

# Which data(set) ontology to use when

Linda van den Brink, Geonovum

Paul van Genuchten, GeoCAT

July 2019, Ispra

@brinkwoman

#DataToBuildOn



SDI.Ne<img alt="SDI.Ne<img alt="GeoCAT logo" data-bbox="391 766 428 828"/>The GeoCAT logo, featuring a stylized 'G' in a blue circle followed by the text 'GEOCAT' in blue. Below it, the text 'GOVERNMENT GEOGRAPHIC DATA PUBLISHING' is written in a smaller font.



Harvest the records to my Open Data Portal. Do you support **CKAN API**?

Can I index your **DCAT** triples or query your repository using **SPARQL**?

Search for datasets to add to my QGIS project. What is your **CSW** endpoint?

Search in a search engine and browse the catalogue in a web browser. You support **schema.org**?

Harvest into library archive, do you support **Dublin Core** over **oai-pmh**?

Catalogue



[SSN](#):Sensor

OGC:[SensorML](#)

[SAREF](#):sensor

INSPIRE:  
[Environmental  
Monitoring Facility](#)

[Schema.org/device](#)

Sensor



# Various dataset ontologies

- ISO19115
- DCAT
- PROV
- Dublin Core
- Schema.org/dataset
- DDI, SDMX, CERIF, VoID



# Which ontology when?

Some ontologies are relevant to specific use cases:

- CSW: Dublin Core, ISO19115
- OAI-PMH: Dublin Core
- Open data: DCAT, Datapackage
- Search engines: Schema.org, DCAT
- Linked Data: DCAT, PROV

How to combine?



# 4 options

- Separate url for each ontology (CSW current)
- Use most relevant ontology for each content encoding (Geo4Web)
- Profile content negotiation
- Describe data using multiple ontologies (W3C)



# Most relevant ontology for each encoding

- XML -> ISO19115
- HTML -> schema.org
- rdf+xml / json-ld / ttl -> DCAT



# Profile content negotiation

- Since the geo4web testbed this draft has been published on the topic, coming out of the dataset exchange wg: <https://www.w3.org/TR/dx-prof-conneg/>
- See also [https://www.w3.org/2016/11/sdsvoc/SDSVoc16\\_paper\\_14](https://www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_14)





# Mix and combine ontologies

Annotate a single property using concepts from multiple relevant ontologies

## EXAMPLE 31

```
<html>
<head>
  ...
</head>
<body prefix="dc: http://purl.org/dc/terms/ schema: http://schema.org/">
  <div resource="/alice/posts/trouble_with_bob" typeof="schema:BlogPosting">
    <h2 property="dc:title">The trouble with Bob</h2>
    ...
    <h3 property="dc:creator schema:creator" resource="#me">Alice</h3>
    <div property="schema:articleBody">
      <p>The trouble with Bob is that he takes much better photos than I do:</p>
    </div>
    ...
  </div>
</body>
</html>
```

# Session Discussion results

SDI.Ne<img alt="SDI.Next logo" data-bbox="225 845 275 895"/>xt

# Group 1

Any transformation from a source model to the target model involves a process in which potentially information is lost or converted using use case specific assumptions. Applying such a transformation should be an explicit action by the user, so it makes sense to create a separate endpoint for the transformation process.

The best way forward is to make sure the ontologies used are aligned on common elements, so various clients will be able to read parts of the model (5th approach). Iso19115 metadata model is maintained outside current web best practices (reuse common ontologies where possible). A strong request to the iso community is formulated for the next version to base the common parts of that standard on common standards. DCAT 2 seems most relevant.

An improvement there is quite far on the horizon. Until that point transformation for specific use cases will be relevant.



# Group 2

Transformations should be applied if it is relevant and not too much effort.

Relevance should be evidence based (experiments).

Why do we have multiple ontologies in the first place.

Any mapping transformation should be applied in an automated way.



# Group 3

The ISO schema is a topic not up for discussion soon, multiple tooling is available for that standard and would have to be rewritten. Various best practices, such as the geodcat-ap mapping, and reports, such as the best practices for spatial data on the web, tested and recommend to provide data in multiple ontologies.

The 4 approaches for profile negotiation are not mutually exclusive. Supplying the preferred ontology for the relevant encoding is a common accepted solution.

If you provide content negotiation, make sure to also add an override mechanism, so you can for example open json in we web browser. A perception is that not many clients support content negotiation yet. The inspire registry currently adds a parameter to request a concept in alternative schema.

It could be interesting to have profile negotiation for INSPIRE datasets, a feature could then easily be requested in the as-is schema as well as the harmonized inspire schema on the same endpoint (dataproviders currently use multiple distributions to facilitate as-is and harmonized).



# Thank you

Linda van den Brink  
@brinkwoman   
#DataToBuildOn  
l.vandenbrink@geonovum.nl

