COOS project proposal 2018

Core Ontology for Official Statistics

Franck Cotton (Insee) & Monica Scannapieco (Istat)

modernistats BY HLG - MOS

Motivation

Most of our standard models have no formal representation

Some core concepts are missing

Some entire domains are missing (provenance)

The models are not integrated in a global framework

Our business semantics is isolated

modernstats by HLG - MOS

Project objectives

Produce and publish a Core Ontology for Official Statistics

- Covering base concepts like statistical activities, actors, products...
- Expressed in OWL
- Articulated with previous work in the Official Statistics domain
- Leveraging existing standard vocabularies
- Gone through public review
- Version-controlled on GitHub
- Referenced in vocabulary catalogs like LOV
- With clear governance

Why OWL?

The best framework for
- Model integration
- Knowledge representation

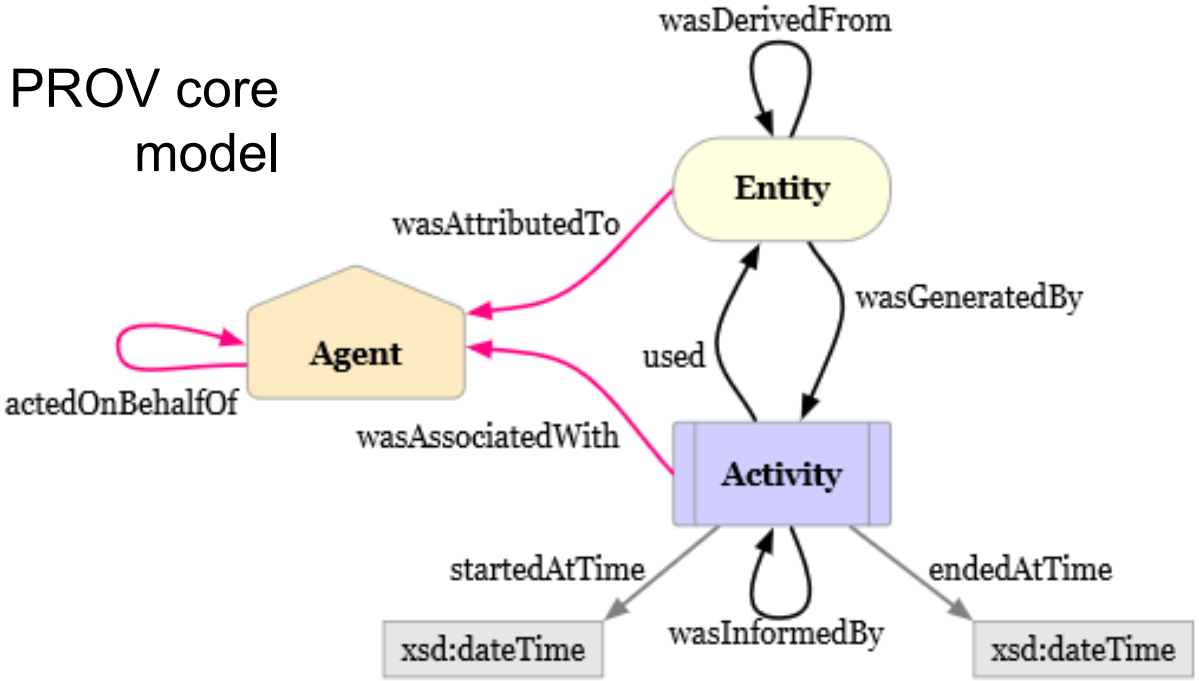
Previous or current work on
- Classifications
- Conceptual models (GSBPM, GSIM, CSPA)
- Data and metadata models (DDI, SDMX)

Gives access to a wealth of existing vocabularies

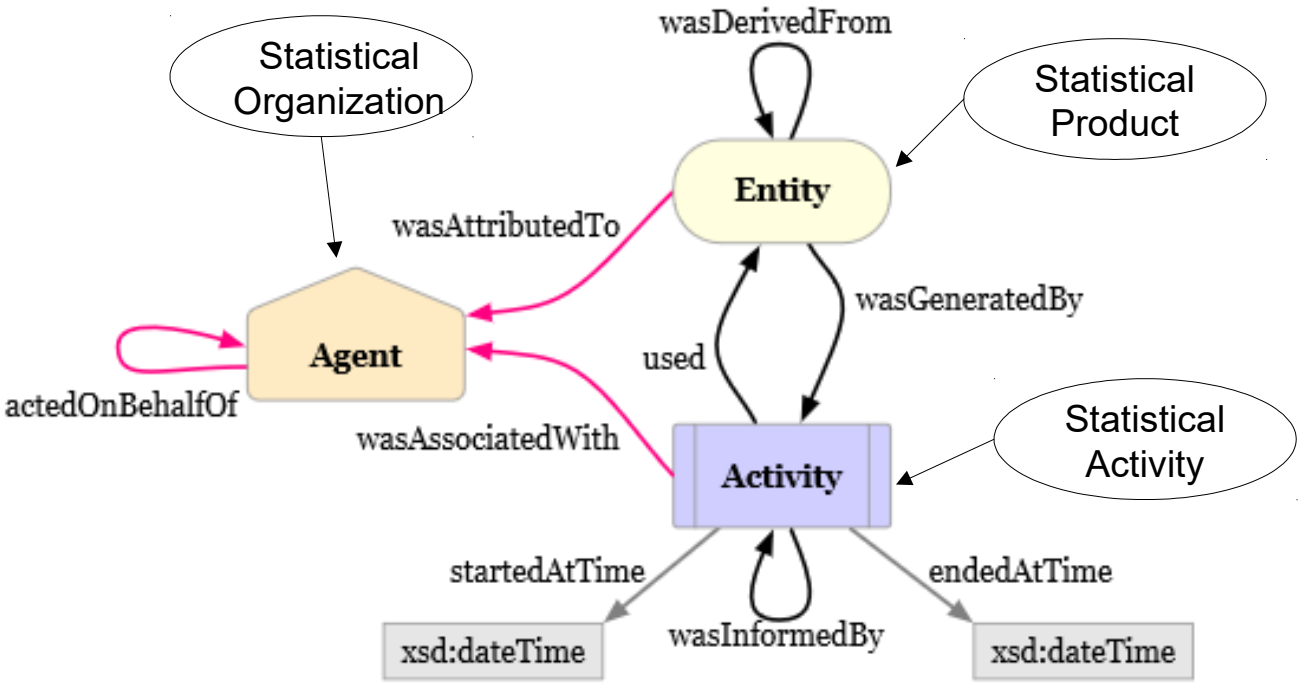
modernstats by HLG - MOS

Why OWL?
Example of PROV

PROV core model



Why OWL?
Example of PROV

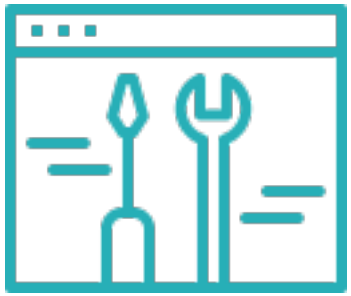


Work Packages

WP1: Produce and publish core ontology

Activities:

- Organize group of authors/reviewers for draft version
- Draft naming policy and ontology
- Liaise with stakeholders (LOS ESSnet, SEMIC, DDI Alliance…)
- Draft ontology guide and use cases
- Organize public review and process feedback
- Finalize ontology and documents
- Present in SemStats workshop
- Seek HLG approval
- Publish and reference

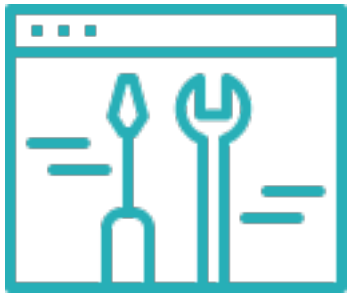


Work Packages

WP2: Adapt existing works to the core ontology

Activities:

- Review GSIM ontology
- Review CSPA ontology
- Review SDMX-MM & SIMS ontology
- Adapt all vocabularies to naming policy
- Publish new versions

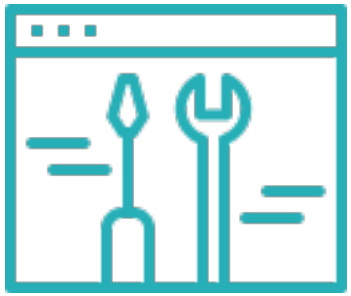


Work Packages

WP3: Define governance of statistical ontologies

Activities:

- Specify management and governance needs
- Liaise with interested bodies
- Draft governance structure and policy document
- Validate and publish document



Deliverables

- A URI policy for UNECE RDF resources
- The Core Ontology
  - OWL/Turtle file in English/French/Spanish
  - validated by public review
  - and referenced in a vocabulary catalog.
- 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

Benefits

A step forward towards common terminology

Global interoperability (e.g. inter-NSI search for products)

Uniform identification of concepts

From clickable models to actionable models

From related models to integrated models

New rich and standard semantics

What the users say

Jenny L., Oslo
"The base ontology work that is proposed will improve the coherence of the ModernStats standards, increase the interoperability with other international, national and branch standards and reduce the costs required for modernisation. I definitely think this work is needed and could save a lot of organizations from fumbling in the dark."

Daniel G., Washington, DC
"Terminology management is fundamental and key to making GSIM work in practice. Rendering terms using semantic technology can greatly enhance their usefulness, especially with tying them to other metadata. I very much want to be part of this."

Monica S., Rome
"Working on a base ontology for ModernStats standards means building bridges between our world and world-wide standards by strenghtening relationships of statistical institutes with public administrations as well as private organizations."

Franck C., Paris
"Wot!?"

modernstats by HLG - MOS