

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

ModernStats World Workshop

21-22 October 2024, Geneva

Enhancing Interoperability and Transparency through Linked Open Data Standards: Lessons Learned from the ESS LOD Community of Practice

Speaker: Christine Laaboudi-Spoiden, Eurostat

Author(s): Luca Gramaglia, Christine Laaboudi-Spoiden (Eurostat)

Abstract

This abstract highlights the ongoing work of the European Statistical System (ESS) LOD Community of Practice (LOD CoP), initiated by Eurostat in April 2023.

The LOD CoP comprises 11 national statistical offices (NSOs) from the ESS (Netherlands, Latvia, Croatia, Hungary, Spain, France, Italy, Finland, Denmark, and Norway), Canada, and the Food and Agriculture Organization (FAO), focusing on developing use cases and recommendations for improving interoperability and transparency in statistical data dissemination. The use cases are based on pilot datasets selected from three domains (EU and Canadian Census, Labour Force Survey (LFS), and High-Value Datasets), covering both EU and non-EU datasets, economic activities (NACE/NAICS) and related but non-identical variables (Sex vs Gender, LFS status vs Activity status).

To address the challenges of enhancing the FAIR principles for the production and dissemination of official statistics, the LOD CoP is working on the following aspects and applying them to the pilot datasets:

- Linking related datasets to their structural metadata by identifying available useful properties in the Data Cube Vocabulary (DCV) or STAT-DCAT and other relevant ontologies. As part of this topic, the LOD CoP aims to identify cases where no common resources or common controlled vocabularies for structural metadata currently exist but could be beneficial in the future.
- Linking statistical classifications and code lists using XKOS, a SKOS extension for representing statistical classifications. The LOD CoP evaluates properties from SKOS/XKOS and other relevant ontologies for their effectiveness in linking classifications, code lists, and concepts/variables.
- Developing a proof-of-concept API for querying statistical classifications from different SPARQL endpoints (ESTAT, FAO, Insee, CBS Statistics Netherlands). The API supports various formats/serializations, including JSON, CSV, XML, RDF/XML, RDF/Turtle, and JSONLD.
- Exploring linking datasets in Open Data Catalogue using DCAT and other ontologies (SIMS/ESMS, Schema.org, DATACITE, COOS) through shared properties. The team identifies valuable properties in the DCAT ontology and other relevant ontologies for linking datasets, as well as cases where common resources/controlled vocabularies are lacking but could be beneficial.

The LOD CoP's outcomes will yield a set of guidelines and recommendations detailing strategies for linking datasets and structural metadata, with practical examples supporting the application of these guidelines.