The Initial SmartOpenData Model

Deliverable D3.2 :: Public

Keywords: INSPIRE, schema, RDF

Linked Open Data for environment protection in Smart Regions
Document Metadata

Contractual Date of Delivery to the EC: July 2014

Actual Date of Delivery to the EC:

Editor(s): Phil Archer W3C/ERCIM

Contributor(s) Martin Tuchyna, Erich Pacola, Tomáš Kliment, Peter Hanečák, Gabriel Lachmann SAŽP
John O’Flaherty, MAC, Tatiana Tarasova, Spaziodati

DOCUMENT HISTORY

<table>
<thead>
<tr>
<th>Version</th>
<th>Version date</th>
<th>Responsible</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>24 July 2014</td>
<td>W3C</td>
<td>Draft ready for review</td>
</tr>
<tr>
<td>0.02</td>
<td>25 July 2014</td>
<td>W3C</td>
<td>Edits following MAC review</td>
</tr>
<tr>
<td>0.03</td>
<td>29 July 2014</td>
<td>W3C</td>
<td>Edits following SAŽP review</td>
</tr>
<tr>
<td>0.04</td>
<td>31 July 2014</td>
<td>W3C</td>
<td>Important citation added for GeoKnow data, p23</td>
</tr>
<tr>
<td>1.0</td>
<td>12 August 2014</td>
<td>TRAGSA</td>
<td>Minor editorial changes</td>
</tr>
</tbody>
</table>

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.
Table of Contents

1 Executive Summary .............................................................................................................. 5
2 Introduction ......................................................................................................................... 6
3 Existing Work .......................................................................................................................... 6
  3.1 The INSPIRE Data Model ............................................................................................. 6
  3.2 The GeoKnow Work on geodata.gov.gr ....................................................................... 10
  3.3 Emerging Work on RDF schemata for INSPIRE .......................................................... 12
  3.4 INSPIRE Code Lists ..................................................................................................... 14
4 The Initial SmartOpenData Model .......................................................................................... 16
  4.1 Namespaces ..................................................................................................................... 16
  4.2 Protected Site .................................................................................................................. 16
  4.3 Identifier ......................................................................................................................... 17
  4.4 Document Citation ......................................................................................................... 17
  4.5 Geographical Name ....................................................................................................... 17
  4.6 Cardinality ...................................................................................................................... 18
5 URI Construction ................................................................................................................... 18
6 Data Example ......................................................................................................................... 19
7 Putting the Model to Work .................................................................................................... 21
8 Future Proofing ..................................................................................................................... 22
9 Conclusion ............................................................................................................................. 22
10 Acknowledgements ............................................................................................................. 23
11 References .......................................................................................................................... 23
12 Listings ................................................................................................................................ 23
  12.1 Sample Protected Site data from GeoKnow ................................................................. 23
  12.2 Sample SPARQL Query against geodata.gov.gr ......................................................... 25
  12.3 Snippet from INSPIRE GCM Ontology ..................................................................... 26
List of Figures

Figure 1 INSPIRE Data Model for Geographic Names ................................................................. 7
Figure 2 INSPIRE Data Model for Protected Sites ................................................................. 8
Figure 3 INSPIRE Protected Sites Simple Application Schema ............................................. 9
Figure 4 INSPIRE Existing Land Use Object ....................................................................... 10
Figure 5 The Initial SmartOpenData Model .......................................................................... 15

List of Tables

Table 1 Namespaces used in the model ................................................................................... 16


1 Executive Summary

The 5 pilots that form the core of the SmartOpenData project are very diverse but the INSPIRE data model and infrastructure provide a common reference point. As far as possible, the project will therefore use INSPIRE as the basis for the data structures in each pilot. However; the INSPIRE guidelines have been defined in terms of XML and GML-based schemas, i.e. using non-Linked Data technologies. Some work has already been done to use Linked Data in the context of INSPIRE, notably in the GeoKnow project. The JRC is itself in the process of making the INSPIRE schemata available as RDF/OWL but this work is not yet complete and most of what has been done is not yet publicly available.

By examining the work done in GeoKnow, and by communicating with the JRC and other individuals involved in their work, SmOD partners have been able to make good progress in creating an Initial Data Model that uses Linked Data techniques. As with all such data models, it can readily be extended to cover particular domains and this will certainly happen in each of the pilots, but the initial model provides a core and an approach that can be built upon.

It is expected that the Initial Data Model this will need to be updated later in the project in the light of two factors:

1. experience of running the pilots;
2. the eventual publication by the JRC of the RDF schemas for INSPIRE.

Such updates are likely to be relatively minor compared with the bigger hurdle being tackled in each pilot which is to adapt and extend existing data infrastructure to make use of Linked Data technologies and realise the benefits that it offers.
2 Introduction

The SmartOpenData project (SmOD) is developing a number of pilots in rural areas of Europe. Each pilot is distinct and needs to draw on very different data sources. However, in order to maximise the potential for reuse of the pilots, and to minimise the effort required for each one, the project partners need to use as many common elements as possible, if not in the data itself then in the data structures and the approach taken. The Linked Open Data (LOD) paradigm is designed to support exactly this kind of situation.

In the European legislative context, geospatial and environmental data is increasingly made available using the INSPIRE data model. Defined using standards like GML and XML, and made available through network services (e.g. via GeoServers) that rely principally on standards developed by the Open Geospatial Consortium, some publishers and users of INSPIRE data are experimenting with Linked Open Data. In particular, using HTTP URIs as identifiers and expressing code lists as SKOS Concept Schemes. Staff at the Joint Research Centre, the section of the European Commission responsible for developing the INSPIRE standards and guidelines, are themselves leading much of this work.

It seems self-evident that the SmOD partners should use the INSPIRE data model too, and that we should draw on those experiments – and that is the intention. However, the model is not yet fully expressed in terms of Linked Data and so the fit between SmOD and INSPIRE is not perfect. Furthermore, each SmOD pilot uses some data that is out of scope for INSPIRE and so it does not offer a full solution.

Therefore, the plan as laid out in this document is that the SmOD pilots will use the INSPIRE data model wherever possible. We will draw on the work done by others to express INSPIRE data as LOD and then extend the model as necessary within each pilot. This will enable SmOD to make its own contribution to the broader community by validating, or suggesting amendments to, others' work as well as to offer experience of extending INSPIRE within the LOD environment.

3 Existing Work

3.1 The INSPIRE Data Model

The INSPIRE Data Model is defined by the JRC and published\(^1\) as a set of components. The Generic Concept Model\(^2\) underpins all the other components but of particular relevance are the data models for each of the Themes. INSPIRE comprises a total of 34 Themes, grouped into three Annexes, and for the initial phase it is the Geographic Names\(^3\) and Protected Sites\(^4\)

\(^1\) http://inspire.ec.europa.eu/index.cfm/pageid/2/list/datamodels
(Simple application schema) themes of Annex I that most interest us in SmOD (see following figures). Based on this experience additional INSPIRE themes will be addressed by the pilots as described in chapter 8.

Figure 1 INSPIRE Data Model for Geographic Names

Figure 2 INSPIRE Data Model for Protected Sites
In addition to these, the SmOD pilots also need to draw on the Land Use theme (Existing Land Use application schema) from Annex III (Figure 4).

These elements of the INSPIRE Data Model are the starting points for SmartOpenData Initial data model. The task ahead is to make these elements available as a Resource Description Framework (RDF) schema so that the pilots can use that schema where appropriate.

---

6. http://www.w3.org/standards/techs/rdf#w3c_all
3.2 The GeoKnow Work on geodata.gov.gr

The three year GeoKnow project\(^7\) began in December 2012, almost a year ahead of SmOD, and has many features in common. It can be thought of as the geospatial extension of the LOD2\(^8\) project with many of the same partners. Whilst SmOD is focused specifically on rural and protected areas, GeoKnow is a general geospatial Linked Data project aiming to develop a toolkit for handling any geospatial data. The work of the GeoKnow project is therefore very relevant to SmOD and we will make extensive use of its outputs.

Of particular relevance to the SmOD data model is GeoKnow’s deliverable 2.7.1 \textit{geodata.gov.gr Geospatial Data as Linked Data} \cite{geoknow}. It describes how the team at the Athena Research and Innovation Centre used the Humboldt Alignment Editor\(^9\), HALE, to first transform Greek government spatial data to the INSPIRE model encoded in GML and then a set of XSLT transformations to convert that into RDF. Listing 12.1 (page 23) shows an example of the data available through geodata.gov.gr.

This approach has several attractive features:

- HALE can be configured to map any input data to INSPIRE conformant GML
- any INSPIRE conformant GML can be readily converted to RDF;
- the hard work has already been done and we can reuse it.

\(^7\) \url{http://geoknow.eu/} \\
\(^8\) \url{http://lod2.eu/} \\
\(^9\) \url{http://www.dhpanel.eu/humboldt-framework/hale.html}
On the downside, there is no underlying data model other than that provided by INSPIRE. It’s a mechanical translation that does not make use of prior work and, crucially, there’s no readily available graphical representation of the schema. This is a very handy tool when writing SPARQL queries and the result of it not being available is that queries against the data are cumbersome and error prone (Listing 12.2).

It is interesting and reassuring to note that the methodology followed by the Athena Research and Innovation Centre broadly matches that proposed by previous work such as the COMSODE project\(^\text{10}\) and the Irish Open Data Publication Handbook\(^\text{11}\). Namely:

1. Development of linked open data plan
   a. Analysis of data sources
   b. Identification of datasets for opening up
   c. Determination of target level of openness
   d. Identification of datasets with restricted access
   e. Effort estimation
   f. Definition of the publication plan

2. Preparation of transformation and publication
   a. Data sources access configuration
   b. Definition of metadata schemas (catalogue record schema and the target data catalogues)
   c. Description of the datasets via metadata
   d. Definition of the approach to the dataset transformation and publication
   e. Design and implementation of the transformation (ETL) procedures
      i. Mapping and transformation of input datasets into the INSPIRE target schemas. Tools: HALE, FME, SnowFlake – GeoKnow used HALE
      ii. Mapping and transformation from GML to RDF. Tools: Altova Map Force etc. i.e. XSLTs as developed and used by GeoKnow
   f. Testing of the transformation (ETL) procedures
   g. Licensing

3. Realisation of the publication
   a. Initial publication of the dataset
   b. Data cataloguing
   c. Dataset maintenance

4. Archiving

5. Use of datasets in pilots

\(^{10}\) http://www.comsode.eu/
3.3 Emerging Work on RDF schemata for INSPIRE

At the time of writing (July 2014) the JRC is completing work on developing a set of RDF schemas that reflect the INSPIRE themes. This work is undertaken via Reusable INSPIRE Reference Platform (ARE"NA, ISA Action 1.17) focused on building bridges from INSPIRE to e-government via RDF and Persistent identifiers (PIDs)\(^\text{12}\). Preliminary results were presented via open webinar on 5\(^{th}\) and 6\(^{th}\) May 2014\(^\text{13}\), where initial context, possibilities and challenges for RDF encoding of INSPIRE content were presented by Danny Vandenbroucke and Diederik Tirry\(^\text{14}\).

Unfortunately much of further work is not yet publicly available but we can glimpse some of it. Three experts were asked to experiment with RDF models based on INSPIRE: Clemens Portele of Interactive Instruments, Linda van den Brink and her colleagues at Geonovum and Stuart Williams of Epimorphics.

Portele’s work includes methods for converting UML diagrams to RDF/ Web Ontology Language (OWL) based on ISO 19150-2 and is publicly available\(^\text{15}\). Van den Brink et al have published some of their work under the PiLOD project\(^\text{16}\) and Stuart Williams has kindly shared some of his work through personal communications with SmOD. Andrea Perego and Michael Lutz at the JRC have also shared some of the emerging work. Although we are not able to include what might become the INSPIRE schemas in this (published) document, and any use of any of this work must be flagged ‘provisional,’ we can review it all and make some observations. It is perhaps worth noting that Clemens Portele, Linda van den Brink and Stuart Williams all know each other well and reference each other’s work so that the ‘three’ may be seen as a coherent whole (all three were at the Linking Geospatial Data workshop\(^\text{17}\) organised by SmOD in March this year).

Clemens Portele has done extensive work on extending schemas and encoding rules for GML\(^\text{18}\) within the Open Geospatial Consortium. Van den Brink et. al referenced this work and followed a similar route as the GeoKnow project’s work on geodata.gov.gr: automated transformations according to relatively simple rules expressed in XSLT, but they make the following observation:

"By using the automatically generated IMRO vocabulary (Dutch standard IMRO (Information Model Ruimtelijke Ordening - spatial planning)) one harmonizing aspect of RDF is not used: no existing RDF vocabularies are used. So imro:naam would get an entry in the IMRO


\(^{13}\)https://joinup.ec.europa.eu/community/are3na/event/rdf-and-pids-location-preliminary-results


\(^{15}\)http://shapechange.net/targets/ontology/uml-rdfowl-19150-2/

\(^{16}\)http://www.pilod.nl/wiki/Boek/BrinkEtAl-GML2RDF

\(^{17}\)http://www.w3.org/2014/03/lgd/report

\(^{18}\)http://www.opengis.net/spec/GML/3.3
vocabulary where it would be more meaningful to map it to rdfs:label. However, the knowledge that imro:naam is in fact an rdfs:label is not available in the UML model and cannot be automatically mapped. In order to improve the UML model for better mapping to RDF one could extend the UML model by annotating the UML attributes that have a special meaning in RDF with a link to their RDF counterpart. If, for example, the imro:naam attribute in the UML model would be get the following annotation (via a tagged value): ‘rdfVocabulary=rdfs.label’ it would be possible to make an optimal link between UML and RDF. We plan to make a proposal for the extension of UML modeling\(^\text{19}\)."

This matches the experience of the Athena Research and Innovation Centre: fully automated transformations are effective but miss out on the advantages that can be gained by developing and using a model that has been designed to maximise reuse of existing vocabularies and data structures.

What we have been able to see so far of the ontologies emerging from the JRC shows that, as one would expect, there are namespaces for the Generic Conceptual Model (GCM) and the individual themes. Perhaps surprisingly for a pan-European effort, the namespaces in the schemata seen so far are within location.data.gov.uk/inspire URI space. One can speculate that this might change before their final publication but it is a reflection of the advanced work already done on INSPIRE-compliant data in the UK, mostly by Stuart Williams\(^\text{20}\) and at the Ordnance Survey by John Goodwin\(^\text{21}\).

The ontologies make use of OWL restrictions to define cardinality and value constraints so that, for example, INSPIRE IDs must have a namespace Id, a local Id and an optional version Id (see Listing 12.3).

To complete the picture, there are upcoming another possible approaches\(^\text{22}\) how to fulfil above mentioned ambitions (e.g. GeoJSON-LD)\(^\text{23}\) and it can be expected that in near future their list will grow. Where possible, SmOD will take them into the consideration to ensure the most effective and easy to deploy options will be provided for future stakeholders willing to extend their spatial content towards the semantic dimension.

\(^{19}\) http://www.gdmc.nl/publications/2013/Geo-data_to_Linked_Data.pdf


\(^{21}\) http://data.ordnancesurvey.co.uk/datasets/os-linked-data

\(^{22}\) http://genuchten.blogspot.sk/2013/07/the-internet-of-things-and-geo-services.html

3.4 **INSPIRE Code Lists**

The INSPIRE infrastructure already provides clear descriptions and persistent URIs as identifiers for many components available from the INSPIRE registry\(^24\). It offers ready access to the application schemas, the feature concept dictionaries, the theme registers and, of particular importance to anyone wishing to work with INSPIRE and Linked Data, code lists. For example, the Designation Scheme code list\(^25\) provides stable URIs for the possible values for this property:

http://inspire.ec.europa.eu/codelist/DesignationSchemeValue/emeraldNetwork
http://inspire.ec.europa.eu/codelist/DesignationSchemeValue/IUCN
etc.

These URIs currently dereference to XML, JSON, ATOM and HTML (through content negotiation). RDF data (using SKOS Concept Schemes) is expected to be added in the near future (The JRC ran a session on this at INSPIRE 2014\(^26\)). Although the latter is not yet available, the URIs are stable and usable today. Where possible work on the implementation of the model can contribute to the ongoing process of testing the new INSPIRE registry, registers and Re3gistry software\(^27\).

---


\(^{26}\) [http://inspire.ec.europa.eu/events/conferences/inspire_2014/pdfs/19.06_5_11.00_Michael_Lutz.pdf](http://inspire.ec.europa.eu/events/conferences/inspire_2014/pdfs/19.06_5_11.00_Michael_Lutz.pdf)

Figure 5 The Initial SmartOpenData Model
4 The Initial SmartOpenData Model

Taking into account the work reviewed in the previous section, it has been possible to create an initial data model (Figure 5). The following sections describe each class in the model, and their properties. Choices made are described and justified.

4.1 Namespaces

<table>
<thead>
<tr>
<th>Schema</th>
<th>Prefix</th>
<th>Namespace</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protected Sites</td>
<td>ps</td>
<td><a href="http://inspire.jrc.ec.europa.eu/schemas/ps/3.0/">http://inspire.jrc.ec.europa.eu/schemas/ps/3.0/</a></td>
</tr>
<tr>
<td>Land Use</td>
<td>lu</td>
<td><a href="http://inspire.jrc.ec.europa.eu/schemas/lu/3.0/">http://inspire.jrc.ec.europa.eu/schemas/lu/3.0/</a></td>
</tr>
<tr>
<td>Dublin Core Terms</td>
<td>dct</td>
<td><a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a></td>
</tr>
<tr>
<td>Friend of a Friend</td>
<td>foaf</td>
<td><a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a></td>
</tr>
<tr>
<td>GeoSPARQL</td>
<td>geo</td>
<td><a href="http://www.opengis.net/ont/geosparql#">http://www.opengis.net/ont/geosparql#</a></td>
</tr>
<tr>
<td>SKOS</td>
<td>skos</td>
<td><a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#</a></td>
</tr>
<tr>
<td>XML Schema</td>
<td>xsd</td>
<td><a href="http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#</a></td>
</tr>
</tbody>
</table>

Table 1 Namespaces used in the model

The INSPIRE namespaces shown in Table 1 are believed to be the ones that the JRC will use but this is not known for certain. The general pattern is clear enough: each of the 34 themes has a short code (gcm, ps etc.) and these appear in the namespace URI. If these change in future, then the final version of the SmOD data Model will reflect this.

Other namespaces used are very well established and one can be very confident that they will not change.

4.2 Protected Site

The Protected Site class is straight from the INSPIRE Data Model. We have followed the GeoKnow project's choice of using geoSPARQL's hasGeometry property to link to a geo:Geometry class that has a property filled by a WKT literal. This gives us an immediate hook with GeoSPARQL, which is implemented in a number of triplestore endpoints, and allows us to reuse the GeoKnow XSLTs directly.
The ps:legalFoundationDate property takes a simple xsd:date as its value (NB: date seems more appropriate than xsd:dateTime since the time of day when a site became protected is going to be irrelevant in all but a trivial minority of cases).

The ps:siteDesignation property is the first in this model to make use of the code lists in the INSPIRE Registry. The value of the property should be a URI that identifies a SKOS Concept in one of the 6 schemes that are specialisations of the generic Designation scheme. INSPIRE defines an enumerated list of possible values for the ps:siteProtectionClassification property and the SmOD model includes that list.

Finally, the SmOD pilots' requirement to include the Land Use theme is covered at its simplest by including the lu:hilucsLandUse property, the value of which should be a URI from the relevant code list in the INSPIRE registry. Since the domain of this property is lu:ExistingLandUse, if this is used then the spatial object can be inferred to be both a Protected Site and an Existing Land Use object.

### 4.3 Identifier

This is the INSPIRE ID class and is defined in the GCM. Its use is mandatory across the INSPIRE Data Model and, again, the GeoKnow XSLT does a good job of generating INSPIRE-conformant data.

### 4.4 Document Citation

Sites that are protected acquire that status through legislation or formal designation at some level and this is reflected in the PS model. The UK work (sensibly) asserts that documents that embody legislation are instances of the widely used foaf:Document class. It is usual to use Dublin Core to provide metadata about documents and the SmOD model does that for the document title, but there are other properties in the INSPIRE model that are more specific and these are reflected.

### 4.5 Geographical Name

Toponymy is never as straightforward as one might hope, and in multilingual Europe it's doubly complicated. The INSPIRE Data Model for Geographic Names is necessarily complex therefore and SmOD needs to handle this in full. The code lists in the registry provide SKOS Concepts/URIs for gn:nativeness, gn:nameStatus, gn:grammaticalGender and gn:grammaticalNumber. For example, there are two possible values for gn:nativeness:

---


Since a single name in a single language can be spelled with different alphabets, it's necessary to refer to the language of the Geographical Name as a separate class, not (just) as a language tag on the gn:text property of the Spelling of Name class. The EU Publications Office maintains a SKOS Concept scheme for languages based on 3 character codes (two are insufficient for their needs) so that, again, there are URIs for every language. The usual way to refer to a resource's language is using the Dublin Core property (dct:language). The range of this property is dct:LinguisticSystem so that one can infer that the SKOS Concept pointed to is also an instance of that class.

4.6 Cardinality

Linked Open Data, and the Semantic Web technologies on which it depends, use the Open World Assumption. That is, no dataset is complete, someone else can always say more about the things you're talking about. One consequence of this is that the concepts of cardinality, mandatory properties, enumerated values and so on are not readily handled. To specify cardinality constraints it's necessary to use OWL constructs as described by Stuart Williams in his work on URIs for location [URIPL]. There are other solutions for checking that a given RDF dataset conforms to a model that includes cardinality constraints, such as SPIN and relevant work is almost certain to begin soon at W3C to develop a standardised RDF validation language. The SmOD project and the JRC are well placed to inform and influence this work.

5 URI Construction

Linked Data depends on persistent HTTP URIs as identifiers for real world objects, concepts as definitions as well as documents. A great deal of work has been done and continues to be done around URI construction. For example, the Publications Office is currently leading the development of a policy on persistent URIs that will apply across the European Institutions. The ISA Programme’s Study on Persistent URIs [PURI] and Stuart Williams’ work on URI patterns for location [URIPL], carried out on behalf of the UK Government’s Linked Data Working Group, are of particular relevance to that work and projects such as SmOD.

The design pattern for spatial objects recommended by the UK Government work is:

http://{domain}{/collection*}{/so}{/class}{inspireLocalId}{/inspireVersionId}

where:

{domain} means the internet domain name.

{/collection*} means any number of segments that provide hints to humans as to the dataset to which the identified object belongs.

{/so} means the literal string /so hinting that this URI identifies a spatial object. Although optional, its use is strongly recommended as it immediately differentiates between the identifier for the spatial object, which is defined as an abstract representation, and /id which would, by convention, identify the real world place and /doc that would identify a document describing the real world place.

{/class} means an optional segment describing the class of spatial object, for example, ProtectedSite.

{inspireLocalId} means the local identifier for the site that SHOULD be included and SHOULD match the value of the gcm:localID property of the gcm:Identifier class.

{/inspireVersionId} means the optional segment giving the version ID. If used, this SHOULD match the value of the gcm:versionId property of the gcm:Identifier class.

Taking the example in Listing 12.1, the URI
http://geodata.gov.gr/id/ProtectedSite/Natura2000GR/GR2530003
might give:
http://geodata.gov.gr/Natura2000GR/so/ProtectedSite/GR2530003

i.e. this might be the collection of Natura 2000 classified sites in Greece, and then after the /so element we get the class (Protected Site) and local ID (GR2530003). However, the circumstance may be that this looks and feels wrong and so one should apply personal judgement. The original URI designed by the GeoKnow project seems correct, except that the type of identifier should be /so, not /id, i.e.

http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/GR2530003

6 Data Example

As an initial validation of the model it is useful to apply it to some real data. Listing 12.1 shows an example of data describing a protected site. Applying the SmOD Initial Data Model produces the following:

@prefix ps: <http://inspire.jrc.ec.europa.eu/schemas/ps/3.0/> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix gn: <http://inspire.jrc.ec.europa.eu/schemas/gn/3.0/> .
@prefix gcm: <http://inspire.jrc.ec.europa.eu/schemas/gcm/3.0/> .
<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/GR2530003> a ps:ProtectedSite;
  geo:hasGeometry
<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/Geometry/GR2530003> ;
  ps:siteName
<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/siteName/GR2530003> ;
  ps:siteProtectionClassification "natureConservation" ;
  ps:siteDesignation
<http://inspire.ec.europa.eu/codelist/Natura2000DesignationValue/proposedSiteOfComm
ityImportance/> ;
  gcm:identifier
<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/Identifier/GR2530003> .

<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/Geometry/GR2530003> a
geo:Geometry ;
  geo:asWKT "MULTIPOLYGON(((400140.625 4194794.5,
    400217.9689 4194590.5, 400697.5624 4194765.5001, 401044.5.939 4194788.5, 401276.5935
    4194635.5, 401587.9998 4194633.4998, 401738.8437 4194454.4999, 401803.2187
    4194631.5001, 402038.125 4194528.4998, 402130.5312 4194325.5001, 402086.2187
    4194121.5002, 401994.1249 4193970.4999, 401906.8438 4193767.4998, 401726.125
    4193615.5001, 401408.5938 4193439.4998, 401111.0935 4193236.4999, 400941.3123
    4193213.4998, 400760.5625 4193062.5, 400457.2812 4193140.4998, 400436.0624
    4192987.2501, 400330.9687 4193191.4998, 400246.1561 4192998.2499, 400071.7502
    4193040.2498, 400111.0314 4192837.2501, 399873.9373 4192584.2499, 399761.1251
    4192763.2498, 399537.2501 4192586.2501, 399250.8123 4192537.2498, 398815.5002
    4192565.2499, 398509.5625 4192643.2501, 398486.9376 4192662.2501, 398265.1877
    4192848.2499, 398033.9373 4193128.2499, 398228.2813 4193432.4999, 398485.3752
    4193660.5001, 398469.3752 4193939.4998, 398666.0624 4194218.5, 398854.5001
    419495.5001, 399082.2185 4194368.4999, 399227.7187 4194570.5002, 399554.5313
    4194594.4999, 399833.8438 4194719.4998, 399959.3436 4194541.5002, 400140.625
    4194794.5))

<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/GeographicalName/GR2530003> a
gn:GeographicalName ;
  gn:nameStatus <http://inspire.ec.europa.eu/codelist/NameStatusValue/official> ;
  gn:sourceOfName "Ministry of Environment, Energy, and Climate Change" ;
  gn:spelling
<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/SpellingOfName/GR2530003> .

<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/SpellingOfName/GR2530003> a
gn:SpellingOfName ;
  gn:text "AKROKORINTHOS"@en ;
  gn:script "Latin" .

<http://geodata.gov.gr/so/ProtectedSite/Natura2000GR/Identifier/GR2530003> a
gcm:Identifier ;
A direct comparison might be considered unfair since the Athena Research and Innovation Centre generated their data at scale using an automated system whereas the example here is a single hand-crafted Turtle instance. Noting that caveat, we proceed with caution.

First of all, the GeoKnow data uses the base namespace whereas SmOD uses gcm. This matches the work passed to SmOD in private by the JRC.

The SmOD data uses URLs that identify SKOS Concepts defined in the INSPIRE Registry wherever possible.

The SmOD data explicitly defines the type of each node in the graph.

It’s also notable that the GeoKnow data includes several properties for which there are no values. This is a result of machine processing rather than the line by line construction of the SmOD data. There are many more terms even in the subset of the INSPIRE Data Model than are used in this sample data. This is almost certainly going to be the case with the pilots, i.e. real world data is rarely complete.

It’s easier to spot errors in the data when one can match it against a model. For example, the GeoKnow data has some errors around the gn:spelling property and the gn:SpellingOfName class.

### 7 Putting the Model to Work

As has been emphasised, the Initial Data Model attempts to capture the elements that are common across the SmOD pilots. In each pilot, the data structure will need to be extended to include domain-specific vocabularies and may require the definition of new terms. When doing this the partners will follow the Best Practices set out by the W3C [LD-BP]. This includes advice on identifying and selecting standard vocabularies that, in the context of SmOD, are likely to include some permutation of EuroVoc, GEMET, AgroVoc, TaxonConcept[^34], EnvThes[^35], GeoSpecies Knowledge Base[^36] and the ISA Core Location Vocabulary[^37]. The authors of the current deliverable will support each of the pilots in developing their data models.

It is worth noting that the datasets used by the projects will, of course, be described by metadata. The DCAT Application Profile[^38] for European data portals, developed under the ISA Programme, is a key reference point for this. Štěpán Kafka of HSRS contributed to the

[^34]: http://www.taxonconcept.org/ontologies/
[^35]: http://www.lter-europe.net/info_manage/EnvThes
[^36]: http://lod.geospecies.org/
[^37]: http://www.w3.org/ns/locn
work carried out by Andrea Perego at the JRC on aligning INSPIRE metadata with the DCAT-AP. Again, this work is not yet finalised at the time of writing but is open to public review\(^{39}\). In addition to this, the metadata potential will be further exploited via possibilities to discover spatial information resources through spatial data infrastructure\(^{40}\) as well as linked data interfaces with mainstream web search engines.

8 Future Proofing

For many partners in the SmartOpenData project, the world of Linked Data, with its graphs, URIs and triples, is alien. The major step in creating the data for each pilot will be to apply and adapt existing deep knowledge of geospatial technologies to this world, use the Registry entries etc. That process in itself will constitute significant progress towards the project's goals.

The data model presented here is essentially our best guess for what will come out of the JRC in the coming months. Although some changes and corrections are all but inevitable, we believe these will be minor. Furthermore, partners are aware of this and will act accordingly. By using only terms from INSPIRE and a few terms from established vocabularies, SmOD is not introducing any new points of failure such as a namespace that would be hard to manage over the long term.

Based on outcomes of initial SmartOpenData model deployments further INSPIRE themes spatial is foreseen to be investigated in order to fulfil expectations of the pilots:

- Land cover
- Habitats and biotopes
- Species distribution
- Area Management/Restriction/Regulation Zones and Reporting Units

9 Conclusion

The SmartOpenData project is embarking on a series of pilots that need a common data structure encoded in RDF. Given the broader context of the project, it seems clear that the starting point for such a model should be INSPIRE. At the time of writing, the European Commission’s JRC is in the process of finalising a set of official RDF schemata for INSPIRE. Whilst these are not fully available, the project partners have been given advanced access to enough of the emerging work to make some well-informed approximations. Communication with many of the relevant individuals has been helpful too.


The result is an initial data model for the project that follows closely the underlying INSPIRE Data Model. The project foresees future changes to this model in the light of experience and the work undertaken in the related project work packages and their deliverables.

Furthermore, the model is open to extension in the domain-specific pilots.

The work of the GeoKnow project in converting data to conformant INSPIRE data in GML and then to RDF is very helpful and shows a scalable, repeatable route for SmOD to follow. The XSLTs produced by the Athena Research and Innovation Centre for GeoKnow deliverable 2.7.1 enable SmOD to hit the ground running. However, by defining an initial data model, SmOD partners can adapt and improve upon those XSLTs for each pilot and, it is believed, create SPARQL queries against the data more reliably.

10 Acknowledgements

Thanks are due to Andrea Perego and Michael Lutz at the JRC, Stuart Williams at Epimorphics, Clemens Portele of Interactive Instruments and Kostas Patroumpas at the Athena Research and Innovation Centre.

11 References


12 Listings

12.1 Sample Protected Site data from GeoKnow

The output from GeoKnow is in RDF/XML. This has been converted to Turtle and slightly reformatted for easier reading. It is a small sample from the data available from the Athena Research and Innovation Centre at
12.2 Sample SPARQL Query against geodata.gov.gr

This is one of the sample queries available from geodata.gov.gr and returns a list if protected sites within 20km of the given point. Without a graphical representation of the schema, writing queries such as this is cumbersome, requiring a good deal of trial and error. What new comers to SPARQL generally find hard is to think in terms of a graph rather than a set of tables and to do that, a picture of the data structure is extremely useful.

PREFIX ps: <http://inspire.jrc.ec.europa.eu/schemas/ps/3.0/>
PREFIX geo:<http://www.opengis.net/ont/geosparql#>
SELECT ?fName ?dist_km
WHERE {
  ?p gn:GeographicalName ?gnName .
  ?spt gn:text ?fName .
  ?fGeom geo:asWKT ?fWKT .
  FILTER (?dist_km < 20)
} ORDER BY ?dist_km
12.3 Snippet from INSPIRE GCM Ontology

This is a snippet from an OWL Ontology currently under review by the JRC for use in its imminent publication of a set of ontologies for the INSPIRE Data Model. It shows the use of OWL restrictions to define the cardinality of properties on the INSPIRE Id class within the Generic Conceptual Model.

gcm:Identifier
  rdf:type owl:Class ;
  rdfs:comment "The class of resources representing the INSPIRE spatial–object identifiers each comprising of a namespaceId, a localId and an optional versionId."@en ;
  rdfs:label "Inspire id"^^xsd:string ;
  rdfs:subClassOf owl:Thing ;
  rdfs:subClassOf
    [ rdf:type owl:Restriction ;
      owl:cardinality "1"^^xsd:nonNegativeInteger ;
      owl:onProperty gcm:localId ] ;
  rdfs:subClassOf
    [ rdf:type owl:Restriction ;
      owl:cardinality "1"^^xsd:nonNegativeInteger ;
      owl:onProperty gcm:namespaceId ] ;
  rdfs:subClassOf
    [ rdf:type owl:Restriction ;
      owl:maxCardinality "1"^^xsd:nonNegativeInteger ;
      owl:onProperty gcm:versionId ] ;
  skos:prefLabel "Inspire id"^^xsd:string .