



This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 603824.

The Final SmartOpenData Model

Deliverable D3.4 :: Public

Keywords: INSPIRE, schema, RDF

An abstract graphic composed of overlapping, semi-transparent blue and grey geometric shapes, resembling a stylized landscape or data structure.

Linked Open Data for environment protection in Smart Regions

Document Metadata

Contractual Date of Delivery to the EC:	May 2015		
Actual Date of Delivery to the EC:	15 June 2015		
Editor(s):	Phil Archer W3C/ERCIM		
Contributor(s)	Tatiana Tarasova (Spaziodati), Martin Tuchyna (SAŽP), Jindřich Mynarz (SAŽP)		
DOCUMENT HISTORY			
Version	Version date	Responsible	Description
0.01	22 April 2015	W3C	Initial model description
0.02	22 April 2015	W3C	Initial feedback incorporated
0.03	24 April 2015	W3C	Further feedback from pilots
0.04	7 May 2015`	W3C	Incorporating corrections & feedback from TT
0.05	17 May	W3C	Initial incorporation of EF theme
0.06	27 May	W3C	Adding Intro section etc.
0.07	5 June	W3C	Inclusion of revisions to EF theme from Spaziodati
1.0	15 June	TRAGSA	Minor editorial changes

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.

Copyright © 2015, SmartOpenData Consortium.

Table of Contents

The Final SmartOpenData Model.....	1
Executive Summary	4
1 Introduction	5
2 From Initial to Final SmartOpenData Model.....	5
3 The Model, Theme by Theme	9
3.1 Namespaces.....	9
3.2 The GCM & Geographical Names	9
3.3 Protected Sites.....	10
3.4 Land Use	12
3.5 Administrative Units	12
3.6 Bio-Geographical Regions.....	13
3.7 Species Distribution	15
3.8 Corine Land Cover.....	16
3.9 Environmental Monitoring Facilities	17
Conclusion	19

List of Figures

Figure 1 The initial SmartOpenData model.....	7
Figure 2 The Final SmartOpenData model.	8
Figure 3 The Natura 2000 And Emerald Bio-geographical Region Classification.....	14
Figure 4: EU-Nomen: 97523, EUNIS: 1023 (Yellowhammer)	15

Executive Summary

SmartOpenData has drawn on the experience of previous work done by the GeoKnow project and under the ARE3NA project within the JRC to produce the beginnings of an RDF vocabulary that mirrors the INSPIRE Data Model. Importantly, the model does not try to replicate the whole of the INSPIRE model, rather, it offers a model that is suitable for use in Linked Data structures, one that is amenable to linking geospatial and environmental data to other sources of data on the Web.

The vocabularies have been installed within the highly stable w3.org namespace and cover the aspects needed for the pilots conducted within SmartOpenData:

- The Generic Concept Model that underpins INSPIRE
- Protected Sites
- Land Use
- Administrative Units
- Bio-geographical Units
- Species Distribution
- Corine Land Cover
- Environmental Monitoring Facilities

These may be extended by other projects wishing to cover further INSPIRE themes. Additional terms that are specific to SmartOpenData have also been defined. These are discussed separately in D3.5.

1 Introduction

SmartOpenData is testing the usefulness of Linked Data to solve a number of different problems related to the rural economy and environment. The pilots are very different in terms of the kind of problems being tackled and so the common aspects are limited. The initial expectation was that the project would develop a core data model of its own, this would be updated towards the end of the project, and each pilot harmonised using the core model as a basis. In practice, the partners decided that a better approach would be to make maximum use of the INSPIRE data model as the basis for interoperability.

As explained in the following section, ideas around how to do this have evolved in the light of new developments and experiences gained. The end result is that SmartOpenData has defined:

1. A set of RDF classes and properties that make use of the aspects of the INSPIRE data model relevant to the pilots. These can be extended to cover more of the INSPIRE themes if necessary.
2. A separate vocabulary that is specific to the SmOD pilots only.

This document describes the first of these; the second, the specific SmOD vocabulary, will be described in D3.5.

2 From Initial to Final SmartOpenData Model

The initial SmartOpenData model was developed over the first half of 2014 and delivered as D3.2. It attempted to capture relevant sections of the INSPIRE data model and render them faithfully as an RDF schema. This general approach matched the work in the GeoKnow project where XSLTs were used to derive RDF data from an input of INSPIRE-compliant XML.

In the final model, much of this has been rejected in favour of a simpler approach that is more in line with 'Linked Data thinking' and no longer attempts to recreate the full scope of INSPIRE in RDF. There are two principal motivations for this:

1. Experience: when creating Linked Data for use in the pilots, a slavish following on INSPIRE proved burdensome and unhelpful. The aim of taking a different approach (Linked Data) must be to gain some benefit from that approach not available from the original (XML/GML), recognising that the original will offer features not available in the derived work.
2. The publication of the Study on RDF and PIDs for INSPIRE by Diederik Tirry and Danny Vandembroucke under ISA Action 1.17: A Reusable INSPIRE Reference Platform (ARE3NA). This report summarised work by three experts: Clemens Portele, Linda van den Brink and Stuart Williams. Some of this work was shared with the SmartOpenData (SmOD) project before publication but the summary report was not available at that time. The summary proved extremely useful to the current project partners. All documents from that work are available at <https://ies-svn.jrc.ec.europa.eu/documents/25> and remain marked as 'for review' although a

conversation with JRC staff suggests that no further work is foreseen on these documents.

One call to action from the ARE3NA work was that the INSPIRE Registry¹ be extended to include SKOS concept schemes in addition to the formats already offered. This has been done and allows SmOD to use the registry's persistent URIs as identifiers for many of the concepts that are important in the current work.

It is this combination of factors that is behind the final model being at once simpler and much more comprehensive than the initial one in its coverage of the INSPIRE themes. This can be seen in the most cursory look at Figure 1 that shows the initial model and Figure 2 that shows the final one. In addition to the basics, the initial model covers just the Protected Sites and one relationship from the Land Usage theme. The final model covers more themes but with the same number of classes. Note in particular that the three classes associated specifically with Geographical Names have disappeared altogether.

¹ <http://inspire.ec.europa.eu/registry/>

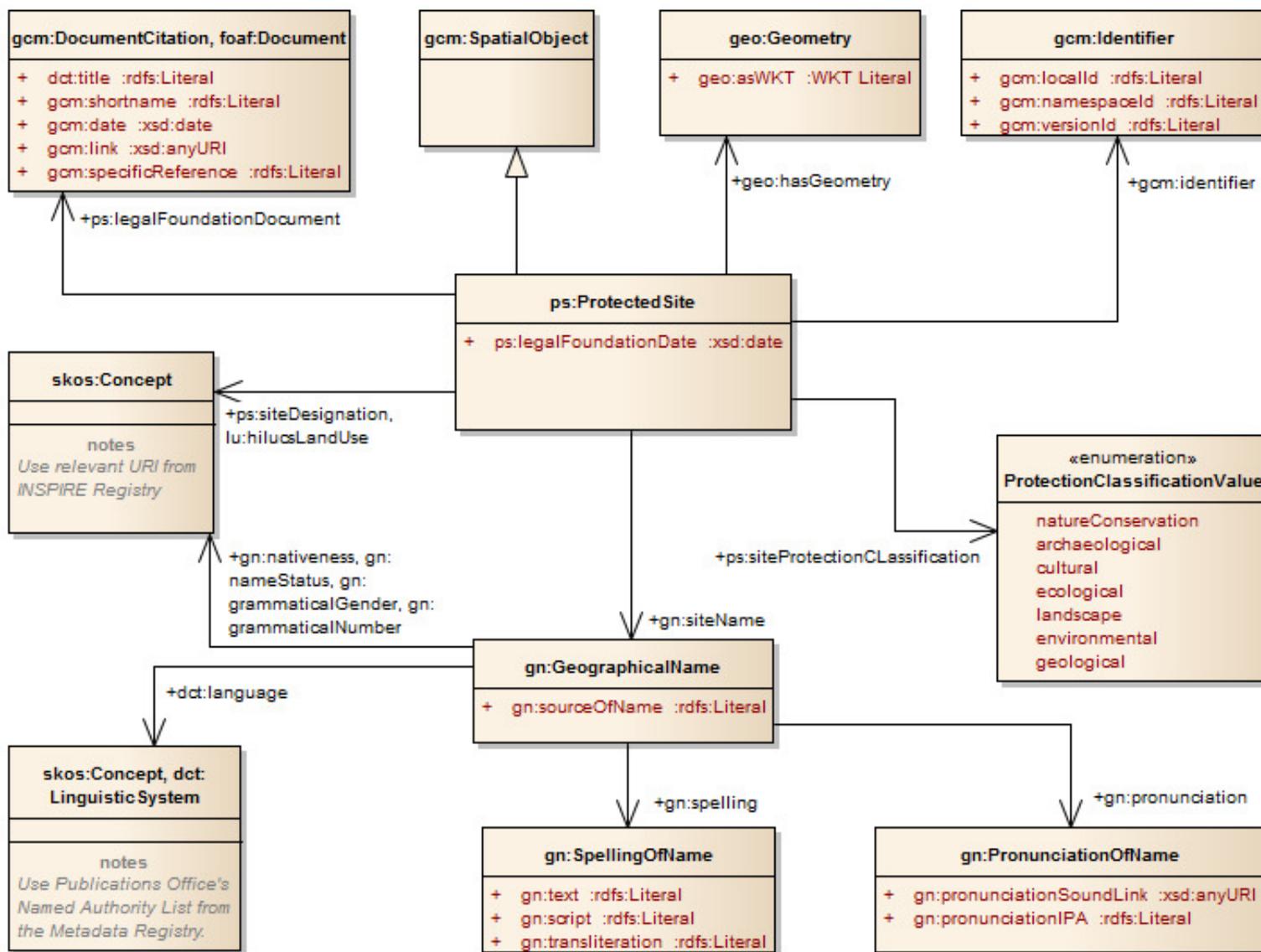


Figure 1 The initial SmartOpenData model tried to capture the full complexity of the INSPIRE model

3 The Model, Theme by Theme

This section describes the decisions made when incorporating each INSPIRE theme in the SmartOpenData model.

3.1 Namespaces

The Study on RDF and PIDs's first recommendation is that RDF namespaces should be aligned with the XML namespaces. This was done in the initial SmOD model so that, for example, the namespace used for Protected Sites was <http://inspire.jrc.ec.europa.eu/schemas/ps/3.0/>. However, the project received anecdotal information from the JRC that the ARE3NA work had effectively stopped for now and that SmOD should not wait for the RDF schemas to be developed and published at those URLs. Therefore the decision was taken to use the most stable and recognisable namespaces available to the project. These will be of the form

<http://www.w3.org/2015/03/inspire/{xx}>

where xx is the INSPIRE theme. The W3C website is extremely stable and these namespaces should be considered as persistent although they are not the product of any W3C working group. If the JRC were to publish its own schemas then the ones created by SmOD would be deprecated in favour of them. If, however, the JRC or other parties wish to extend the vocabularies hosted at W3C then this would be possible, particularly through the Locations and Addresses Community Group that is co-chaired by Andrea Perego and Michael Lutz of the JRC, together with Frans Knibbe of Geodan.

3.2 The GCM & Geographical Names

The most visible change from the initial model to the final one is the elimination of the Geographical Names theme. The full INSPIRE model supports the provision of multiple spellings of place names, using multiple scripts, linked to audio files for pronunciation and more. Initial work by Spaziodati on the Portuguese-Spanish pilot used the model that mirrored this. However, the result was a lot of complexity in the data with many of the properties unused and several unnecessary blank nodes. The Study on RDF and PIDs's recommendation in section 3.2.12 is to simply use `rdfs:label` (a string with an optional language tag). This simplification risks losing some of the rich data that *might* be available in some situations but is absent in the SmOD pilots. Making this change was inline with the overall recommendation of 'thinking Linked Data' and made the data in the pilots much easier to work with at a stroke.

The SmOD partners have also used GeoSPARQL's `geo:SpatialObject` class in preference to creating a new class of `gcm:SpatialObject`. The definition of `geo:SpatialObject` is: "The class Spatial Object represents everything that can have a spatial representation. It is superclass of feature and geometry."² This very general class

² <http://www.opengis.net/doc/IS/geosparql/1.0>, p6.

therefore fits very well within INSPIRE and the SmOD pilots and there is no gain in defining an INSPIRE-specific class of the same name.

3.3 Protected Sites

Class	<code>ps:ProtectedSite</code>
Object Properties	<code>ps:legalFoundationDocument</code> <code>ps:siteDesignation</code> <code>ps:siteProtectionClassification</code> <code>ps:isManagedBy</code>
Data type property	<code>ps:legalFoundationDate</code>
Concept schemes used	INSPIRE Registry, Protection Classification
External vocabularies used	FOAF, ORG

In the initial model, Protected Sites are defined as a sub class of `gcm:SpatialObject` with a single data type property of `ps:legalFoundationDate` that has a range of `xsd:date`. In the final model we use `geo:SpatialObject`. Protected Sites are defined by a document that details the relevant protection, this might be legislation but is more usually some sort of order or notice. Using `ps:legalFoundationDocument` to link to a class describing a document, such as the `gcm:DocumentCitation` class is quite awkward from a Linked data point of view. The natural thing to do is simply to link to the document itself and that document will be an instance of the well used `foaf:Document` class. The `ps:legalFoundationDocument` property has this as its range. But such documents may not be available online (or their URL unknown) and so a method of referring to offline documents needs to be provided. SmOD simplifies the various properties of the INSPIRE `gcm:DocumentCitation` class (title, shortname, date etc.) down to the `dcterms:bibliographicCitation` property. Where a legal foundation document is not available online, a blank node will be created in the graph with this property that gives a reference to the actual document. Where the document does exist online it will be linked to directly.

The Protected Sites theme is the first of many in the SmartOpenData model that makes use of the SKOS concept schemes published in the INSPIRE Registry.

The `ps:siteDesignation` property can point to one or more of the specialisations of the `Designation Value code list type` (<http://inspire.ec.europa.eu/codelist/DesignationValue/>). These are:

- IUCN Designation (7 possible values)
- National Monuments Record Designation (18 possible values)
- Natura2000 Designation (5 possible values)
- Ramsar (1 possible value)

- Unesco Man and Biosphere Programme Designation (1 possible value)
- UNESCO World Heritage Designation (3 possible values).

So for example, if a site were designated as an area of special conservation under Natura 2000, the value of the `ps:siteDesignation` property would be

<http://inspire.ec.europa.eu/codelist/Natura2000DesignationValue/specialAreaOfConservation>

The provision of SKOS Concepts schemes makes this easy and avoids this or any other project writing its own version of designation schemes like Natura 2000. However, the Registry does not provide SKOS concepts schemes for all aspects of INSPIRE or some of the closely related data models. For example, the `ps:siteProtectionClassification` property in the Protected Sites theme takes one of 7 enumerated values:

- natureConservation
- archaeological
- cultural
- ecological
- landscape
- environment
- geological

In XML-centric systems these would be provided as strings but in Linked Data, they are better rendered as SKOS concepts so that they can be pointed to via their URI, with multilingual labels etc.

A very simple SKOS concept scheme was created to provide such URIs for the values in the ProtectedClassificationValue enumeration at

<http://www.w3.org/2015/03/inspire/ProtectionClassification#>

replacing the namespace URI with the prefix `pspc`, each of the terms in the list can be referred to as `pspc:natureConservation`, `pspc:archaeological` etc. Note that the lower camel case capitalisation has been preserved from the original, rather than the more usual practice in Linked Data of naming classes using title case.

The SmOD data model makes use of the Protected Sites Simple data model from INSPIRE but takes one extra class and relationship from Protected Sites Full, namely `ps:isManagedBy`. This has a range of `foaf:Agent` to keep the vocabulary as general as possible but it is expected that in practice, `org:Organization` (or one of its sub classes) will be used. `org:Organization` is a sub class of `foaf:Agent`. In some cases `foaf:Group` or even `foaf:Person` will be better, both of these are also sub classes of `foaf:Agent`. As well as basic information like the organisation's name, the ORG ontology³ is recommended as it has the following features:

- posts and roles,

³ <http://www.w3.org/TR/vocab-org/>

- it can model multiple locations for a single organisation,
- it can model changes in an organisation's name.

The latter aspect matches the Responsible Agency class in Protected Sites Full that includes properties for recording the beginning and end of the agency's lifespan.

3.4 Land Use

Class:	lu:ExistingLandUseObject
Object Property:	lu:hilucsLandUse

In the same way that the INSPIRE Registry is used as a source of SKOS concepts as value for the `ps:siteDesignation` property, the `lu:hilucsLandUse` property can link a Spatial Object to one of the values from the code list at <http://inspire.ec.europa.eu/codelist/HILUCSValue/>. The domain of `lu:hilucsLandUse` is the `lu:ExistingLandUseObject` and so systems can, at least in theory, infer that any Spatial Object that has a `lu:hilucsLandUse` property is also an instance of `lu:ExistingLandUseObject`.

3.5 Administrative Units

Class	au:AdministrativeUnit
Object Properties	au:nationalLevel au:country
Data type property	au:nationalCode
Concept schemes used	INSPIRE Registry, MDR (countries)

From a SmartOpenData perspective, the Administrative Units theme is very simple. The `AdministrativeUnit` class itself is defined as a sub class of the `geo:SpatialObject` class so it inherits `rdfs:label` as the property for its name and the usual means of providing boundary information (via `geo:Geometry`). Administrative Units typically have a national code associated with them and this is provided as a string value for the `au:nationalCode` property which is defined as a sub property of `skos:notation`.

The INSPIRE Registry provides a SKOS concept scheme for the Administrative Hierarchy Level (<http://inspire.ec.europa.eu/codelist/AdministrativeHierarchyLevel/>) with URIs for the 6 levels as

<http://inspire.ec.europa.eu/codelist/AdministrativeHierarchyLevel/1stOrder/>,
<http://inspire.ec.europa.eu/codelist/AdministrativeHierarchyLevel/2ndOrder/> etc.

These URIs are the value for the `au:nationalLevel` property.

Finally, SmOD can make use of the Metadata Registry (MDR) provided by the European Publications Office as a source of URIs as values for `au:country`⁴. This URI set provides the names of all countries in the world in all official languages of the EU and follows a predictable pattern, based on a country's ISO 3166 3 character code:

<http://publications.europa.eu/resource/authority/country/FIN>

<http://publications.europa.eu/resource/authority/country/GBR>

etc.

The downside of using these URIs is that, for now, they are not resolvable. The Publications Office is known to be working on making them so but at the time of writing they are not following Linked Data principles – something the Publications Office is very aware of.

3.6 Bio-Geographical Regions

Class	<code>br:Bio-geographicalRegion</code>
Object Properties	<code>br:regionClassification</code> <code>br:regionClassificationLevel</code>
Concept Schemes Used	INSPIRE Registry, EEA Codelist for bio-geographical regions, Europe 2011

INSPIRE recognises 4 regional classification schemes within this theme:

- Environmental Stratification Classification
- Marine Strategy Framework Directive Classification
- Natura 2000 And Emerald Bio-geographical Region Classification
- Natural Vegetation Classification

The Natura 2000 And Emerald Bio-geographical Region Classification is the one of most interest for SmartOpenData. The European Environment Agency maintains this list and publishes it in a variety of formats⁵.

Code	Name	Region	pre_2012
alpine	Alpine Bio-geographical Region	Bio-geographical Region	ALP
Anatolian	Anatolian Bio-geographical Region	Bio-geographical Region	ANA

⁴ <http://publications.europa.eu/mdr/authority/country/>

⁵ <http://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-1/codelist-for-bio-geographical-regions/codelist-for-bio-geographical-regions/>

Code	Name	Region	pre_2012
arctic	Arctic Bio-geographical Region	Bio-geographical Region	ARC
atlantic	Atlantic Bio-geographical Region	Bio-geographical Region	ATL
blackSea	Black Sea Bio-geographical Region	Bio-geographical Region	BLS
boreal	Boreal Bio-geographical Region	Bio-geographical Region	BOR
continental	Continental Bio-geographical Region	Bio-geographical Region	CON
macaronesian	Macaronesian Bio-geographical Region	Bio-geographical Region	MAC
marineAtlantic	Marine Atlantic Region	Marine Region	MATL
marineBaltic	Marine Baltic Region	Marine Region	MBAL
marineBlackSea	Marine Region Black Sea	Marine Region	MBLS
marineMacaronesian	Marine Macaronesian Region	Marine Region	MMAC
marineMediterranean	Marine Mediterranean Region	Marine Region	MMED
Mediterranean	Mediterranean Bio-geographical Region	Bio-geographical Region	MED
pannonian	Pannonian Bio-geographical Region	Bio-geographical Region	PAN
steppic	Steppic Bio-geographical Region	Bio-geographical Region	STE

Figure 3 The Natura 2000 And Emerald Bio-geographical Region Classification

An RDF vocabulary at <http://rdfdata.eionet.europa.eu/eea/biogeographic-regions2011.rdf> provides URIs for each of the classifications in the form

`http://rdfdata.eionet.europa.eu/eea/biogeographic-regions2011/{code}` and these can be used as values for the `br:regionClassification` property. Since this is not published as a SKOS Concept scheme per se, the range of `br:regionClassification` is undefined. The property could therefore also be used to link to Concept schemes for this or any of the other regional classification schemes. In common with `lu:hilucsLandUse`, the domain of `br:regionClassification` is defined. In this case the domain is `br:Bio-geographicalRegion`, which allows systems to infer that a `geo:SpatialObject` with the property is also an instance of its

subclass, `br:Bio-geographicalRegion`. This is not shown in Figure 2 to aid readability.

The `br:regionClassificationLevel` property links directly to the concept scheme in the INSPIRE Registry at <http://inspire.ec.europa.eu/codelist/RegionClassificationLevelValue/> that provides URIs for the 4 possible values of International, Local, National and Regional in the form <http://inspire.ec.europa.eu/codelist/RegionClassificationLevelValue/{code}> where {code} is the terms from the list all in lower case.

3.7 Species Distribution

Classes	<code>sd:SpeciesDistributionUnit</code> <code>sd:Species</code>
Object Properties	<code>sd:eunisSpeciesCode</code> <code>sd:occurrenceCategory</code> <code>sd:hasSpecies</code>
Data type properties	<code>sd:eunomenID</code> <code>smod:eunomenPage</code>
Concept Schemes Used	INSPIRE Registry, EUNIS/EEA, EU-NOMEN



Figure 4: EU-Nomen: 97523, EUNIS: 1023 (Yellowhammer)

The Species Distribution theme provides a framework to support detailed information about population densities, counting methodologies etc. For SmOD, and again, when 'thinking Linked Data,' it is sufficient to use a simpler model.

The `sd:SpeciesDistributionUnit` class uses `sd:hasSpecies` to link to a class that represents any species of interest. This is equivalent to INSPIRE's Species Name Type. Species can be identified in multiple ways.

The European Environment Agency maintains its European Nature Information System, EUNIS⁶, as a URI set for species and serves the data in a HTML or RDF/XML using content negotiation. The URIs are of the form

<http://eunis.eea.europa.eu/species/{species No}>

so that, for example, the yellowhammer is identified by

<http://eunis.eea.europa.eu/species/1023>.

⁶ <http://eunis.eea.europa.eu/>

The data returned from the EUNIS system is comprehensive, providing the species' vernacular name in the official languages of the EU and equivalent identifiers from many other schemes. SmOD defines the domain of both `sd:eunisSpeciesCode` and `sd:occurrenceCategory` as `sd:Species`. The range of `sd:occurrenceCategory` is SKOS Concept and the INSPIRE Registry provides the relevant concept scheme at <http://inspire.ec.europa.eu/codelist/OccurrenceCategoryValue/> but for `sd:eunisSpeciesCode`, the range is `eunis:SpeciesSynonym`, the type defined in the EUNIS data.

One of the other identifiers included in the EUNIS data is the EU-Nomen identifier which is present in some of the data used in the SmOD pilots. This identifier can be included directly in SmOD data using the `sd:eunomenID` property, which is defined as a subProperty of `skos:notation`. The literal value, e.g. 97523, is typed as such. A further property, `smod:eunomenPage`, links the species to its EU-Nomen Web page, e.g.

<http://www.eu-nomen.eu/portal/taxon.php?GUID=urn:lsid:faunaeur.org:taxname:97523>

This page, indeed EU-Nomen, is not Linked data friendly since non-URI identifiers are used and the associated information is only available as a Web page, not as RDF. The EUNIS system uses the `eunis:sameSynonymFaEu` property to provide the EU-Nomen species number and the property is defined as having a domain of `eunis:SameSynonym` and range of literal. This is semantically close enough to define `sd:eunomenID` as a subproperty of this as well as of `skos:notation`. The two together provide the detailed semantics we need – that the literal value is of a specific type and that that value can also be matched against values of the `eunis:sameSynonymFaEu` property.

As with any class, properties like `rdfs:label` may be used to give the name of the species as a string literal if needed.

3.8 Corine Land Cover

Classes	None
Object Property	<code>lc:corineLandCover</code>
Schemas Used	Corine Land Cover

The Corine Land Cover taxonomy is defined by EIONET⁷ and published as a set of Web pages. At the time of writing, the partners understand that plans are in place to publish it as a SKOS Concept scheme but that has not yet happened. Therefore, a scheme was created and published at <http://www.w3.org/2015/03/corine>. The definition text for each concept is taken from the EIONET pages and served in HTML, RDF/XML and Turtle. The RDF serialisations also include labels and definitions in Spanish and Slovak as well as English, the latter taken from SAŽP's website at <http://www.SAŽP.sk/slovak/struktura/ceev/DPZ/CLC2000/corine/english.html>. SAŽP also

⁷ <http://sia.eionet.europa.eu/CLC2000/classes>

supplied the RGB colours associated with each Corine Land Cover type and these are used in the HTML page.

An issue to highlight in this work is the choice of identifiers for each CLC type. These are usually in available data as three digit numbers, sometimes with, sometimes without separating dots (i.e. 111 or 1.1.1). It proved much easier therefore to use these numbers in the URIs than to use the names to create URIs like <http://www.w3.org/2015/03/corine#ContinuousUrbanFabric>. However, XML, and therefore RDF, requires that class names begin with either a letter or an underscore to each class begins with 'clc' – something that easy be inserted when processing input data.

It should be noted that Clemens Portele created a vocabulary for recording Corine Land Cover datasets at <http://portele.de/ont/inspire/lcv#>. This captures the full complexity of the original model that, again, goes beyond what is required for SmOD. The equivalent property to `lc:corineLandCover` in Portele's work is `lcv:class` defined thus:

```
lcv:class a owl:ObjectProperty ;
  rdfs:comment "The range is a type for which no RDF
              representation is known:
              LandCoverClassValue"@en ;
  rdfs:range owl:Class ;
  skos:definition "The assignment of a land cover class
                  through a classification code
                  identifier"@en ;
  skos:notation "class"^^xsd:NCName ;
  skos:prefLabel "class"@en ;
  skos:scopeNote "The identifier, eg 1, 1.1.2, ... (for CORINE
                  LC classes) allow to access to the value and the definition or
                  narrative description of the corresponding class."@en.
```

Rather than create the missing SKOS Concept scheme, Portele defines a general object property (confusingly called 'class') that has a range of `owl:Class`. SmOD would like to refer to this work but to keep the data simple, the `lc:corineLandCover` property is defined as a sub property of `lcv:class` and has a range of `skos:Concept`.

3.9 Environmental Monitoring Facilities

Class	<code>ef:EnvironmentalMonitoringFacility</code>
Object properties	<code>ef:mediaMonitored</code>
Datatype properties	<code>ef:specialisedEMFType</code> <code>ef:purpose</code>
Schemes used	INSPIRE Registry

The EF vocabulary was created based on the INSPIRE Environmental Monitoring Facilities (EMF) theme⁸. The scope includes the monitoring facilities and the observations linked to them. Observations vary between use cases and the specific details of the uses within SmOD are discussed in D3.5. However, it is worth noting here that the RDF Data Cube vocabulary⁹ has been used extensively as this provides a method for recording statistical hypercube data in RDF. Examples of the data cube structures used (for the Italian and Portuguese-Spanish pilots) can be seen at in the relevant Github repository¹⁰.

One of the specific cases for using the EF theme is to combine Protected Sites from Natura2000¹¹ and various water measurements from Waterbase (Lakes¹² Rivers¹³ and UndergroundWaters¹⁴). Several possible user queries were defined¹⁵ that ARPA wanted to address in the pilot, some of which involved further datasets. These queries dictated the requirements for the model as follows:

- the model must allow for defining locations of monitoring stations (to link them to protected sites);
- the model must allow for representing measurements of hazardous substances in water by monitoring stations (to determine polluted water):
 - what substances were found;
 - concentration of the substances;
 - unit of measurement used to measure them.

To satisfy these requirements, the following classes and properties were defined.

`ef:EnvironmentalMonitoringFacility` is a spatial entity that collects or processes data about real-world objects whose properties (physical, chemical, biological or other aspects of environmental conditions) are observed or measured. `ef:EnvironmentalMonitoringFacility` is defined as a sub-class of `geo:SpatialObject`.

`ef:specialisedEMFType` provides categorisation of EMF, such as platform, site, station, sensor, etc. INSPIRE has a codelist for this¹⁶ however, it is empty and therefore adds little value. For the ARPA pilot, the need is to represent the EMF type for humans, not to

⁸ http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_EF_v3.0.pdf

⁹ <http://www.w3.org/TR/vocab-data-cube/>

¹⁰ <http://smod-fp7.github.io/>

¹¹ <http://www.eea.europa.eu/data-and-maps/data/natura-5>

¹² <http://www.eea.europa.eu/data-and-maps/data/waterbase-lakes-10>

¹³ <http://www.eea.europa.eu/data-and-maps/data/waterbase-rivers-10>

¹⁴ <http://www.eea.europa.eu/data-and-maps/data/waterbase-groundwater-10>

¹⁵ <https://docs.google.com/spreadsheets/d/1J0LyaAEqje-bqUyagOSa9TGZJ9RjnBof2DE6IKWwkMU/edit?usp=sharing>

¹⁶ <http://inspire.ec.europa.eu/codelist/SpecialisedEMFTypeValue/>

integrate/link it with similar datasets and so `ef:specialisedEMFType` is a datatype property that takes a literal value.

The code list for 'purpose', defined in SMOD as `ef:purpose`, is also empty¹⁷, so, again, SMOD defines this as a datatype property.

The INPSIRE registry does, however, include a SKOS Concept Scheme that offers values for the `ef:mediaMonitored` property¹⁸ (air, biota, etc.). This object property therefore has a range of `skos:Concept` and a domain of `ef:EnvironmentalMonitoringFacility`.

Conclusion

SmartOpenData operates in a very active field, benefiting from work in other projects and making its own contributions. The project has created a set of simple vocabularies, driven by real world use cases, that are true to the INSPIRE Data Model and that can be re-used and extended by others as driven by their use cases.

¹⁷ <http://inspire.ec.europa.eu/codelist/PurposeOfCollectionValue/>

¹⁸ <http://inspire.ec.europa.eu/codelist/MediaValue/>