Project Proposal: Statistical Data Governance Framework to Achieve Data Interoperability

Prepared by the Blue Skies Thinking Network / Supporting Standards Group/ Juan Muñoz

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

1. 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.
2. 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	
☒	Reduced costs
☒	Increased efficiency
☒	Reduced risks
☒	New capabilities to meet user needs

5 Which key priorities in the HLG-MOS Strategic Framework does the proposed project relate to?	
☐	Take cost out of our organizations to reinvest in more value-added areas
☐	Explore new areas collectively and leverage each other's' research investments in specific areas
☒	Provide whole of government data ecosystems based on international standards, for better estimates in key policy areas
☒	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 timetable

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 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