

Meeting of the UN-GGIM Expert Group on the integration of statistical and geospatial information –Lisbon 24 May 2015

Background document for session 6

Metadata interoperability based on DCAT

Date: 12/5/2015

Version: 0.1

Authors: PETRI

Revised by:

Approved by:

Public:

Reference Number:

Document History

Version	Date	Comment	Modified Pages
0.1	12/05/15	Document created by PETRI	All

1 PURPOSE OF THE DOCUMENT

This document responds to the request from the Expert Group on the integration of statistical and geospatial information to progress the topic of data integration and interoperability between statistical and geospatial information by harmonising existing metadata standards.

This paper proposes DCAT-AP as a candidate to achieve interoperability between two existing standards (SDMX) for statistics and ISO19115/19119 for geospatial information.

2 BACKGROUND

Several metadata standards have been developed and implemented in the geospatial and statistical communities (INSPIRE, ISO19115/19119, SDMX, DDI) but so far no common or bridging standard has been proposed.

As part of the INSPIRE implementation process, a first attempt to map SDMX and INSPIRE metadata elements and thus achieve interoperability was made by the JRC in 2010, with mixed results. The report found many communalities between statistical and geospatial metadata standards but also many discrepancies and ambiguities in encodings, code lists and also definition of elements. Thus the findings of the direct mapping exercise were not definite but the study concluded that full convergence of either standard to reach direct interoperability with the other was probably difficult to achieve due to the different requirements of the statistical and geospatial community. As a possible way out, it was suggested developing a bridging standard supporting the discovery use case across data portals.

It should be noted that communalities in terms of metadata elements also extend with other standards implementing the discovery use case of information in data catalogues, most importantly with Dublin Core and as a result a generic Dublin Core based metadata standard was probably a good candidate for such a bridging standard.

In the meantime the concept of Open Data Portals has significantly gained grounds and a corresponding metadata standard building on Dublin Core and describing data for these portals has been developed by the W3C: DCAT. Relatively fast the potential of DCAT to connect geospatial and statistical metadata across portals was discussed in both communities in Europe and related projects using e-government and interoperability programs have been launched.

3 OVERVIEW OF STANDARDS AND THEIR APPLICATION PROFILES

The **Statistical Data and Metadata eXchange (SDMX)** initiative was launched in 2001 by seven organisations working on statistics at the international level: the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat, the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), the United Nations Statistical Division (UNSD) and the World Bank. These seven organisations act as the sponsors of SDMX. The stated aim of SDMX was to develop and use more efficient processes for exchange and sharing of statistical data and **metadata** among international organisations and their member countries. To achieve this goal, SDMX provides standard formats for data and **metadata**, together with content guidelines and an IT architecture for exchange of data and metadata.

ISO 19115:2003, ISO 19139 and 19115:2014¹ are standard of the International Organization for Standardization (ISO), defines how to describe geographical information. ISO 19139 provides the XML Schema implementation schema for ISO 19115. ISO19115:2014 supersedes ISO19115:2003. At the time of writing this document, no XML binding for ISO19115:2014 has been defined yet (expected 2015).

DCAT² is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. Its basic use case is to enable cross-data portal search for data sets and make public sector data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of datasets among data portals. By using DCAT to describe datasets in data catalogues, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogues. It further enables decentralized publishing of catalogues. DCAT is widely used as a metadata specification in Open Data portals. DCAT is a W3 recommended specification.

¹ ISO (International Organization for Standardization), "ISO 19115:2003 Geographic information -- Metadata," 2003. [Online]. Available: http://www.iso.org/iso/catalogue_detail?csnumber=26020.

² W3C, "Data Catalog Vocabulary (DCAT)," 2014. [Online]. Available: <http://www.w3.org/TR/vocab-dcat/>. [Accessed 2015]

The DCAT Application profile for data portals in Europe (DCAT-AP)³ is a specification based on the Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe.

The application profile is a specification for metadata records to meet the specific application needs of data portals in Europe while providing semantic interoperability with other applications on the basis of reuse of established controlled vocabularies (e.g. EuroVoc) and mappings to existing metadata vocabularies (e.g. Dublin Core, SDMX, INSPIRE metadata, etc).

To further improve the data exchange between community specific metadata vocabularies and open data portals, the European Commission has launched the development of corresponding application profiles of DCAT.

GeoDCAT-AP is an extension of the DCAT-AP for describing geospatial datasets, data series, and services. This extension will provide a RDF syntax for the union of metadata elements of ISO19115 based metadata standards including the ISO 19115/19119:2003 Core.

At this stage ISO19115:2003 will form the basis for the development of GeoDCAT-AP, but a later revision might also take into account the freshly adopted standard ISO19115:2014, currently lacking concrete implementations.

GeoDCAT-AP-AP will make it easier to share descriptions of spatial datasets between spatial data portals and general data portals, and thus help increase public and cross-sector access to such high value datasets. The datasets could include:

- Datasets on the INSPIRE Geoportal.
- Datasets on national SDIs.
- General geospatial datasets.

4 ROADMAN FOR ON-GOING STANDARD DEVELOPMENT

In February 2015, the ISA Programme of the European Commission started the revision process of the DCAT Application Profile. The revised application profile will be submitted for public review in April 2015.

³ European Commission, ISA Programme, “DCAT Application profile for data portals in Europe,” [Online]. Available: <https://joinup.ec.europa.eu/node/63567>

In March 2015, the ISA Programme of the European Commission started working on a GeoDCAT-AP extension. The working group is expected to propose a draft in May 2015, followed by a public review. GeoDCAT-AP is expected to be endorsed by the end of 2015.

The StatDCAT-AP project has just been discussed between European bodies responsible for DCAT and Eurostat as one of the main sponsors of SDMX in Europe. A concrete work plan has not been established yet but based on experiences from GeoDCAT-AP it can be expected that the development of a SDMX interoperable extension of DCAT-AP will require about one year.

5 PROPOSAL

It should be noted first that the discussion around interoperable metadata standards is restricted to the implementation of the publication of metadata in data portals and their discovery including federated search across data portals.

Given their widespread implementation by public authorities, and their support by standardisation bodies like the ISO and OGC, it is suggested to limit the task of finding an interoperability path for statistical and geospatial metadata to SDMX for the statistical and ISO19115 for the geospatial community.

It is further suggested to propose DCAT-AP as the overarching vocabulary to achieve interoperability between statistical and geospatial metadata. This would have the advantage that geospatial and statistical resources following this approach would automatically be ready for publication in Open Data Portal implementing DCAT. The recent G8 Charta on open data explicitly makes reference to geospatial and statistical information as important contributions to open data. Many of the G8 countries will adopt DCAT as the standard for their national Open Data Portal, and the EU will do the same.

Under this working assumption the main challenge will be to develop interoperability paths from ISO19115 and SDMX to DCAT-AP respectively. This can be either done by mapping ISO and SDMX elements directly to DCAT-AP elements or by using GeoDCAT-AP and StatDCAT-AP.

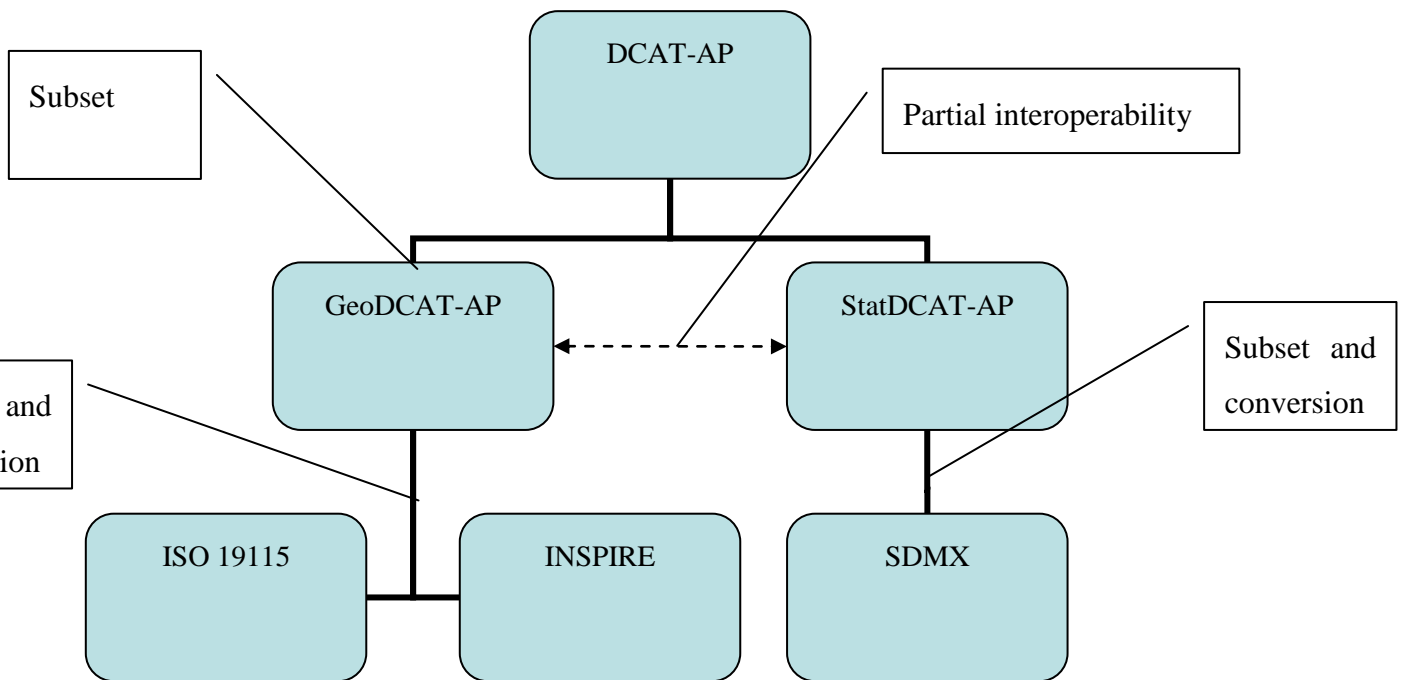


Figure 1: Relationship between geospatial and statistical metadata standards and DCAT.

As outlined earlier, several interoperability and e-government activities at EU level are already working on the conceptual and technical challenges to convert geospatial and statistical metadata into DCAT. The working groups are composed of leading specialists and therefore high quality deliverables taking into account all risks and side-effects can be expected.

From a technical perspective conversion rules to RDF syntax would allow Member States to maintain their collections of ISO19115 and SDMX-relevant datasets, while at the same time publishing these collections on DCAT-AP-conform data portals. A lossless conversion to RDF syntax allows additional metadata elements to be displayed on general-purposed data portals, provided that such data portals are capable of displaying of additional metadata elements. Furthermore, data portals frequently are complemented by a triple store, making that the full set of GeoDCAT-AP and StatDCAT-AP metadata can be queried through a SPARQL endpoint.

It can be expected that the European Union projects dealing with the extensions of DCAT-AP to GeoDCAT-AP and StatDCAT-AP will also provide the tools to carry out the conversion, including detailed documentation. It can also be expected that providers of catalogue software will implement conversion features into their products. As an example the developers of GeoNetwork have already included an ISO19115 to DCAT converter into their update plans.

6 NEXT STEPS

The Expert Group is invited to discuss the proposal of using DCAT-AP and its thematic implementations GeoDCAT-AP and StatDCAT-AP and assess the approach from a global perspective. If approved by the working group, the proposal should be proposed to the Committee of Experts for endorsement.

The members of the expert group are also invited to participate in the review process of DCAT-AP and GeoDCAT-AP. This involvement will be facilitated by Eurostat.

Last the members of the working group are invited to participate in the development of StatDCAT-AP. Again this involvement will be facilitated by Eurostat.