023 HLG-MOS work programme:

(a) Extension of the Data Governance Framework to Achieve Data Interoperability (Mexico);

(b) ModernStats Carpentries (phase 2 of the Meta Academy project) (OECD/Canada);

(c) Robotic Process Automation in Official Statistics (Hungary);

(d) Cloud for Official Statistics (Ireland).

34. Other activities concluded by the Network in 2022 were:

(a) **Cloud for official statistics**: BSTN convened a session on Cloud for official statistics in Belgrade, Serbia on 29th June where the potential demand for international collaboration on this topic was identified. The session helped to build consensus about which elements of cloud adoption and use would benefit most from a project. It was followed by a further cloud session at the ONS Data Science Campus International Collaboration event in Newport, UK, on 12th July. This helped to further scope the areas that a project should cover and identified Ireland as a volunteer to lead the development of the project proposal that was presented to the HLG-MOS workshop.

(b) **Rapid surveys**: A session was organized at the July Newport event to further identify opportunities of rapid surveys. Different national experiences from Canada, Ireland, Serbia and the Netherlands such as business confidence surveys, COVID impact surveys, and public perception surveys, were shared and discussed. It was suggested that multi-country

surveys could share data in a cloud-based system to allow easy comparison of their results. Based on the examples presented on non-probabilistic surveys, conducted via a social media platform, a BSTN activity proposal was prepared.

(c) **Digital twins:** Work on this topic was accelerated in another sprint session at the Newport event. Several implementations of using remote sensing technologies fixed to infrastructure were presented. A possible vision was discussed for integrating separate infrastructure-related digital twins into a data-sharing ecosystem to improve the availability of information and decision making on topics such as environmental monitoring, resource utilization and interaction and knock-on effects between different types of infrastructure at times of disruption to these systems. This led to a BSTN activity proposal for 2023.

35. The group continues to be ready to evaluate activity and project proposals submitted from the statistical community. More details can be found on the public BSTN wiki page.

C. Capability and Communication Group

36. The Capability and Communication Group is responsible for aspects of Human Resource (HR) Management and Training as well as Communication in Statistical Organisations. HR and communication departments were strongly involved in the Covid-19 response. After an initial focus on the various challenges the pandemic posed, the emphasis in 2022 was shifted to the legacy of the pandemic.

37. During 2022, 52 colleagues from 22 organisations participated in the group and its task teams and subgroups. The group meets once a month and the different task teams and subgroups meet on monthly bases or if needed, more frequently. The group is led by Anna Borowska from Statistics Poland and Maria Hurley from CSO Ireland. Each task team has a chair assigned. The various activities are grouped into three streams:

(a) **Ethics management**: In addition to the two surveys that were conducted in 2021, another clarification survey was sent to the respondents of the earlier surveys. The survey was focusing on "business ethics", that is ethics as a cross-cutting element throughout the organisation and representing a key for a qualitative performance improvement. Preliminary results of the surveys were presented at the Workshop and can be viewed at: https://unece.org/statistics/events/HRMT2022;

(b) **Market research, digital marketing, and communication strategies**: A well-developed brand and reputation are a key strategic asset in promoting the value of Official Statistics and combatting the impact of disinformation on our societies. The task team developed guidelines on brand management, marketing and communication for Statistical Organisations and the guidelines were finalised and published on the Brand and Reputation Management wiki: https://statswiki.unece.org/x/cRGxEw.

(c) **Future work, future workplace, and future skills:** The future of work team focused on three areas of work: toolkits for blended working environment, for employers, employees and managers; reaching youth (as future employees, users of official statistics and data source for official statistics) and job of the future. The team sent a questionnaire to countries, focusing on issues such as wellbeing and blended working environment, how blended working is defined in different countries, how has learning and development evolved, etc. Results of the questionnaire and outputs of the teams could be viewed on the 'Human resources and more...' wiki: https://statswiki.unece.org/x/UgAMDQ.

38. The group was also responsible for organizing the **Workshop on Human Resources Management and Training** (11-13 October, Brussels, Belgium). The workshop was organised jointly with the European Free Trade Association (EFTA) and UNECE and was attended by 60 participants form 28 countries and organisations. The workshop consisted of four sessions: (i) WHY we need to re-frame NSI Work, People and Workplaces; (ii)WHAT have we tried, what have we learned from recent Crisis Response Initiatives?; (iii) HOW we build NSIs for the Future – (a) Organisation, culture & ethics; (b) People; (c)Training and skills; (iv)Where to from here? The workshop also had a special session for the EFTA partner countries in the afternoon of 13 October. Presentations and other documents are available from the UNECE website: https://unece.org/statistics/events/HRMT2022.

D. Supporting Standards Group

39. The Supporting Standards Group is responsible for the maintenance and development of the ModernStats models. These are modernisation models and standards developed by the HLG-MOS such as GAMSO, GSBPM, GSIM and CSPA[1]. The goal of the Group is to develop, enhance, integrate, promote, support and facilitate implementation of the range of standards needed for statistical modernisation.

40. The Group was chaired by Zoltán Vereczkei from Hungary and supported by the UNECE secretariat. Beside the monthly plenary expert group meetings, the task teams met on a frequent basis. Each task team has a chair assigned. The main group consists of twenty experts from twelve NSOs and four International Organisations. Many more colleagues collaborated in the various task teams.

41. The main activities and outcomes of the Group and its Task Teams in 2022 were:

(a) **GSIM revision**: After an initially planned soft update, it grew into a full review in 2022. Feedback from the community was collected on possible improvements of GSIM. There is a strong relationship with the work of the Core Ontology for Official Statistics and issues identified were shared. For details, please check our GitHub group at https://github.com/UNECE/GSIMRevision/. The activity will be finalized in the first quarter of 2023 after which a public review and release will follow. Release of the final results is planned for June 2023.

(b) **Core Ontology for Official Statistics (COOS)**: The work aims to develop an integration model for the core set of ModernStats standards backed by elements of well-known standard vocabularies. It defines a conceptual integration framework to provide semantic coherence across these models based on a common vocabulary of terms, definitions and a well-defined set of inter- and intra-model relationships formalized in RDF/OWL, using standards vocabularies, e.g., SKOS, PROV, DCAT, DC, ORG. The new version of COOS was ready, and a public review ended on 25 November. Details are available on GitHub: https://linked-statistics.github.io/COOS/coos.html. In line with the roadmap, activities will continue in 2023.

(c) **GSBPM task:** the GSBPM 'task' Task Team started in the fall of 2021. Tasks are the lower-level activities added at the national level to adopt the higher-level generic activity descriptions at the GSBPM sub-process level. Based on examples from eight countries, a finer level of activities of the GSBPM was developed. This can help other countries in implementing GSBPM. Outputs and further details are available from the GSBPM wiki: https://statswiki.unece.org/x/c4OuFQ.

(d) **SDMX/DDI/GSBPM**: the work is aimed at providing a description for each GSBPM sub-process both from SDMX and DDI point of view. This will help our users to understand how SDMX and DDI can help to implement the logical elements that are described by the GSBPM and GSIM (using the output of our previous Task Team on linking GSBPM and GSIM). The work will continue throughout 2023.

42. **ModernStats World Workshop 2022**: The workshop was hosted by Statistics Serbia on 27-29 June in Belgrade. It provides a platform for users of ModernStats models to exchange experiences and lessons learned, and to progress on development and maintenance of the models. The discussion and group activities were very lively, and lot of input was received from the community during the workshop in many different forms (direct comments, group discussions, brainstorming, ModernStats clinic etc.). The workshop conclusions (mainly suggestions for future work, issues, needs) are being followed up by the Supporting Standards Group. Workshop material, including the meeting report, are available from: https://unece.org/statistics/events/MWW2022.

[1] Generic Activity Model for Statistical Organisations, Generic Statistical Business Process Model, Generic Statistical Information Model, Common Statistical Production Architecture.

43. Due to lack of resources and the need for the group to focus on other activities first, the work on GSBPM overarching processes and on CSPA capacity building did not commence in 2022. Both will be postponed to 2023.

E. Machine Learning Community Group

44. The ONS-UNECE Machine Learning Group is an international platform for research collaboration, knowledge exchange and capability building on machine learning for official statistics. It brings together statisticians, data scientists and academics from 45 different countries around the world to explore how ML can improve statistical output and be integrated successfully into production. It was set up as a community of practice to follow-up on the successful 2019-2020 HLG-MOS Machine Learning Project. The community is led by the ONS Data Science Campus with support from UNECE. The objectives include providing a platform to facilitate the creation, development and implementation of research projects and skill-building activities that meet the global statistical community's needs. Further, to build and engage a strong machine learning community by sharing resources and good practice, exchanging ideas and experiences, and keeping abreast of developments in the field. If offers open, shareable and easily accessible resources to the community, and facilitates machine learning capacity building for official statistics.

45. Over 400 colleagues from 45 countries participated in the community. There is a strong demand for knowledge sharing and capacity building and high attendance at meetings. The work is structured around five work streams that aimed to address different issues that arise when using machine learning for official statistics.

46. The work in 2022 focused on moving from proof of concept to production. Based on 34 activity proposals and further discussions, the Theme Groups were formed around the following topics:

(a) Text classification;

(b) Imagery analysis;

(c) Automatic Identification System data;

(d) Modelling;

(e) Web scraping data;

(f) Quality of training data;

(g) Model retraining;

(h) IT infrastructure.

47. Other key areas focused on ethics, quality of training data, research collaboration and capability development. Progress was accelerated through sprints, notably those held at the UK Data Science Campus during an international event in Newport on 12-14 July. Capacity building took place through activities such as Coffee and Coding sessions and through the various Theme Groups that were set up. The work was shared through various communication and outreach activities such as web and wiki sites, discussion forums, guides, papers, a YouTube channel and various presentations and a Webinar (on 30 November).

48. To get an overview of all of the achievements of the Machine Learning Community Group, the final report of the ONS-UNECE Machine Learning Group 2022, as well as all reports of the Theme Groups, material presented at the monthly meetings and in the Coffee and Coding Sessions and the sprints held in Newport, are available from the ML wiki: https://statswiki.unece.org/x/34AqF.

IV. Meetings and workshops

49. In response to calls for action to reduce climate change and facilitate a broader audience every second year, the Executive Board and the UNECE secretariat decided to change the cycle of in person expert meetings and workshops to 24 months. In the in-between

years, online expert meetings and workshops can be organized. In 2022, the following expert meetings and workshops were held (in chronological order):

(a) **HLG-MOS meeting and Chief Statisticians Sprint** (10 March 2022, online): As the customary HLG-MOS meeting the weekend ahead of the Statistical Commission in New York could not be held due to COVID-19 restrictions, the members of the HLG-MOS met online. The work programme was approved, and there was a strong acknowledgement among the Chief Statisticians of the output produced in 2021. The sprint resulted in a revamped modernisation framework and updated priority topics. The Executive Board was tasked with continuing to oversee the progress and to assure that activities and projects will be aligned with the mission and vision statement and the updated priorities of the HLG-MOS.

(b) **Expert meeting on Dissemination and Communication of Statistics** (13-15 September, online): It was attended by 156 participants from 36 national statistical organisations and nine international organisations as well as from academia. Twenty-two presentations were given in three sessions on: (i) Brand, reputation, trust – Who we are, What we stand for, Our place in society; (ii) Market research, digital marketing and communication strategies; and (iii) Statistics made easy. There were plenary discussions at the end of the meeting discussin lessons learned and topics for future work. Topics identified and all meeting documents are available from the UNECE website: https://unece.org/statistics/events/DissComm2022.

(c) **Expert meeting on Statistical Data Editing** (3-6 October, online): The expert meeting focused on sharing experiences on modernisation of data editing and statistical production, including the use of administrative data and the use of Machine Learning and AI, as well as quality. Additionally, lightning talks and several collaborative activities to identify topics for future work, were held. Presentations and the outcome of the meeting are available from the meeting site: https://unece.org/statistics/events/SDE2022.

(d) **Expert meeting on Statistical Data Collection** (26-28 October, Rome, Italy): The meeting received strong interest and support, with more than 210 registrations. Due to the post-emergency prevention measures for COVID-19 risk management in Italy, the meeting venue could only accommodate around 50 persons and a hybrid meeting was arranged. A total of 53 participants from 20 countries attended the meeting in-person and around 60 participants joined the meeting remotely. The focus was on managing multimode and rapid data collection, improving the experience of respondents from initial contact through to completion of the survey, and responsive design that leverages data collected in the early stages of a survey to inform data collection design in the later stages. Sixty contributions were submitted and thirty-four of them were selected by the Organisation Committee. Presentations, papers, and outcomes are available from the meeting website: https://unece.org/statistics/events/DC2022;

(e) **HLG-MOS Modernisation Workshop and Seminars** (22-24 November, Geneva, Switzerland, and 30 November, online). The main event was attended in person by 55 representatives from 27 organisations and about 70 persons listened in online. The main aim of this workshop was to present the activities and output from the various groups, projects and workshops under the 2022 HLG-MOS work programme. Another major goal was to present the plans for the 2023 work programme. The proposals were discussed by the experts in small groups and projects were prioritized and suggestions for scoping were made. All meeting documents are available from the UNECE meeting website (https://unece.org/statistics/events/HLG2022). As a side event, three seminars were organized to share the work of two 2022 HLG-MOS projects and of the Machine Learning Community:

- **Meta Academy Seminar** (24 November, in person). The seminar consisted of two interactive sessions. The first focused on the outputs of work package 2 of the Meta Academy project on Git and version control, while the second was a brainstorming to identify gaps in and barriers to training of staff.

- **Input Privacy-Preservation project Webinar** (24 November, hybrid). The project outputs in private set intersection, private machine learning and multi-party private

computing-as-a-service, were presented. It concluded with possible applications and additional work that needs to be done for bringing them into production.

- **Machine Learning Group Webinar** (30 November, online). In this joint ONS-UNECE event, the work under the various works streams was presented. The focus was on various applications (web scraping, text classification, imagery data, AIS maritime data), quality of training data, model retraining and infrastructure.

V. Collaboration, communication and coordination

50. **Collaboration**: The work of HLG-MOS is facilitated by a large number of wiki spaces, web pages, GitHub spaces and other platforms to collaborate and share information. For example, the UNECE secretariat is managing and maintaining about fifty different public and restricted wiki sites available for collaborative purposes or for sharing output from HLG-MOS activities.

51. **Communication**: To further communicate the work of HLG-MOS, various news articles and reports to intergovernmental bodies were prepared. Moreover, the UNECE HLG-MOS secretariat responded to several requests, from within and outside the UNECE region, for receiving our publications, sharing information, presenting work or for assistance. Further information is available from the main HLG-MOS online wiki portal. The UNECE Secretariat can be contacted as well for further information.

52. **Publications**: most of the output of the work done under HLG-MOS is shared publicly on the above mentioned HLG-MOS wikis and UNECE website. In 2020, it was decided to include one publication related to statistical modernisation in the formal UNECE publication programme. As mentioned above, in 2022, the main output of the Synthetic Data project, was published as the '*Synthetic Data for Official Statistics: A Starters Guide'*. As we had the opportunity to be included in a formal digital publication, the output of the Machine Learning project was also compiled into an official UNECE publication (pdf format)[2]. This allows us to further showcase the work created under HLG-MOS.

53. **International Coordination**: The work was also coordinated with other International Organisations working on the modernisation of official statistics (e.g., Eurostat, OECD, and United Nations Statistics Division. Coordination was assured and linkages were made with international activities in similar areas. The most effective way of doing this was to make sure we have cross-membership between the various groups. The UNECE Secretariat and the Executive Board also regularly meet with representatives of other international modernisation initiatives to further coordinate and align the various activities.

[2] https://unece.org/statistics/publications/machine-learning-official-statistics