Spatial Data URI Design for Linked Data and W3C

Inspire 2014 Conference, Aalborg
June 20th 2014

Esa Tiainen, Kai Koistinen
National Land Survey (NLS) of Finland
Contents

• Introduction of the Finnish approach
• Inspire specifications
• Linking spatial and non-spatial data
• Challenges for URI set-up and management
• URI service pilot
Introduction of the approach

• The initial goal was envisioned to successfully search data instances by concepts
  – Data model driven approach - E.g. harmonised concepts in INSPIRE data specifications and beyond (controlled vocabularies)
  – This would enable e.g. generating re-usable INSPIRE data products required by national law (instead of establishing numerous data products with several data providers updating many of them in parallel)
  – To implement a framework for Linked Data; which is to address the INSPIRE requirement “to fit INSPIRE IDs to the national framework of identifiers”.

• Conclusion: An ontology approach was a precondition
  – Started with URI design for concepts
  – We want to manage not only PIDs but semantics as well
  – Further conclusion: URI management applicable as national solution
  – Initialization of national recommendation to public administration on unique identifiers (HTTP URI) for spatial data
Related Ontologies

• Finnish spatial data ontology (PTO/SDO)
  1. Application ontology generated from correspondent concepts of the INSPIRE schema; theme, spatial object types and attributes
  2. A thesaurus of keywords in the national spatial data metadata

• The Finnish national ontology Finto (SKOS)
  – Based on the generic national thesaurus

• Further vision
  – To enable the discovery of INSPIRE as well as other spatial data, through the national ontology service besides national Geoportal and INSPIRE services
  – Realization of full potential of INSPIRE and spatial data at large
  – Operational Linked Data infrastructure employing http URI’s
Fig 3 Reflected to Finnish national and spatial data Ontology (PTO)

- Application ontology corresponds i.e. to an ontology generated from a data product specification according to an INSPIRE data specification

- In practice, Domain ontology is the Finnish spatial data ontology (PTO) embedded in the national Finto-ontology SKOS (Finnish Ontology and Thesaurus Service, www.finto.fi)
Spatially enabled semantic data search

• In the following slide, Fig. 1, a detailed search method is presented
  – If the performance of the automated search process presented is feasible, a spatially ”conscious” semantic search is possible linking spatial data schemas and search terms with ontology and metadata through URIs and RDF

  – The search is more simple with direct pre-annotation of controlled vocabulary concepts to the data object instances – and using the same method but preprocessed to connect URI concepts to the (spatial) data instances
    ▪ Pre-processing may be feasible for faster performance or usability issues
### Fig 1: Data entity discovery and retrieval – annotations on metadata and schema level

<table>
<thead>
<tr>
<th>User (evacuation plan for a water polluting emission)</th>
<th>Location based search application</th>
<th>Catalogue service schemas)</th>
<th>Discovery service</th>
<th>Ontology service</th>
<th>Data sources providing service interfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task to warn/report bodies responsible of swimming places and harbours</td>
<td>Search terms “Hydrology protection” and “emission” with location description / boundaries</td>
<td>Retrieval of concepts for terms “Hydrology protection” and “emission” using SDO</td>
<td>Retrieval of metadata from Discovery service and schemas with object classes from Catalogue service using concept URI:</td>
<td>Provides URIs of concepts related to search terms “Hydrology protection” and “emission”</td>
<td>Hydrological observations and sites</td>
</tr>
<tr>
<td><strong>Query results re search terms:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Water resource areas</td>
</tr>
<tr>
<td><strong>Hydrology protection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ground water areas</td>
</tr>
<tr>
<td>Dataset: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Swimming places</td>
</tr>
<tr>
<td>Service: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Harbours</td>
</tr>
<tr>
<td><strong>Emission</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Dataset: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Service: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td><strong>Ground water</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Dataset: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Service: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td><strong>Swimming places</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Dataset: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Service: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td><strong>Harbours</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Dataset: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Service: ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
</tbody>
</table>

- REST/SPARQL
- CSW
- SDO
- WFS
- WMS

- Provides URIs of concepts related to search terms “Hydrology protection” and “emission”
- Provides datasets and services related to “Hydrology protection” and “emission” and with data within the location requested
- Provides schemas and their object classes related to “hydrology protection” and “emission”
- Spatial data instances related to “Hydrology protection” and “emission”
Current stage

• Preparing national recommendation to public administration on unique identifiers (HTTP URI) for spatial data
  
  • Use cases e.g.
    – Re-use of data employing URIs
    – Open data delivery and applications
    – Data update or synchronization with URIs
    – Representation of download service interfaces (as standard URI responses)
  
• Piloting – Proof of concept
  
  – CSIRO PIDService as a resolution service
    • Redirections: paikkatiedot.fi -pilot
      – URL: http://paikkatiedot.fi/pidsvc/index2.html
    • Pre-pilot URI-service (http://inspire-hy.fgi.fi/index_en.html)
  
  – Re-use of INSPIRE data
    • Data discovery through ontology/catalogue service
    • Re-use of INSPIRE and non-INSPIRE spatial data
URI-pattern

A. **Generic for spatial data** -/so/

1)  
http://paikkatiedot.fi/so/{namespace}/{localId}[/{versionId}]  
– centralized and common URI management and service, or

2)  
http://{register}.fi/so/{namespace}/{localId}[/{versionId}]  
- URI-service by the individual data provider, especially regarding large municipalities
  - {namespace} = data source, i.e. dataset identifier in national spatial data metadata
  - {register} = URI management body i.e Helsinki for municipality, or authority
  - e.g. to avoid minting URI’s to different thematic domains within a single dataset

B. **Specification for inspire data:**  
http://{paikkatiedot}.fi/so/{namespace}/[{theme}/{class}]{localId}[/{versionId}]

C. **Linking URI-components**

/so/ for spatial objects
/id/ for real world objects
  - mandatory – linked to /so/, or both ways as necessary for application use
/doc/ for documents
  etc
Issues on identification and INSPIRE ID

• Both the INSPIRE spatial object entity (instance) and the original spatial object (source object) correspondent to the INSPIRE instance have URIs of their own:

• The URI path of the INSPIRE instance has parts: /{theme}/, /{class}/, but the spatial source object URI is opaque.
  – Basically a non-semantic URI has the advantage that it functions like a surrogate (key) and thus effective for anonymous use, facilitating data protection and security.
  – Note that one or several INSPIRE spatial objects, e.g. in different INSPIRE themes, may consist subsets of a single spatial source object.

• Real world entity referred with a /id/-URI refers both to the /so/-URI of the spatial source object (instance) and /so/-URI(s) of the INSPIRE /so/-instance(s).

• The organization responsible of the source object shall create the /id/-URI. If the INSPIRE instance is created before the /id/-URI of the source object then it shall be created by the organization responsible of the INSPIRE spatial object and /so/-URI.

• Practically always the organization responsible of the spatial source object is responsible of the INSPIRE spatial object (this is the case in Finland).
Linking non spatial with spatial

• In URI design presented
  – /id/-URI for the real world entity is always referring to one or several /so/-URI
  – Any non spatial data and information can be linked to /id/-URIs as well

• /id/-URI connects non-spatial data URIs to spatial data URIs
  – In national ontology service semantically equivalent concepts from different domains can be linked together enabling to identify the /id/-URI of the spatial data domain paikkatiedot.fi (“spatialdata.fi”)

• Spatial data can be brought to any non-spatial through /id/-URI
Challenges for URI set-up and management – Consistent rules at the EU level

• At the EU level, we would expect that there would come a technical guidance or high-level recommendations to establish “cornerstones” for pan-European interoperability on issues such as
  • standard URI response patterns (minimum for cross-border use)
  • re-use or configuration of register services such as INSPIRE schemas or codelists etc. for reasoning in order to manage semantics
  • standard representation of life-cycle rules (minimum for cross-border use)
  • further definition/specification of operational roles of /so/- and /id/-components in URI-linking and response patterns
    – linking method between /so/- and /id/- one-way or both ways, which is our conclusion.
  • representation of real-world object; whether /id/-component is feasible for that as common practice, similarly to ISA recommendation
Thank you for attention

Pilot for URI service:
http://inspire-hy.fgi.fi/index_en.html

Esa Tiainen  esa.tiainen@nls.fi
Kai Koistinen  kai.koistinen@nls.fi
URI service pilot


- Pilot made by Lassi Lehto from Finnish Geodetic Institute (FGI)

- INSPIRE Hydrography data from National Land Survey of Finland, Finnish Environment Institute and city of Tampere

- URI links for exploring the spatial objects related to the INSPIRE Hydrography theme
  - No connection from spatial objects to ontologies on this pilot
  - Piloting for ontology utilization is in progress
Unique Object Identifiers in Hydrography Theme

Below a few exemplary links that can be used to start exploring the spatial objects related to the INSPIRE Hydrography theme (feature classes WatercourseLink and StandingWater). Please note that the responses containing /so/ and /doc/ URI-based identifiers are actually encoded as an XML structure, and are converted to HTML pages containing traditional <a>-links by the browser, as defined in the linked stylesheet. In this pilot version object serialisations have been brought under the /doc/ identifier, and the available encoding alternatives are presented by the extension (//gml, //html). Two successive slashes are used to reserve room for the possible object version identifier. The pilot also contains redirections to the Lake Wiki service of the Finnish Environment Institute (in case of the StandingWater features). The experimental html encoding of the spatial object contains an OpenLayers-based map window, in which the feature is visualised as a vector object on top of the raster base map provided by the National Land Survey of Finland. If you follow the link ‘Linkitetty paikkatietokohde’ (relatedHydroObject in the INSPIRE HY schema) in the HTML view, you’ll notice how the hydro objects are linked together in the background HY database (lakes, StandingWater, are linked with the linear pseudo-watercourses, WatercourseLink, that relate to the lake and vice versa).

In the background there is an INSPIRE-compliant HY Download Service and a service database that combines together data sets of the National Land Survey of Finland and the Finnish Environment Institute. The database and the service has been developed in a joint project of the mentioned data providers, the Finnish Ministry of Agriculture and Forestry, and the Finnish Geodetic Institute. Some content is also attached from the service of the city of Tampere (the URIs containing the string 'YV_PIENVEDET').

In this pilot the domain http://inspire-hy.fgi.fi is used as an exemplary base domain of the HY theme.

You can also start by clicking on a lake in a map viewer (the biggest, very complicated lakes will most probably make your browser freeze, though).

http://inspire-hy.fgi.fi/id/hy/WatercourseLink/uuid:1005072502646
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/uuid:1010100104794
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/uuid:1012051400632
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/uuid:101203800910
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/uuid:1007050903644
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/uuid:1005072500892

A complete data set might also have an own URI-based identifier (dereferencing this will return URIs of all the contained spatial objects):

http://inspire-hy.fgi.fi/so/hy/WatercourseLink
http://inspire-hy.fgi.fi/so/hy/StandingWater
http://inspire-hy.fgi.fi/so/hy/YV_PIENVEDET
Unique Object Identifiers in Hydrography Theme

Below a few exemplary links that can be used to start exploring the spatial objects related to the INSPIRE Hydrography theme (feature classes WatercourseLink and StandingWater). Please note that the responses containing /so/ and /doc/ URI-based identifiers are actually encoded as an XML structure, and are converted to HTML pages containing traditional <a> -links by the browser, as defined in the linked stylesheet. In this pilot version object serialisations have been brought under the /doc/ identifier, and the available encoding alternatives are presented by the extension (//gml, //html). Two successive slashes are used to reserve room for the possible object version identifier. The pilot also contains redirections to the Lake Wiki service of the Finnish Environment Institute (in case of the StandingWater features). The experimental html encoding of the spatial object contains an OpenLayers -based map window, in which the feature is visualised as a vector object on top of the raster base map provided by the National Land Survey of Finland. If you follow the link ‘Linnitetty paikkatietokehde’ (relatedHydroObject in the INSPIRE HY schema) in the HTML view, you’ll notice how the hydro objects are linked together in the background HY database (lakes, StandingWater, are linked with the linear pseudo-watercourses, WatercourseLink, that relate to the lake and vice versa).

In the background there is an INSPIRE-compliant HY Download Service and a service database that combines together data sets of the National Land Survey of Finland and the Finnish Environment Institute. The database and the service has been developed in a joint project of the mentioned data providers, the Finnish Ministry of Agriculture and Forestry, and the Finnish Geodetic Institute. Some content is also attached from the service of the city of Tampere (the URIs containing the string 'YV_PIENVEDET').

In this pilot the domain http://inspire-hy.fgi.fi is used as an exemplary base domain of the HY theme.

You can also start by clicking on a lake in a map viewer (the biggest, very complicated lakes will most probably make your browser freeze, though).

http://inspire-hy.fgi.fi/id/hy/WatercourseLink/umoa10.100572502646
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/umoa10.101010104794
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/umoa10.1012051400632
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/umoa10.101203800890
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/umoa10.1007050505644
http://inspire-hy.fgi.fi/id/hy/WatercourseLink/umoa10.100572500899

A complete data set might also have an own URI-based identifier (dereferencing this will return URIs of all the contained spatial objects):

http://inspire-hy.fgi.fi/so/hy/WatercourseLink
http://inspire-hy.fgi.fi/so/hy/StandingWater
http://inspire-hy.fgi.fi/so/hy/YV_PIENVEDET
List of related URIs – spatial objects etc.

http://inspire-hy.fgi.fi/so/hy/WatercourseLink/uoma10.1005072500899
http://inspire-hy.fgi.fi/so/hy/YV_PIENVEDET/YV_PIENVEDET.849
http://inspire-hy.fgi.fi/so/hy/YV_PIENVEDET/YV_PIENVEDET.850
List of related URIs – spatial objects etc.

http://inspire-hy.fgi.fi/so/hv/WatercourseLink/uoma10.1005072500899
http://inspire-hy.fgi.fi/so/hv/YV_PIENVEDET/YV_PIENVEDET.849
http://inspire-hy.fgi.fi/so/hv/YV_PIENVEDET/YV_PIENVEDET.850
Different representations of spatial object

Different representations of spatial object

HTML representation of spatial object

hy-n: WatercourseLink

Kohdetunniste:
http://inspire-hy.fgi.fi/so/hy/WatercourseLink/uoma10.1005072500899

Ominaisuudet:
Paivitetty: 2008-03-05
Pituus: 549.4 m
StandingWater <-> WatercourseLink

Kohdetunniste:

Ominaisuudet:
Päivitetty: 2011-04-12
Pinta-ala: 0.03 ha

Kohdetunniste:
http://inspire-hy.fgi.fi/so/hy/WatercourseLink/uoma10.1010100104794

Ominaisuudet:
Päivitetty: 2011-04-12
Pituus: 33.5 m
Map user interface for exploring spatial objects

http://inspire-hy.fgi.fi/uidview/
Thank you for attention

Esa Tiainen    esa.tiainen@nls.fi
Kai Koistinen  kai.koistinen@nls.fi