GeoDCAT-AP
Cross-sector sharing and re-use of geospatial metadata

Michael Lutz, Andrea Perego

INSPIRE 2016

Barcelona, 30 September 2016
INSPIRE, PSI & Open Data

EU activities and regulations

- Communication on Open Data (COM(2011)882)

- Revision to the Decision governing re-use of Commission's documents (2011/833/EU)

- Revision to the Directive on re-use of Public Sector Information (2013/37/EU)

- EU Programme on Interoperability Solutions for European Public Administrations (ISA)
The European Data Portal

- A single access point for the cross-border discovery of datasets available from EU data portals
- It will harvest any kind of dataset metadata
- The **geospatial component** plays an important role in the underlying infrastructure
- Complemented with additional services – as an **EU gazetteer**
- Based on harmonised and cross-domain metadata interchange formats
What is GeoDCAT-AP

- Geospatial extension to DCAT-AP (DCAT application profile for data portals in Europe)
  - DCAT-AP is a metadata profile meant to provide an interchange format for data portals operated by EU Member States
  - DCAT-AP is based on and compliant with the W3C Data Catalog (DCAT) vocabulary
- Developed in the framework of the EU Programme “Interoperability Solutions for European Public Administrations” (ISA)
- GeoDCAT-AP is meant to provide a DCAT-AP compliant representation for the set of metadata elements included in
  - INSPIRE metadata
  - The core profile of ISO 19115:2003
INSPIRE & GeoDCAT-AP: Why?

Agree upon a common RDF representation
- RDF is increasingly being used as an alternative representation of INSPIRE metadata
- Without a harmonised INSPIRE-to-RDF mapping, metadata interoperability is lost

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
- INSPIRE metadata are harvested and published also on the European Data Portal, which uses DCAT-AP as a metadata interchange format
GeoDCAT-AP: Objectives

- The GeoDCAT-AP specification does not replace the INSPIRE Metadata Regulation nor the INSPIRE Metadata Technical Guidelines based on ISO 19115:2003 and ISO 19119

- Its purpose is to give owners of geospatial metadata the possibility to achieve more by providing an additional RDF syntax binding

- Its basic use case is to make spatial datasets, data series, and services searchable on general data portals, thereby making geospatial information better searchable across borders and sectors
GeoDCAT-AP: Current status

- Final specification (**GeoDCAT-AP 1.0**) released in December 2015:
  https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/geodcat-ap-v10

- **Reference implementation** (XSLT-based):

- **GeoDCAT-AP implementations**, including **CSW-based** ones, are already available:
  https://joinup.ec.europa.eu/node/144843
GeoDCAT-AP API & Sandbox

GeoDCAT-AP API

Proof-of-concept of the implementation of GeoDCAT-AP using the standard CSW interface, and supporting multiple RDF serialisations (including HTML+RDFa) and HTTP content negotiation

http://geodcat-ap.semic.eu:8890/api/

INSPIRE GeoDCAT-AP Sandbox

Faceted browser / SPARQL endpoint for records harvested from the INSPIRE Geoportal and transformed into GeoDCAT-AP

GeoDCAT-AP API

**Output Schema**: DCAT-AP


Output format: RDF/XML

**Usage notes**

Copy & paste the URL of a file or of a CSW request returning ISO 19139 records.

Supported CSW request types: GetRecords, GetRecordById.

Supported CSW outputschema: http://www.isotc211.org/2005/gmd

**NB**: The current version of the API supports only CSW calls using the GET HTTP method.

A description of the GeoDCAT-AP API is available on the API's Stash repository.
The GeoDCAT-AP API

- The main objective of this prototype is to provide a working example on how GeoDCAT-AP can be supported without changing the existing infrastructure, based on INSPIRE / ISO 19115 metadata and CSWs.

- The GeoDCAT-AP API provides also an example on how to enable traditional HTTP functionalities in CSWs, as content negotiation, allowing a better integration with non-geospatial services and APIs.

- Another key objective is to show how the existing catalogue infrastructure can be used to publish metadata in a way that increases their visibility on the Web, by following standards as HTML+RDFa and Search Engine Optimisation (SEO) techniques.

- Notably, these are some of the issues addressed by the Geonovum testbed “Spatial Data on the Web”:

http://geo4web-testbed.github.io/topic4/
GeoDCAT-AP: Open issues

- Lack of vocabularies able to model some information – in particular:
  - reference systems
  - spatial / temporal resolution
  - data quality

- Limited use of global & persistent identifiers (in particular, HTTP URIs) in the original metadata records. Some consequences:
  - impossible to (safely) implement incremental harvesting
  - impossible to unambiguously identify resources referred to from metadata (keywords from reference vocabularies, responsible organisations, licences, etc.)
  - faceted search (esp., language neutral) cannot be effectively implemented
GeoDCAT-AP: Ongoing work

Although the GeoDCAT-AP WG is formally closed, their members are contributing, on a volunteer basis, to carry on additional implementation activities. These include:

- Alignment of INSPIRE spatial data themes, ISO topic categories and MDR data themes (i.e., the ones used in DCAT-AP) – This work is documented here:

- Testing and enhancing the mappings defined in GeoDCAT-AP. This work will be used as a basis for possible future revisions to GeoDCAT-AP
Another issue concerns how to model dataset distributions available via services / APIs (WMS, WFS, WCS, as well as SPARQL endpoints)

This work is carried out in the framework of the DCAT-AP Implementation Guidelines WG, and may lead to a standardised and interoperable representation for any kind of service-based data access – not limited to the geospatial platform

The current proposal is to use OpenSearch to describe the service and the request parameters

For geospatial services, the OpenSearch document can be automatically generated from a GetCapabilities document, but a standardised mapping need to be designed
Profile-based content negotiation

- This is basically about **HTTP content negotiation based also on “profiles”** – e.g., being able to request (meta)data in a given schema (ISO 19115, Dublin Core, etc.), and not only in a given format (XML, RDF, etc.)

- CSW (but also by OAI-PMH) already supports the ability to choose the output schema with a specific request parameter. However, the rationale is defining a **standardised approach for any type of service run on the Web**

- This issue might be addressed in the framework of the W3C/OGC Spatial Data on the Web WG, or by future work at W3C on DCAT – on this topic, see also the “Smart Descriptions & Smarter Vocabularies” (SDSVoc) Workshop: 
  https://www.w3.org/2016/11/sdsvoc/
Conclusions

- Promoting the use of **global, persistent and (ideally) resolvable identifiers, in form of HTTP URIs**, is beneficial to the geospatial infrastructure itself, and enables a better integration with other data sources and service platforms.

- Making geospatial services more Web-friendly would increase the re-use of geospatial data. This includes support to **HTTP content negotiation**, service output provided also in formats optimised for Web consumption and discovery (e.g., **HTML with embedded metadata**).

- **Standardised GetCapabilities to OpenSearch mappings** would facilitate the use of geospatial services by general purpose clients.

- **Profile-based HTTP content negotiation** would enable Web clients to use functionalities already supported by geospatial services.
Thanks for your attention!

michael.lutz@jrc.ec.europa.eu

andrea.perego@jrc.ec.europa.eu
For more information

- GeoDCAT-AP 1.0 specification
  https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/geodcat-ap-v10

- GeoDCAT-AP implementations
  https://joinup.ec.europa.eu/node/144843

- GeoDCAT-AP XSLT

- GeoDCAT-AP API (demo)
  http://geodcat-ap.semic.eu:8890/api/

- INSPIRE GeoDCAT-AP Sandbox