

GeoDCAT-AP: Representing geographic metadata by using the “DCAT application profile for data portals in Europe”

Andrea Perego, Vlado Cetl, Anders Friis-Christensen, Michael Lutz

European Commission, Joint Research Centre (JRC), first-name.last-name@ec.europa.eu

Abstract: GeoDCAT-AP is a metadata profile aiming to provide an RDF-based representation of geospatial metadata compliant with the *DCAT application profile for European data portals* (DCAT-AP), specifically designed to enable the sharing of geospatial metadata, in particular those available via the INSPIRE infrastructure. In this paper, we provide an overview of GeoDCAT-AP, and we outline the lessons learnt, in particular with respect to the issues concerning the cross-domain integration of the geospatial platform with other data infrastructures.

Disclaimer: *The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.*

1 Introduction

GeoDCAT-AP (European Commission, 2015a) is a metadata profile for the RDF-based representation of geospatial metadata, in particular those available via the INSPIRE infrastructure and, more in general, ISO 19115 metadata records. Developed in the framework of the EU Programme “Interoperability solutions for public administrations, businesses and citizens” (ISA²),¹ GeoDCAT-AP is defined as a domain-specific extension of the “DCAT application profile for European data portals” (DCAT-AP). DCAT-AP (European Commission, 2015) is an application profile of the W3C Data Catalog Vocabulary (DCAT) (W3C, 2014), used as a cross-domain and cross-platform metadata interchange format for data catalogues operated in the EU.

The final specification of GeoDCAT-AP (version 1.0) has been released in December 2015, and it is currently implemented in a number of geospatial and general-purpose data catalogues (European Commission, 2015b), one of them being the European Data Portal,² the EU-wide data catalogue released in November 2015, providing a single discovery and access point for data published in the EU.

Besides providing an overview of GeoDCAT-AP, in this paper we illustrate the lessons learnt and the issues concerning the cross-domain integration of the geospatial platform with other data infrastructures. Moreover, we will present the methodology

¹ <https://ec.europa.eu/isa2/>

² <http://data.europa.eu/europeandataportal/>

adopted to facilitate the implementation of GeoDCAT-AP on existing geospatial data catalogues, along with the existing implementation experiences. Finally, we will provide an outline of the activities under-way to enhance the GeoDCAT-AP specification, and to facilitate its cross-domain integration with the existing data infrastructures.

2 Background

The motivation behind the development of GeoDCAT-AP is twofold:

1. Agree upon a common RDF representation of INSPIRE metadata

The Resource Description Framework (RDF) (W3C, 2014a) is a Semantic Web technology which is increasingly being used in Europe for describing government data, and to provide an alternative representation of INSPIRE metadata. Without a harmonised RDF mapping of INSPIRE metadata, interoperability would be lost.

2. Facilitate cross-sector sharing of INSPIRE metadata

INSPIRE metadata are already being harvested by and published in cross-domain data catalogues at the national and/or regional level, as well as by the European Data Portal, which uses DCAT-AP as a metadata interchange format. The availability of INSPIRE metadata in an DCAT-AP compliant format would ensure their consistent sharing across sectors and platforms.

For example the co-existence of GeoDCAT-AP with another DCAT profile for statistical data (StatDCAT-AP, see parallel paper on StatDCAT by Pellegrino) is a further step towards a standard mapping between RDF, statistical standards and geospatial standards to improve interoperability and to integrate data production systems.

It is important to note that the GeoDCAT-AP specification does not replace the INSPIRE Metadata Regulation (European Commission, 2008) nor the INSPIRE Metadata Technical Guidelines based on ISO/TS 19139:2007 (European Commission, 2017). Its purpose is to give owners of geospatial metadata the ability to share them on general data portals, thereby making geospatial information better searchable across borders and sectors, in a way that would not require dramatic changes in their infrastructures.

3 Methodology

The mappings defined in GeoDCAT-AP have the primary objective of providing a harmonised DCAT-AP representation of geospatial records, covering all the elements of INSPIRE metadata and of the core profile of EN ISO 19115:2003.

A revision of GeoDCAT-AP to extend its scope to other metadata elements and align it with the latest version of ISO 19115 (ISO, 2014) could be part of future activities. Notably, a preliminary analysis on the required changes has been included in appendix to the GeoDCAT-AP specification (European Commission, 2015a).

It is worth mentioning that one of the key criteria followed during the development of GeoDCAT-AP was to base it as much as possible on existing practices. The objective was to build upon related work on the RDF-based representation of geospatial metadata, avoid a negative impact on existing implementations, and facilitate adoption.

The defined mappings have been organised into two groups, referred to as *core* and *extended*. More precisely, the former group (core) includes mappings for INSPIRE / EN ISO 19115:2003 metadata elements already supported by DCAT-AP, and it is meant to enable the harvesting and re-use of geospatial metadata records through DCAT-AP-conformant applications and services, including data portals and APIs. By contrast, the latter group (extended) includes all the mappings defined in GeoDCAT-AP, and addresses those use cases requiring an RDF representation covering also those metadata elements of INSPIRE and the core profile of EN ISO 19115:2003 more specific to geospatial data (e.g., coordinate reference systems, spatial resolution, as well as services).

An overview of the INSPIRE / ISO 19115 elements covered in GeoDCAT-AP is provided in Table 1. In the table, the “DCAT-AP” column denotes the “core” group of mappings defined in the GeoDCAT-AP, whereas the “GeoDCAT-AP” column denotes the “extended” group. For more details, we refer the reader to the GeoDCAT-AP specification (European Commission, 2015a).

Table 1 INSPIRE / ISO 19115 metadata elements covered in GeoDCAT-AP.

INSPIRE	ISO 19115:2003 Core	DCAT-AP	GeoDCAT-AP
Metadata point of contact	Metadata point of contact		Yes
Metadata date	Metadata date stamp	Yes	Yes
Metadata language	Metadata language	Yes	Yes
	Metadata character set		Yes
	Metadata file identifier		Yes
	Metadata standard name		Yes
	Metadata standard version		Yes
Resource title	Dataset title	Yes	Yes
Temporal reference - Date of creation / publication / last revision	Dataset reference date	Partially – creation date not included	Yes
Resource abstract	Abstract describing the dataset	Yes	Yes
Resource language	Dataset language	Yes	Yes

INSPIRE	ISO 19115:2003 Core	DCAT-AP	GeoDCAT-AP
Topic category	Dataset topic category		Yes
Geographic bounding box	Geographic location of the dataset (by four coordinates or by geographic identifier)	Yes	Yes
Character encoding	Dataset character set		Yes
Temporal reference - Temporal extent	Additional extent information for the dataset (vertical and temporal)	Yes	Yes
Lineage	Lineage	Yes	Yes
Spatial representation type	Spatial representation type		Yes
Encoding	Distribution format	Yes	Yes
Spatial resolution	Spatial resolution of the dataset		Yes
Responsible organisation	Dataset responsible party	Partially – only 3 of the 11 responsible party roles are supported	Yes
Resource locator	On-line resource	Yes	Yes
Coordinate reference system; Temporal reference system	Reference system		Yes
Conformity		Yes	Yes
Resource type		Partially – only datasets, series and discovery services	Yes
Spatial data service type			Yes
Keyword		Partially – only for datasets and dataset series	Yes
Coupled resource			Yes
Unique resource identifier		Yes	Yes
Conditions for access and use		Yes	Yes
Limitations on public access		Yes	Yes
Maintenance information - Maintenance and update frequency		Yes	Yes
Data quality – Logical consistency – Topological consistency – Conformance results			Yes
Data quality – Logical consistency – Conceptual consistency – Conformance results			Yes
Data quality – Logical consistency – Domain consistency – Conformance results			Yes

4 Implementation

The activities concerning the design and implementation of GeoDCAT-AP have been carried out in parallel in order to ensure the feasibility and soundness of the adopted solutions. Notably, one of the main requirements taken into account was the ability to implement GeoDCAT-AP on top of the existing geospatial catalogues, limiting as much as possible the impact on the underlying infrastructure.

The GeoDCAT-AP reference implementation addresses this issue by providing mechanisms that can be easily integrated on top of standard geospatial catalogue services. More precisely, they provide the ability to serve GeoDCAT-AP metadata along with INSPIRE / ISO 19115 records, without changing the service interface, and by supporting in addition HTTP content negotiation.

The GeoDCAT-AP reference implementation, available as open source, is documented in (European Commission, 2015b), along with third party implementations carried out by different organisations across Europe, as well as by companies maintaining widely used catalogue platforms, as GeoNetwork and CKAN.

5 Lessons learnt

The development and implementation activities of GeoDCAT-AP have demonstrated that the real challenges for cross-sector interoperability are not technical, but rather concern governance issues.

In particular, there are two aspects which are most relevant:

1. *There is a lack of standards and best practices on how to model in RDF some important aspects of spatial information.*

Relevant examples include spatial / temporal resolution, coordinate reference systems, and data quality. These gaps can only be addressed by a close collaboration between stakeholders and standard bodies across the geospatial and Semantic Web domains.

2. *The use of global and persistent identifiers (in particular, HTTP URIs) is not common practice in geospatial data.*

This situation limits the exploitation of one of the key features of Semantic Web technologies, i.e., the ability to link data. However, persistent and global identifiers would be beneficial to the geospatial platform itself, especially when data and services are part of a federated architecture—as it is the case in INSPIRE.

These issues do not prevent the effective exploitation of GeoDCAT-AP, and they can be solved on an *ad hoc* basis. However, the widespread adoption of best practices to

address them would be beneficial to the interoperability of geospatial data and metadata across sectors and platforms.

6 Contribution to standard bodies

GeoDCAT-AP has been brought to the attention of standard bodies, in particular the Open Geospatial Consortium (OGC) and the World Wide Web Consortium (W3C), in order to be consolidated with further review and to contribute the identified issues—including those outlined in Section 5.

In particular, some of the solutions defined in GeoDCAT-AP for modelling spatial and data quality information have been integrated in the W3C Data Quality Vocabulary (DQV) (W3C, 2016) and in the “Spatial Data on the Web Best Practices” specification (W3C/OGC, 2017) developed by the joint W3C/OGC Spatial Data on the Web Working Group. Moreover, discussions are under-way in OGC about the possibility of adopting GeoDCAT-AP as a community standard or a best practice. Finally, GeoDCAT-AP is listed among the reference specifications in the charter of the recently launched W3C Data Exchange Working Group (W3C, 2017), whose objectives include a revision to the DCAT vocabulary based on implementation and deployment experiences.

7 Conclusions

The GeoDCAT-AP specification is complemented by a suite of tools that, without requiring dramatic changes in the existing infrastructure, enables owners of geospatial metadata to publish an alternative and harmonised representation of INSPIRE / ISO 19115 records which facilitates their sharing and re-use across sectors and platforms.

Although, in its current version, GeoDCAT-AP does not provide a full coverage of ISO 19115 and it is not based on the latest version of this standard, future revisions may address these gaps based on requirements gathered from implementors and stakeholders. This feedback collection process, which started right after the release of the GeoDCAT-AP specification, has already resulted in refinements of the GeoDCAT-AP implementation to cover additional requirements—e.g., modelling multilingual metadata records—and in side activities, in particular, the definition of semantic alignments across geospatial thesauri and general purpose ones.³

It is worth noting that the development and implementation of GeoDCAT-AP have been also beneficial in identifying the critical issues for cross-sector interoperability of geospatial metadata, which cannot be addressed only by technical solutions, but require actions also in the current data governance practices.

³ The consolidated version of some of these alignments is maintained in the Metadata Registry of the EU Publications Office: <http://publications.europa.eu/mdr/eurovoc/>.

References

- European Commission (2008) “Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata (Text with EEA relevance)”, OJ L 326, 4.12.2008, p. 12–30, <http://data.europa.eu/eli/reg/2008/1205/oj>
- European Commission (2015) “DCAT application profile for data portals in Europe. Version 1.1”, <https://joinup.ec.europa.eu/node/146653>
- European Commission (2015a) “GeoDCAT-AP: A geospatial extension for the DCAT application profile for data portals in Europe. Version 1.0”, <https://joinup.ec.europa.eu/node/139283>
- European Commission (2015b) “GeoDCAT-AP Implementations”. <https://joinup.ec.europa.eu/node/144843>
- European Commission (2017) “Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007”, <http://inspire.ec.europa.eu/id/document/tg/metadata-iso19139>
- ISO (2003) “ISO 19115:2003: Geographic information – Metadata”, <https://www.iso.org/standard/26020.html>
- ISO (2014) “ISO 19115-1:2014: Geographic information – Metadata – Part 1: Fundamentals”, <https://www.iso.org/standard/53798.html>
- W3C (2014) “Data Catalog Vocabulary (DCAT)”, <https://www.w3.org/TR/vocab-dcat/>
- W3C (2014a) “RDF 1.1 Concepts and Abstract Syntax”, <https://www.w3.org/TR/rdf11-concepts/>
- W3C (2016) “Data on the Web Best Practices: Data Quality Vocabulary”, <https://www.w3.org/TR/vocab-dqv/>
- W3C (2017) “Dataset Exchange Working Group Charter”, <https://www.w3.org/2017/dxwg/charter>
- W3C/OGC (2017) “Spatial Data on the Web Best Practices”, <https://www.w3.org/TR/sdw-bp/>