

Background and rationale for a Core Ontology for Official Statistics at Insee

Guillaume Duffes (INSEE, France)

email

Abstract

Insee was involved in the development of the GSBPM, GSIM and CSPA OWL ontologies in 2016 and 2017.

In parallel, Insee started during the last two years to implement a reference and quality metadata system based on the European Single Integrated Metadata Structure (SIMS) that will not only take over from the legacy system (DDS) but also standardise areas that were missing from DDS.

The framework is RDF (Resource Description Framework), the main pillar of the Semantic Web.

The SIMS will be used as the main metadata report structure both for national and European/international purposes. The national version of the SIMS is enriched with a couple of additional items requested internally. The SIMS is underpinned by the SDMX information model.

Whereas the RDF flavour of the SIMS metadata report would be at the end of the production chain, it is essential to build a coherent and sound RDF/OWL representation of the conceptual model backing it. The main classes are culled from the SDMX information model related to reference metadata.

However, the ontology is supplemented by additional classes that could be deemed as “core statistical” objects (e.g. “StatisticalOperation” or “StatisticalOperationSeries”). This ontology is the one that will be reviewed and enhanced by the common semantic model and vocabulary for official statistics task team.

Insee is going to start developing a user interface to manage and produce the metadata reports based on the aforementioned Core Ontology for Official Statistics.