Modernisation Projects

2023 Modernisation Projects
(For project updates, please see Modernisation updates)

Data Governance for Interoperability Framework 2023
### CONTEXT
With statistical organisations increasingly engaging with new data sources and accelerating efforts in sharing and re-using data, the governance and management of data have become crucial. The interoperability across different data assets and metadata can greatly facilitate data exchange and help statistical organisations address new data needs (e.g., through data integration). Unfortunately, a large part of information in many statistical organisations is managed and governed in silos, making information semantically and synthetically non-interoperable. There is not enough work done on the governance side (e.g., policy, process, capability) which is indispensable to institutionalise the interoperability across the entire organisational journey.

### PROJECT OBJECTIVES
The main goal of the Statistical Data Governance Framework project is to produce a document describing a reference framework containing the main elements needed to implement a governance program focused on achieving data interoperability. This framework will provide the ability to create, exchange, and use data while preserving its meaning and context independently from a given system or a set of systems. The purpose of this project is to develop a framework describing a set of data governance elements, recommendations, and guidelines to achieve statistical information interoperability.

The main output of the project will be a document containing as its main sections:

- Glossary of core terms that could facilitate the communication and collaboration in the fields of data governance and interoperability
- A framework describing a set of data governance elements required to achieve statistical interoperability which includes (but not limited): organisational elements; data and metadata management; business and legal considerations; data quality; data analysis and dissemination needs; documentation and transparency; capabilities, culture, and skills; and Information Technologies aspects
- Recommendations and guidelines on how to start achieving interoperability in statistical organisations and national statistical systems; in particular how we can apply the existing models and standards (e.g., GSBPM, GSIM, CSPA, SDMX, DDI, CSDA, COOS, FAIR) to achieve interoperability

Participation is open to staff from statistical organisations and others interested in Official Statistics. Please contact the UNECE secretariat if you wish to participate in the project.

### Cloud for Official Statistics 2023

#### CONTEXT
Many NSOs are adopting Cloud approaches to a varied degree. It can directly contribute to modernising statistical production, and complements themes explored previously/currently under the HLG, such as Big Data, privacy-preserving techniques and governance, and may have synergies with current work on Data Science and Machine Learning. NSOs will benefit from informed approaches.

#### PROJECT OBJECTIVES
The objective is to develop a set of guidelines and recommendations, across multiple themes, to assist each statistical organisation on their cloud adoption journey. The initial five themes that were identified are:

1. **Common set of Considerations needed relating to the procurement of Cloud Services, assessing new user intellectual property, migration to another provider / vendor lock-in / exit strategy / terms and conditions.**
2. **Understanding the behavioural nudges needed to adopt cloud.** This theme will review indigenous / minority people’s perspectives on Cloud, public perception, Data Sovereignty, challenges relating to convincing an organisation’s executive board to approve the use of Cloud Services, the impact of cloud use on the official statistics brand.
3. **The types of Cloud service models and services which exist, and which are suitable for organisations in which context.** Topics for consideration include Infrastructure as a Service, Platform as a service, Software as a Service, Hybrid Cloud, Public Cloud, Private Cloud
4. **Explore the security and privacy considerations relating to the use of Cloud which may enhance or inhibit its adoption across statistical organisations.**
5. **The skillsets needed for the utilisation Cloud.** Topics for review will include staff retraining efforts needed, the challenge for public sector organisations in a competitive marketplace for Cloud skills, and how could knowledge be shared between organisations.

Participation is open to staff from statistical organisations and others interested in Official Statistics. Please contact the UNECE secretariat if you wish to participate in the project.

### ModernStats Carpentries 2023

#### CONTEXT
The ModernStats Carpentries project draws from the lessons learned in the context of the 2022 Meta Academy project (see the short Annex explaining the main takeaways from the Meta Academy project, including a definition of the gaps and potential capabilities that would make for a Meta Academy, and how the Carpentries initiative is well positioned to support them).

#### PROJECT OBJECTIVES
The purpose of the project is to pilot a partnership with the Carpentries organization to create the ModernStats Carpentry. The Carpentries are a non-profit organization, registered in the US, funded by membership and workshop fees, and grants from donors. Their vision is to be the "leading inclusive community teaching data and coding skills." In order to engage with the Carpentries, the HLG-MOS and/or member organizations will need to pay a membership fee; however, in the context of a ModernStats Carpentry, participating organisations could organise as many trainings as they wish (within national context or in the context of a cross-national initiative) at no fee; also, all Carpentries contents and training materials (data samples, codes, documents, etc.) are open and free under CC license, stored in the open Github platform, and can be reused at no cost

The Carpentries business model addresses several of the needs identified in the Meta Academy project in the following ways:

- A common understanding of the training needs, a shared methodology or pedagogic approach to create learning content;
- A forum or community for ‘academy managers’ or ‘trainers’
- A forum and method to ensure training content and delivery evolve with the industry.

WP1 will focus on repurposing existing Carpentries content for select key personas within statistical agencies as well as exploring how to put traditional official statistics courses into the Carpentries framework. WP2 will explore membership, collaboration and organizational models between the HLG-MOS and the Carpentries.

Participation is open to staff from statistical organisations and others interested in Official Statistics. Please contact the UNECE secretariat if you wish to participate in the project.
Data Governance for Interoperability Framework 2022

CONTEXT

The main goal of the Statistical Data Governance Framework project is to produce a document describing a reference framework containing the main elements needed to implement a governance program focused on achieving data interoperability. This framework will provide the ability to create, exchange, and use data while preserving its meaning and context independently from a given system or a set of systems.

PROJECT OBJECTIVES

The objective is to increase the value of the statistical information by establishing connections between the data from different domains. The project will aim at reducing costs by creating a way to effectively reuse information and tools as well as improve the information products and services adding the capacity to create a new generation platform of systems and tools that will enhance the analysis and dissemination of statistics. In this way, the framework can meet the emerging and more complex needs of our users while at the same time improving data and metadata quality by making it more transparent, manageable and comparable.

- WP1: Establishing a data governance body
- WP2: Structuring and using the existing models and standards
- WP3: Identifying core aspects to be covered during GSBPM phases and sub-processes
- WP4: Guide to implement transversal platforms for data interoperability and concept-driven integrated information systems

Participation is open to staff from statistical organisations and others interested in Official Statistics. Please contact the UNECE secretariat if you wish to participate in the project.
Meta-Academy for the Modernization of Official Statistics 2022

CONTEXT

Moving from innovation to implementation keeps being a major challenge. 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, which will ultimately unleash the creation and use, at scale, of open digital assets to boost the National Statistical Office (NSO) upskilling necessary for modernization.

PROJECT OBJECTIVES

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.

- WP1: Benchmarking
- WP2: Co-create capacity building content
- WP3: Finalizing the framework for virtual learning, co-creating and reusing content

Participation is open to staff from statistical organisations and others interested in Official Statistics. Please contact the UNECE secretariat if you wish to participate in the project.

---

Input Privacy-preserving Techniques 2022

CONTEXT

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 allows for analysing or integrating external data sources and producing statistics without revealing the microdata to the external partner. 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.

PROJECT OBJECTIVES

The objective is to expand and continue the existing collaboration between the involved participating organizations and to further explore and broaden the applicability of input privacy preservation techniques. This will allow NSOs to become part of or leading in data ecosystems by allowing the use of private data between NSOs and, more generally, between organizations.

- WP1. Deepening practical experiments
- WP2. Document use cases and provide guidelines for implementation
- WP3. Create user community

Participation is open to staff from statistical organisations and others interested in Official Statistics. Please contact the UNECE secretariat if you wish to participate in the project.
CONTEXT

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.

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.

PROJECT OBJECTIVES

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 is divided into four work packages, with the scoping work already completed through the Working Group on Synthetic Data.

Objectives

- WP1: Use cases for synthetic data
- WP2: Recommended methods for creating synthetic data
- WP3: Utility and Disclosure Risk Measures
- WP4: Experimenting with the recommendations
Due to staffing shortages at UNECE and the Covid19 pandemic, this project was on hold until 1 August.

CONTEXT
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. 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 ‘privacy by design’ concept. National Statistical Organizations (NSOs) are used to apply techniques for enforcing privacy by design on the output side, however, NSOs have still to invest on dealing with privacy protection on the input side, in a complementary but distinct way with respect to output privacy preservation investments.

PROJECT OBJECTIVES
The first objective of the project is to scope the goals and work packages and to prevent duplication by identifying the state-of-the art and current activities in the area (WP0.) Initially, the project proposal was divided into four work packages. The approach is iterative and modular in a way that more mature techniques can be tested with PoCs to speed up their adoption and additional techniques could be added as new work packages and strengthen each other if we do them jointly.

- WP1. Documenting statistical use-cases relevant for application of privacy-preserving techniques
- WP2. Secure Multiparty Computation (SMC) methods
- WP3. Homomorphic Encryption (HE) methods
- WP4. Identify opportunities for operationalization of methods and sharing of solutions

During the initial stage (WP0), these might be further scoped.
CONTEXT
The interest in the use of Machine Learning (ML) for official statistics is rapidly growing. For the processing of some secondary data sources (including administrative sources, big data and Internet of Things) it seems essential to look into opportunities offered by modern ML techniques, while also for primary data ML techniques might offer added value, as illustrated in the ML position paper mentioned above. Although ML seems promising there is only limited experience with concrete applications in the UNECE statistical community, and some issues relating to e.g. quality and transparency of results obtained from ML still have to be solved.

PROJECT OBJECTIVES
Based on mutual interest and building on existing national developments, the objective of the project is to advance the research, development and application of machine learning techniques to add value to the production of official statistics. To achieve this objective the Machine Learning (ML) will aim to:

- Investigate and demonstrate the value added of ML in the production of official statistics, where "value added" is increase in relevance, better overall quality or reduction in costs.
- Advance the capability of ML to add value to the production of official statistics.
- Advance the capability of national statistical organisations to use ML in the production of official statistics.
- Enhance collaboration between statistical organisations in the development and application of ML.

The objectives will be attained by:

- Conducting pilot studies in ML solutions in: (a) common statistical processes (classification and coding; edit and imputation); and (b) the use of alternate data sources (imagery or big data; sentiment and web).
- Researching and experimenting approaches to inform users on the quality of ML solutions, notably on accuracy
- Identifying best practices in the development and application of ML solutions, including organisational aspects
- Conducting the activities in groups of national organisations

STRATEGIC COMMUNICATION PROJECT OVERVIEW
CONTEXT
Official statistics are operating in a competitive and challenging environment – one that has changed significantly over the last twenty years. For traditional users of official statistics their values and importance is undisputed. Yet for the average citizen the digital and social media revolutions have meant that more and more people have instantaneous access to various data sources, outside official statistics. The 24/7 news cycle is reality, trust in government is decreasing and the fake news phenomenon is growing.

Now more than ever, timely and relevant data and stories produced by statistical organizations are essential to healthy democratic societies as they remain the only independent, impartial, trusted and reliable source of official statistics. For official statistics to be beneficial to society, policy debate, and decision-making they must be known, understood, communicated and used.

PROJECT OBJECTIVES
The objectives for the project are to provide statistical offices with:

- support in the development a strategic approach to communication and increase their capacity to review and renew their communication approach, methods and processes;
• with tools to increase their visibility, relevance and brand recognition; and
• tools to take a proactive approach to managing issues and reputation.

The outputs of the project will focus on enabling statistical offices to modernize their communications at the strategic level and help organizations look at communications strategies in a broader risk management and business continuity context. They include:

• Defining skillsets of a professional communication programme and organizational options for the strategic communication function within the statistical organization;
• Developing a Communication Maturity Model, including metrics and a description of how to use the model and examples of how the model can be used;
• Developing guidelines to create a communications strategy and its implementation plan (including examples);
• Developing the branding options that are most relevant for statistical organizations; and
• Establishing an issues management process including guidance and tools to support statistical organizations in times of issues or crisis management.

DATA ARCHITECTURE PHASE 2 PROJECT OVERVIEW

CONTEXT

Statistical organisations deal with many different data sources – each with their own set of characteristics. Statistical organisations need to find, acquire and integrate data from both traditional and new types of data sources in an ever increasing pace and under ever stricter budget constraints, while taking care of security and data ownership.

The 2017 HLG-MOS Data Architecture project developed the first version of the Common Statistical Data Architecture (CSDA). This Reference Architecture is a template for NSOs in the development of their own Enterprise Data Architectures.

The project will focus on providing a more robust version of the Common Statistical Data Architecture as a result of validation against a number of use-cases and integration with the outcomes from other related groups. It will also provide guidance on implementing the architecture.

PROJECT OBJECTIVES

The objectives of this project are:

• To complete the development of the Common Statistical Data Architecture, testing the reference architecture defined in 2017 against other use-cases
• To apply and validate the Data Architecture against the outcomes from other groups like UN-GWG, Data Integration project and groups working on statistical ontologies.
• To provide guidelines to support statistical organisations in using the Common Statistical Data Architecture.

Click for Output

Linked Statistical Metadata

CONTEXT

HLG-MOS has been jointly developing common models and vocabularies to prevent each organization from developing their own using different vocabularies for the same concepts. Linked open metadata provides the next step. Instead of each organization having to maintain and update their individual vocabularies, this would be made available and managed in a centralized way. This not only reduces costs but also prevents discrepancies in structural and reference metadata and semantic heterogeneity.

PROJECT OBJECTIVES

The main objective of the project is to demonstrate the usefulness of linked metadata for the statistical community and to acquire hands-on experience in that field. It is proposed to fulfill this objective by

Click for Output
constructing two concrete examples of linked metadata-based information systems: one aimed at improving the way that we disseminate core structural metadata, the other at supporting the advancement of the HLG vision by creating an harmonized and semantically enhanced information system grouping the main CSPA models and standards in a coherent and machine-actionable form. This will be achieved through three Work Packages:

- WP 1: Build a dissemination system for core structural metadata
- WP 2: Build an information system supporting the HLG vision
- WP 3: Project evaluation and sustainability plan

Click for Output

CONTEXT

There are many new opportunities created by data sources such as Big Data and Administrative data. These sources have the potential to provide more timely, more disaggregated statistics at higher frequencies than traditional survey and census data.

It is clear that NSOs are challenged by the capacities needed to incorporate new data sources in their statistical production process while at the same time companies have appeared exploiting these new sources to provide alternative statistics. If official statistics can’t find an answer to this, we are at risk of losing our unique position. We can, however, join forces and keep or even increase our value proposition by providing relevant, reliable and comparable data of high quality. NSOs are particularly well placed to integrate data from various sources and to use them to satisfy the needs of policy makers and other partners for data. It is thus time to intensify our efforts and commence working on it within the framework of an HLG project.

PROJECT OBJECTIVES

The main objectives of the project are to gain experience in data integration by pooling resources in joint practical activities and to translate experiences into general recommendations for data integration and to provide initial guidance for a quality framework. It will consist of these Work Packages:

- WP0: Data sets for common approaches
- WP1: Integrating Survey and Administrative Sources
- WP2: New data sources (such as big data) and traditional sources
- WP3: Integrating geospatial and statistical information
- WP4: Micro-Macro integration
- WP5: Validating Official Statistics
- WP Package A: Synthesis Lessons Learned from new working methods

Click for Output

CONTEXT

There are many new opportunities created by data sources such as Big Data and Administrative data. These sources have the potential to provide more timely, more disaggregated statistics at higher frequencies than traditional survey and census data.

It is clear that NSOs are challenged by the capacities needed to incorporate new data sources in their statistical production process while at the same time companies have appeared exploiting these new sources to provide alternative statistics. If official statistics can’t find an answer to this, we are at risk of losing our unique position. We can, however, join forces and keep or even increase our value proposition by providing relevant, reliable and comparable data of high quality. NSOs are particularly well placed to integrate data from various sources and to use them to satisfy the needs of policy makers and other partners for data. It is thus time to intensify our efforts and commence working on it within the framework of an HLG project.

PROJECT OBJECTIVES

For 2017, the project proposes to develop an online, adaptive, practical guide to Data Integration for Official Statistics which supports successful data integration projects; using lessons learnt within the project and in related work. Furthermore, to undertake more joint experiments in high priority practical interest areas. The project has identified a number of areas where working together should bring faster results than working alone. The following activities were identified:

- WPA: Develop an online, adaptive, practical guide to Data Integration for Official Statistics
- WP0-5: Further work on joint experiments in priority areas:
  - WP0: Data sets for common approaches
  - WP1: Integrating Survey and Administrative Sources
  - WP2: New data sources (such as big data) and traditional sources
  - WP3: Integrating geospatial and statistical information
  - WP4: Micro-Macro integration
  - WP5: Validating Official Statistics
- Align approaches for applying new data sources to integrated price measurement (WP0 and WP2)
- Create synthetic datasets for sandbox experiments (WP0)
- Develop practical guidance for integrating survey, administrative data and big data (including case studies) (WP1 and WP2)
- Develop practical guidance on integrating geospatial and statistical information (WP3)
- Develop practical guidance on using additional sources to validate official statistics (WP5)

Click for output

CONTEXT

To built on the momentum gained during the 2014 project a common shared Sandbox Computing environment was proposed to engage in collaborative research activities using various Big Data sources. Continuation of the experiments started in 2014 will allow to consolidate the technical skills. It will allow to test the production of multi-national statistics only basing on Big Data sources in a common environment.

PROJECT OBJECTIVES

The main goals of the 2015 project are:
PROJECT OBJECTIVES

The project has three main objectives:

• To identify, examine and provide guidance for statistical organizations to act upon the main strategic and methodological issues that Big Data poses for the official statistics industry.
• To demonstrate the feasibility of efficient production of both novel products and ‘mainstream’ official statistics using Big Data sources, and the possibility to replicate these approaches across different national contexts.
• To facilitate the sharing across organizations of knowledge, expertise, tools and methods for the production of statistics using Big Data sources.

CONTEXT

An statistical industry architecture will make it easier for each organization to standardize and combine the components of statistical production, regardless of where the statistical services are built. The Common Statistical Production Architecture (CSPA) is a framework about Statistical Services to create an agreed top level description of the ‘system’ of producing statistics which is in alignment with the modernization initiative. CSPA provides a template architecture for official statistics, describing:

• What the official statistical industry wants to achieve
• How the industry can achieve this, i.e. principles that guide how statistics are produced
• What the industry will have to do, compliance with the CSPA

PROJECT OBJECTIVES

To implement CSPA in practice by creating CSPA-compliant services that can be shared between processes and organizations (including resolving any specific licensing issues). To develop the resources necessary to support CSPA implementation, including training materials, and the proposed catalogue of services and other artefacts. To further test the applicability of the GSIM, and, if necessary, to suggest further refinements to that model for a possible future revision. The desired project outcomes are:

• Interoperability in Official Statistics through the sharing of processes and components
• Ability to find real/genuine collaboration opportunities
• Ability to make international decisions and investments
• Sharing of architectural/design, knowledge and practices

CONTEXT

A review of the 2014 CSPA project has identified that the technical implementation governance and support is a significant area for improvement, the AWG is proposing a HLG project for 2015 which would see the expansion of the role of the governance and support offered by AWG to cover implementation and the establishment of a Technical Coordination Committee to support NSI’s and NSO’s who are developing or implementing CSPA compliant statistical services.

PROJECT OBJECTIVES

The project has three main objectives:

• To extend the governance and support offered by the AWG to the implementation of CSPA compliant statistical services.
• To establish and maintain a new Technical Coordination Committee which will provide full technical guidance to implementing organizations and put in place technical implementation communities.
• To facilitate the transitioning of CSPA governance from HLG project governance arrangements to the Modernization Committee for Production and Methods, currently this is an identified risk.