

Business Case for the Core Ontology for Official Statistics

This business case was prepared by Supporting Standards Group and is submitted to the HLG-MOS for their approval.

Type of Activity			
<input type="checkbox"/>	New project	<input type="checkbox"/>	New activity
<input type="checkbox"/>	Extension of existing project	<input checked="" type="checkbox"/>	Extension of existing activity

Purpose

As more and more statistical offices turn to semantic standards in order to formalize their data and metadata, it is time to build on our international core models to establish common foundations on which the different works can develop in a coherent way. This should be done using a formal framework that allows model integration, interoperability, activation and globally unique identification.

Initial work on a base OWL vocabulary for Official Statistics has begun. It suggests formal representations for the core concepts and models used in the official statistics domain, and builds on a set of well-known OWL vocabularies, which favors interoperability with similar initiatives in other domains (finance, legal, industry, etc.).

This activity aims to consolidate this material in order to create a common semantic model and vocabulary for official statistics. If successful, the activity's output will form the base for the next generation of HLG standards.

Description of the activity

Various works had been conducted in order to represent the GSBPM, the GSIM, parts of CSPA and the SDMX metadata model in OWL (see <http://ceur-ws.org/Vol-1551/article-06.pdf>, <http://ceur-ws.org/Vol-1654/article-03.pdf>, <http://ceur-ws.org/Vol-1654/article-06.pdf> and <https://linked-statistics.github.io/SDMX-Metadata/sdmx-metadata.html> respectively). They show that OWL provides an excellent tool for the integration of the models into a coherent and interoperable framework. Additional work was pursued at Insee in order to create a base ontology covering the core notions of statistical or support activities, as well as statistical organizations and products.

In 2019, the task team identified core concepts that are cutting across different models. The team produced definitions of these core concepts, properties and the specification of the links between this base ontology and the external vocabularies. Building on this work, further work is needed:

1. Develops the links with vocabularies in other models (e.g. CSDA capabilities). Articulate this base ontology with more sectorial and in-depth works existing on GSIM, CSPA or the SIMS model for quality metadata (based on the SDMX metadata model). This could lead to marginal evolutions of these models. The identification of the resources defined in these works will also be adapted in order to align with the naming policy defined in 2019.
2. Establish the management and governance of these different initiatives. The management activity should be tightly integrated with the management of the current HLG-MOS standards, but also deal with the connection to external vocabularies and corresponding standardization bodies.

The activity will produce the following deliverables:

- A URI policy for UNECE RDF resources
- The Core Ontology as an OWL/Turtle file in English/French/Spanish, validated by public review and referenced in LOV.

- An ontology description and users guide (English)
- New versions of previous works on GSIM, CSPA and the SDMX metadata model
- A document on the governance

The task team has drafted the version 1 of the ontology and completed internal review. The task team seeks extension of the activity with proposed timeline as below:

- January: Start of public review
- February-March: Work on articulation with work on GSIM and CSPA
- April-May: Work on governance
- June-July: Final version 1 of the ontology
- August-September: Document on governance
- October-November: Submission to HLG-MOS for adoption

Alternatives considered

The following alternatives were considered:

1. Do nothing: In this alternative, different NSIs will invest in duplicate work which will produce incoherent results. No globally recognized semantic modelling or naming policy for the base semantics of official statistics will be available for the community to build on.
2. Realize the activity outside of UNECE: In this alternative, the work is carried out under a different umbrella (standards organization like the W3C or the DDI Alliance, international interoperability frameworks like the ISA2 European program, etc.). In this case, there is a great risk that the results will not be clearly articulated with the HLG standard models, and in particular that it will be hard to coordinate their evolutions with those of the UNECE standards. There is also a clear risk that the results will not be fully accepted by the UNECE or its members.

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

Production of official statistics is the core business of statistical organisations and standard-based production is fundamental prerequisite for modernising organisations to be agile and innovative. The proposed activity will provide a formal representation for concepts, activities and products in the statistical organisations that could integrate the HLG-MOS models and facilitate collaboration with key stakeholders in government, academia and industry. The activity is tightly linked to other HLG activities related to standards as well as the 2016 HLG project: Implementing Modernstats Standards: Linked Open Metadata and the 2017 HLG project: Data Architecture project which prescribes the use of OWL for the semantic layer of the architecture, and of linked metadata in particular in the "metadata management" and "provenance and lineage" capabilities.

Proposed start and end dates

Start: January 2020

End: December 2020