

CONFERENCE OF EUROPEAN STATISTICIANS

For discussion and  
recommendations

Meeting of the 2020/2021 Bureau  
Geneva (online), 17-18 February 2021

Item III (f) of the Provisional  
Agenda

## High-Level Group for the Modernisation of Official Statistics: 2021 Work Programme

Prepared by the Secretariat

The document provides an overview of the planned work of the High-Level Group for the Modernisation of Official Statistics (HLG-MOS) in 2021.

*The Bureau is invited to provide recommendations on the work under HLG-MOS to be undertaken in 2021.*

### I. 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 HLG-MOS is open to all who are willing to contribute. 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 has to reflect these changing needs, while at the same time continue to support or further develop previous outputs. To guarantee an optimal outcome, a process combining top-down and bottom-up approaches is followed.

2. Annually, the chair of HLG-MOS sends out to all members of the Conference of European Statisticians (CES) an open call for submission of project and activity proposals for the next year<sup>2</sup>. Since 2019, proposals can also be submitted to the Blue Skies Thinking Network (BSTN) at any time during the year. After an initial evaluation by BSTN, the Executive Board of HLG-MOS provides feedback and selects proposals that will be considered as projects or flagged for other types of follow-up. At the annual Workshop on the Modernisation of Official Statistics<sup>3</sup> at the end of November, these proposals are presented and discussed by experts that are 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. The Executive Board then discusses and further refines the proposals, and the selected projects and activity proposals are finally submitted for endorsement by HLG-MOS.

---

<sup>1</sup> Refer to Annex 2 for the structure of HLG-MOS.

<sup>2</sup> Due to the Covid-19 pandemic, in 2020 no open call was sent out. However, BSTN kept welcoming proposals.

<sup>3</sup> The participants of the workshop are the experts that were active under the various projects and groups under the auspicious of HLG-MOS in 2020 and other experts sent by NSOs of CES member countries as well as members of HLG-MOS and the Executive Board.

3. This document outlines the work program that resulted from the discussions at the Workshop on the Modernisation of Official Statistics (held online on 18-19 November 2020), and includes subsequent further refinements by the Executive Board. The paper provides summary information to the community on the work that will take place so they can participate in the activities that are of most benefit to them.

## II. HLG-MOS projects in 2021

4. The following two projects were selected by HLG-MOS for 2021: i) Input Privacy-Preserving Technics Project and ii) Practical Guide to the Use of Synthetic Data Project. Both projects are in the area of confidentiality and data protection. As the focus and approach differs between the two projects, specifically a focus on input versus a focus on output, it was decided that the projects should not be merged. However, to create synergy and exchange ideas, regular joint meetings will be held. The difference between the two projects will have to be clarified for example by sharing clear use cases from the projects showing the different aspects they address. The end-to-end privacy cycle from input to output will also be discussed.

5. **Input Privacy-Preservation Project:** The project on input privacy-preserving techniques was selected for the 2020 HLG-MOS work programme. However, due to the Covid-19 pandemic, activities could only start in the second half of 2020 with a small group and limited scope. Lead by the project manager, this group further scoped the project proposal and started an inventory of use cases from participating NSOs. In 2021, the focus will be on progressing the work and gathering experiences from generalized use cases. Technical experts from outside official statistics will now be invited to join the project. The following work packages have been identified:

- **WP1: Document, generalize and prioritize the use-cases:** New participants will bring new use-cases that will be added to the inventory started in 2020. Generalized use-case might be defined and a subset of the most prominent use-cases will be identified. Reference scenarios that are relevant for different types of source data and/or different NSOs will be drafted and prioritized for follow-up action in WP2.
- **WP2: Set up and test the use-cases:** Selected types of use-cases will be tested against different techniques, including in edge computing. For use-cases for which a broader spectrum of technological solutions is available, we will launch a wider call for proposed solutions from external technical audiences. Use-case testing and comparison of different solutions based on benchmark conditions (e.g. with synthetic test data) will be conducted. The technology aspect will be considered as well as organizational, business and legal aspects. Due to the complexity of this work package, it might be structured into parallel sub-streams.
- **WP3: Lessons learned and guidelines:** The experiences from WP1 and WP2 will be used to make clear recommendations on which techniques are most suitable in which cases and where such techniques are not applicable. Although the focus will be on technical and methodological aspects of implementation in official statistics, lessons learned and recommendations with respect to organizational, business, and legal aspects will be documented as well.

6. The project will be coordinated with the UN Global Working Group Task Team on Privacy Preserving Techniques, related Eurostat activities, the International Working Group for Machine Learning (continuation of the 2020 HLG-MOS Machine Learning Project) and the 2021 Expert Meeting on Statistical Data Confidentiality.

7. **Practical Guide to the Use of Synthetic Data Project:** Building on the success and network of the Blue Sky Thinking Network's Working Group on Synthetic Data, the aim of the practical guide to Synthetic Data project is to develop a hands-on guide for creating and using synthetic data primarily for data protection and disclosure control. The target audience of this guide includes NSOs as well as their clients such as academia, the private sector and

the general public. The guide will focus on how to use synthetic data in practical applications, considerations for implementation, and important aspects to share with users. This guide can serve as the foundation for future standards as synthetic data is more broadly adopted within NSOs and by their users. The emphasis will be on creating synthetic data sets from new data sources. The project will be divided into four work packages, with the scoping work already completed through the 2020 BSTN Working Group on Synthetic Data:

- **WP1: Use cases for synthetic data:** The methods and measures of synthetic data are highly dependent on the reason for using synthetic data. The Working Group on Synthetic Data has identified four broad categories of uses for synthetic data: public data release, testing analysis, training and testing technology. This work package will detail the use case categories and highlight their specific methodology and measure requirements.
- **WP2: Recommended methods for creating synthetic data:** This work package will gather an inventory of methods for creating synthetic data both currently in use and in research. The core outcome of this work package will be the assessment of the inventory of methods and recommendations on the methods suitable for different use cases of synthetic data.
- **WP3: Measuring the analytical value and/or disclosure risk of synthetic data sets:** This work package will gather an inventory of utility and disclosure risk measures. This work will outline the circumstances in which the measures apply, associate them with the use case categories and provide clear explanations of what the measures signify and how they should be interpreted by those that create the synthetic data sets and those that use them.
- **WP4: Experimenting with the recommendations:** Since the guide focuses on practical applications of synthetic data, this final work package will test the recommendations of the guide with real life scenarios of synthetic data. Activities will include pilot projects from NSOs, presentations geared towards user needs and hands-on events such as hackathons or training workshops.

8. The work of the HLG-MOS projects is normally led by project managers that are either assigned by a member of HLG-MOS (in kind) or contracted by UNECE using the HLG-MOS Trust Fund. Project managers have already been identified for both projects. As neither are available full time and given the nature of the two projects, lead substantive experts might be assigned. The UNECE wiki and web conference are used to monitor the progress of the projects and UNECE provides additional administrative and secretariat support.

9. Additional details on the two selected projects can be found in Annex 1.

### III. Modernisation Groups

10. Modernisation Groups are organized around common themes and the selected activities are normally carried out through Task Teams. Groups have frequent virtual plenary meetings and Task Teams have monthly virtual meetings. The groups may have additional meetings or organise sprint workshops to expedite the work. The chairs of the groups provide monthly updates to the Executive Board.

11. HLG MOS has mandated the Executive Board to regularly reflect on the structure and areas of work of the Modernisation Groups and evaluate whether the groups are aligned with the key priorities identified by HLG MOS. It had already 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 HLG-MOS, it was decided not to set up a new group in the first half of 2021. The HLG-MOS Executive Board will monitor the situation and might come with proposals in the second half of 2021.

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

#### A. Capability and Communication Group

13. The Modernisation Group on Capability and Communication focuses on the organisational changes and the communication challenges necessary to support modernization in statistical organizations. There were many activity proposals in this area that were all considered relevant. Rather than selecting only some, they were grouped into three thematic areas. This will create synergy and mitigate against risks of a mismatch between proposed topics and availability of experts in the group. The group will set up task teams to consider the following substantive topics during their monthly web conferences:

- **Future of work, future workplace and future skills** – The purpose of this proposed activity is to develop a framework/operating model on skills, capabilities and training that is future proofed and can be adopted in any NSO. This model will include skills and capabilities which will be complementary to the technical skills applicable to roles including statistical, human resources, IT etc. The task team will explore successful best practices, current models and practices adopted before and during Covid-19 which can be adopted across NSOs.
- **Ethical leadership as part of culture evolution** – The goal is to define a common vocabulary and framework, and give concrete suggestions (a kind of Handbook) to support NSOs' leadership in managing ethics and integrity as part of the Corporate Social Responsibility and an organisational culture aligned with this. A short poll will be held to identify how many and which NSOs have implemented policies, procedures and programmes on ethics management. This will be used to collect feedback, insights and examples. Additionally, it is proposed to set up a collaborative platform and/or organise a meeting to share experiences and ideas.
- **Role of market research, digital marketing & communication strategies and tools in crisis communication and in promoting public engagement in surveys** – The pandemic has highlighted how important agile and responsive communication strategies are to allow NSO's to manage emerging and fast-changing situations. It is a follow-up on the Strategic Communication Framework to address the crisis communication issues that emerged during the pandemic. The key objective is to explore how NSO's can use digital communication to better position themselves to respond to rapidly evolving communication and marketing requirements.
- The Capability and Communication Group organises the biennial **Human Resources Management and Training workshops**. In 2021, the group will start organising the 2022 Workshop.

#### B. Supporting Standards Group

14. 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. Currently, this group is chaired by Hungary.

15. The group has plenary web conferences every month. The group may have additional meetings or organise a sprint workshop to progress the work, as necessary.

16. In addition to the plenary meeting, separate task teams will work on the following topics:

- **Linking GSBPM and GSIM** – This extension of a 2020 activity will continue to map the information objects of GSIM to the processes and sub-processes in GSBPM. The

activity will create a model for GSIM-GSBPM mapping and prepare relevant documentation, including the clickable version.

- **Core Ontology for Official Statistics** – This extension of an existing activity aims to create a common semantic model and vocabulary for official statistics. The core concepts that are cutting across different models were identified. The focus will now be on developing the links with vocabularies in other models and establish the management and governance of these different initiatives.
- **Updating GSIM** – This activity aims to incorporate experiences from countries that have started to use GSIM in their organization. Based on these experiences, a soft update of GSIM will be done to clarify ambiguous points, fix minor errors and update the GSIM documents.
- **Application of GSBPM for Geospatial Information** – The activity will provide guidance on how to use GSBPM with geospatial information.
- **GSBPM task** – The activity will compile examples of lower-level GSBPM activities (“tasks”), identify commonalities among these and summarise the findings. The output can be used as a supplementary GSBPM document to support countries using GSBPM in more detailed level and can be used as an input for next GSBPM revision.

17. The Sharing Tools Group provided implementation support for the Common Statistical Production Architecture (CSPA). In 2021, it will be merged with the Supporting Standards group. Several activities on CSPA will be bundled to be covered by one task team:

- **CSPA** – Due to the ever-changing environment, CSPA needs continuous updating. Furthermore, the outcome of the ESSnet project Implementing Shared Statistical Services (I3S)<sup>4</sup> that plans to finish by mid-2021 needs to be integrated. Making the CSPA document interactive and developing better communication material are other pending activities. Now that CSPA is added to the Supporting Standards group, the group will review whether the services are sufficiently described, how GSIM and GSBPM are used in CSPA, and evaluate whether further work on CSPA Logical Information Model is needed. The work is planned to start in the second half of the year.

18. By decision of the HLG-MOS Executive Board, all workshops, except for the HLG-MOS Modernisation Workshop, will take place with 24-month intervals<sup>5</sup>. During 2021, the Supporting Standards Group will set up an Organisation Committee to prepare 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.

### C. Blue Skies Thinking Network

19. The Blue Skies Thinking Network is the ideas factory of the ModernStats community. It is led by an innovation manager and consists of a core group of around twelve members from NSOs and international organisations. The core team should have both in-depth and broad knowledge of innovation related aspects to facilitate the work, and sufficient heterogeneity to allow for a variety of views. The core group draws from a flexible pool of resources to assist in the evaluation of proposals.

20. The group can also set up temporary activities to follow-up on project proposals that were not selected or activity proposals that were not assigned to a Modernisation Group. In

---

<sup>4</sup> The project is part of the Eurostat’s European Statistical System Network program. More information is available from the CROS platform: [https://ec.europa.eu/eurostat/cros/content/implementing-shared-statistical-services\\_en](https://ec.europa.eu/eurostat/cros/content/implementing-shared-statistical-services_en)

<sup>5</sup> The main reason is the lack of resources at the UNECE Management and Modernisation Unit that is supporting the work of HLG-MOS

2021, the Network will elaborate further on the following topics (other topics might be added throughout the year):

- **Network data** (feasibility study): The immediate goal of this activity is to determine the methodological feasibility of deriving sufficiently valid inferences about networks from data aggregates. If methodologically feasible, suitable business applications will be chosen that are of interest to all participants, and network data providers and analysts will be consulted on the feasibility of taking this forward.
- **Covid-19 Hotspot Joint Biosecurity Centre Platform:** The aim is to investigate how to set up a secure flexible analytical single platform for big data and data science collaboration. At ONS, the Covid-19 pandemic has shown the benefits of such an approach which merits investigating further.
- **User Research for Official Statistics:** The activity focuses on user-oriented research in an international context. It tries to find answers to questions such as: Which are the users' problems we are trying to solve? What is our level of understanding of users and user needs? Are we at edge of user research techniques available and used on the market? Further ideas for joint activities include developing a 'statistical personas' database, a research study on data journalists, and a toolkit to improve visibility in search engines.
- **Rapid survey systems:** The COVID-19 outbreak imposed difficulties in regular data collection while policy makers needed more and different information. Successful implementations of rapid survey systems to collect timely and often new data of good quality without overburdening reporting units (such as set up by Statistics Serbia), are worth sharing and common elements can be identified.
- **From experimentation to implementation in official statistics:** This activity will explore the environments of participating NSOs, the methods, technology and the means they have used to make data science and analytics a part of the culture within official statistics. The goal is to share experiences, lessons learned and open a discussion on ways to collaborate internationally to facilitate adopting new methods and technologies within statistical organizations.
- **Microdata for understanding declining response rates:** NSOs across the world are facing the challenge of declining response rates, while demand for data is growing. Face-to-face interviewing will remain for a long-time to come and it is necessary to share experiences how to deliver an effective and efficient service. The aim is to take research already done further<sup>6</sup>.

21. Other topics that were brought to BSTN for follow-up are: i) using the Facebook survey on Covid-19 related symptoms and behaviour, ii) the trust people have in data governance, iii), the market versus government discussion related to providing data and statistics, iv) data access (how can you get data and how can you protect it) and how this fits into national data strategies, and v) processing platforms that have the right libraries of algorithms to treat use cases from the UNECE past and current projects.

22. Depending on the interest of members and experts that are willing to get involved, topics might be worked out in detail or only receive a quick scan. BSTN will also monitor developments with respect to Statsbots and the International Working Group for Machine Learning that was set up as a pilot case of the Collaboration Platform concept proposed by BSTN in 2020.

23. HLG-MOS members will continue to bring the Network to the attention of their national innovation groups and stimulate generation of ideas for proposals to be submitted to

---

<sup>6</sup> Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys (Journal of Official Statistics, Vol. 36, No. 3, 2020, pp. 469–487, Annemieke Luiten, Joop Hox, and Edith de Leeuw)

the Network. The Network will also actively engage with the statistical community to identify potential topics that merit further consideration.

## **IV. Other activities and expert meetings**

### **A. Other activities supported by HLG MOS**

24. The StatsBot activity will continue under the guidance of OECD. A more sustainable funding model is sought and interested countries are invited to join the effort (Canada and the Netherlands are already involved). BSTN and the Executive Board will follow the work, make suggestions and provide support and input where possible.

25. The interdisciplinary nature of Machine Learning and the need for acquiring specific skills was recognized. This led to the broad agreement that knowledge sharing through communities of practice and international organizations is valuable. It was therefore decided to continue the work and create an ML community for official statistics as a follow-up of the project. The Data Science Campus at the UK Office for National Statistics Office stepped forward to support this initiative. Regular updates will be provided to the Executive Board that will provide strategic support and recommendations.

### **B. Expert Meetings**

26. As in the past, under the auspices of HLG-MOS several Expert Meetings will be organized in substantive areas. The focus is 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 the meetings are senior and middle-level managers. The meetings are organised by the UNECE secretariat in collaboration with Organizational Committees. The programme of the meetings will be aligned with the HLG-MOS activities where needed.

27. Due to the uncertainty with the Covid-19 pandemic, no meetings are planned for the first half of 2021. For the second half of 2021, the following physical meetings are being considered:

- Expert Meeting on statistical data dissemination and communication
- Expert Meeting on statistical data collection
- Expert Meeting on statistical data confidentiality

28. HLG-MOS or the Executive Board can also decide to organise short, focused ad-hoc online workshops in emerging areas, such as the Covid-19 response workshops organized in 2020.

## **V. Monitoring progress and coordination**

29. The work of the Groups and Projects is reported on a monthly basis to the Executive Board of 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. There is also a conscious effort to align and coordinate all activities with other international initiatives like the programs under the European Statistical System and the UN Global Platform. The modernisation updates are made available to the wider public every two months at the ModernStats wiki (<https://statswiki.unece.org/x/QY0HBg>).

**Annex 1****2021 Project Proposals****Business Case for Practical Guide to Synthetic Data**

<b>Type of Activity</b>			
<input checked="" type="checkbox"/>	New project	<input type="checkbox"/>	New activity
<input type="checkbox"/>	Extension of existing project	<input type="checkbox"/>	Extension of existing activity
<i>Projects are undertaken by separate project teams. Projects are expected to produce a significant contribution to achieving the HLG-MOS vision</i>		<i>Activities are undertaken by Modernisation Groups. These activities produce smaller, more detailed outputs to help achieve the HLG-MOS vision</i>	
<i>See here for more details: <a href="https://statswiki.unece.org/x/nwEzCw">https://statswiki.unece.org/x/nwEzCw</a></i>			
<b>Purpose</b>			
<p>Data has become a valuable commodity, providing information for statisticians, economists, and data scientists to generate more timely and granular insights. National statistical offices (NSOs) are striving to provide greater transparency and openness and so are looking to expand safely sharing of data, expertise and best practices both internally as well as with external partners. In addition, different types of users are increasingly searching for quality data sets to support testing, evaluation, education and development purposes. These aspects provide more value to users and bring the need to uphold data integrity and confidentiality to the forefront.</p> <p>The demands for timely, integrated data compiled from ever-growing sources of increased complexity, along with the unequivocal commitment to trusted data protection call for a modernized, interoperable approach to mobilizing these large and complex data sources. Synthetic data can be a solution to providing rich data while respecting integrity and confidentiality imperatives.</p> <p>Synthetic data can find its roots in edit and imputation methods, however synthetic data uses are becoming broader and increasing in complexity. New methods are emerging for generating and evaluating confidentiality of synthetic data, and more guidance is needed to maximize utility while ensuring confidentiality. Utility is seen as the value that a data set brings to a particular usage, for example, systems testing or model testing. Increasingly, utility also encompasses the desire that distributions or results of the synthetic data closely approximate those found in the real data. For example, extracting more detailed insights using increasingly big data sets requires new methods such as machine learning and modeling techniques. The integrity of these new methods requires that, as much as possible, data sets (both those from survey or non-survey sources) preserve the structure, characteristics and often the distributions of the original data as much as possible. However, the more closely the synthetic data set emulates the real data, the higher the risk that confidential information in the original data set could be disclosed. As governments become more open with their data and work, the confidentiality aspect remains a top priority. Once properly explored and understood, synthetic data can play an important role in the way that NSOs share data while maintaining public trust.</p> <p>Synthetic data is a relatively new topic, particularly its use by NSOs. A better understanding of not only the theoretical methods of how to create synthetic data are needed, but an international consensus on practical applications and best practices is required for consistency,</p>			



transparency and comparability within statistical agencies, and with users in academia and the private sector. Additionally, in order for synthetic data to be a viable option for NSOs to distribute and disseminate microdata, clear methods of communication on the utility and risk of using this type of data must be well available to stakeholders and users. In order for synthetic data to be used to its potential, an international consensus on practical methods and uses must be achieved.

### Description of the project

Building on the success and network of the Blue Sky Thinking Network's Working Group on Synthetic Data, 'the practical guide to synthetic data' project sets out to develop a hands-on guide for creating and using synthetic data primarily geared towards data protection and disclosure control. The target audience of this guide includes NSOs as well as their clients such as academia, the private sector and the general public. The guide will focus on how to use synthetic data in practical applications, considerations for implementation, and important aspects to share with users. This guide can serve as the foundation for future standards as synthetic data is more broadly adopted within NSOs and by their users.

The project will be divided into four work packages, with the scoping work already completed through the Working Group on Synthetic Data.

**WP1: Use cases for synthetic data:** The methods and measures of synthetic data are highly dependent on the reason for using synthetic data. The Working Group on Synthetic Data has identified four broad categories of uses for synthetic data: public data release, testing analysis, training and testing technology. This work package will detail the use case categories and highlight their specific methodology and measure requirements.

**WP2: Recommended methods for creating synthetic data:** This work package will gather an inventory of methods for creating synthetic data both currently in use and in research. The core outcome of this work package will be the assessment of the inventory of methods and recommendations on the methods suitable for different use cases of synthetic data.

**WP3: Measuring the analytical value and/or disclosure risk of synthetic data sets:** This work package will gather an inventory of utility and disclosure risk measures. This work will outline the circumstances in which the measures apply, associate them with the use case categories and provide clear explanations of what the measures signify and how they should be interpreted by those that create the synthetic data sets and those that use them.

**WP4: Experimenting with the recommendations:** Since the focus of this guide is on practical applications of synthetic data, this final work package will test the recommendations of the guide with real life scenarios of synthetic data. Activities will include pilot projects from NSOs, presentations geared towards user needs and hands-on events such as hackathons or training workshops.

### Alternatives considered

The work to build consensus and understanding of creating and using synthetic data could continue as a Blue Sky Thinking Network (BSTN) working group, however with membership over 25 participants, spanning 9 national statistics offices, academia and the private sector, this initiative has outgrown the scope of the BSTN. In addition, the target audience for synthetic data and the planned products are beyond the current methods of BSTN communication. A

formal project would provide the proper scope, oversight and communications for the intended deliverables.

**How does it relate to the HLG-MOS vision and other activities under the HLG-MOS?**

The Synthetic Data project relates to all HLG-MOS visions and values by creating a collaborative initiative to promote sound methods and practices of synthetic data while upholding statistical integrity of those methods and the confidentiality of the data in question.

**Proposed start and end dates**

**Start:** January 2021

**End:** December 2021

## Business Case for Input Privacy-preserving Techniques

Type of Activity			
<input type="checkbox"/>	New project	<input type="checkbox"/>	New activity
<input checked="" type="checkbox"/>	Extension of existing project	<input type="checkbox"/>	Extension of existing activity
The project is an extension of the 2020 Input Privacy-Preserving Techniques project sponsored by the HLG-MOS.			
For an overview of the progress of the current IPP project, refer to the meeting documents: <a href="https://statswiki.unece.org/display/IPPP/Meeting+Documents">https://statswiki.unece.org/display/IPPP/Meeting+Documents</a>			
Purpose			
<p>Statistical organizations are more and more investing on becoming part of a data ecosystem where they acquire and integrate data from multiple sources and provide richer statistical products.</p> <p>In this scenario, the issue of privacy preservation is particularly relevant: the more sources are acquired and integrated, the higher are the risks of disclosing information violating individual privacy rights. Hence, from a legislative perspective there are indications to take privacy into account throughout the whole data treatment process, through the ‘<b>privacy by design</b>’ concept.</p> <p>National Statistical Organizations (NSOs) are used to apply techniques for enforcing privacy by design on the <b>output side</b>, i.e. when publishing aggregated statistical data for dissemination purposes and when sharing microdata for research purposes with statistical disclosure control (SDC) and other output privacy-preserving techniques.</p> <p>However, NSOs have still to invest on dealing with privacy protection on the <b>input side</b>, in a complementary but distinct way with respect to output privacy preservation investments<sup>7</sup>.</p> <p>Different classes of techniques can be used to deal with input privacy<sup>8</sup>. Among them <b>Secure Multiparty Computation (SMC)</b> and <b>Homomorphic Encryption (HE)</b> play a relevant role.</p> <p>These methods are particularly suitable for use in a non-trusted environments such as access to private data, interconnectivity of highly sensitive data for the purpose of scientific research, data analytics in Cloud and AI. The goal of this project is to investigate statistical use cases that require protection on the input side, assess and determine applicability of selected classes of techniques for main scenarios, identify opportunities for sharing across statistical community and create community of practice across statistical organizations and external partners (academia, private sector).</p>			

<sup>1</sup> F. Ricciato, A. Bujnowska, A. Wirthmann, M. Hahn, E. Barredo-Capelot, A reflection on privacy and data confidentiality in Official Statistics, ISI 2019.

<sup>8</sup> UN Handbook on Privacy-Preserving Computation Techniques,  
<http://publications.officialstatistics.org/handbooks/privacy-preserving-techniques-handbook/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf>

### Description of the activity

After a late start, a number of use cases in the field of input Privacy-preserving techniques were inventoried and documented in 2020. In 2021, the focus will be on gathering experiences from generalized use cases. External technical audiences will be called for help. The following work packages have been identified:

- **WP1: To document, generalize and prioritize the use-cases:** New participants will bring new use-cases that will be added to the inventory and possibly generalized. A subset of most prominent use-cases will be identified. Reference scenarios that are relevant for different types of source data and/or different NSOs will be drafted and prioritized for follow-up action in WP2.
- **WP2: Setup and test the use-cases:** For selected use-cases (or groups thereof) that are amenable to be implemented with different technology approaches, set up benchmarking instances and compare different technological solutions. For use-cases for which a broader spectrum of technological solutions, is available, we will launch wider call for proposed options from external technical audiences. Use-case testing and comparison of different solutions based on benchmark conditions (e.g. with synthetic test data) will be conducted, considering the technology aspect jointly with organizational, business and legal aspects. Due to the complexity of this WP, it will be possible to structure WP2 into distinct sub-streams running in parallel.
- **WP3: Lessons learned:** The experiences from WP1 and WP2, will be used to make clear recommendations of which techniques are most suitable in which cases and where such techniques are not applicable. Although the focus will be on technical and methodological aspects of implementation in official statistics, lessons learned and recommendations with respect to organizational, business, and legal aspects will be documented as well.

## Annex 2

### 2021 Structure of HLG-MOS

