GMLID.EU

data access point for short URLs.

Harmonized access to harmonized data using INSPIRE WFS download services

Prepared by: Sorin RUSU: 06.SEP.T.2017  GIS@TEAMNET.RO

Authors: Sorin RUSU, Iurie Maxim, Daniel Cocanu
The big question?

What if INSPIRE were simple, clear, intuitive and easy for users?

It is possible using short URLs in combination with smart IDs and some technical manipulation of existing services
INSPIRE implementation considerations

More and more MS and responsible authorities are moving towards providing Network Service access to their data. Vector data should be available as WFS, it just makes more sense. There is still no Google for searching in INSPIRE, not really. There are very large, very cumbersome datasets available. Users usually require small subsets of the main dataset provided. Despite the 1000s pages of documentation, there is no simple intro to INSPIRE for end users.
**Problems facing using INSPIRE**

Conceptual framework and a prototype to make INSPIRE work for end users

<table>
<thead>
<tr>
<th>What we want?</th>
<th>The state of the art now?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Data sets available in INSPIRE are not googlable</td>
</tr>
<tr>
<td></td>
<td>INSPIRE GEOPORTAL &gt; 2 search systems &gt; Dataset’s MD files (interlinked with WFS/WMS MD files) &gt; WFS (different server completely w/o standard service endpoints) &gt; Parametrize requests to the Service &gt; GML</td>
</tr>
<tr>
<td>Intuitive</td>
<td>Know about MD files + difference about DS MD and WS MD</td>
</tr>
<tr>
<td></td>
<td>Add GetCapabilities files, start learning about WFS/WMS KVP requests, POST requests are the worst, learn WFS/WMS specific parameters and values</td>
</tr>
<tr>
<td></td>
<td>Fail a lot in a lot of the above points, strive to get the data, Try not to go mad in the process of just getting some data</td>
</tr>
<tr>
<td>Easy</td>
<td>This has to change!</td>
</tr>
</tbody>
</table>
Scope and specific objectives

A simple **logic-based, clear & straight-forward way to** get to *MD, Services endpoints* and *GML data*, that is intuitive for users & can work within INSPIRE diversity

- One domain to bring them all together
- One rule for standardizing Providers and Datasets
- Greatly simplify getting to the GML data
- Integrate everyone (DS MD + WS MD + server endpoints + cross theme datasets + filtering data)
Getting a dataset now AU

Go to INSPIRE geoportal > Go to search engine > Start searching for keywords

Find the DS MD file for RO:
- If GetCapabilities present for the WFS:
  - Go to the WFS’s GetCapabilities endpoint
  - Request Data from WFS with GetFeatures
  - Get a whole bunch of GML (usually all the data)
  - Start conversion of the data and then filter to get the AU needed

Find the DS MD file for HU:
- If GetCapabilities present for the WFS:
  - Go to the WFS’s GetCapabilities endpoint
  - Request Data from WFS with GetFeatures
  - Get a whole bunch of GML (usually all the data)
  - Start conversion of the data and then filter to get the AU needed

If step 4 or 5 fails, go to the proxy browser in the geoportal
Start searching for Download Services that refer AU, filter by country
Get one or more WS MD > might lead to WFS GetCapabilities endpoint
- See what endpoint is responsive or not
- Repeat steps 4b-4e and 5b-5e
If this fails, contact the national geoportal, and get a reply if contact page exists

Try unsuccessfully not to lose hope!
Examples of short URLs to get data from the Romanian National Cadastre Agency from an ArcGIS Server endpoint

UAT – administrative units dataset comprised of AU, AB, NUTS, condominium

Get MD for the UAT DS: 
gmlid.eu/RO/ANCPI/UAT/MD
Get MD for the UAT WFS: 
gmlid.eu/RO/ANCPI/UAT/MD/WFS
Get the WFS endpoint for the UAT DS: 
gmlid.eu/RO/ANCPI/UAT/WFS
Get the Capabilities for the WFS: 
gmlid.eu/RO/ANCPI/UAT/WFS/Capabilities
Get the GML data for entire UAT DS: 
gmlid.eu/RO/ANCPI/UAT/GML
Get the GML data for the AU DT from the DS: 
gmlid.eu/RO/ANCPI/UAT/AU/GML

PADS – protected areas dataset comprised of PS, AU, BR, NP

Get MD for the PADS DS: 
gmlid.eu/RO/ENV/PADS/MD
Get MD for the PADS WFS: 
gmlid.eu/RO/ENV/PADS/MD/WFS
Get the WFS endpoint for the PADS DS: 
gmlid.eu/RO/ENV/PADS/WFS
Get the Capabilities for the WFS: 
gmlid.eu/RO/ENV/PADS/WFS/Capabilities
Get the GML data for entire PADS DS: 
gmlid.eu/RO/ENV/PADS/GML
Get the GML data for the PS DT from the DS: 
gmlid.eu/RO/ENV/PADS/PS/GML

Examples of short URLs to get data from the Romanian National Cadastre Agency from a Geoserver endpoint

Teamnet
Straight-forward Logic in all URLs

BASE_URL = DOMAIN / COUNTRY / DS PROVIDER / DATA SET /
BASE = http://gmlid.eu
COUNTRY = registry of countries: RO, HU, FR, BE, ES, LT, CY
    Based on already established registries
DS PROVIDER = intl. abbreviation of DS Provider:
    No existing single registry, however not that difficult to create one
        Romania Ministry of Environment – ENV
        Romanian National Cadastral Agency – ANCPI
        Société nationale des chemins de fer français – SNCF
DS Name = abbreviated name of the provider’s dataset
    should be the the same as DS MD RS Identifier element:
        ENV – Protected Areas Data Set (PADS)
        ANCPI – Unități Administrativ Teritoriale (UAT)
Once we have a BASE_URL the true magic can start happening:

**BASE_URL / MD** -> DS Metadata Document
**BASE_URL/MD/WFS** -> DS WFS Metadata Document
**BASE_URL/MD/WMS** -> DS WMS Metadata Document
**BASE_URL/WFS** -> DS WFS endpoint
   (we can start KVP requests here, **WFS?request=Get [...]**)
**BASE_URL/WFS/Capabilities** -> WFS’s GetCapabilities endpoint
**BASE_URL/WMS** -> DS’s WMS endpoint
   (we can start KVP requests here, **WMS?request=Get [...]**)
**BASE_URL/WMS/Capabilities** -> WMS’s GetCapabilities endpoint
**Straight-forward Logic in all URLs (3)**

Getting the GML - BASE_URL / Data Theme /Format

- abbreviated INSPIRE annex data theme
- getting the actual data as GML, possibly other formats
- use &count=, and &startindex= WFS parameters

**BASE_URL/GML**

- gmlid.eu/RO/ENV/PADS/GML > elements from PS, AU, NP, BR
- limit number of GML elements: BASE_URL/GML/count
  - gmlid.eu/RO/ENV/PADS/GML/10 > just 10 elements from each DT
- limit number of GML elements and specify start index: BASE_URL/GML/count/index
  - gmlid.eu/RO/ENV/PADS/GML/10/507 > just 10 elements from each DT, start at 507th element in each DT if there

**BASE_URL/DT/GML**

- gmlid.eu/RO/ENV/PADS/PS/GML
- limit number of GML elements: BASE_URL/DT/GML/count
  - gmlid.eu/RO/ENV/PADS/PS/GML/10 > just 10 elements from PS
- limit number of GML elements and specify start index: BASE_URL/DT/GML/count/index
  - gmlid.eu/RO/ENV/PADS/PS/GML/10/507 > just 10 elements from each DT, start at 507th element in PS
The need for smart identifiers in our data

Case 1: ENV/PADS/AU
http://gmlid.eu/RO/ENV/PADS/AU/AR
gml:id = RO.ENV.PADS.AU.AR
inspireID/localID = AR
inspireID/namespace = http://gmlid.eu/RO/ENV/PADS/AU/
gml:identifier = http://gmlid.eu/RO/ENV/PADS/AU/AR

The gml:id reflects the namespace+localID from the INSPIRE ID element, as well as providing the short URL in the gml:identifier element. The element AR can be traced back to the the corresponding DT, DS, Provider, Country.

Case 2: ANCPI/UAT/AU
http://gmlid.eu/RO/ANCPI/UAT/AU/auAdmUnitS.4
gml:id = auAdmUnitS.4
inspireID/localID = 1.29
inspireID/namespace = RO.Ancpi.AU

Problem: the gml:id value is used in the identifier link that should generate the element, and there is no tie in with INSPIREID
Data with meaningful identifiers

BASE_URL/DT/IDENTIFIER

Getting down to the element (feature) level through INSPIRE local ID
IDs matter in INSPIRE – everything should be unique
   Might as well make it intelligent and meaningful
   Randomly generated IDs help no one, users or data providers
An ID of E21DBDFF-4FE3-4FB7-BE08-DD55A4D635EC is functional, yet meaningless,

Getting individual elements from PS or AU:
N2K sites – keep their names ROSCI0135, ROSPA0082
   gmlid.eu/RO/ENV/PADS/PS/ROSCI0135
natural protected sites are standardized: RONPA0022
   gmlid.eu/RO/ENV/PADS/PS/RONPA0022

Getting individual elements from DT with lifecycles (versionID in INSPRIREID):
Latest Continental BR: gmlid.eu/RO/ENV/PADS/BR/CON
Previous Continental BR > gmlid.eu/RO/ENV/PADS/CON/2016-08-31
Subsets in the data by DT categories

BASE_URL/DT/QUERY/CATEGORY

On the fly predefined subsets in our data based on registry values for GML data:
For PS to get natura2000 protected sites, we use the registry value or abbreviation:
gmlid.eu/RO/ENV/PADS/PS/Q/natura2000 or gmlid.eu/RO/ENV/PADS/PS/Q/N2K
or getting a subset of the previous designation:
gmlid.eu/RO/ENV/PADS/PS/Q/siteOfCommunityImportance
gmlid.eu/RO/ENV/PADS/PS/Q/SCI

From AU to get data based on their au:nationalLevel:
gmlid.eu/RO/ENV/PADS/AU/Q/3rdOrder
or filter based on NationalLevelName
gmlid.eu/RO/ENV/PADS/AU/nationalLevelName/Comuna
or add count, start index to the previous filter
gmlid.eu/RO/ENV/PADS/AU/nationalLevelName/Comuna/2/5
What is behind the short URLs

URL rewrites (we used NGINX) are essential for meaningful short URLs

to turn: http://inspire.biodiversity.ro/metadate/PS_SetDeDateSpatiale_MD.xml
into: http://gmlid.eu/RO/ENV/PADS/MD

and to turn:
into http://gmlid.eu/RO/ANCPI/UAT/MD

to turn
http://inspire.biodiversity.ro/PADS/PADS/wfs?version=2.0.0&request=GetCapabilities
into http://gmlid.eu/RO/ENV/PADS/WFS/Capabilities
taking
http://geoportal.ancpi.ro/arcgis/rest/services/AU/AU_Download/GeoDataServer/exts/InspireFeatureDownload/service?VERSION=2.0.0&SERVICE=WFS&REQUEST=GetCapabilities
into http://gmlid.eu/RO/ANCPI/UAT/WFS/Capabilities
What is behind the short URLs

to turn WFS GetFeature requests with Stored Queries from:
into http://gmlid.eu/RO/ENV/PA.../GML/10

turning the very long and complex:
into a cleaner http://gmlid.eu/RO/ENV/PA.../PS/Q/natura2000

to go from
to: http://gmlid.eu/RO/ENV/PA.../AU/Q/Muncipiul/10/25
GetFeatureByINSPIRE ID Stored Query

<wfs:QueryExpressionText isPrivate="false" language="urn:ogc:def:queryLanguage:OGC-WFS::WFS_QueryExpression" returnFeatureTypes="${featureType}">
  <wfs:Query wfs:typeNames="${featureType}">
    <fes:Filter><fes:And>
      <fes:PropertyIsEqualTo>
        <fes:ValueReference>${featureType}/${inspireId}/base:Identifier/base:namespace</fes:ValueReference>
        <fes:Literal>${namespace}</fes:Literal>
      </fes:PropertyIsEqualTo>
      <fes:PropertyIsEqualTo>
        <fes:ValueReference>${featureType}/${inspireId}/base:Identifier/base:localId</fes:ValueReference>
        <fes:Literal>${localId}</fes:Literal>
      </fes:PropertyIsEqualTo>
    </fes:And></fes:Filter>
  </wfs:Query>
</wfs:QueryExpressionText>

http://gmlid.eu/RO/ENV/PADS/PS/ROSCI0135
GetFeatureSingleCriterion Stored Query

<wfs:QueryExpressionText isPrivate="false" language="urn:ogc:def:queryLanguage:OGC-WFS::WFS_QueryExpression" returnFeatureTypes="${featureType}">
  <wfs:Query wfs:typeNames="${featureType}">
    <fes:Filter>
      <fes:PropertyIsEqualTo>
        <fes:ValueReference>${propertyPath}</fes:ValueReference>
        <fes:Literal>${propertyValue}</fes:Literal>
      </fes:PropertyIsEqualTo>
    </fes:Filter>
  </wfs:Query>
</wfs:QueryExpressionText>

Examples:
Conclusions & summing up

- Homogenous access to MD, WS, GML is easily achievable
- Clear, straight-forward access to data through short URL’s is possible while also using existing INSPIRE infrastructure
- WFS GML data provides a wide array of options
- Stored Queries are severely neglected in current TG
- We must address the problem of meaningless identifiers in our data
- Results of this prototype are easily repeatable and scalable
- Might move INSPIRE to a more integrated SDI
- Help users reach data!
Thank you!

any comments or questions?