GeoDCAT-AP
Use cases and open issues

Andrea Perego

SDSVoc 2016

Amsterdam, 30 November 2016
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 in a nutshell

**Purpose**
Providing access to cross-border EU data to be used in support to
- EU environmental policies
- Policies or activities which impact on the environment

**Governance & scope**
- Comprehensive legal & technical framework for data & service interoperability
- Cross-sector thematic scope

**Process**
Development & revision process involving experts & stakeholders from all EU Member States (public & private sectors, research institutions)

**Current status**
~150K data sets from EU Member States, discoverable through the INSPIRE Geoportal
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:
  

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


- **GeoDCAT-AP implementations**, including *CSW*-based ones, are already available:

  [https://joinup.ec.europa.eu/node/144843](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 output schema: 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

- Limited use of **global & persistent identifiers** (in particular, HTTP URIs) in the original metadata records
- Lack of common practices on how to model some information in RDF
- Some cases:
  - Service / API-based data access
  - Data quality, spatial / temporal reference systems, spatial / temporal resolution
- The way metadata are published online does not support a standardised mechanism to fetch metadata based on their profile – an important feature for federated / decentralised harvesting
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.)
- Geospatial catalogues services, aka CSWs (but also other ones as 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
- Use cases include cross-platform and cross-domain harvesting, ability to retrieve records in a “format” fit for a given application, etc.
## Conneg in the GeoDCAT-AP API

### Request (query string)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>outputSchema</td>
<td>The GeoDCAT-AP profile to be used for the transformation</td>
</tr>
<tr>
<td>outputFormat</td>
<td>The RDF serialisation to be returned. If missing, HTTP conneg is used</td>
</tr>
</tbody>
</table>

### Response (HTTP Link headers)

<table>
<thead>
<tr>
<th>Relation type</th>
<th>Type</th>
<th>Target URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>derivedFrom</td>
<td>application/xml</td>
<td>Source document URL</td>
</tr>
<tr>
<td>profile</td>
<td>Media type of the document returned by the API</td>
<td>Profile URI</td>
</tr>
<tr>
<td>self</td>
<td>Media type of the document returned by the API</td>
<td>URL of the document returned by the API</td>
</tr>
<tr>
<td>alternate</td>
<td>Media types of the alternative RDF serialisations supported by the API</td>
<td>The URL of the document, encoded with the relevant RDF serialisation, as would be returned by the API</td>
</tr>
</tbody>
</table>
## Conneg in the GeoDCAT-AP API

### Request (query string)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>outputSchema</code></td>
<td>The GeoDCAT-AP profile to be used for the transformation.</td>
</tr>
<tr>
<td><code>outputFormat</code></td>
<td>The RDF serialisation to be returned. If missing, HTTP conneg is used.</td>
</tr>
</tbody>
</table>

### Response (HTTP Link headers)

<table>
<thead>
<tr>
<th>Relation type</th>
<th>Target URI</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>derivedFrom</code></td>
<td>Source document URL</td>
</tr>
<tr>
<td><code>profile</code></td>
<td>Media type of the document returned by the API</td>
</tr>
<tr>
<td><code>self</code></td>
<td>Media type of the document returned by the API</td>
</tr>
<tr>
<td><code>alternate</code></td>
<td>Media types of the alternative RDF serialisations supported by the API.</td>
</tr>
</tbody>
</table>

The URL of the document, encoded with the relevant RDF serialisation, as would be returned by the API.

Not enough to know which are the alternative profiles.
A turnaround, not very efficient

Response (HTTP Link headers)

<table>
<thead>
<tr>
<th>Relation type</th>
<th>Type</th>
<th>Target URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>alternate</td>
<td>Media types of the alternative profiles and RDF serialisations supported by the API</td>
<td>The URL of the document, represented with the relevant profile, and encoded with the relevant RDF serialisation, as would be returned by the API</td>
</tr>
</tbody>
</table>

This requires retrieving all the alternative representations to find out the one you are looking for (which may be not available...)
Data, services & APIs

A dataset distribution pointing to view service (WMS)
Data, services & APIs

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```xml
<!DOCTYPE WMS_Capabilities [ 

xmlns:inspire_common="http://inspire.ec.europa.eu/schemas/common/1.0" 

  <Service>
    <Name>WMS</Name>
    <Title>Global Marine Information System</Title>
    <Abstract>Dkm resolution MERIS KD490 datasets</Abstract>
  
  <KeywordList>
    <Keyword vocabulary="ISO">infoMapAccessService</Keyword>
    <Keyword>meris</Keyword>
    <Keyword>ocean color</Keyword>
    <Keyword>satellite observations</Keyword>
    <Keyword>marine environment</Keyword>
    <Keyword>coastal environment</Keyword>
    <Keyword>sea water protection</Keyword>
    <Keyword>marine monitoring</Keyword>
    <Keyword>GIS digital format</Keyword>
  </KeywordList>
  <OnlineResource>
  </OnlineResource>

  <ContactInformation>

  <ContactPersonPrimary>
```
Data, services & APIs

• This issue concerns how to model **dataset distributions available via services / APIs** (WMS, WFS, WCS, as well as SPARQL endpoints)

• Requirements:
  1. Denote distributions as pointing to a service / API, and not directly to the actual data.
  2. Provide a description of the API / service interface

• About point (1), a possible approach, used at JRC, is the following:
  • Specify whether the access / download URL of a distribution points to data or to a service / API (dct:type)
  • In the latter case, include the specification the service/API conforms to (dct:conformsTo)
Example from the JRC Data Catalogue

```
a:Dataset a dcat:Dataset;
   dcat:distribution [ a dcat:Distribution ;
      dct:title "GMIS - WMS (9km)"@en ;
      dct:description "Web Map Service (WMS) - GetCapabilities"@en ;
      dcat:accessURL <http://gmis.jrc.ec.europa.eu/webservices/9km/wms/meris/?dataset=kd490> ;

  # The distribution points to a service

  # The service conforms to the WMS specification
  dct:conformsTo <http://www.opengis.net/def/serviceType/ogc/wms> ] .
```
API description

• This issue was discussed in the framework of the DCAT-AP Implementation Guidelines

• The proposal was to use **OpenSearch** to describe the service and the request parameters

• This solution is general-purpose – i.e., can be applied to any type of service / API

• For geospatial services, the OpenSearch document can be automatically generated from a GetCapabilities document
Modelling data quality

- “Data quality” in its broadest sense - including
  - Fit-for-purpose
  - Data precision / accuracy
  - Compliance with given quality benchmarks, standards, specifications
  - Quality assessments based on data review / users’ feedback

- A number of vocabularies provide possible solutions, but harmonized best practices are missing

- Also, it is to be seen how to deal with existing gaps, e.g.:
  - Spatial / temporal resolution (data granularity)
  - Quality assessments expressed with quantitative test results
  - Quality assessments based on users’ feedback
# Existing approaches in ISA & W3C vocs

<table>
<thead>
<tr>
<th>Data quality aspect</th>
<th>Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata conformance with a metadata standard</td>
<td>DCAT-AP / W3C DQV</td>
</tr>
<tr>
<td>Data conformance with a given data schema/model</td>
<td>DCAT-AP / W3C DQV</td>
</tr>
<tr>
<td>Data conformance with a given reference system (spatial or temporal)</td>
<td>GeoDCAT-AP / W3C DQV</td>
</tr>
<tr>
<td>Data conformance with a given quality specification / benchmark</td>
<td>GeoDCAT-AP / W3C DQV</td>
</tr>
<tr>
<td>Associating data with a quality report</td>
<td>StatDCAT-AP / W3C DQV</td>
</tr>
<tr>
<td>Spatial / temporal resolution</td>
<td>W3C DQV</td>
</tr>
<tr>
<td>Data quality assessments expressed with quantitative test results</td>
<td>W3C DQV</td>
</tr>
<tr>
<td>Data quality assessments via users’ feedback</td>
<td>W3C DUV+DQV</td>
</tr>
</tbody>
</table>
Thanks for your attention!

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

- GeoDCAT-AP 1.0.1 specification
  https://joinup.ec.europa.eu/node/154143/

- 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