

## High-Level Group for the Modernisation of Official Statistics: 2022 Work Programme

*SUMMARY - The annual work programme is mandated and approved by the thirteen Chief Statisticians that form the HLG MOS. The Executive Board oversees the implementation and steers the work on a continuous basis, while UNECE provides the secretariat support for all work. In 2022, the work program consists of the following:*

*I. Three projects:*

- a. Input Privacy-Preserving Techniques (IPPT) phase 2*
- b. Data Governance for Interoperability Framework Project*
- c. Meta-Academy for the Modernization of Official Statistics*

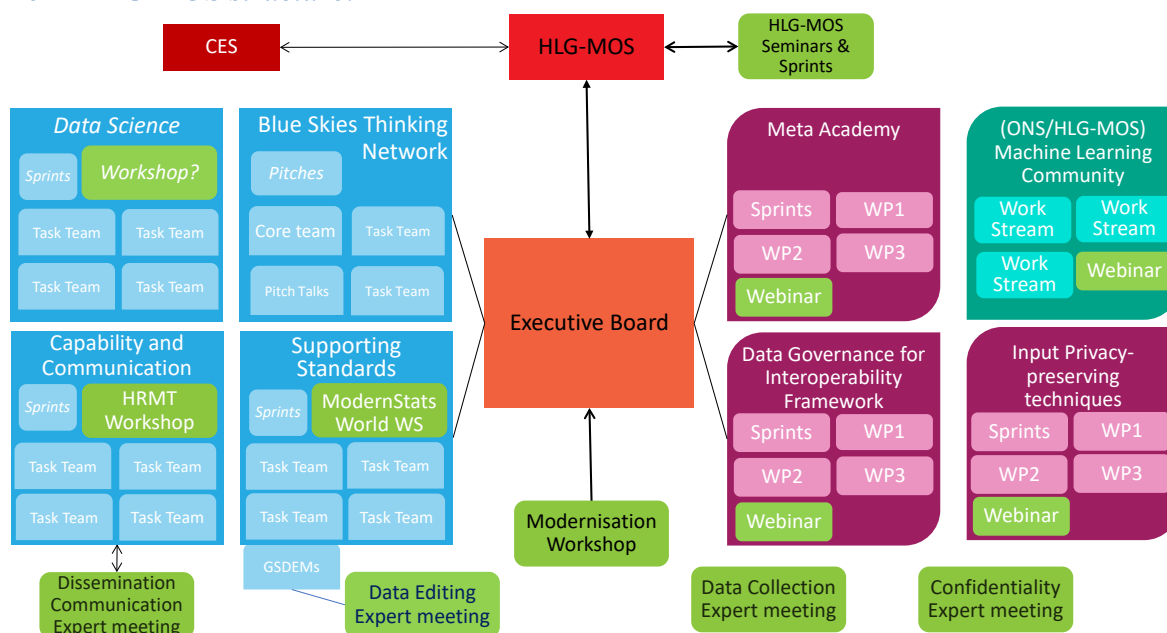
*II. Three groups, one network and a community:*

- a. Supporting Standards Group*
- b. Capability and Communication Group*
- c. Applying Data Science and Modern Methods Group (new)*
- d. Blue Skies Thinking Network*
- e. ONS/UNECE Machine Learning Community*

*III. Three workshops and three expert group meetings:*

- a. ModernStats World Workshop*
- b. HLG-MOS Modernisation Workshop*
- c. Workshop on Human Resources Management and Training for statistical organisations*
- d. Expert Meeting on Statistical Data Dissemination and Communication*
- e. Expert Meeting on Statistical Data Collection*
- f. Expert Meeting on Statistical Data Editing*

*2022 HLG-MOS structure:*



## Introduction

1. The High-Level Group for the Modernisation of Official Statistics (HLG-MOS)<sup>1</sup> provides a collaborative platform for experts in statistics to develop strategies and solutions in a flexible and agile way. The work of the HLG-MOS is open to all who are willing to contribute to the advancement and modernization of statistical concepts, standards and business processes. HLG-MOS must ensure that it remains adaptive to the changing environment and shifting landscape of challenges and opportunities. Its strategic vision is therefore regularly updated and new priorities are set. The annual work program endeavours to reflect these changing needs, while at the same time continues to support or further develops previous outputs. To guarantee an optimal outcome, a process combining top-down and bottom-up approaches for inputs is followed.

2. The HLG-MOS is a unique platform that has made several important contributions to the modernisation of official statistics. For example, GSBPM and GSIM were developed and are now global standards. GAMS0, CSPA, CSDA and GSDEM are other models that originate from the HLG-MOS<sup>2</sup>. Projects in areas such as Big Data, Data Integration, Strategic Communication and Machine Learning spearheaded the implementation of new technologies, methods, and other capabilities in statistical organisations. Due to its success, the HLG-MOS mode of working is now being replicated in other modernisation initiatives. Many colleagues have contributed to the output and every year, 200-300 colleagues actively participate in the programme activities while thousands of colleagues benefit indirectly by collaborating with members or by participating in the expert meetings, workshops and webinars organized under the HLG-MOS work programme<sup>3</sup>. As experiences and all output developed under the HLG-MOS are made available for public use, the real impact is even larger and increasing continuously.

3. All members of the Conference of European Statisticians (CES) are invited to submit project and activity proposals for consideration by the HLG-MOS. Proposals should be aligned with the mission, long-term vision and short- and mid-term priority areas identified by the HLG-MOS. The Blue Skies Thinking Network (BSTN) can also be requested to evaluate and strengthen ideas and proposals. The Executive Board of the HLG-MOS provides feedback and selects proposals that will be considered as projects or flagged for other types of follow-up. During the annual Workshop on the Modernisation of Official Statistics at the end of November, these proposals are presented and discussed by experts working on national and HLG-MOS modernization activities. Through small group discussions, project proposals are evaluated and ranked, and suggestions are made for follow-up on activity and other proposals. The Executive Board then discusses and further refines the proposals and the selected projects and activity proposals are finally submitted for endorsement by the HLG-MOS.

4. This document outlines the work program that was created as a result of the discussions at the Workshop on the Modernisation of Official Statistics held online on 15-16 November 2021 and includes subsequent further refinements by the Executive Board. This paper provides summary information to the community on the work that will take place to enable their participation in the activities that are of greatest benefit to them.

---

<sup>1</sup> Refer to Annex 1 for the structure of the HLG-MOS.

<sup>2</sup> Generic Activity Model for Statistical Organisations, Common Statistical Production Architecture, Common Statistical Data Architecture, and Generic Statistical Data Editing Model.

<sup>3</sup> The impact and reach is also apparent as the HLG-MOS output dominate the list of top consulted content on the wiki and web pages of the UNECE Statistical Division.

## HLG-MOS Projects 2022

5. Exceptionally, three projects were selected by the HLG-MOS for 2022<sup>4</sup>: i) Input Privacy-Preserving Techniques project extension, ii) Project proposal for Data Governance Framework to Achieve Data Interoperability and iii) Meta Academy for the Modernization of Official Statistics.

6. The three projects aim to improve data interoperability, secure sharing of data, and training and capability development. These areas of focus are enablers for new national data stewardship roles, for the sharing of data, tools, and capabilities and for jointly creating learning and training needed for continuous modernisation and implementation of new tools and technologies. The outcome of the projects will allow accessing external data sources, for example held by other national and international agencies, and integration of data sources without direct access to sensitive information. It will provide guidance to setup a framework to access, share and combine all data produced throughout the organisation in an easy way and with the metadata needed to understand the data using common standards. Additionally, it will show how training initiatives can be shared and co-created to develop the capabilities needed to do so.

### Input Privacy-Preservation Techniques Project

7. The 2021 project on input privacy-preserving techniques (IPPT) proved that such techniques can play an important role in making external data sources accessible when there are confidentiality concerns. This approach allows for analysing or integrating external data sources and producing statistics without revealing the microdata to the external partner. At the end of 2021, it was concluded that a continued collaboration was needed to further develop the performed experiments and to better understand the environment that is required for IPPT as well as to get a better understanding of the methodological challenges. The community will benefit from developing a handbook with implementation guidelines and examples of practical applications. The following work packages have been identified:

- **WP1: Deepening practical experiments:** with real examples and more complex case studies such as cross-NSO and external data providers cases to access resources that are hitherto inaccessible due to privacy constraints.
- **WP2: Document use cases and provide guidelines for implementation:** based on the 2021 and additional experiments that will be conducted under WP1, guidelines for successful implementation of specific techniques in practical use cases, will be developed. This includes guidelines on the environment and infrastructure needed for successful implementation, including the requirements of the (external) data provider. A broader perspective on what is needed to successfully implement the use cases will be obtained through a consultation survey.
- **WP3: Create user community:** To continue to share experiences and test remote access and integration of sources held by different partners, as well as to share the lessons learned and to train other colleagues and assist NSOs in implementing the techniques, a community of experts on input privacy preservation techniques will be created.

8. The work of the project will be coordinated with the UN Global Working Group Task Team on Privacy Preserving Techniques and activities of Eurostat in this area. The objective is to achieve a broader level of engagement, including with academia and the private sector.

---

<sup>4</sup> In 2021, the two projects were supported by in-kind project managers from Statistics Canada and Statistics Netherlands. As a consequence, pandemic in 2020 and 2021, limited resources were needed and the HLG-MOS Trust Fund (2019-2022) has sufficient resources left to cover financial needs for three projects while the UNECE unit supporting the work is now again fully staffed and able to provide the secretarial support required.

### Data Governance for Interoperability Framework Project

9. The main goal of the Data Governance for Interoperability Framework project is to produce a document describing a reference framework containing the main elements needed to implement a governance program where the focus is on achieving data interoperability in National Statistical Offices. This framework will provide the ability to create, exchange, and use data where its meaning and context is preserved independently from a given system or a set of systems. It is an important first step for creating a Statistical Data Governance Framework<sup>5</sup>.

10. The objective is to increase the value of the statistical information by establishing connections between the data from different domains. This will allow comparison, reuse, and integration of data from different sources. By creating a way to effectively reuse information and tools, costs will be reduced. The capacity to create a new platform of systems and tools that enhance the analysis and dissemination of statistics, will improve the information products and services. In this way, the framework can meet the emerging and more complex needs of our users while at the same time improve data and metadata quality by making it more transparent, manageable, and comparable.

11. The project will take a realistic approach to obtaining a solid result and focus on data interoperability. The aim is to create a document describing a reference framework to achieve data interoperability that will include recommendations and guidelines around the following four components:

- Establishing a data governance body inside the statistical organisations to agree on conceptual decisions that are necessary for data interoperability.
- Structuring and using the existing models and standards produced by the ModernStat program and by other relevant projects such as SDMX, DDI, and the Semantic Web as components of this framework.
- Identifying core aspects that need to be covered during the phases and sub-processes described by the GSBPM to ensure that reliable data interoperability will be achieved by the statistical projects.
- Describing how to implement transversal platforms supporting data interoperability and set up concept-driven integrated information systems that have shared concepts and classifications, provide common views, and are supported by common reusable information services, applications, and tools.

12. For each component, a task team will be set up. The project will be aligned to the earlier work developed under the HLG-MOS such as the ModernStats models (GAMSO/GSBPM, GSIM) and other important standards widely used in the statistical community (DDI, SDMX) and the Common Statistical Data Architecture and other outcome of the 2017-2018 Data Architecture project. The HLG-MOS Supporting Standards Group and the SDMX working group will be consulted where relevant. The participation of experts on data governance, ModernStats Models, SDMX, and DDI, as well as of methodologists, statisticians, and ICT people, will be fundamental to develop each of the sections of the document.

### Meta-Academy for the Modernization of Official Statistics

13. The purpose of the Meta-Academy for the Modernisation of Official Statistics is to remove barriers to co-creation of training and reuse of content at an international level. This has the power to

---

<sup>5</sup> Initially, the project proposal referred to Statistical Data Governance Framework, but given the focus on data interoperability, the name was changed to better reflect the scope.

unleash the creation and use of open digital assets and training material to boost the skills at National Statistical Office (NSO) that are necessary for modernization. This project intends to raise the standards of virtual learning on topics necessary for the modernization of statistics but that are missing or that are inconsistent with that available from academic, commercial, or in-house sources. The meta-academy project sets out to create a benchmark to better map existing initiatives and offerings in order to better coordinate efforts, reduce duplication and fill in training gaps. This project will facilitate sharing of skills strategies, as well as catalogues of contents and pedagogical artefacts, and more generally good practices and standards in that space, so that scopes for reuse or co-creation in learning capabilities can be more easily and more systematically promoted and leveraged by all NSOs.

14. This project has three main work packages that target creation of synergies between existing initiatives, identification of gaps in topics and methods development of training and experimentation with co-creation and training in a virtual world:

- **WP1 benchmarking exercise** among existing initiatives for shared good practices, to repurpose existing available materials and identifying co-creation to fill any gaps. This exercise will start off with brainstorming sessions and presentation series to scan existing initiatives in other international organisations, NSOs, academia and the private sector with the aim to identify how much of the landscape we would like to analyse in this exercise. This next step will be an in-depth analysis of a selection of relevant initiatives as a benchmark to then co-design a final framework. The full delivery of the WP1 may be subject to the onboarding of external expertise, the precise scope of which will be better known in the early stages of the process when participants will decide on the scope and cost and their contribution.
- **WP2 co-create capacity building content:** an experiment to co-create capacity building content by NSOs on the topic of Git and version control. This is an area that broadly impacts modernization of statistics from basic programming to advanced data science, however it gets little attention in international forums. Through a series of workshops and online collaboration a “git manual” will be completed capturing best practices in the management of algorithms and code bases by statisticians and data scientists in the official statistics context. This collaboration will involve experts leading the development of algorithm design and sharing good practices in their organisations. The expected outputs for this work package include co-creating a manual covering usage and best practices in git and version control. This will form the basis for an HLG-MOS pilot course on git and version control for official statistics. In addition, this experience will provide real-time co-creation lessons learned on frameworks, opportunities and methods developed from work package one.
- **WP3 finalizing the framework for virtual learning, co-creating and reusing content:** formalizing concepts, skills mapping, learning outcomes and journeys and content creation best practices. The work package will also address the best delivery method for content and co-creation such as a central platform or network.

15. There is an existing starting point with the current collaboration between Statistics Canada and OECD on creating a git manual. This project will consolidate and provide continuity to efforts across the HLG-MOS, UNECE and other associated groups, such as the UN Big Data Group, on creating content and frameworks for virtually training of key aspects necessary for the modernization of statistics. This project will leverage existing frameworks such as the HLG-MOS Capabilities Development and Training Framework<sup>6</sup> and the UNECE Methodology Framework as well as existing

---

<sup>6</sup> <https://statswiki.unece.org/download/attachments/218890322/Statistical%20Training%20Framework%20-%20January%202018%20.pdf>



networks such as the Blue Skies Thinking Network, and other frameworks developed at the national and international level<sup>7</sup>.

16. Additional details on the three projects can be found in Annex 2. The work of the HLG-MOS projects is normally led by project managers that are either assigned by a member of the HLG-MOS (in-kind) or contracted by UNECE using the HLG-MOS Trust Fund. Projects can also be led by a project manager or a substantive lead. Work Package, task teams or other specific activities are normally led from chairs chosen from the project members. The UNECE wiki and web conference are used to monitor the progress of the projects and UNECE provides additional administrative and secretariat support.

## Modernisation Groups

17. Modernisation Groups are organized around common themes and the selected activities are normally organized through Task Teams. The groups select an overall chair and, in most cases, chairs are assigned to the various task teams. The chairs of the groups provide monthly updates to the Executive Board. Groups have frequent (monthly) virtual plenary meetings and Task Teams typically meet virtually as well at least once a month. The groups may have additional meetings or organise sprint workshops to expedite the work. The UNECE wiki and other platforms are used to collaborate and coordinate the work. Secretariat support is always provided by UNECE.

18. The HLG-MOS has mandated the Executive Board to regularly reflect on the structure and the areas of work of the Modernisation Groups and to evaluate whether the groups are aligned with the key priorities identified by the HLG MOS. It had been decided that in 2021, the Sharing Tools Group would be merged with the Supporting Standards Group. Due to the uncertainty about the COVID-19 pandemic and lack of resources at the Unit supporting the work of the HLG-MOS, it was decided not to set up a new group 2021. As offices are back to (the new) normality and UNECE is now fully staffed again, for 2022, a 4<sup>th</sup> modernisation group could again be established.

19. The planned activities of the Modernisation Groups are briefly described below. More detailed descriptions of all activities can be found on the HLG-MOS site: <https://statswiki.unece.org/x/lwF-EQ>.

### Applying Data Science and Modern Methods Group

20. The HLG-MOS tasked the Executive Board to setup a new group if the resources and the COVID-19 pandemic permits. After considering various options, there was unanimous support for setting up a group on applying data science and modern methods.

21. The group will consider and make proposals yearly on how to develop, implement, promote, support, and facilitate the implementation of data science and modern methods initiatives needed for statistical modernization of business processes. The proposals will be largely driven by the interests expressed across the HLG-MOS community. The group will coordinate its work with related activities under the HLG-MOS, such as the Machine Learning Community or the Blue Skies Thinking Network as well as with other major international initiatives.

22. The main activities and outputs the group will be working on will be decided early 2022 in a scoping exercise by the group members and in coordination with the Executive Board. Examples of potential activities that will be considered are:

---

<sup>7</sup> For example, activities implemented under The Task Team on Training, Competencies and Capacity Development of the UN Committee of Experts on Big Data and Data Science for Official Statistics, and the ESSnet Trusted Smart Statistics – Web Intelligence Network, or Statistical Training Framework based on the GSBPM of CSO Ireland or the Statistics Poland Data Science Academy and other national initiatives.

- Identifying data science initiatives and new methods that support modernising existing processes.
- Identifying new needs that could be addressed by NSOs by means of data science initiatives.
- Providing support to implement such data science initiatives and modern methods including:
  - i. Developing and publishing supporting materials to help implementing and use, such as case studies, good practices, etc.
  - ii. Organising workshops and training to promote and ensure consistent use of the HLG-MOS supported data science initiatives and modern methods;
  - iii. Managing the periodic reviews of the new data science initiatives and methods to measure their impacts.
- Following developments in data science and statistical methods where relevant to the modernization of official statistics and of interest to the HLG-MOS community.
- Providing input as needed to HLG-MOS projects through the Executive Board.

23. The target composition of the group will be a balanced mix of experts in the field of data science and practitioners/methodologists already involved in the setting up of innovative processes and infrastructure for Official Statistics. The list of members will be confirmed by the HLG-MOS, taking due account of the representation of different regions, specialisms, and stakeholder groups. The Group will collaborate closely with the relevant activities of the ONS/UNECE Machine Learning Community, the Blue Skies Thinking Network as well as with the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD) and other related international activities.

### Capability and Communication Group

24. The Modernisation Group on Capability and Communication focuses on the organisational changes and the communication challenges necessary to support modernization in statistical organizations. The COVID-19 pandemic has put a lot of strains on the HR and communication departments of NSOs. From the start, the group has been adjusting their work programme into assisting statistical offices to cope with the changing working arrangements and the need for more extensive internal and external communication. The group will setup task teams to consider the following substantive topics during their monthly web conference meetings:

- **Future of work toolkits** – The "next normal" at the workplace is conditioned by a substantial evolution regarding work from home (WFH) and flexible work policies. In order to help countries to transition to new working arrangements, in 2021 a framework of three toolkits was created, for manager, employer and employee. The next step will be to create an on-line repository of tools, examples, and other material to assist NSOs in implementing new flexible working arrangements and to provide guidance in how to use the toolkits. It is not only about providing potential solutions, but the goal is also to raise awareness of issues that might arise and need to be addressed in the next normal working arrangements.
- **The Job of the Future** – New generations of employees look different at life, work, and their job. It is and will become harder for NSOs to attract and retain new people to by simply offering them a job and a salary. At the same time, the nature of the work of NSOs is changing. Statistics need to be created faster, be timelier and answer questions of immediate concern. Cross-departmental and multi-disciplinary/multi-generational teams are often needed to achieve this. The team will explore research in this area and investigate what this means for NSOs and develop solutions, best practices from NSOs and other institutions and academic research to provide guidance and support towards these changes.

- **Reaching Young People** – The purpose of the activity is to gain insight, develop best practices and universal standardized methods on how to reach young people more effectively. NSOs have an official task to inform policymakers and citizens to stimulate data use and statistical literacy. Young people are the future generation and often use non-traditional (media) sources and speak their own language. With current communication strategies NSIs are increasingly struggling to reach them effectively. This activity aims to gain more insights on this group and facilitate NSOs to engage young people in statistics in the following three dimensions: young people as a data source, as a data user and as an employee.
- **Ethics** – The activity continues the work started in 2021. Surveys held and further in-depth analysis by the task team, identified that ethics management to improve performance and the organisation in general and ethics management of statistical production and dissemination processes (data ethics) are areas of specific interest. Respondents would welcome setting up a strategic framework, presentation of case studies, a collaboration platform and training programmes. The goal of the activity is therefore to define a common vocabulary, a framework, and give concrete suggestions (a sort of Handbook) to support NSOs’ leadership in real-work-type situations and how to deal with potential behavioural dilemmas in areas such as data processing, personal data protection, and conflict of interests.
- **Strategic Communication Framework follow-up** – After expanding the Strategic Communication Framework (SCF) in 2021 to include the role of brand and reputation management, marketing and crisis communication in meeting the challenges facing modern NSOs, the task team will continue to add case studies in other areas identified at the Expert Meeting on Dissemination and Communication of Statistics, for example, how to measure the impact of our communications or strategies to tackle and anticipate disinformation. The group will also contribute to communication aspects of the other activities of the modernisation committee.

25. The Capability and Communication Group will organise the (biennial) **Human Resources Management and Training workshop** in 2022 and contribute to the organisation of the Expert meeting on Dissemination and Communication of Statistics. The group will furthermore collaborate and create synergy with the 2022 HLG-MOS Meta Academy for Official Statistics project and the new group on data science. The group is jointly chaired by Ireland and Poland.

### Supporting Standards Group

26. The Supporting Standards Group provides support for the implementation of the “ModernStats” models (GAMSO, GSBPM, GSIM and since 2021, CSPA) through a range of activities which include development, promotion, and maintenance of the models. As the HLG-MOS is the custodian of these models that have a global reach, continued support is essential. In 2022, the Supporting Standards Group prioritized the following activities:

- **GSIM review** – In mid-2020, the Group created a GSIM task team for the “soft updates” of the model. However, as the work of the task team progressed, it became apparent it would require more than “soft updates”. The work on Linking GSBPM and GSIM and on Core Ontology, in particular, have provided new perspectives on how GSIM, should be used in conjunction with GSBPM and GAMSO. To make all ModernStats models work better together and to properly implement GSIM, several key GSIM objects and relationships need to be re-assessed and modified, leading to a full model revision. The task team will be divided into sub-teams that will work on GSIM Groups in parallel and the new version will be shared for feedback and a global consultation at the CES at a later stage.



- **Core Ontology for Official Statistics phase 2** – In 2021, Core Ontology task team delivered key outputs such as the ontology specification, a governance document, a URI policy, and an OWL ontology. In order to support the community to use these outputs and because there is still a lot of potential in further developing the model, the task team will focus on supporting the use of the outputs by providing use cases, the operationalization into an RDF database, establishing more links to external models and vocabularies (DDI-CDI etc.) and providing user guides and explaining and promoting the ontology.
- **GSBPM Task** – The task team started in September 2021. The team will complete the review of examples from countries that have added tasks to their national version of GSBPM and develop a proposal for the community for the task level for all GSBPM phases and sub-processes under a common set of principles (e.g., coding, granularity, minimality, description format). The task team also identifies issues, proposals that will feed into the future revision of the GSBPM.
- **SDMX-DDI-GSBPM** – The complex landscape of standards that have different scopes and strengths often confuses users as to which standards are used for which stage of production process. Complexity has increased with newer versions offering more options to the organizations on how to use them in the production processes individually or together. The task team will develop a high-level view of where and how SDMX and DDI can be used across data life cycle based on GSBPM and identify SDMX and DDI artefacts used for GSBPM sub-processes and provide real-life use cases how it can be used in the production process and statistical life cycle. The work of the Linking GSBPM and GSIM will provide a basis to link GSBPM and SDMX/DDI. Collaboration will be sought with the SDMX and DDI community.
- **GSBPM overarching processes** – The Linking GSBPM and GSIM activity in 2021 identified several very important lessons learned and inconsistencies with the models. To improve interoperability, additional work is needed on the GSBPM. This will build on earlier work on clarifying GAMS0 and overarching processes (OPs) in GSBPM. The task team will work on the breadth and depth of the OPs in the GSBPM and will complete their descriptions in the GSBPM and further break down the OPs at level 2 of the GSBPM (overarching sub-processes). This will result in an easier way to further clarify the relationship between GSBPM and GAMS0 and the outcome will directly feed into the future GSBPM revision.
- **CSPA capacity building** – CSPA has too often remained disconnected from other ModernStats models in the past, having CSPA work and people under the same umbrella will be helpful to have clearer idea on interrelation of the models and their integration. The purpose of this activity is twofold: to (re)build the community behind CSPA by involving expertise and build CSPA capacity to provide support for the whole community and for the future development of CSPA. Processing all the available material and defining a roadmap for the future development of the model are necessary prerequisites for CSPA support.

27. The group will also collaborate and coordinate work with the Statistical Data Governance Framework project and start an internal discussion on how CSDA is interrelated with other ModernStats models and can be reviewed considering recent developments. The Supporting Standards Group will organise the 4th **ModernStats World Workshop** in 2022. The workshop aims to progress work on development and maintenance of the ModernStats models and provide a ModernStats models user platform. The group is chaired by Hungary.

### Blue Skies Thinking Network

28. The Blue Skies Thinking Network is the ideas factory of the ModernStats community. It is led by an innovation manager and consist of a core group comprised of around twelve members from

various NSOs and International Organisations. The core team aims at having in-depth as well as broad knowledge of innovation related aspects to facilitate the development of the work and of sufficient heterogeneity to allow for a variety of expertise and views. The core group draws from a flexible pool of resources to assist in the evaluation of proposals.

29. The group members and invitees hold regular pitch talks to test new ideas. The network can also setup temporary activities to follow-up on promising topics or project proposals that were not selected. In 2022, the Network will elaborate further on these topics (other topics might be added throughout the year):

- **Community on metadata and data virtualization**
- **Digital twins**
- **Follow-up of the Joint Biosecurity Centre and Rapid Survey System**
- **Microdata for understanding declining response rates** (postponed from 2021).
- **Mobile survey data collection for climate change.**
- **Nowcasting.**

30. During the year, the BSTN will identify new potential topics and at any time ideas can be submitted by colleagues from the community of official statistics for evaluation. HLG-MOS members will continue to bring the Network to the attention of their national innovation groups and actively stimulate idea-generation for proposals to be submitted to the Network. The Network will actively engage with the statistical community to identify potential topics that merit further consideration. The BSTN will collaborate with the other modernisation groups to further identify and evaluate new topics. Additionally, it will coordinate with the new group on Applying Data Science and Modern Methods on overlapping and joint activities. Depending on the interest of members and expertise needed, sub-groups might be setup around specific topics for a quick scan or a more detailed analyse.

## ONS/UNECE Machine Learning Community 2022

31. After a very successful first year, thanks to the leadership and ongoing commitment of ONS, the community will continue in 2022. Support for this community is largely provided by a dedicated team at the data science campus of ONS. Other countries are invited to consider providing additional support. The ambition for 2022 is to be even bolder, bigger, and more collaborative. A strong focus will be on moving from proof of concept to production. Ethics and Quality of training data are other key areas. There will be strategic guidance from lead data scientists.

32. The 2022 setup and activities were steered by an evaluation and future direction survey and from additional proposals received during the 2021 ONS/UNECE Machine Learning Community Webinar. The 2022 ML Community will have the following structure and planned activities:

### *Knowledge Sharing*

- **Global Data Squad:** will bring together members of different countries in an active research collaboration that will conduct research to advance the statistical community's understanding of how ML technologies can be used. It brings together members of different countries in an active research partnership to network, exchange resources and good practice, and share ideas and experiences and provides hands on experience that builds the capacity of ML Group members to apply ML technologies in their day-to-day work. Project teams are already planned on AIS Shipping Data and on an ML Training Curriculum. Additional topics will be decided by the members.
- **Workstreams:** will provide a platform for knowledge exchange, research development and networking focussed on specific areas of the machine learning process. They will enable

members to share new research findings, get and provide feedback and allows members to keep abreast of the latest developments as well as to get access advice and expertise, and exchange ideas and experiences.

- **Hackathon(s)**: to produce solutions to common problems in an agile and challenging setting.

#### *Capacity Building*

- **Monthly Forum**: for the community to share new research and knowledge, connect and network, and hear strategic views.
- **Lightning Talks**: to introduce a subject, project, or idea and to get feedback on it.
- **Data Science Leads**: regular speaking slots at the monthly forum for example from national data science leads

#### *Communications*

- **Coffee and Coding**: expert tutorial on technical aspects of ML processes
- **Exchange Scheme**: to exchange skills between ML community members
- **Online Community**: a platform for data scientists to exchange knowledge and ideas

33. All the above activities depend on the collaboration of members and organisations and the sharing of resources. Depending on whether additional resources will be obtained, a newsletter, website, library of training material, a Yammer network and external engagements at related conferences will be considered. The work will be coordinated with the Applying Data Science and Modern Methods group as well as with the Capability and Communication group (on ethics, capabilities and training) and with the Global Working Group (GWG) on Big Data and other international activities in the area.

## HLG-MOS Expert meetings

34. As in the past, under the auspices of the HLG-MOS several Expert Meetings<sup>8</sup> will be organized in several substantive areas. By decision of the HLG-MOS Executive Board, all expert meetings (and workshops) will have a 24-month cycle for in-person meetings<sup>9</sup>. In the alternate year, online meetings can be organized. These can have a more specific focus on emerging issues and on related HLG-MOS projects and activities. As all meetings were online in 2020 and 2021, the intention is to have most meetings in 2022 in-person, if the pandemic permits.

35. The focus is always on innovative developments and modernization. Another key output of these events is the identification of areas for future work and collaboration between countries. The target audience for these expert meetings includes senior and middle-level managers and they are organised by the UNECE secretariat in collaboration with expert Steering Committees. Due to the continuing uncertainty with the COVID-19 pandemic, no meetings are planned for the first half of 2022. For the second half of 2022, the following meetings are being considered:

- Expert Meeting on Statistical Data Dissemination and Communication
- Expert Meeting on Statistical Data Collection
- Expert Meeting on Statistical Data Editing

---

<sup>8</sup> For internal reporting/programming purposes, Workshops were renamed to Expert Meeting from 2021 onward

<sup>9</sup> The Expert meeting on Statistical Confidentiality already had a 24-month cycle, while the Expert meeting on Statistical Data Editing had an 18-month cycle

36. These meetings are organized by Steering Committees and the programme will be aligned with the HLG-MOS activities where needed. The HLG-MOS Executive Board can also decide or the HLG-MOS can request to set up short, focused ad-hoc online workshops in emerging areas, for example, the Covid-19 response workshops organized in 2020.

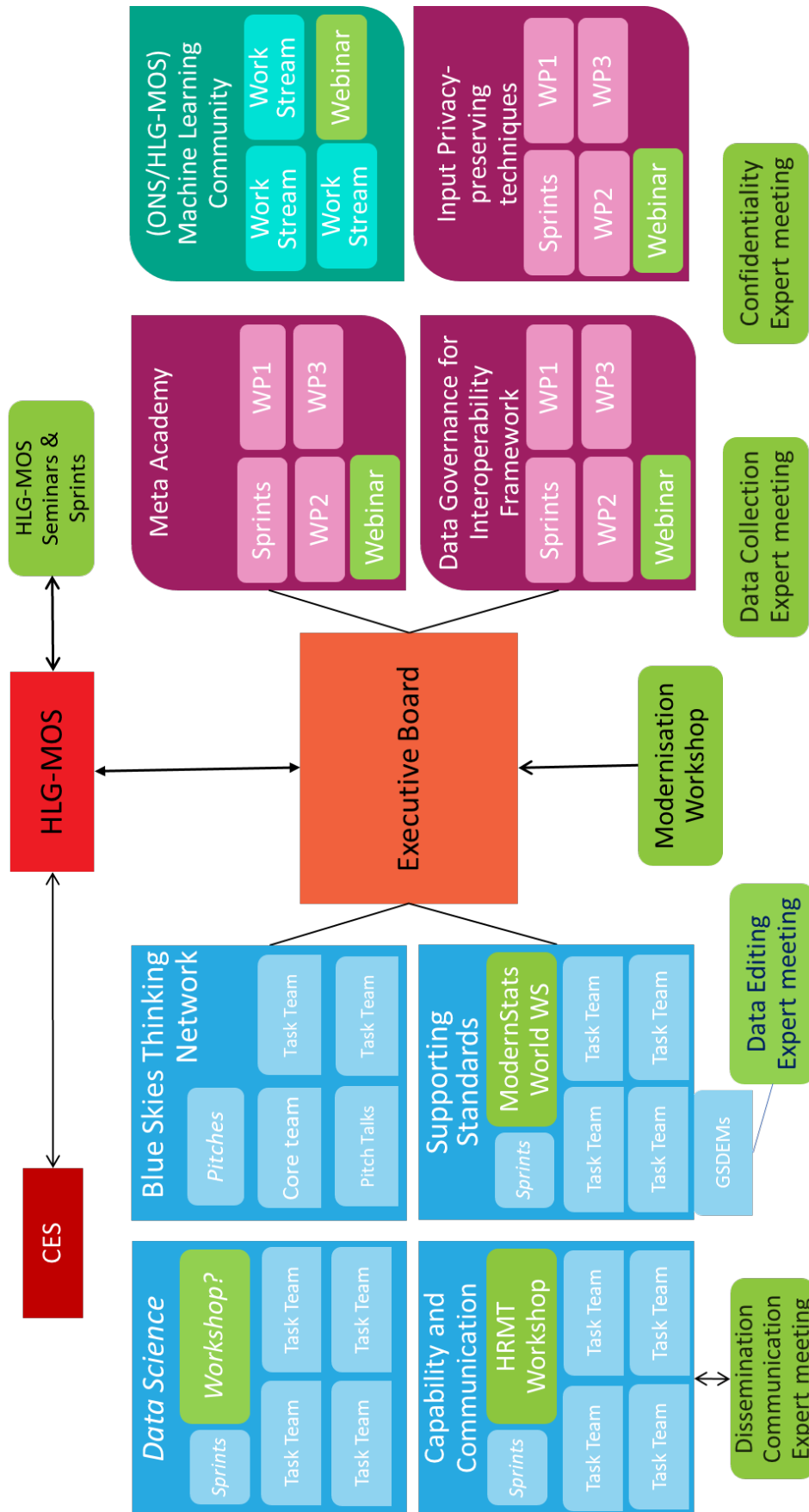
### Monitoring Progress, Coordination and Participation

37. The work of the Groups and Projects is reported on a monthly basis to the Executive Board of the HLG-MOS. The Executive Board discusses the modernization updates and evaluates the progress together with the chairs of the groups and the project managers in their monthly meeting. If needed, the work program is adjusted. The modernisation updates are made available to the wider public every two months at the [ModernStats wiki](https://statswiki.unece.org/x/QY0HBg) (<https://statswiki.unece.org/x/QY0HBg>).

38. There is also a conscious effort to ensure that all activities are continuously aligned and coordinated with other international initiatives like the programs under the European Statistical System and the UN Global Platform. This can be achieved by exchange of information, coordination, and collaboration and, in most cases, by achieving joint membership.

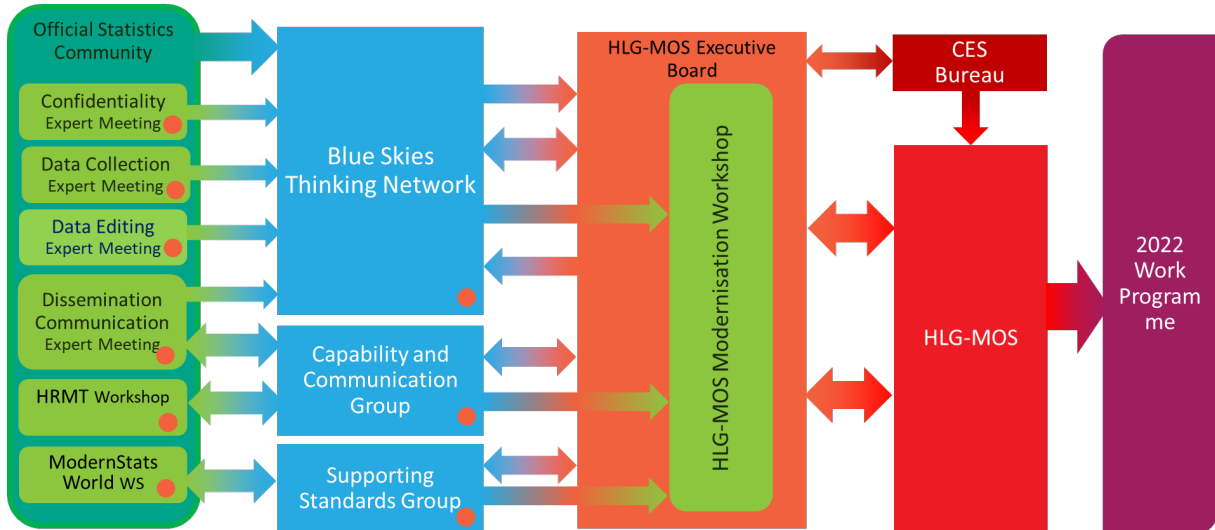
39. Participation in the activities of the HLG-MOS is open to staff of national and international institutes working in official statistics. Active participation contributions from academia and relevant public and private sector organisations that are willing to contribute to modernising official statistics, is welcomed, and actively sought.

### Annex 1: 2022 Structure of the HLG-MOS





## Annex 2: HLG-MOS Innovation Pipeline



## Annex 3: 2022 Project Proposals

### Business Case for Input Privacy-preserving Techniques (project extension)

<b>1 Purpose</b>	
Continuing the existing collaboration between the involved participating organizations to further explore and broaden the applicability of input privacy preserving techniques due to the need to become part of data ecosystems, use of private data between NSOs and, more generally, between organizations.	
<b>2 Project description</b>	
<p>The input privacy preserving technique project has been running for over a year now. A number of important milestones have been achieved during this time. A template was developed to describe use cases and this was then applied in practice. This has resulted in a common picture of what kind of use cases are involved. The described use cases have been generalized. Two prototype use cases have been jointly described and subsequently realized as a test. The experiences formed the basis for a public consultancy that has yet to start.</p> <p>In the discussion during the project it became clear that much more is needed to make input privacy preservation techniques suitable for application at NSOs. On the one hand, it concerns a joint deepening of the experiments carried out. On the other hand, an initial consultation on what is needed to make these use cases a success.</p> <p>The latter relates to the environment in which these use cases will run. The environment has to be intended not only within a single NSO but rather cross-NSOs and, more in general, cross-organizations.</p> <p>Another factor is that such an environment is expensive, so it can take a long time to come into existence.</p> <p><b>The project outputs will be:</b></p> <ol style="list-style-type: none"> <li>1. Next level of knowledge and experience what it takes to access resources that are hitherto inaccessible due to privacy rules.</li> <li>2. Documented use cases that transcend NSOs with pilot implementations.</li> <li>3. Community of experts on the theme of input privacy preservation techniques.</li> </ol>	
<b>3 Alternatives considered</b>	
<ol style="list-style-type: none"> <li>1. End the project this year and document the results of the project with advice on possible next steps. Then leave it to the NSOs/organizations involved to continue working together themselves, if there is interest.</li> </ol>	
<b>4 Expected Benefits</b>	
<input checked="" type="checkbox"/>	Reduced costs
<input checked="" type="checkbox"/>	Increased efficiency
<input type="checkbox"/>	Reduced risks
<input checked="" type="checkbox"/>	New capabilities to meet user needs
<b>5 Which key priorities in the HLG-MOS Strategic Framework does the proposed project relate to?</b>	
<input checked="" type="checkbox"/>	Take cost out of our organizations to reinvest in more value-added areas
<input checked="" type="checkbox"/>	Explore new areas collectively and leverage each other's' research investments in specific areas
<input type="checkbox"/>	Provide whole of government data ecosystems based on international standards, for better estimates in key policy areas
<input type="checkbox"/>	Renew our governance and operating processes
<b>Justification:</b>	
Exploiting the potential of new sources, which we do not have direct access to for privacy reasons, or the exploitation of sources in different countries, requires further research into techniques for preserving the privacy of inputs. The project has given a lot of insight that much is still needed.	

What we need is a Secure Private Computing-as-a-service to process information from sources that may not be shared among themselves. Such an environment requires much more than technology. Public trust is needed. While certainly not the goal of the project, input from legal experts, privacy advocates, civil rights activists is a possible by-catch of the public consultation. At the same time, the use of these types of techniques, where sources are no longer accessible, poses methodological challenges that we have not seen before.

An extension is highly recommended in order to be able to continue the collaboration between experts that originated in the UNECE project. And it contributes to the UN Task Team for the UN Handbook by providing practical examples.

#### **6 How does the proposed project relate to other activities under the HLG-MOS?**

This project brings one long-term goal of the HLG-MOS closer to modernize the production of statistics when it comes to using resources instead of surveying.

It also potentially makes it possible to produce statistics that are not yet possible, such as trade between countries that requires resources from both countries.

The aim of the project is not to replace other efforts, such as those of the UN Privacy Preserving Techniques Task Team, but rather to collaborate with them in activities that leverage work of both teams such as providing use cases and concrete requirements and use testing environment facilitated by the UN Privacy Preserving Techniques Task Team.

#### **7 Proposed timetables**

The first phase of the project focuses on continuing the existing tracks with deepening of the tracks:

- Private set intersection
- Private machine learning

The mini use cases are further elaborated. The track private machine learning also explores methodological challenges.

With the track Organize a public consultancy, the consultation is first completed and based on the results of the consultation it is determined which next steps are beneficial.

For the rest of the year, the working group will focus on those areas where more in-depth research is needed to make input privacy preservation techniques a success.

Participation of this group in the HLG-MOS workshops will yield activities that stimulate thinking as a community that shares things from the start. Participation of this group in the HLG-MOS workshops will bring activities that stimulate thinking as a community that shares things from the start.

#### **8 Expected resources and costs**

These are the expected resources needed for 2022:

- Experts in statistical methodologies, privacy preservation techniques, information management and information technology
- Space in UNECE's Wiki to facilitate community
- Virtual meeting facilities to support monthly meetings
- Resources to organize a physical sprint
- Resources to organize a physical workshop
- Support to the UN/ECE HLG-MOS Secretariat to help coordinate the group

## Business Case for Statistical Data Governance Framework

### 1 Purpose

To create a document describing a data governance reference framework focused on achieving data interoperability.

### 2 Project description

The main goal of the project is to produce a document describing a reference framework containing the main elements to implement a governance program focused on achieving data interoperability. Although a governance framework may cover other aspects related to the whole data management lifecycle, a realistic project must take a very cautious approach to get a solid result.

#### The project outputs will be:

1. A document describing a reference framework to achieve data interoperability that will include recommendations and guidelines about:
  - Establishing a data governance body inside the statistical organisations to agree on conceptual decisions that are necessary to get data interoperability.
  - How to structure and use the existing models and standards produced by the ModernStat program and by other relevant projects like SDMX and DDI as components of this framework.
  - Core aspects that need to be covered during the phases and sub-processes described by the GSBPM to ensure that reliable data interoperability will be achieved by the statistical projects.
  - Recommendations about how to implement transversal platforms supporting data interoperability and able to set up concept-driven integrated information systems that have shared concepts and classifications, provide common views, and is supported by common reusable information services, applications and tools.

### 3 Alternatives considered

2. The data governance ecosystem described in the World Development Report 2021 published by the World Bank, the CES task force on data stewardship, the European Statistical System, the Data Stewardship project approved by the fifty-second session of the Statistical Commission is related but complementary efforts that can be put all together to improve the data management of the statistical organisations on different aspects.
3. Doing nothing. If the project does not go on there will still be opportunities to collaborate but these efforts will still be needed as data interoperability is a characteristic needed to provide a new generation of services to fulfil the emerging demands coming from the users of official statistics.

### 4 Expected Benefits

<input checked="" type="checkbox"/>	Reduced costs
<input checked="" type="checkbox"/>	Increased efficiency
<input checked="" type="checkbox"/>	Reduced risks
<input checked="" type="checkbox"/>	New capabilities to meet user needs

**5 Which key priorities in the HLG-MOS Strategic Framework does the proposed project relate to?**

<input type="checkbox"/>	Take cost out of our organizations to reinvest in more value-added areas
<input type="checkbox"/>	Explore new areas collectively and leverage each other's' research investments in specific areas
<input checked="" type="checkbox"/>	Provide whole of government data ecosystems based on international standards, for better estimates in key policy areas
<input checked="" type="checkbox"/>	Renew our governance and operating processes

**Justification:**

There is not a single definition of data governance, but in general terms, we can say that this term is focused on ensuring the quality, integrity, security, and usability of the data that is collected and managed by an organisation during its whole lifecycle.

Data governance focuses on making data easily accessible to all users in a reliable way, providing the most value of it while taking care of the standards and regulations that are relevant. To set a data governance environment in statistical organisations we need to set up a framework to structure four main elements:

- A governing body
- Standards, models, and guidelines
- Processes
- IT Infrastructure and tools

As data governance is a concept applied to many aspects of data management during its whole lifecycle it would be unrealistic to try to take care of all those aspects at once, so the proposed project will be focusing on just one important aspect for statistical offices: interoperability of the statistical data and metadata.

A data governance framework helps to improve the following elements of the organisational data environment:

- Data discovery and assessment
- Data classification and organization
- Data catalogue and metadata management
- Data quality management
- Data access management
- Data monitoring, auditing and tracking
- Data protection

Data interoperability is the ability to create, exchange and use data conserving its meaning and context independently of a given system or a set of them. A framework can be used to set a favourable environment to gain this ability.

Statistical offices perform projects to produce information related to facts from different domains. It is very common for these projects to create particular definitions to model the data, metadata and classifications they use and produce, creating something that we may conceive as conceptual silos.

Information Technologies have made it possible to put together information from different sources and implement data warehouses, data lakes, banks of indicators and other kinds of dissemination systems that we qualify as multi-domain or as integrated. The specialized data structures designed to feed those systems add complexity to make the information flow from one source to the other. It is very common to have transformation tools and processes to make the information fit in the systems. But the information is still being born isolated. Maybe we still creating a lot of silos, with the difference that they have become bigger, more complex, and are mixed in the same bags.

It doesn't matter which kind of advanced technologies we implement. If we don't share common concepts and classifications to describe statistical data and metadata we will not be able to build real integrated multi-domain statistical systems or data lakes. Consequently, we will be falling short in our



intention to deliver reliable and easily accessible statistical information to all users, providing the most value of it while taking care of the standards and regulations that are relevant.

If we want to implement data interoperability, then we need to agree on common concepts and classifications, and for doing this, data governance is a condition.

Conceptual and technical interoperability will provide the elements to implement architectures in statistical organisations to provide information not just from separate facts, but also to present the linked trends and patterns stated by a complex reality where the measures are conceptually connected in time and space. With this new approach, we expect to be ready to provide a new generation of information services, to satisfy emerging and complex information needs, being able to provide the information and tools to answer complex questions like “Which has been the effect of population growth and development of industrial activities in the environment?”

The aim is to get the following benefits from this project:

- Increase the value of the statistical information by establishing connections between the data from different domains
- Reduce the costs by creating a way to effectively reuse information and tools
- Improve the information products and services adding the capacity to create a new generation platform of systems and tools to enhance the analysis and dissemination of statistics satisfying complex emerging needs from the users
- Improve data and metadata quality by making it more transparent, manageable and comparable.

## 6 How does the proposed project relate to other activities under the HLG-MOS?

As formerly described, the project is completely aligned to the application of the work developed by the HLG-MOS groups and provides an excellent opportunity to integrate the models of the ModernStats program with other important standards in the statistical community environment on a framework that will demonstrate the specialisation of each of them and how they can be structured to support relevant use cases.

## 7 Proposed timetables

First Quarter of 2022: Constitute the group and agree on the structure of the document.

Second and Third Quarter of 2022: Produce the contents of each of the four components of the framework. The participation of experts on data governance, experts on the HLG-Models, experts on SDMX, experts on DDI, methodologists, statisticians and ICT people will be fundamental to develop each of the sections of the document. Periodic communication between the tasks teams will be needed to ensure that the document keeps coherence and aligned with the objectives.

Fourth Quarter of 2022: Will be dedicated to integrating, reviewing and making the final edition of the framework.

## 8 Expected resources and costs

The following are the expected resources required for 2022:

- Experts on data governance, experts on the HLG-Models, experts on SDMX, experts on DDI, methodologists, statisticians and ICT people
- Space in the UNECE’s Wiki to facilitate the collaborative work
- Virtual conference facilities to support periodic meetings of the group and its task teams
- Resources to organise a physical sprint (depending on the pandemic circumstances)
- Resources to organise a physical workshop (depending on the pandemic circumstances)
- Support from the UNECE’s HLG-MOS Secretariat to help in the coordination of the group

## Business Case for Meta-Academy for the Modernization of Official Statistics

### 1 Purpose

The purpose of the meta-academy is to remove barriers to co-create training and reuse content on an international level, which will ultimately unleash the creation and use, at scale, of open digital assets to boost National Statistical Offices (NSOs) upskilling necessary for modernization. This project intends to raise the standards of virtual learning on topics necessary for the modernization of statistics but are missing or inconsistent from academic, commercial or in-house offerings. The meta-academy project sets out to create a benchmark to better map existing initiatives and offerings in order to better coordinate efforts, reduce duplication and fill in training gaps. This project will facilitate sharing of skills strategies, as well as catalogues of contents and pedagogical artefacts, and more generally good practices and standards in that space, so that scopes for reuse or co-creation in learning capabilities can be more easily and more systematically spotted and leveraged by all NSOs.

### 2 Project description

This project has three main work packages that target to create synergies between existing initiatives, identify gaps in topics and methods of training and experiment with co-creation and training in a virtual world.

**Work package 1** will focus on the benchmarking exercise among existing initiatives for shared good practices, repurpose existing available material and identifying co-creation of any gaps. This exercise will start off with brainstorming sessions and presentation series to scan existing initiatives in other international organisations, NSOs, academia and the private sector. The brainstorming and presentation series aim to filter down how much of the landscape we would like to analyse in this exercise. The work package will then proceed with an in-depth analysis of a selection of relevant initiatives as a benchmark to then co-design a final framework.

The project is an opportunity to connect multiple initiatives at international level:

- Global capacity building initiatives focusing on certain specific areas of skills or techniques: UNSD big data task force, SDMX sponsors – or specific topics (think of P21 Academy on gender statistics);
- Regional initiatives such as capacity building undertaken by Eurostat in the ESS context or other regional agencies (UNESCAP, AfDB...);
- Sectorial initiatives with an important statistical component – especially driven by UN agencies (ITCILO, FAO eLearning academy, UNICEF,...) or the Development Banks;
- Countless initiatives with a large international resonance, led by NSOs or national cooperation agencies (think, for example, of Norway's DHIS2 initiative on health statistics).

Initiatives at national levels (especially those in association with the Academic world – for example, Switzerland's FSO upskilling initiative in association with the EPFL; or ONS' government data scientist programme) should be mapped and analysed in that context, in order to identify transposable approaches.

It is also crucial to position the experience of NSOs with the private sector offering (global e-learning platforms such as Coursera, EDX, BaseCamp... as well as niche players specialised in the field of data science) so as to clarify the scope, specific to official statistics, that should specifically invest in by NSOs in complement to private sector.

This benchmarking extends beyond topics of training and will assess virtual training frameworks in place or being proposed. Interviews will be conducted with initiative leads to map skills

strategies and capacity building vehicles, their potential complementarities, and propose good practices to make them more reusable or discoverable. The output for this work package will be a report of finding and an HLG-MOS workshop.

**Work package 2** will be an experiment to co-create capacity building content by NSOs on the topic of Git and version control. This is a suitable use case because it is an area that broadly impacts modernization of statistics from basic programming to advanced data science, however it gets little attention in international forums; there is private sector material, but they do require an extra layer of adaption and coaching to match the official statistics context; and there is an existing starting point with the current collaboration between Statistics Canada and OECD on creating a Git manual. This work package will hold a series of workshops and online collaboration to complete a “git manual” capturing best practices in the management of algorithms and code bases by statisticians and data scientists in the official statistics context. This collaboration will involve experts leading the development of algorithms design and sharing good practices in their organisations. The expected outputs for this work package include a git manual for official statistics and one HLG-MOS pilot course on version control for official statistics. In addition, this experience will provide real time co-creation lessons learned on frameworks, opportunities and methods developed from work package one.

**Work package 3** will finalize the framework for virtual learning, co-creating and reusing content by formalizing concepts, skills mapping, learning outcomes and journeys and content creation best practices. This work package will also address the best delivery method for content and co-creation such as a central platform or network.

### 3 Alternatives considered

1. There are existing training initiatives that are underway, however there is no coordination, consistency and common standards between initiatives;
2. Existing initiatives at an international level are either sectorial in nature (e.g. labour statistics) or focus on specific horizontal practices (e.g. SDMX or big data techniques);
3. Existing initiatives at a national level are rarely advertised beyond national boundaries and may be inaccessible due to language barriers;
4. The private sector and academia offer obviously great upskilling opportunities that should be leveraged, however failing to match some dimensions that may be specific to official statistics;
5. If nothing occurs, then NSOs will be duplicating efforts and paying for their own training, either in house or commercial offerings, with varying levels of quality and outcomes – when, in essence, needs are very similar and significant knowledge spill-overs and efficiency gains could be achieved that are only marginally tapped into today.

### 4 Expected Benefits

<input checked="" type="checkbox"/>	Reduced costs
<input checked="" type="checkbox"/>	Increased efficiency
<input type="checkbox"/>	Reduced risks
<input checked="" type="checkbox"/>	New capabilities to meet user needs

### 5 Which key priorities in the HLG-MOS Strategic Framework does the proposed project relate to?

<input checked="" type="checkbox"/>	Take cost out of our organizations to reinvest in more value-added areas
-------------------------------------	--

<input checked="" type="checkbox"/>	Explore new areas collectively and leverage each other's' research investments in specific areas
<input type="checkbox"/>	Provide whole of government data ecosystems based on international standards, for better estimates in key policy areas
<input type="checkbox"/>	Renew our governance and operating processes

Justification:

In 2020 the Blue Skies Thinking Network ran an activity called *Experimentation to Implementation* to determine if there are common areas that the international community struggle with to move innovations to production. Through this activity, capacity building was recognized as a strategic priority and an area the international community can take concrete steps to address. There are many international and national 'data academies' that are duplicating efforts with no common standards. Much of the existing training is limited to online webinars without pedagogic architecture, certification strategy, or coaching /peer-learning. There are no common standards in terms of technical formats (including discoverability or catalogue) or pedagogic concepts (learning outcomes, self-assessment) that could be leveraged in the official statistics space. COVID-19 has increased the need for quality virtual training. The online component of capacity building and academies at NSOs are struggling to adopt suitable virtual frameworks.

#### 6 How does the proposed project relate to other activities under the HLG-MOS?

This project will consolidate and provide continuity to efforts that the HLG-MOS, UNECE and other associated groups, such as the UN Big Data Group, on creating content and frameworks for virtually training of key aspects necessary for the modernization of statistics. This project will leverage existing frameworks such as the HLG-MOS Capabilities Development and Training Framework and the UNECE Methodology Framework as well as existing networks such as the Blue Skies Thinking Network, as well as frameworks developed by other agencies or at national level<sup>10</sup>.

#### 7 Proposed timetables

Work packages 1 and 2 will run concurrently in 2022. While the benchmarking of existing training activities is taking place in the first half of 2022, collaborators will put together a Git manual for official statistics.

Once the Git manual has been completed, members of work package 2 will collaborate with those of work package one to create a framework for virtual learning and implement the framework in practice with the course on Git and version control at the end of 2022.

#### 8 Expected resources and costs

The following are the expected resources required for 2022:

- This work package would require a funded resource with pedagogical experience (in-kind or, more likely, \$60kUSD to be co-funded by interested stakeholders);
- Participants from international and national training initiatives (here, we may think to open the initiative to stakeholders that are not usually participating in UNECE works);
- Possibly, participants from the academic and private sector world (especially learning specialist focused on data topics);
- Experts on Git and version control;
- Space in the UNECE's Wiki to facilitate communication;
- Support from the UNECE's HLG-MOS Secretariat to help in the coordination of the group.

<sup>10</sup> Think of FAO's [e-learning methodologies and good practices](#); or ONS' work on [the government data scientist](#).





