

Pilot implementation of Identifiers using URIs and own URI service

Tuuli Pihlajamaa

INSPIRE Conference 2016 / 26 Sep. 2016

Terms used in presentation

- **URI's:** *“A Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or physical resource.”* (<https://tools.ietf.org/html/rfc3986>)
- **URI service:** *data provider's URI service*
- **RDF:** *Resource Description Framework, “One of the key ingredients of Linked Data, and provides a generic graph-based data model for describing things, including their relationships with other things.”* (<http://linkeddata.org/faq>)
- **Turtle** - Terse RDF Triple Language: *“Turtle is a plain text format for serializing RDF data.”* (<http://linkeddatabook.com/editions/1.0/#htoc17>)
- **SPARQL:** *query language for RDF*
- **JHS 193:** *“JHS 193 Unique identifiers of the geographic information”, a national recommendation for Public Administration in Finland*

Background for the pilot

— INSPIRE Generic Conceptual Model, Annex H, Implementation of Identifiers using URIs in INSPIRE:

“The INSPIRE Directive, in Art. 8(2), requires the adoption of a common framework for the unique identification of spatial objects to support interoperability of spatial data across Europe. ---The Directive requires that identifiers under national systems can be mapped to that common framework.”

— Guidelines for the encoding of spatial data, Annex C, Resource identifiers:

- **Recommendation 12** Identifiers of resources should be URIs in the “http” scheme.
- **Recommendation 13** While it is not strictly a requirement that the resource can be accessed via the HTTP protocol using its http URI, it is strongly recommended that this is the case.
- **Recommendation 15** URIs of spatial objects should be persistent http URIs and include the namespace and the local identifier part of the INSPIRE identifier, if available.

— In Finland there is centralized dereferencing service

- Hosted by National Land Survey of Finland
- URI's for real-world phenomena /id/, spatial objects /so/, definitions /def/ and documentation /doc/
- URI's for spatial objects (INSPIRE and other)
- Data providers determine local-ID's of inspire-ID's / URI's for so's and (id's)
- Requests to URI's -> data provider's own URI service
- URI service of the data provider will return the documentation related to the spatial object, in requested format

Statistics Finland's INSPIRE services and data

- **View services** (WMS) and **download services** (WFS and ATOM-feed) are available since May 2013 (http://www.stat.fi/tup/rajapintapalvelut/inspire_aineistot_en.html)
- **Annex III**
- Statistical Units:
 - Municipality-based statistical units
 - Grid net for statistics 1 km x 1 km
- Population Distribution:
 - Population by municipality-based units
 - Population grid data 1 km x 1 km
- Production and industrial facilities:
 - Production and industrial facilities
- Utility and governmental services:
 - Educational institutions
- **Data is not yet harmonized to INSPIRE data schemas**

Aim of the pilot

- 1. Example with the unique identifiers** of geographic information
(finding data, machine readability)
- 2. Forming unique identifiers** of geographic information
(based on the data source)
- 3. Input for harmonizing spatial data** to INSPIRE schema
(understanding TG's)
- 4. Steps towards Linked data**
 - Both http URI's of spatial objects (so) and own URI service are piloted
(sample data municipalities, LAU2)
 - Requests to /so/-URI's in own URI service will be dereferenced (*WFS-service, RDF-server*)
 - Pilot will be done in intranet (*Later dereferencing from domain paikkatiedot.fi*)

Tasks in the pilot

A) Creating pilot URI's for INSPIRE- spatial objects (so)

- Example URI's for each spatial object (a municipality) of certain time, here year 2014.
- In JHS 193 is given recommendations about URI's:
 - URI's for INSPIRE-harmonized data:
`http://paikkatiedot.fi/so/resourceIdentifier/theme/class/localid/(versionid)`
- In the pilot there won't be created new localid's in database, since the localid's will be generated from the information already provided in GML data.
- URI's in the pilot are examples. Final URI's will be decided, when the INSPIRE data products are defined and there are life-cycle rules defined for them.

B) Building URI service in intranet

- Based in technical description and examples provided in JHS 193 appendix 2.
 - Establishing of Geoserver and sample data
 - Establishing Apache server and creating redirecting rules in .htaccess-file
 - Establishing of Jena Fuseki server
 - Creating xslt-transformation file
 - Uploading Turtle-format result to Jena Fuseki.

What has been done in pilot

Technical infrastructure for piloting has been created:

- Geoserver is running and there is example data, Municipalities 2014.
- First xslt-file has been created.
 - Based on that and Espoo-GML was created a Turtle upload sentence.
- Apache Jena Fuseki server is running and some example data was uploaded there in Turtle format.
- Apache server is running and dereferencing can be done.
 - JSON and GML requestes are dereferenced to Geoserver.
 - `.../doc/1000520/049/EPSG:3067.json` -> returns data from Espoo as json
 - `.../doc/1000520/049/EPSG:3067.gml2` -> returns data from Espoo as GML
 - Other requests, than JSON and GML, are dereferenced to RDF-server
 - `.../doc/1000520/Espoo_2014` will be dereferenced as a SPARQL request to Jena Fuseki server that returns the result in outputformat `application/sparql-results+json`

Xslt-file for generating Turtle

```
<?xml version="1.0" encoding="UTF-8"?>

<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:fo="http://www.w3.org/1999/XSL/Format"
xmlns:tilastointialueet="http://geoserv.server.stat.fi/geoserver/tilastointialueet"
xmlns:gml="http://www.opengis.net/gml/3.2" >

<xsl:output method="text" indent="yes"/>

<xsl:template match="/">

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#&gt;;
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#&gt;;
@prefix dc: <http://purl.org/dc/elements/1.1/&gt;;
@prefix jhs: <http://paikkatiedot.fi/jhs-skeema#&gt;;

<xsl:for-each select="//tilastointialueet:kunta4500k">

<http://192.168.197.43/doc/1000520/<xsl:value-of select="gml:name"/>_<xsl:value-of select="tilastointialueet:vuosi"/>&gt;a jhs:Doc;

</xsl:for-each>

</xsl:template>

</xsl:stylesheet>
```


Espoo-GML

- with reference to xslt-file

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="muunnosespoo.xsl"?>
<wfs:FeatureCollection xmlns:wfs="http://www.opengis.net/wfs/2.0"
xmlns:tilastointialueet="http://geoserv.server.stat.fi/geoserver/tilastointialueet"
xmlns:gml="http://www.opengis.net/gml/3.2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
timeStamp="2016-09-12T07:38:15.750Z"
xsi:schemaLocation="http://geoserv.stat.fi/geoserver/tilastointialueet
http://geoserv.server.stat.fi:8080/geoserver/tilastointialueet/wfs?service=WFS&version=1.0.0&request=DescribeFeatureType&typeName=tilastointialueet%3Akunta4500k
http://www.opengis.net/gml/3.2 http://schemas.opengis.net/gml/3.2.1/gml.xsd http://www.opengis.net/wfs/2.0 http://schemas.opengis.net/wfs/2.0/wfs.xsd">
<wfs:boundedBy><gml:Envelope srsDimension="2" srsName="http://www.opengis.net/gml/srs/epsg.xml#3067">
<gml:lowerCorner>361652.1139 6668349.4413</gml:lowerCorner>
<gml:upperCorner>379983.85 6693775.3773</gml:upperCorner></gml:Envelope></wfs:boundedBy>
<wfs:member><tilastointialueet:kunta4500k gml:id="kunta4500k.fid-7b5f3fa4_1569c52cec2_319a">
<gml:name>Espoo</gml:name><gml:boundedBy><gml:Envelope srsDimension="2" srsName="http://www.opengis.net/gml/srs/epsg.xml#3067">
<gml:lowerCorner>361652.11390000023 6668349.441299999</gml:lowerCorner>
<gml:upperCorner>379983.84999999996 6693775.3773</gml:upperCorner></gml:Envelope></gml:boundedBy>
<tilastointialueet:gid>12</tilastointialueet:gid>
<tilastointialueet:vuosi>2014</tilastointialueet:vuosi>
<tilastointialueet:kunta>049</tilastointialueet:kunta>
<tilastointialueet:nimi>Espoo</tilastointialueet:nimi>
<tilastointialueet:namn>Esbo</tilastointialueet:namn>
</tilastointialueet:the_geom>
<gml:MultiSurface srsDimension="2" srsName="http://www.opengis.net/gml/srs/epsg.xml#3067">
<gml:surfaceMember><gml:Polygon srsDimension="2"><gml:exterior><gml:LinearRing>
<gml:posList>375773.69770000037 6691596.517000001 377355.89109999966 6680366.1788 379983.84999999996 6681677.614 379820.9347000001 6671252.395 379079.1303000003 6671248.31540001
<gml:interior><gml:LinearRing><gml:posList>377668.70289999999 6678096.224300001 374093.29019999997 6680830.363500001 370465.29789999966 6679358.1347 370728.19589999997 6676834.31
```

RDF in Turtle-format to be uploaded in Jena Fuseki:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
@prefix dc: <http://purl.org/dc/elements/1.1/>  
@prefix jhs: <http://paikkatiedot.fi/jhs-skeema#>
```

```
<http://192.168.197.43/doc/1000520/Espoo_2014>a jhs:Doc;
```

Uploading RDF-Turtle in Jena Fuseki

```
graph: default

1 @prefix jhs: <http://paikkatiedot.fi/jhs-skeema#> .
2 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
3 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
4 @prefix dc: <http://purl.org/dc/elements/1.1/> .
5
6 <http://192.168.197.43/doc/1000520/Espoo_2014>
7   a      jhs:Doc .
8
```

Added 1 triple

SPARQL query

SPARQL query

To try out some SPARQL queries against the selected dataset, enter your query here.

EXAMPLE QUERIES

Selection of triples

Selection of classes

PREFIXES

rdf

rdfs

owl

xsd

SPARQL ENDPOINT

http://192.168.197.56:3030/demo/query

CONTENT TYPE (SELECT)

JSON

CONTENT TYPE (GRAPH)

Turtle

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 SELECT ?subject ?predicate ?object
3 WHERE {
4   ?subject ?predicate ?object
5 }
6 LIMIT 25
```



QUERY RESULTS



Table

Raw Response



Showing 1 to 1 of 1 entries

Search:

Show 50 entries

subject	predicate	object
1 <http://192.168.197.43/doc/1000520/Espoo_2014>	rdf:type	<http://paikkatiedot.fi/jhs-skeema#Doc>

Rewrite rules in Apache .htaccess-file

- Dereferencing to Geoserver:

```
RewriteRule ^doc/1000520/([a-zA-Z0-9_]+)/([a-zA-Z0-9:]+)/([a-zA-Z0-9]+)$  
http://geoserv.server.stat.fi/geoserver/tilastointialueet/ows?service=WFS&version=1.0.0&request=GetFeature  
&typeName=tilastointialueet:kunta4500k&outputformat=$3&srsName=$2&Filter=<Filter><PropertyIsEqualTo  
><PropertyName>kunta</PropertyName><Literal>$1</Literal></PropertyIsEqualTo></Filter>
```

- Dereferencing to Jena Fuseki RDF-server:

```
RewriteRule ^doc/1000520/([a-zA-Z0-9_-]+)$  
http://192.168.197.56:3030/esbo/query?query=prefix+rdfs:+<http://www.w3.org/2000/01/rdf-  
schema>prefix+owl:+<http://www.w3.org/2002/07/owl>SELECT+?subject+?predicate+?object+WHERE+{?s  
ubject+?predicate+?object}LIMIT+25
```

Answers from the servers, Fuseki

- Path in Apache `http://192.168.197.43/doc/1000520/Espoo_2014` is forwarded to Jena Fuseki as a SPARQL query
- Fuseki returns answer (with subject, predicate and object), in format `application/sparql-results+json`

```
{
  "head": {
    "vars": [ "subject" , "predicate" , "object" ]
  } ,
  "results": {
    "bindings": [
      {
        "subject": { "type": "uri" , "value": "http://192.168.197.43/doc/1000520/Espoo_2014" } ,
        "predicate": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "object": { "type": "uri" , "value": "http://paikkatiedot.fi/jhs-skeema#Doc" }
      }
    ]
  }
}
```

Answers from the servers, GML

-Path in Apache <http://192.168.197.43/doc/1000520/049/EPSG:3067.gml2> gives the answer in GML-format

```
-<wfs:FeatureCollection xsi:schemaLocation="http://geoserv.stat.fi/geoserver/tilastointialueet http://geoserv.server.stat.fi:8080/geoserver/tilastointialueet/wfs?service=WFS&version=1.0.0&request=DescribeFeatureType&
typeName=tilastointialueet%3Akunta4500k http://www.opengis.net/wfs http://schemas.opengis.net/wfs/1.0.0/WFS-basic.xsd">
  -<gml:boundedBy>
    <gml:null>unknown</gml:null>
  </gml:boundedBy>
  -<gml:featureMember>
    -<tilastointialueet:kunta4500k fid="kunta4500k.fid-7b5f3fa4_1569c52cec2_53b3">
      <tilastointialueet:gid>12</tilastointialueet:gid>
      <tilastointialueet:vuosi>2014</tilastointialueet:vuosi>
      <tilastointialueet:kunta>049</tilastointialueet:kunta>
      <tilastointialueet:nimi>Espoo</tilastointialueet:nimi>
      <tilastointialueet:namn>Esbo</tilastointialueet:namn>
      <tilastointialueet:name>Espoo</tilastointialueet:name>
    -<tilastointialueet:the_geom>
      -<gml:MultiPolygon srsName="http://www.opengis.net/gml/srs/epsg.xml#3067">
        -<gml:polygonMember>
          -<gml:Polygon>
            -<gml:outerBoundaryIs>
              -<gml:LinearRing>
                -<gml:coordinates decimal="." cs="," ts=" ">
                  375773.6977,6691596.517 377355.8911,6680366.1788 379983.85,6681677.614 379820.9347,6671252.395 379079.1303,6671248.3154 376530.1592,6670478.6122 372621.7409,6668509.3846 3711
                  370212.7045,6669548.9544 369303.0659,6670438.5921 368242.629,6670403.5777 361652.1139,6677734.3961 362624.1514,6685325.923 363235.963,6688883.655 365362.8603,6688879.3454 369
                  375773.6977,6691596.517
                </gml:coordinates>
              </gml:LinearRing>
            -<gml:outerBoundaryIs>
              -<gml:innerBoundaryIs>
                -<gml:LinearRing>
                  -<gml:coordinates decimal="." cs="," ts=" ">
                    377668.7029,6678096.2243 374093.2902,6680830.3635 370465.2979,6679358.1347 370728.1959,6676834.314 372095.2654,6674731.13 375618.0986,6674415.6524 377668.7029,6678096.2243
                  </gml:coordinates>
                </gml:LinearRing>
              </gml:innerBoundaryIs>
            -<gml:Polygon>
              -<gml:polygonMember>
                </gml:MultiPolygon>
            </tilastointialueet:the_geom>
          </tilastointialueet:kunta4500k>
        </gml:featureMember>
      </wfs:FeatureCollection>
```

Answers from the servers, JSON

-Path in Apache <http://192.168.197.43/doc/1000520/049/EPSG:3067.json> gives the answer in JSON-format

```
{ "type": "FeatureCollection", "features": [ { "type": "Feature", "id": "kunta4500k.fid-7b5f3fa4_1569c52cec2_53b4", "geometry": { "type": "MultiPolygon", "coordinates": [ [ [ [ [375773.69770000037, 6691596.517000001], [377355.89109999966, 6680366.1788], [379983.8499999996, 6681677.614], [379820.9347000001, 6671252.395], [379079.1303000003, 6671248.315400001], [376530.15919999965, 6670478.612199999], [372621.7408999996, 6668509.3846], [371112.3487, 6668349.441299999], [370212.7045, 6669548.954399999], [369303.0658999998, 6670438.5921], [368242.6289999997, 6670403.5777], [361652.11390000023, 6677734.3961], [362624.15139999986, 6685325.923], [363235.96300000045, 6688883.654999999], [365362.8602999998, 6688879.3454], [369134.56799999997, 6693775.3773], [375773.69770000037, 6691596.517000001] ] ] ] ] ], "geometry_name": "the_geom", "properties": { "gid": 12, "vuosi": 2014, "kunta": "049", "nimi": "Espoo", "namn": "Esbo", "name": "Espoo" } } ], "crs": { "type": "EPSG", "properties": { "code": "3067" } } }
```


Next steps in pilot

1. Transformation file needs to be fixed

- Despite of using encoding UTF-8 in the GML-data, there was a problem with the Turtle transformation when there were Scandinavian letters (å, ä, ö) in the source data.

2. After that, **transformation could be done online** in transformation service, utilizing data directly from Geoserver.

3. Pilot URI's need to be tested more

- Espoo_2014 works fine since the name and year can be picked from the data.
- But how to pick only part on the string from the data (for example year from validityPeriod)

4. Enriching the RDF data needs to be done

- For example adding INSPIRE code list for ontology

5. Fixing the rewrite rules

- Instead of using QUERY, could be used CONSTRUCT, to give the result in RDF graph format

Conclusions

Input for INSPIRE work:

- Piloting has helped to see both challenges and benefits of harmonizing data into INSPIRE schema.
- It has helped to ask relevant questions
 - E.g. questions regarding the URI's and the data has helped to understand better the Technical Guidelines.

Input for statistics:

- Http URI's and own URI service for geographical data can be utilised when creating Linked data solutions for statistics in wider aspect
- Created solution can also function as an example of unique identifiers for statistics.

Useful links

- INSPIRE Generic Conceptual Model:

http://inspire.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4.pdf

- Guidelines for the encoding of spatial data:

http://inspire.ec.europa.eu/documents/Data_Specifications/D2.7_v3.3.pdf

- JHS Recommendation 193 (in Finnish):

<http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/193/full>



Thank you!

Statistics Finland 

Tuuli Pihlajamaa tuuli.pihlajamaa@stat.fi
INSPIRE Conference 2016 / 26 Sep. 2016