REPORT ON THE WORK OF THE HIGH-LEVEL GROUP FOR THE MODERNISATION OF OFFICIAL STATISTICS

Prepared by UNECE and the HLG-MOS Executive Board

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting of the 2021/2022 Bureau
Geneva, 23-24 February 2022

ECE/CES/BUR/2022/FEB/8
28 January 2022

For decision

Item III (e) of the Provisional
Agenda

REPORT ON THE WORK OF THE HIGH-LEVEL GROUP FOR THE MODERNISATION OF OFFICIAL STATISTICS

Prepared by UNECE and the HLG-MOS Executive Board

The paper provides a summary of the outcomes of the work of the High-Level Group for the Modernisation of Official Statistics (HLG-MOS) in 2021. Addendum 1 to the current document describes the plans for 2022.

The Bureau is invited to review and approve the outcomes of work in 2021 and plans for 2022 (presented in ECE/CES/BUR/2022/FEB/8/Add.1).

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 2021.

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

3.   In 2021, the HLG-MOS work programme consisted of two projects, three Expert Groups (each with additional task teams) and several expert meetings and workshops organized by steering committees. A schematic overview of the groups, network, projects, and workshops active under the HLG-MOS in 2021 is presented in figure 1.

4.   The ongoing COVID-19 pandemic continued to pose a challenge but impacted mainly organization of in-person meetings. In the end, only one expert meeting was organized with in-person participation. In-person sprint sessions could not be organized, which somewhat delayed the delivery of the projects outputs. On the other hand, there have been some positive effects as more participants were proficient at working together remotely, which had been the general working mode under the HLG-MOS already before the pandemic. In total, Modernization Groups, Task Teams, Steering Committees and projects had over 250 members from over 100 different organisations. Meetings and workshops were attended by over a thousand colleagues. This number was slightly lower than in 2020 but with one group less and fewer expert meetings, workshops[1] and sprints, the participation per activity was higher.

---

[1] Due to the pandemic and staff shortages at UNECE, the Executive Board had decided that expert meetings and workshops would have a 24-month cycle.
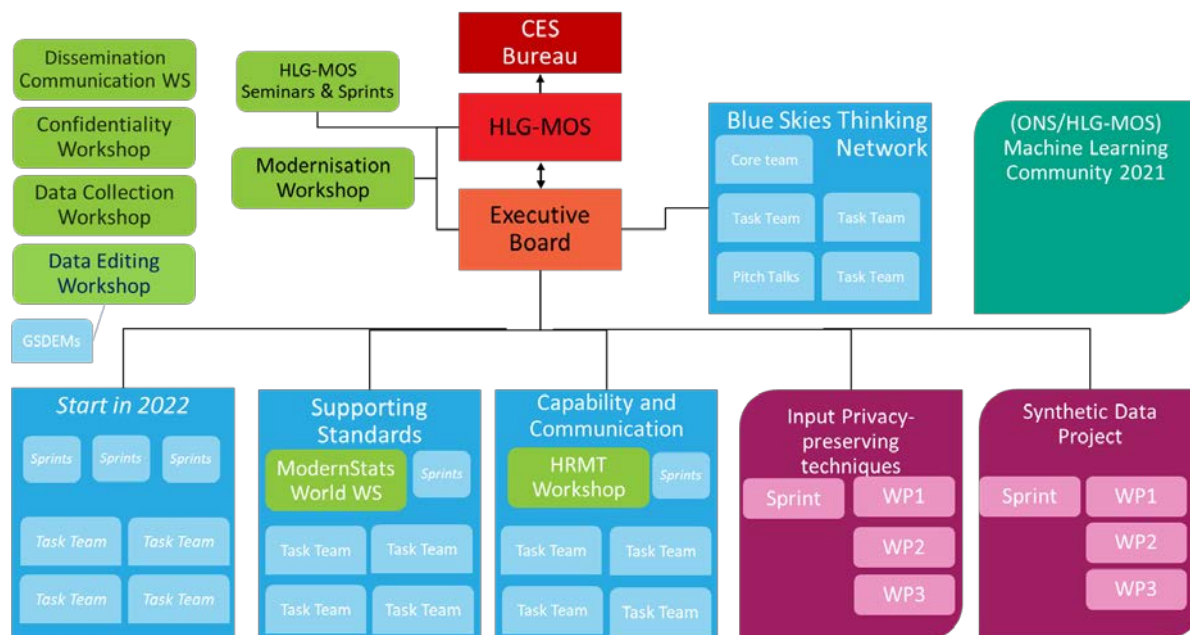
Figure 1 HLG MOS Structure in 2021

A.   Executive Board

5.   The Executive Board (EB) oversaw the work continuously and met every month, including with project managers and the chairs of the groups. The proposed work programme for 2021 already included measures to mitigate the impact of the COVID-19 pandemic and staff shortages at UNECE, such as the reduced frequency of Expert Meetings so no further adjustments were needed.

6.   The monthly virtual meetings alternate between modernisation updates and strategic discussions. At any time, chairs and project managers can request assistance or interventions from the EB. The EB members are also assigned as 'champions' of specific activities and 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 also responsible for organising the HLG-MOS Modernisation Workshop. Monthly progress updates of the groups and projects are available here: Modernisation Updates.

II.   PROJECTS

7.   For 2021, two projects were selected: the Input Privacy-Preservation Techniques and the Synthetic Data Guide project.

B.   Input Privacy-Preservation project

8.   The Input Privacy-Preservation project was delayed from 2020 due to the pandemic. Statistics Netherlands made available Dennis Ramondt as a part-time project manager. In total, 24 colleagues from nine 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. Scoping exercises already undertaken in 2020 helped in this.

9.   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;
>
> b) Private machine learning;
>
> c) Organization of a public consultative survey.

10.   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 to be done on the combined data set without revealing the micro data or other than agreed information of the counterpart.

11.   The private machine learning 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 practises 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 NSO's gathering data from individuals (sensors) to predict their activities (time use and well-being surveys) was tested successfully.

12.   Results were shared at a well-received webinar and preparations were made for a public consultative survey on secure private computing-as-a-service. As it was felt that extending the project would lead to significant benefits, a proposal for continuation was submitted to HLG-MOS and accepted.

C.   Synthetic Data Guide project

13.   The Synthetic Data Guide project was a follow-up to the work on synthetic data started under the Blue Skies Thinking Network in 2020. The project was led by Kate Burnett-Isaacs from Statistics Canada[2]. 50 colleagues from 25 institutions, including from academia and the private sector, participated in the project. The overall project group met on a monthly basis and Work Package and sub teams met frequently. These meetings included several presentations of methodological and practical work undertaken by the members.

14.   The main aim of the project was to produce a practical guide to synthetic data for NSOs. Synthetic data can be a solution to providing rich data while respecting integrity and confidentiality. This helps NSOs in their attempt to provide greater transparency and openness.

15.   The Guide presents theoretical methods to create synthetic data and provide an international consensus on practical applications and best practices to promote consistency, transparency, and comparability within and across statistical agencies, as well as among users in academia and the private sector. It provides coherent guidance to decision makers working at any level in NSOs so that they can determine if synthetic data is the right solution to their data disclosure problem. The guide is intentionally designed for practical application; it is not an exhaustive textbook. Resources to support further exploration of technical concepts are highlighted in the guide. It includes a practical decision tree to choose your method based on the properties of your original data and the desired properties of your synthetic data.

---

[2] She was assisted by a technical committee consisting of Christine Task (Nexus Research), Gillian Raab (University of Edinburgh), Kenza Sallier (Statistics Canada) and Allistair Ramsden (StatsNZ).

16.	The guide was presented in a dedicated public online webinar on 17 November as a side event of the 2021 HLG-MOS Modernisation Workshop (15-16 November 2021). A data challenge is planned in early 2022 (24-28 January 2022). Based on the evaluation of the challenge, the final version of the Guide will be prepared and is expected to be released in March 2022. The guide is included in the UNECE publication programme for 2022 as well. Preliminary results and the draft of the Guide are already publicly available (https://statswiki.unece.org/x/UQTUE).

## III.	MODERNISATION GROUPS

17.	In 2021, there were three Modernisation Groups under the HLG-MOS structure: the Supporting Standards Group, the Capability and Communication Group, and the Blue Skies Thinking Network. The Groups coordinate their activities between themselves, but also with the projects and other international activities in similar areas (for example at Eurostat, OECD, and the United Nations Statistical Division). The Group chairs report to and discuss issues with the Executive Board on a monthly basis. The start of a new modernisation group replacing the Sharing Tools Group (merged with Supporting Standards Group) was postponed.

### A.	Blue Skies Thinking Network

18.	The Blue Skies Thinking Network (BSTN) is the idea 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 how new innovations to production process 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 Strategic Framework can be considered. To allow for new innovations, space is also be given to out-of-the-box thinking.

19.	The Network is led by the innovation manager Barteld Braaksma from Statistics Netherlands and supported by UNECE. The core expert group consists of around fifteen members that met at least once a month. Additionally, several pitch talks sessions are organised to identify new topics. The pitch talks sessions are mini sprints in which members and outsiders can present briefly an idea or a project in the area of modernisation, which are subsequently discussed by the core group and invited experts.

20.	In 2021, the Network identified and discussed the following topics:

   a)	The power of network data: feasibility studies (Daniel Elazar, ABS);

   b)	COVID-19 hotspot joint biosecurity centre platform (Eric Deeben, ONS);

   c)	User research for official statistics (Eric Anvar, OECD);

   d)	Rapid survey system (Branko Josipović, SORS);

   e)	From experimentation to implementation in official statistics (Kate Burnett, Statistics Canada);

   f)	Microdata for understanding falling response rates (Gary Dunnet, StatsNZ);

   g)	Facebook-based COVID survey (Barteld Braaksma, Statistics Netherlands).

21.	The 'From experimentation to implementation in official statistics' activity (topic (e) above) led to organization of a session on this topic at the ISI World Statistics Congress by the

BSTN members. Feedback and further discussions led to a project proposal for the 2022 HLG-MOS work programme: 'Meta Academy for the Modernization of Official Statistics'. Based on a topic discussed in 2020, the BSTN developed another project proposal (in collaboration with the HLG-MOS Supporting Standards Group): 'Data Governance Framework to Achieve Data Interoperability'. As a result of 'The power of network data feasibility study' (topic (a)), ABS, ONS, Statistics Netherlands and a few universities are now exploring collaboration on a Supply Chain Network Disruption Model. The joint biosecurity centre platform (topic 20 (b)) and a rapid survey system (item (d)) topics will be discussed further in 2022.

22.   The HLG-MOS Executive Board oversees the work and can request assessment of topics that might be relevant for the modernisation of statistics. The group continues to be ready to evaluate activity and project proposals submitted from the statistical community. More details can be found on the Blue Skies Thinking Network wiki page.

B.   Supporting Standards Group

23.   The Supporting Standards Group is responsible for the maintenance and development of models and standards developed by the HLG-MOS such as GAMSO, GSBPM, GSIM and CSPA[3]. The goal of the Group is to develop, enhance, integrate, promote, support, and facilitate implementation of the range of standards needed for statistical modernisation. The Group has been chaired by Zoltán Vereczkei from Hungary and supported by the UNECE secretariat. In 2021, five task teams worked under the Group. The main Group consists of twenty experts from twelve NSOs and four international organisations. Many more colleagues collaborated in the task teams. Plenary group meetings took placed monthly and, in addition, the five Task Teams met on a frequent basis.

24.   The main activities and outcomes of the Group and its Task Teams in 2021 were:

a) **Linking GSBPM and GSIM**: an integrated view for GSBPM and GSIM was created. This helps to better understand the relationship between GSBPM and GSIM by describing for GSBPM sub-processes the input and output specifications in terms of GSIM objects. It is an important tool to move forward with the implementation of ModernStats models (especially GSIM after GSBPM or the two models together);

b) **GeoGSBPM**: The Geospatial view of the Generic Statistical Business Process Model operationalizes the Global Statistical Geospatial Framework (GSGF) principles throughout the production process. The GSGF provides principles that supports the production of harmonized and standardized geospatially enabled statistical data. GeoGSBPM describes geospatial-related activities and considerations using the framework of the GSBPM;

c) **GSIM update**: it was originally planned to be a "soft update" to clarify ambiguous points and fix minor errors identified during other activities of the group such as on Linking GSBPM and GSIM and Core Ontology work. However, it was soon concluded that a more thorough review was needed. The activity will continue as a full GSIM Review in 2022.

d) **Core Ontology for Official Statistics (COOS)**: The work aims to develop a formal framework that allows interoperability, activation, and globally unique identification of the various ModernStats models (GSBPM, GSIM, GAMSO, CSDA). The core ontology suggests formal representations for the core concepts used in our domain.

---

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

The COOS Task Team has so far made available the ontology specification, a governance document, a URI policy, and OWL ontology (currently under expert review). Details are available on GitHub: https://linked-statistics.github.io/COOS/coos.html. Activity will continue in 2022;

e) **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. Many national examples have now been collected, which can help other countries in implementing GSBPM. The work will continue in 2022 with the main aim to identify common tasks to be considered for inclusion in the next version of GSBPM;

f) **ModernStats Governance:** The Group developed guidance on governance policies for the various ModernStats models and other outputs from the HLG-MOS related to supporting the use of these models. It clarifies revision frequencies, revision procedures, and revision principles as well as versioning principles and ownership;

g) **ModernStats Usage Survey**: regular surveys about the usage of the ModernStats models are undertaken. In the 2021 survey, 45 countries responded. Results show that GSBPM is widely implemented, while the use of GAMSO, GSIM, and CSPA is more limited. More results can be found here: https://statswiki.unece.org/x/rQixEw.

C. Capability and Communication Group

25. The Capability and Communication Group is responsible for aspects of human resources (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 2021 was shifted to the legacy of the pandemic.

26. In 2021, 46 colleagues from 18 organisations participated in the Group and its task teams and subgroups. The Group is led by Anna Borowska from Statistics Poland and Maria Hurley from CSO Ireland. The Group and its task teams and subgroups met on monthly bases or if needed, more frequently. The activities of the Group are grouped into three streams:

a) **Ethics Management**: the activity includes ethical leadership and covers both business ethics as well as data ethics. In early 2021, a short survey was conducted amongst NSOs to investigate if NSOs have already put in place policies related to the key principles of ethics management and ethical leadership. It was followed by second survey to collect information about the existence of practices on ethics management and ethical leadership both on organisational processes (business ethics) as well as production and research (data ethics). Results are available from the report on survey ethics management.

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 a guidelines document on the role of brand and reputation management, marketing, and crisis communication in meeting the challenges facing modern national statistical organisations and international statistical agencies. The guide is accompanied by use cases that will be shared on the HLG-MOS wikis. All material will be available from https://statswiki.unece.org/x/cRGxEw (to be available in late February 2022).

c)   **Future work, future workplace, and future skills:** the pandemic has strongly accelerated changes in how we work and what skills we need. For any NSO in a transition to blended working, the task teams under this stream devised three Employee, Manager and Employer Toolkits to serve as a framework to structure local and individual policies and solutions and to adapt to local and individual needs. They are designed to examine considerations arising in the transition to a Blended or Hybrid Working Model and to explore the challenges, needs and opportunities for the Employer, the Employee, and the Manager. They help to raise questions that require solutions, identify common areas of concern or needing attention and suggest high level actions to help, and finally, they serve as internal conversation starters.

D.   Machine Learning Group 2021

27.   The Machine Learning Group is the continuation of the HLG-MOS 2019-2020 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 research to modernise official statistics, building capacity in machine learning and sharing knowledge. It is community driven, and all activities are led by members. It is open to all official statistical bodies and accessible to various levels of expertise. Resources are available to the wider community as the aim is to offer open, shareable, and easily accessible resources to the community and facilitate machine learning capacity building for official statistics.

28.   Nearly 250 members from over 30 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.

a)   **Work Stream 1 - From Idea to Valid Solution**: The pilot studies are conducted to assess the added value of ML in various thematic areas: coding and classification, edit and imputation the use of imagery data, modelling and route optimization. A study conducted on the replication experience highlighted that benefits of sharing theses ML projects;

b)   **Work Stream 2 - From Valid Solution to Production**: The work stream explored the issues around the operationalization of machine learning solutions and consists of three activities from IMF and INEGI (Mexico). Reports are available from the ML wiki space: https://statswiki.unece.org/x/eAB-EQ.

c)   **Work Stream 3 - Ethical Consideration in the Use of ML for Research and Statistics**: This high-level guidance explored ethical considerations associated with the use of machine learning techniques for research and statistical purposes. The guidance is not exhaustive, but aims to assist and support analysts, researchers, data scientists, and statisticians navigating the ethical issues surrounding machine learning-based projects;

d)   **Work Stream 4 - Model Retraining**: identified the circumstances under which an ML model should be retrained in order to maintain the predictive power and quality of the model;

e)   **Work Stream 5 - Quality Framework for Statistical Algorithm**: explores the dimensions of Quality Framework for Statistical Algorithm (QF4SA) that was developed under the Machine Learning project in a consolidated project to analyse an output based on a set of standard metrics and procedures.

29.   The results of the research collaboration in the five workstreams and the 18 research projects, as well as learning and training resources and other material is (or will be made) available from the communities' pages: https://statswiki.unece.org/x/eAB-EQ. The work has also been shared at "Coffee and Coding" sessions organized by the Group, which served to share knowledge as well as to get feedback.

IV.   MEETINGS AND WORKSHOPS

30.   In response to the COVID-19 pandemic and staff shortages at UNECE, the Executive Board and the UNECE secretariat decided to change the cycle of expert meetings and workshops to 24 months. In 2021, the following expert meetings and workshops were held:

a) **HLG-MOS meeting** (29 January, online): As the customary HLG-MOS meeting the weekend ahead of the Statistical Commission in New York could not be held, the Chief Statisticians that are member of the HLG-MOS met online. The work programme was approved, and the Chief Statisticians very much appreciated the output that had been produced in 2020. All participants were thanked for the great work that was produced, despite the very challenging circumstances. 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 priority topics of the HLG-MOS. The Executive Board was further requested to make proposals for a new Modernisation Group and for project and activity proposals for 2022;

b) **Expert Meeting on Statistical Data Collection (27-30 September, online):** This online event was attended by 151 participants from 30 countries. It included 32 presentations. These were allocated and presented in four substantive sessions: (i) Adapting Data Collection in a Crisis Context; (ii) Managing and Modernizing Data Collection; (iii) Multimode and Integration of Data Sources; and (iv) Respondents Care. Small group discussions identified several areas for future work, for example new techniques and IT for automated collection (machine-to-machine, system2system, APIs, web portals), multi-dimensional aspects of quality, and adaptive survey design. All abstracts, papers, presentations, and other output from the workshop are available at the UNECE website (https://unece.org/statistics/events/DC2021);

c) **Expert Meeting on Dissemination and Communication of Statistics** (11-14 October, online): It was attended by 151 participants from 33 countries. Twenty-seven contributions were submitted. It consisted of the following sessions: (i) Digital age; (ii) Broadening audiences; (iii) Statistics explained; and (iv) Tailoring communication to specific audiences. During the small-group discussions, topics for future work were identified. The most favoured were 'How to engage with and recruit the younger, digital native, generations', 'How to measure success and the impact of our communication', 'Explore unconventional communication platforms/tools such as comics, TikTok and Spotify', 'Communication capabilities, skills, profiles, experts we need', and 'Strategies to tackle and anticipate disinformation'. Other topics identified and all meeting documents are available from the UNECE website: https://unece.org/statistics/events/DissComm2021;

d) **HLG-MOS Modernisation Workshop and Webinars** (15-19 November, online). The main event was attended by 148 representatives from 41 countries. Similar to this paper, the main aim of this workshop was to present the activities and output from the various groups, projects, and workshops under the 2021 HLG-MOS work programme. Another major goal was to present and discuss the plans for the 2022

work programme. All meeting documents are available from the UNECE meeting website (https://unece.org/statistics/events/HLG2021). As a side event, three Webinars were organized to share the work of the two 2021 HLG-MOS projects and the Machine Learning Community:

- **Synthetic Data Webinar** (17 November, online). The presentation of the Synthetic Data Guide was attended by 73 participants from 24 countries;

- **IPP Webinar** (18 November, online). The project outputs were presented as well as follow-up plans such as a consultation on secure private computing-as-a-service. It was attended by 70 participants representing 21 countries;

- **Machine Learning Group Webinar** (19 November, online). In this joint ONS-UNECE event, the work under the various works streams and the future work plans were presented. 279 participants came from 48 countries.

e) **Expert Meeting on Statistical Data Confidentiality** (1-3 December, Poznan, Poland): It was initially planned as an in-person meeting. However, due to the worsening of the pandemic, most participants had to join online. It was hosted by Statistics Poland and the Poznań University of Economics and Business and co-organized with Eurostat. In total 169 participants from 32 countries, ten international organisations, and sixteen universities. Thirty participants were able to attend in person. 35 innovative contributions and the work of the two 2021 HLG-MOS projects were presented in six sessions. The future ideas that received the most support were: 'Automated output checking: a dream or reality?', 'How to relate formal risk measures to the perception of risk', 'Reconsider the rules for frequency tables' and 'Suppression or perturbation?'. All abstracts, papers, presentations, and other output from the workshop are available at the UNECE website (https://unece.org/statistics/events/SDC2021).

V. OTHER ACTIVITIES AND OUTPUT

31. **Publications**: most of the output of the work done under the HLG-MOS is shared publicly on the HLG-MOS wikis and UNECE website. In 2021, two official UNECE publications related to the work on modernization were prepared. The outputs of the Strategic Communication Framework project were combined into a printed publication[4], while the outputs of the Machine Learning project were compiled into an electronic publication[5].

32. The work of the HLG-MOS is facilitated by a large number of wiki sites and web pages. In total, the UNECE secretariat is managing and maintaining over fifty different public and restricted wiki sites available for collaborative purposes or for sharing output from HLG-MOS activities. Various news articles and reports to intergovernmental bodies were prepared (e.g. for the Conference of European Statisticians or the United Nations Secretariat). Moreover, the UNECE HLG-MOS secretariat responded to various requests for publications, information, presentation of work or requests for assistance from within and outside the UNECE region. Several presentations on the work of the HLG-MOS were given at national and international events.

33. The work was also coordinated with other international organisations working on the modernisation of official statistics (e.g. Eurostat, OECD and UNDESA). Linkages were made

---

[4] https://unece.org/info/Statistics/pub/357507
[5] https://unece.org/statistics/publications/machine-learning-official-statistics

with international activities in similar areas through cross-membership between the various groups.

34.   Further information is available from through the <u>main HLG-MOS online portal</u>. The <u>UNECE Secretariat</u> can be contacted as well for further information.

* * * * *