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

Review of geographic resources metadata and related metadata standards

Deliverable D3.1 :: Public

Keywords: Metadata, ontologies, vocabulary, geospatial, W3C, OGC, RDF, OWL

Linked Open Data for environment protection in Smart Regions

Table of Contents

Contractual Date of Delivery to the EC:		March 2014	
Actual Date of Delivery to the EC:			
Editor(s):		Maris Alberts, IMCS UL	
Contributor(s):		Karel Charvat (HSRS), Phil Archer (W3C), Otakar Čerba (HSRS), Maris Alberts (IMCS), Peteris Bruns (IMCS), Premysl Vohnout (HSRS), Stepan Kafka (HSRS), Pavel Vlach (University of West Bohemia), Jhon O'Flaherty (MAC), Jesus Mª Estrada Villegas (TRAGSA)	
DOCUMENT HISTORY			
Version	Version date	Responsible	Description
0.01	28 December 2013	IMCS	Initial draft, TOC, call for contributons from others.
0.02	1 January 2014	HSRS	Contributions on metadata chapters.
0.03	13 January 2014	HSRS	Geographic schemas
0.04	20 January 2014	IMCS	General chapters
0.05	13 February 2014	IMCS,HSRS	TOC changes
0.06	20 February 2014	IMCS	Serialization libraries
0.07	11 March 2014	HSRS	Metadata LOD standarts
0.08	19 March 2014	IMCS, HSRS	Document merge, and last updates.
0.09	20 March 2014	IMCS	Typo clean-up, document updates from contributor comments.
0.10	21 March 2014	MAC	Typo clean-up.
0.11	24 March 2014	W3C	Initial results from the Linking Geospatial Data Workshop
1.0	25 March 2014	TRAGSA	Format changes

Introduction

This deliverable contains review of existing geographic resource metadata and related metadata standards. The report contains review of available technology implementations to provide existing data serialization and publishing.

The description of the LOD vocabularies is done in two parts - Geographic schemas (chapter 3), which define above all geometry and (types of) features, and Geo-domain vocabularies (chapter 4) working mainly with attributes (descriptive vocabularies).

1 General LOD standards

1.1 Resource Description Framework – RDF

The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. Nowadays it is a standard model for data interchange on the Web. It has come to be used as a general method for conceptual description or modelling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats.

The RDF data model is similar to classic conceptual modelling approaches such as entity–relationship or class diagrams, as it is based upon the idea of making statements about resources (in particular web resources) in the form of subject-predicate-object expressions. These expressions are known as triples in RDF terminology.

RDF is not like the tabular data model of relational databases. Nor is it like the trees of the XML world. Instead, RDF is a graph. Therefore RDF swaps object for subject that would be used in the classical notation of an entity–attribute–value model within object-oriented design; object (city), attribute (name) and value (“Berlin”).

RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schema over time without requiring all of the data consumers to be changed.

RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”). This simple model allows structured and semi-structured data to be mixed, and shared across different applications.

For RDF content description two data modelling languages can be used: RDFS (Resource Description Framework Schema) and OWL (Web Ontology Language). **RDFS and OWL are also called RDF vocabularies in which are defined the terms and concepts used in RDF**. Two or more Semantic Web applications that share data must agree on a common used concepts and terminology, hence vocabularies are used in RDF.

1.1.1 Resource Description Framework in Attributes - RDFa

RDFa is a W3C Recommendation that adds a set of attribute-level extensions to HTML, XHTML and various XML-based document types for embedding rich metadata within Web documents. The RDF data-model mapping enables its use for embedding RDF subject-predicate-object expressions within XHTML documents. It also enables the extraction of RDF model triples by compliant user agents.[1]

Working with RDFa it is important to remember, that RDFa 1.1 automatically includes popular vocabularies and it is not necessary to define them. The RDFa Core Initial context includes the vocabularies listed at http://www.w3.org/2011/rdfa-context/rdfa-1.1. The content of the vocabulary prefixes, to be included in the RDFa 1.1 Default Profile, is

1 http://en.wikipedia.org/wiki/RDFa

defined based on the general usage of those vocabularies on the Semantic Web. This general usage is established using search crawl data, courtesy of Sindice and of Yahoo![2]

1.2 Modelling languages

RDFS - Resource Description Framework Schema

RDFS (Resource Description Framework Schema) is the most basic and widely used schema language of the Semantic Web technology's. It is lightweight and very easy to use and get started with. Many of the most popular RDF vocabularies are written in RDFS.

OWL - Web Ontology Language

OWL (or Web Ontology Language) is the ontology (think "schema") language of the Semantic Web. Although OWL is a modelling language in the classical sense, it has many advantages compared to the modelling languages that came before it. The Web Ontology Language is a family of knowledge representation languages or ontology languages for authoring ontologies or knowledge bases. The languages are characterized by formal semantics and RDF/XML-based serializations for the Semantic Web. OWL is endorsed by the World Wide Web Consortium (W3C) and has attracted academic, medical and commercial interest.[3]

1.3 RDF notation and serialization formats

With RDF, there is not a single way to represent it. Instead there are several valid serializations for RDF data[4]:

- RDF/XML. The most popular and widely suported serialization format of RDF is Extended Markup Language (XML), also originally introduced by first RDF specification. This is simply RDF represented as valid XML. This was originally proposed and used due to the plethora of existing tools that could parse and store XML. RDF/XML is verbose and somewhat difficult to read and write as a human, though it can be read and written by just about any RDF tool, so you'll see it around. It's usually not the best serialization to use.
- N-Triples. N-Triples is a very basic RDF serialization. Its key feature is that only one triple exists per line so that it's very quick to parse and Unix command-line tools can easily manipulate it. It's also highly compressible, so large, public RDF sources like DBpedia often publish data in N-Triples form.
- RDFa (RDF embedded in HTML). You can embed RDF data within normal Web pages by using RDFa. This is a very powerful technique that has been used by major companies such as Best Buy.

2 http://www.w3.org/2010/02/rdfa/profile/data/

3 https://www.cambridgesemantics.com/semantic-university/rdfs-vs.-owl

4 https://www.cambridgesemantics.com/semantic-university/rdf-nuts-and-bolts

- Notation3 (N3 for short). This is a largely legacy serialization that was originally proposed by Tim Berners-Lee in 1998. It extended RDF with a form of first-order logic that was never very popular. It is included here for completeness.
- Turtle (subset of N3). If you're writing RDF today, you're probably writing it in Turtle. Turtle is significantly more compact than RDF/XML, more readable than N-Triples, and lacks the first-order logic extensions from Notation3. Furthermore, the SPARQL query language expresses RDF queries in almost exactly the same way.
- TriG. TriG is Turtle but with support for named graphs. It's the de facto standard for serializing RDF with named graphs.
- JSON/JSON-LD JSON-LD is an official Web Standard and W3C Recommendation since 16 January 2014[5]. JSON is a useful and widely used data serialization and messaging format in the Web and other environments. The JSON-LD specification defines, a JSON-based format to serialize Linked Data. The syntax is designed to easily integrate into deployed systems that already use JSON, and provides a smooth upgrade path from JSON to JSON-LD. It is primarily intended to be a way to use Linked Data in Web-based programming environments, to build interoperable Web services, and to store Linked Data in JSON-based storage engines. The JSON-LD serialization examples can be found in the web page http://json-ld.org/playground/index.html .
- Support for the serialization of geospatial data on the basis of the JSON-LD specification has been started with work on GEOJSON-LD. Vocabulary (http://geojson.org/vocab) and the source https://github.com/geojson/geojson-ld .

1.4 Vocabularies

On the Semantic Web, vocabularies define the concepts and relationships (also referred to as “terms”) used to describe and represent an area of interest. Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms. In practice, vocabularies can be very complex (with several thousands of terms) or very simple (describing one or two concepts only).

There is no clear division between what is referred to as “vocabularies” and “ontologies”. The trend is to use the word “ontology” for more complex, and possibly quite formal collections of terms, whereas “vocabulary” is used when such strict formalism is not necessarily used or only in a very loose sense. Vocabularies are the basic building blocks for inference techniques on the Semantic Web.

Definitions from Linked Data Glossary[6]:

> 119. Vocabulary
>
> A collection of "terms" for a particular purpose. Vocabularies can range from simple such as the widely used RDF Schema, FOAF and Dublin Core Metadata Element Set to

5 http://www.w3.org/TR/2014/REC-json-ld-20140116/

6 http://www.w3.org/TR/2013/NOTE-ld-glossary-20130627/

complex vocabularies with thousands of terms, such as those used in healthcare to describe symptoms, diseases and treatments. Vocabularies play a very important role in Linked Data, specifically to help with data integration. The use of this term overlaps with Ontology.

120. Vocabulary Alignment

The process of analyzing multiple vocabularies to determine terms that are common across them and to record those relationships.

There exist several sources of vocabulary repositories in the Web, for example:

Linked open Vocabularies (LOV)

URL: http://lov.okfn.org/dataset/lov/

Provider: Linked open Vocabularies

Resource: http://lov.okfn.org/endpoint/lov

License: Creative Commons CC BY 3.0

Linked open Vocabularies (LOV) contains a repository of RDFS and OWL ontologies of different themes. There are listed vocabularies and individually described by metadata, classified by vocabulary spaces, interlinked using the dedicated vocabulary VOAF. Linked Open Vocabularies endpoint is a central information point about vocabularies.

Schemas

URL: http://schema.org/

Provider: schema.org

Resource: http://schema.rdfs.org/

License: Terms of service

Schemas.org provides a collection of schemas, i.e., html tags, that web-masters can use to markup their pages in ways recognized by major search providers. Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right web pages.

There are several techniques to determine common terms of two or several vocabularies [7]:

- Alignment techniques for generating exact-matches:
  - Baseline technique
    - Plain string matching
    - Ignores ambiguous matches
  - Lexical technique

7 Combining Vocabulary Alignment Techniques, Anna Tordai, Jacco van Ossenbruggen, 2009

    - Matches terms and uses lemmatization and compound splitting
    - Returns all (possibly ambiguous) matches
  - Structural technique
    - Uses the structure of vocabularies
    - Uses lexical measures, lemmatization and distance metrics
- Manual evaluation

1.5 RDF Query languages

The predominant query language for RDF graphs is SPARQL Protocol and RDF Query Language (SPARQL). It is a query language and a protocol for accessing RDF designed by the W3C RDF Data Access Working Group. There also exist other query languages - RDQL, SeRQL and XsRQL. [8]

SPARQL is an RDF query language, that is, a query language for databases, able to retrieve and manipulate data stored in Resource Description Framework format. As a query language, SPARQL is "data-oriented" in that it only queries the information held in the models, there is no inference in the query language itself. SPARQL allows for a query to consist of triple patterns, conjunctions, disjunctions, and optional patterns. [9]

GeoSPARQL is a geographic query language for RDF data. Is an emerging standard within the Open Geospatial Consortium (OGC). Its intent is to provide a standard way to express and query spatial elements in RDF, so that users can exchange data easily, and triple store implementers can have a standard format for indexing. GeoSPARQL defines a vocabulary for representing geospatial data in RDF, and it defines an extension to the SPARQL query language for processing geospatial data. In addition, GeoSPARQL is designed to accommodate systems based on qualitative spatial reasoning and systems based on quantitative spatial computations.

There are three key classes in the GeoSPARQL ontology:

- geo:Feature – A thing that can have a spatial location; i.e., a park or a monument etc.;
- geo:Geometry – A representation of a spatial location; i.e., a set of coordinates;
- geo:SpatialObject – A superclass of both Features and Geometries.

2 Metadata LOD standards

Metadata standards are developed in order to support the usage of uniform descriptions of data that are managed within specific communities. Many different metadata standards were developed and used according to each community's needs (medicine, libraries, agriculture, statistics, ...). This report is focused only on :

8 A Comparison of RDF Query Languages, Kevin Hutt, 21st Computer Science Seminar

9 http://jena.apache.org/tutorials/sparql.html

- Common standards used in the Semantic web
- European common metadata initiatives from the LOD perspective
- Geospatial metadata

2.1 Metadata usage in semantic web

Metadata plays an important role in the Semantic Web because in a distributed environment the resource description is crucial for good interpretation. The metadata became part of some Semantic web data models.

2.1.1 Dublin Core metadata (ISO 15836)

The Dublin Core metadata Initiative[10] gives the common basis for many further metadata activities. The Dublin Core Metadata Element Set has a basic vocabulary of fifteen properties for use in resource description[11]. These elements are:

1. Contributor
2. Coverage
3. Creator
4. Date
5. Description
6. Format
7. Identifier
8. Language
9. Publisher
10. Relation
11. Rights
12. Source
13. Subject
14. Title
15. Type

The listed resource types are: Collection, Dataset, Event, Image, Interactive Resource, Moving Image, Physical Object, Service, Software, Sound, Still Image, Text.

The DCMI Metadata Terms[12] extends the basic element set with many other elements in /terms/ namespace, e.g. - Abstract, Access Rights, Alternative Title, Bibliographic Citation, Contributor, Has Part, Is Part Of, Is Required By, Replaces, Spatial Coverage, Temporal

10 http://dublincore.org/

11 http://dublincore.org/documents/dces/

12 http://dublincore.org/documents/dcmi-terms/

Coverage etc.

Many of these elements are designed to maintain links between resources, like 'Is Part Of', 'Has Part', etc. So users are encouraged not to use literal values for them but links.

The RDF representation of the Dublin Core metadata standard is available.

Starting in 2000, the Dublin Core community focused on "application profiles". The idea was that metadata records would use Dublin Core together with other specialized vocabularies to meet particular implementation requirements.

There are 4 levels of metadata interoperability characterized:

1. Shared term definitions: Based on natural-language definitions. The metadata systems are closed (librarian systems)
2. Formal semantic interoperability: Shared formal model by RDF in Linked data model.
3. Description Set syntactic interoperability: Applications are compatible with the Linked Data model and, in addition, share an abstract syntax for validatable metadata records, the "description set".
4. Description Set Profile interoperability: the records exchanged among metadata-using applications follow, in addition, a common set of constraints, use the same vocabularies, and reflect a shared model of the world.

The levels 3 and 4 are rather experimental[13].

2.1.2 OWL metadata

OWL headers contains a set of DC elements to provide ontology metadata , example - geonames ontology header (part)

```
<owl:Ontology rdf:about="http://www.geonames.org/ontology">
  <dcterms:title xml:lang="en">The Geonames ontology</dcterms:title>
  <dcterms:description xml:lang="en">The Geonames ontologies provides elements of description for geographical
features, in particular those defined in the geonames.org data base</dcterms:description>
  <adms:relatedDocumentation rdf:resource="http://www.geonames.org/ontology/documentation.html"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/3.0/"/>
  <dcterms:creator>
    <foaf:Person rdf:about="http://data.semanticweb.org/person/bernard-vatant">
      <foaf:name>Bernard Vatant</foaf:name><foaf:page rdf:resource="http://blog.hubjects.com"/>
      <foaf:workplaceHomepage rdf:resource="http://www.mondeca.com"/>
    </foaf:Person>
  </dcterms:creator>
  <dcterms:contributor>...
```

13 http://dublincore.org/

2.1.3 Dataset description used in LOD

Semantic sitemaps and VoiD - provide descriptions of published dataset on the server.

2.1.3.1 Semantic sitemaps

Semantic sitemaps provide descriptions of datasets to be accessible for web crawlers etc[14]. Typically it is represented by XML file sitemp.xml in the root directory of web site.

Example:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
        xmlns:sc="http://sw.deri.org/2007/07/sitemapextension/scschema.xsd">
  <sc:dataset>
    <sc:datasetLabel>Example Corp. Product Catalog</sc:datasetLabel>
    <sc:datasetURI>http://example.com/catalog.rdf#catalog</sc:datasetURI>
    <sc:linkedDataPrefix slicing="subject-object">http://example.com/products/</sc:linkedDataPrefix>
    <sc:sampleURI>http://example.com/products/widgets/X42</sc:sampleURI>
    <sc:sampleURI>http://example.com/products/categories/all</sc:sampleURI>
    <sc:sparqlEndpointLocation slicing="subject-object">http://example.com/sparql</sc:sparqlEndpointLocation>
    <sc:dataDumpLocation>http://example.com/data/catalogdump.rdf.gz</sc:dataDumpLocation>
    <sc:dataDumpLocation>http://example.org/data/catalog_archive.rdf.gz</sc:dataDumpLocation>
    <sc:dataDumpLocation>http://example.org/data/product_categories.rdf.gz</sc:dataDumpLocation>
    <changefreq>weekly</changefreq>
  </sc:dataset>
</urlset>
```

2.1.3.2 Vocabulary of Interlinked Datasets - VoiD

VoiD (Vocabulary of Interlinked Datasets) is the de facto standard vocabulary for describing Linked Data sets. It is very similar to Semantic Sitemaps but the information is represented in RDF (turtle notation). It enables descriptions of the dataset as whole as some parts of it (subsets)[15].

Example (dbpedia):

```
:DBpedia a void:Dataset;
    dcterms:title "DBPedia";
    dcterms:description "RDF data extracted from Wikipedia";
    dcterms:contributor :FU_Berlin;
    dcterms:contributor :University_Leipzig;
    dcterms:contributor :OpenLink_Software;
    dcterms:contributor :DBpedia_community;
    dcterms:source <http://dbpedia.org/resource/Wikipedia>;
    dcterms:modified "2008-11-17"^^xsd:date;
    .
:FU_Berlin a foaf:Organization;
    rdfs:label "Freie Universität Berlin";
    foaf:homepage <http://www.fu-berlin.de/>;
```

2.1.3.3 Provenance metadata

The published RDF should contain at least Dublin Core dc:creator, dc:publisher and dc:date to keep track of the data originator. Using URIs for their identification rather than

14 http://sw.deri.org/2007/07/sitemapextension/

15 http://www.w3.org/2001/sw/interest/void/

literal names is recommended[16].

A more complex approach is the Open Provenance model. The vocabulary uses the terms as Agents, Artifacts and Processes[17].

2.1.3.4 Lincenses, Waiverers, data norms

For clarity for consumers the licence or waiver statements should be always included in Linked Data on the web. Creative commons licenses should be used for copyrighted materials if possible[18].

Example CC BY-SA RDF file: http://creativecommons.org/licenses/by-sa/3.0/rdf

Example - putting licence into RDF:

```
@prefix cc: <http://creativecommons.org/ns#> .
...
cc:license <http://creativecommons.org/licenses/by-sa/3.0/> .
...
```

For non-copyrightable material the Waiver vocabulary is intended to be used[19].

2.1.4 DCAT application profile for data portals in Europe

The DCAT Application profile for data portals in Europe (DCAT-AP)[20] is a very new specification based on the Data Catalogue vocabulary (DCAT) [21]for describing public sector datasets in Europe. Its basic use case is to enable cross-data portal searching for data sets and making public sector data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of data sets among data portals.

Mandatory classes for the profile are:

- Agent (e.g. Organisation)
- Category (Subject of Dataset)
- Category scheme (Controlled vocabulary the theme comes from)
- Catalogue (Repository that hosts the dataset)
- Literal (Literal value)
- Resource (Anything described by RDF])

16 Tom Heath and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

17 Open Provenance model: http://open-biomed.sourceforge.net/opmv/ns.html

18 Creative Commons. http://creativecommons.org

19 WAIVER: A vocabulary for waivers of rights. http://vocab.org/waiver/terms/.html

20 DCAT-AP: https://joinup.ec.europa.eu/asset/dcat_application_profile/description

21 DCAT: http://www.w3.org/TR/vocab-dcat/

E.g. for all datasets mandatory properties are title and description, recommended contact point, dataset distribution, keyword/tag, publisher theme/category

2.1.5 CKAN domain model

CKAN is an open source system for publishing data on the web. It provides tools to streamline publishing, sharing, finding and using data.

The published metadata schema are proprietary and provide some mapping to DC elements. Also adding proprietary user defined elements is allowed. Harvesting of OGC CSW 2.0.2 + ISO 19139 spatial metadata standard is available with extensions.

CKAN metadata profile:

- id: unique id
- name (slug): unique name that is used in urls and for identification
- title (dc:title): short title for dataset
- url (home page): home page for this dataset
- author (dc:creator): original creator of the dataset
- author_email:
- maintainer: current maintainer or publisher of the dataset
- maintainer_email:
- license (dc:rights): license under which the dataset is made available
- version: dataset version
- notes (description) (dc:description): description and other information about the dataset
- tags: arbitrary textual tags for the dataset
- state: state of dataset in CKAN system (active, deleted, pending)
- resources: list of resources
- groups: list of groups this dataset is a member of
- “extras” - arbitrary, unlimited additional key/value fields

2.2 Geospatial Metadata standards

Spatial data is often very complex with complicated life cycle. In the geospatial community very complex meta data is required and maintained for spatial dataset and services description.

2.2.1 ISO TC 211 metadata standards

These standard are part of the 19xxx family geospatial standards developed by TC211.

- ISO 19115 was issued in 2003 and from that time was adopted by many SDI (USA, Australia, European Union). It replaced other standards used before, such as ANZLIC, FGDC and CEN. A revision was made in 2006. The standard contains UML models of metadata objects and tabular description of these elements, code-lists and methodology for creating profiles (subsets of existing elements or adding new elements to the profiles). The structure is a very complex hierarchy organizationed. [22]

- ISO 19119 complements the metadata standard for service specific metadata elements. The main part of the metadata element setr is shared with ISO 19115. [23]
- ISO 19139 defines XML representation of these metadata. The XSD are available here: http://schemas.opengis.net/iso/19139/20070417/gmd/metadataEntity.xsd [24]

These standards are not compatible with DCMI or other semantic web activities. There is no exact mapping between ISO and DC standards, but there is the possibility to map ISO more structured elements to DC elements but not the opposite direction. Also only literal values may be mapped to DC because there is (almost) no URI form in ISO.

Examples:

http://one.geology.cz/catalogue/libs/cswclient/cswClientRun.php?serviceName=one&format=application/xml&id=4a9fa6f0-5e04-425b-921e-0e3c0a01080d

2.2.2 Open Geospatial Consortium (OGC) standards

OGC developed many geospatial standards (GML,KML, WMS, WFS, CSW, SWE, etc.)[25]. Some of these were also adopted by ISO TC211.

The Catalogue service for web (CSW) is intended for metadata cataloguing. The service defines a wrapper for metadata exchange, list of queryables, transaction support etc.

Several profiles are supported[26], the default Dublin Core (ISO 15836) metadata element set is mandatory. Catalogues may provide output mapping from ISO TC211 to DC to support this profile, but there are no exact rules for 1:1 mapping.

2.2.3 INSPIRE metadata profile

The INSPIRE metadata profile is defined by the INSPIRE Metadata regulation[27] [28]. It

22 ISO 19115: 2003, Geographic information – Metadata

23 ISO 19119:2005, Geographic information – Services

24 ISO/TS 19139:2007 defines Geographic MetaData XML (gmd) encoding, an XML Schema implementation derived from ISO 19115.

25 http://www.opengeospatial.org/standards/is

26 OGC 07-006, OGC CSW, OGC™ Catalogue Services Specification, version 2.0.2 (Corrigendum Release 2). http://www.opengeospatial.org/standards/cat

27 Corrigendum to Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata (OJ L 326, 4.12.2008)

28 Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata (Text with EEA relevance)

defines a basic set of metadata elements for spatial datasets and services.

The 3 levels of complexity are distinguished according the usage:

- Discovery metadata (The smallest set - only for searching the resources of interest)
- Evaluation metadata (Enables assessment of usability for a particular case)
- Use metadata (Most complex, documentation for use)

The INSPIRE metadata profile concerns mainly the discovery metadata. INSPIRE Data specifications define additional metadata elements with regard to each theme's requirements, especially quality parameters, feature catalogue citation, coordinate reference systems and encoding (data formats and reference to data theme structure description). Feature-level metadata may also be used if necessary.

Technical implementation[29] is based on ISO 19115/19119/19139 Metadata profile. The metadata sharing will be provided by INSPIRE discovery services[30] compatible with OGC CSW 2.0.2 AP-ISO 1.0 profile [31]

So there is no Semantic web approach at this moment, but the RDF (DCAT-AP) representation of INSPIRE metadata is in preparation. Also it is planned, that a XSL template will be available for automatic translation of INPSIRE metadata XML to RDF.

2.2.4 INSPIRE metadata and DCAT-AP

A proposal for the alignment of INSPIRE metadata with the DCAT Application Profile is under preparation by the Joint Research Centre of the European Commission, in the framework of Action 1.17 of the EU ISA Programme. The status of the document is unstable and the working draft may dramatically change before final release. The profile uses INSPIRE registry (http://inspire.ec.europa.eu/registry/) together with DCAT defined classes for the mapping. This profile is very important because it mediate the bridge between the INSPIRE and other European portals. It is intended to implement this profile and test its usability and stability in the scope of SmartOpenData project.

Example (part of the file):

```
<rdf:Description rdf:about="http://someURI">
  <foaf:primaryTopicOf rdf:resource="metadataURI"/>
  <!-- Resource language -->
  <dct:language rdf:resource="http://publications.europa.eu/resource/authority/language/DEU"/>
  <!-- Resource title -->
  <dct:title xml:lang="en">Forest / Non-Forest Map 2006</dct:title>
  <!-- Resource abstract -->
  <dct:description xml:lang="en">Pan-European Forest / Non Forest Map with target year 2006, Data
Source: Landsat ETM+ and Corine Land Cover 2006, Classes: forest, non-forest, clouds/snow, no data;
Method: automatic classification performed with an in-house algorithm; spatial resolution: 25m. In addition,
```

29 Guidelines based on EN ISO 19115 and EN ISO 19119 for Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata

30 Commission Regulation (EC) No 976/2009 of 19 October 2009 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards the Network Services

31 Technical Guidance for the implementation of INSPIRE Discovery Services, 07.11.2011.


```
the forest map 2006 is extended to FTYPE2006 to include forest types (broadleaf, coniferous forest) that
are mapped using MODIS composites.</dct:description>
  <!-- Resource type -->
  <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Dataset"/>
  <dct:type rdf:resource="http://inspire.ec.europa.eu/codelist/resource-type/series"/>
  <!-- Resource locator -->
  <dcat:landingPage rdf:resource="http://someurl.org"/>
  <!-- Unique resource identifier -->
  <dct:identifier
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">namespace/code</dct:identifier>
  <!-- Topic category -->
  <dct:subject rdf:resource="http://inspire.ec.europa.eu/codelist/topic-category/geoscientificInformation"/>
  <!-- Keywords corresponding to an INSPIRE spatial data theme -->
  <dcat:theme rdf:resource="http://inspire.ec.europa.eu/theme/landCover"/>
  <!-- Keyword from another controlled vocabulary -->
  <dcat:theme>
    <skos:Concept>
      <rdfs:label xml:lang="en">coniferous forest</rdfs:label>
      <skos:inScheme>
        <skos:ConceptScheme>
          <rdfs:label xml:lang="en">GEMET - Concepts, version 2.4</rdfs:label>
          <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema-datatypes#date">2010-01-
13</dct:issued>
        </skos:ConceptScheme>
      </skos:inScheme>
    </skos:Concept>
  </dcat:theme>
  <!-- Geographic bounding box -->
  <dct:spatial>
    <dct:Location>
      <locn:geometry rdf:datatype="http://www.opengis.net/rdf#GMLLiteral">
        <gml:Envelope srsName="http://www.opengis.net/def/crs/OGC/1.3/CRS84">
          <gml:lowerCorner>-10.58 34.56</gml:lowerCorner>
          <gml:upperCorner>34.59 70.09</gml:upperCorner>
        </gml:Envelope>
      </locn:geometry>
    </dct:Location>
  </dct:spatial>
```


3 Geographic Schemas

3.1 Geospatial Resource Description Framework

Geospatial Resource Description Framework (GRDF) represents an experiment connecting Geography Markup Language (GML) and ontological structures. GRDF is coded in OWL-DL and refer to RDF. The model is based on two main elements - Feature (based on ISO 10919) and Geometry. There could also another parts such as Style, Topology, CRS or TimeObject.[32]

```
<rdf: Description about="#VECTOR.VECTOR.
HYDRO_STREAMS_CENSUS_line">
<app:hasObjectID>11070
<grdf:LineString srsName="http://.../TX83-NCF">
<grdf:coordinates>2533822.17263276,7108248.
82783879 ...
</gml:coordinates>
</grdf:LineString>
</rdf:Description >
```

3.2 Geographically Encoded Objects for RSS feeds (GeoRSS)

URL	http://www.georss.org/
License	Creative Commons Attribution-ShareAlike 2.5 License[33]

GeoRSS (Geographically Encoded Objects for RSS feeds) is a method of describing and pinpointing the physical location of Web content. It provides a way for including geographic reference information in RSS 2.0, RSS 1.0 or Atom 1.0 feeds via specific encoding with enough simplicity and descriptive power. This encoding is brief and simple, but extensible and compatible with more sophisticated encoding standards such as GML. GeoRSS feeds are also designed to be used by geographic software such as map generators.

Nowadays, there are two types of GeoRSS encoding, Simple and GML.

3.2.1 GeoRSS Simple

GeoRSS-Simple is a very lightweight format which can be used very quick and easily add to existing feeds. Out of geometries, GeoRSS supports the basic shapes: point, line, box or polygon. It also supports feature type and relationship tags as well as the ability to encode elevation and radius. The default coordinate reference system for GeoRSS is WGS-84 latitude/longitude. [34]

An example of GeoRSS-Simple Point representation with elevation:

```
<georss:point>45.256 -71.92</georss:point>
<georss:elev>313</georss:elev>
```

Another example of GeoRSS-Simple additional properties:

```
<georss:point>45.256 -110.45</georss:point>
<georss:featuretypetag>city</georss:featuretypetag>
<georss:relationshiptag>is-centered-at</georss:relationshiptag>
<georss:featurename>Podunk</georss:featurename>
```

32 Alam, A., Khan, L., Thuraisingham, B. 2008. Geospatial Resource Description Framework (GRDF) and Security Constructs. ICDE Workshop 2008.

33 http://creativecommons.org/licenses/by-sa/2.5/

34 http://searchsoa.techtarget.com/definition/GeoRSS

3.2.2 GeoRSS GML

GeoRSS GML can be thought of as a simple GML Application Profile. Thus, GeoRSS GML supports a broader range of features (such as geometry, topology, units of measure) than GeoRSS-Simple. It also supports different coordinate reference systems apart from WGS-84 latitude/longitude. [35]

GeoRSS GML Point representation example:

```
<georss:where>
<gml:Point>
<gml:pos>45.256 -71.92</gml:pos>
</gml:Point>
</georss:where>
```

3.3 GeoSPARQL

URL	http://www.opengeospatial.org/standards/geosparql
Provider	Open Geospatial Consortium
License	The license is listed in the OGC GeoSPARQL document[36]

GeoSPARQL is a standard for representing and querying of geospatial linked data for the Semantic Web from the Open Geospatial Consortium. It attempts to solve the problems with the heterogeneous and incompatible implementations for representing and querying spatial data. It achieves this by defining an ontology that follows the existing standards from the OGC with regard to spatial indexing in relational databases. The GeoSPARQL specification consists of three components – the definition of a vocabulary to represent features, geometries and the relationships between them; a set of spatial functions for use in SPARQL queries; and a set of query transformation rules.

The ontology for features representing geometries is fundamental to being able to build and query spatial data. It includes a class ogc:SpatialObject with two subclasses – ogc:Feature and ogc:Geometry. A geometry is the spatial extension of feature and can have an RDF literal that stores the geometry data.

GeoSPARQL contains two different ways to represent geometry literals: Well Known Text (WKT) and Geography Markup Language (GML). It also provides different OWL classes for the geometry hierarchies associated with both of these geometry representations. This provides classes for many geometry types such as point, polygon, curve, arc and multicurve. The ogc:asWKT and ogc:asGML properties link the geometry entities to the geometry literal representations.

35 OPEN GEOSPATIAL CONSORTIUM INC. OGC White Paper - An Introduction to GeoRSS: A Standards Based Approach for Geo-enabling RSS feeds. [online]. 2006. http://www.opengeospatial.org/pt/06-050r3

36 OPEN GEOSPATIAL CONSORTIUM INC. OGC GeoSPARQL – A Geographic Query Language for RDF Data. [online]. 2012. http://www.opengeospatial.org/standards/geosparql

GeoSPARQL also includes a way to ask for relationships between spatial entities. These come in the form of topological binary properties between the entities and geospatial filter functions.[37]

3.4 NeoGeo Vocabulary

URL	http://geovocab.org/
Resource	http://geovocab.org/

The NeoGeo Vocabulary project make efforts to develop a common vocabulary for the representation of geodata. The NeoGeo Vocabulary is based on the GML Simple Features Profile[38]. Simple geometries are described explicitly in RDF. By the aggregation of simple geometries arise more complex geometries. It allows reasoning and querying on these geometries.

3.4.1 Vocabulary

Everything is represented as an RDF resource in the NeoGeo vocabulary. For instance, each node of polygon is represented by the RDF. These nodes comprise together an RDF Collection. So it's possible to identify points which are shared between shapes. This approach also allows referencing meaningful points by giving them a specific URI. [39]

Points are represented as decimal degrees in the WGS 84 CRS and they are described by reusing the W3C Geo predicates.

```
ex:point_1 rdf:type ngeo:Point .
ex:point_1 geo:lat "52.516262" .
ex:point_1 geo:long "13.377717" .
```

A LineString resource is a collection of points, represented by an RDF Collection. Polygons, as shown the following example, are described by using PolygonResources. Polygons are closed areas described by its bounding LinearRings elements, a particular case of LineString resources. A polygon has always exactly one exterior boundary and may have a number of interior boundaries.

```
_:polygon rdf:type ngeo:Polygon ;
ngeo:exterior [
rdf:type ngeo:LinearRing ;
ngeo:posList (
[ geo:lat "-29"; geo:long "16" ]
[ geo:lat "-28"; geo:long "33" ]
[ geo:lat "-34"; geo:long "27" ]
[ geo:lat "-35"; geo:long "19" ]
[ geo:lat "-29"; geo:long "16" ]
)
```

37 http://www.ssec.wisc.edu/meetings/geosp_sem/presentations/GeoSPARQL_Getting_Started%20-%20KolasWorkshop%20Version.pdf

38 http://geovocab.org/

39 http://geovocab.org/doc/neogeo.html


```
] ;
ngeo:interior [
rdf:type ngeo:LinearRing ;
ngeo:posList (
[ geo:lat "-29.5"; geo:long "27" ]
[ geo:lat "-28.5"; geo:long "28.5" ]
[ geo:lat "-29.5"; geo:long "29.5" ]
[ geo:lat "-31"; geo:long "28" ]
[ geo:lat "-29.5"; geo:long "27" ]
)
] .
```

As was mentioned, the NeoGeo Vocabulary supports the definition of composite geometries (such as MultiPoint, MultiLine String, MultiPolygon), which represent a set of simple geometries.

3.5 W3C Basic Geo

URL	http://www.w3.org/2003/01/geo/
Provider	World Wide Web Consortium
Resource	http://www.w3.org/Consortium/
Licence	http://www.w3.org/Consortium/Legal/2002/copyright-documents-20021231

The W3C Basic Geo Vocabulary is a RDF vocabulary for representing latitude/longitude position of the data geographical location, using WGS84 as a reference datum. It is used within RDF documents, as well as a namespace within non-RDF XML documents, for instance RSS 2.0 or Atom. There are many publishers, applications and services using the format, such as DBpedia.

As is mentioned in the official document about W3C Basic Geo, it does not attempt to address many of the issues covered in the professional GIS world, but it provides just a few basic terms that can be used in RDF when there is a need to describe position. The reason for using RDF is its capability for cross-domain data mixing.[40]

40 http://www.w3.org/2003/01/geo/

Examples:

The vocabulary defines a class 'Point', whose members are points. Points can be described using the 'lat' and 'long' properties:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
<geo:Point>
<geo:lat>55.701</geo:lat>
<geo:long>12.552</geo:long>
</geo:Point>
</rdf:RDF>
```

This is an example of geo-coding with RSS 1.0:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" xmlns="http://purl.org/rss/1.0/">
...
<item rdf:about="http://example.com/geo">
 <title>An example annotation</title>
 <link>http://example.com/geo</link>
 <description>Just an example</description>
 <geo:lat>26.58</geo:lat>
 <geo:long>-97.83</geo:long>
</item>
...
</rdf:RDF>
```

3.6 Location Core Vocabulary

URL	https://joinup.ec.europa.eu/asset/core_location/description
Provider	EU ISA Programme Core Vocabularies Working Group
License	European Commision / ISA Open Metadata Licence

The ISA Programme Location Core Vocabulary[41] is a simplified, reusable and extensible data model that offers a minimum set of classes and properties for describing any place by its name, address or geometry. The vocabulary is primarily designed to aid the publication of data that is interoperable with the EU INSPIRE Directive. The vocabulary is integrated with the Business and Person Core Vocabularies of the EU ISA Programme.[42]

The data is published along with a human readable mapping from the terms used by the publisher to terms defined in an official document (conceptual conformance), or is published using the RDF classes and properties (RDF conformance) as well as using the XML elements and types (XML conformance) specified in an official document and its associated RDF/XML namespace document.[43]

The geometry property may have as its range either a literal (WKT, GML or KML) or a geometry class (e.g, ogc:Geometry, geo:Point, schema:GeoCoordinates).

41 https://joinup.ec.europa.eu/sites/default/files/egovernment-core-vocabularies.pdf

42 http://www.w3.org/ns/locn

43 https://joinup.ec.europa.eu/asset/core_location/description

The Geometry Class (locn:geometry) signifies the notion of geometry at a conceptual level and can be encoded in different formats (such as WKT, GML, KML, RDF etc.). Conceptually, the class includes coordinates, an identifier for the coordinate reference system and the geometry type (point, line or polygon). There is no need for defining any of these properties in the RDF or XML schemas as, in practice, geometry encodings all include these properties and the locn:geometry property is sufficient as can be seen in the following example (written in RDF turtle).

GML:

```
<ex:a> a locn:Location ;
locn:geometry
"<gml:Point srsName="http://www.opengis.net/def/crs/OGC/1.3/CRS84">
<gml:coordinates>-0.001475, 51.477811
</gml:coordinates></gml:Point>"^^ogc:GMLLiteral .
RDF (WGS84 lat/long):
<ex:a> a locn:Location ;
locn:geometry [ a geo:Point;
geo:lat "51.477811"; geo:long "-0.001475" ] .
```

3.7 Address Schema

URL	http://schema.org/PostalAddress
Provider	schema.org
Resource	http://schema.rdfs.org/
Licence	Terms of service

The schemas (data model) published in web page schema.org includes also model for description of addresses. This schema (as well as other schemas from schema.org) could be downloaded as RDF file (with various syntaxes), CSV or JSON. The schema contains properties (including inherited properties), data types and descriptions. From the LOD viewpoint the interconnection to another standards is very important. Schemas from schema.org (including addresses) are related to AddressVocab, dbPedia, Dublin Core or WordNet.

4 Geo-domain vocabularies

This section contains information on various kinds of vocabularies which are involved in geographical and spatial data domain. The chapter mentions just selected examples from the huge set of these semantic tools. Further useful vocabularies can be found in documents such as Linked Open Vocabularies or Vocabulary and Dataset (Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets). For example there is an Accommodation Ontology, that can be used for tourism purposes, BIO: A vocabulary for biographical information, Vocabulario de Resultados Electorales (ontology of election results) or Ontology for Meteorological sensors.

4.1 AgroVOC

URL	http://aims.fao.org/standards/agrovoc/about
Provider	Food and Agriculture Organization
Resource	http://aims.fao.org/standards/agrovoc/functionalities/download
License	http://aims.fao.org/standards/agrovoc/functionalities/download

According to its official web page AGROVOC is a controlled vocabulary covering all areas of interest to the Food and Agriculture Organization (FAO) of the United Nations, including food, nutrition, agriculture, fisheries, forestry, environment etc. It is published by FAO and edited by a community of experts. To date, AGROVOC contains over 32,000 concepts organized in a hierarchy, each concept may have labels in up to 22 languages: Arabic, Chinese, Czech, English, French, German, Hindi, Hungarian, Italian, Japanese, Korean, Lao, Persian, Polish, Portuguese, Russian, Slovak, Spanish, Thai, Turkish. Four more language versions are under development (Malaysian, Moldavian, Telugu, Ukrainian). AGROVOC is formalized as a RDF/SKOS-XL linked dataset, accessible through a SPARQL endpoint, and available for download in various formats. You may also include AGROVOC in your specific application through Web services, and search or browse AGROVOC through its official website.

AGROVOC data contains several geographic objects (e.g. countries or rivers). Is possible to browse data online[44], search a term in the web client[45], download data or use web services[46]. The Linked Data version (RDF/SKOS-XL) is stored in the Allegrograph triple store. Data is accessible to machines through a SPARQL endpoint, and to humans by means of a HTML pages generated with Pubby. LOD data is interconnected to dbPedia, EuroVoc, GEMET, FAO Geopolitical Ontology and other resources.

4.2 GEneral Multilingual Environmental Thesaurus (GEMET)

URL	http://www.eionet.europa.eu/gemet/about
Provider	European Environment Agency
Resource	http://www.eionet.europa.eu/gemet/rdf?langcode=en
License	Creative Commons Attribution License

Detailed information on GEMET is provided in the web page About GEMET - GEneral Multilingual Environmental Thesaurus[47]. From the viewpoint of spatial data the theme Geography[48] is the most important. It contains definitions, translations, links to Wikipedia, AgroVoc or EuroVoc and typical thesaurus information (near terms, broad terms etc.) for geographic concepts (e.g. border or volcano) and also concrete objects (e.g. Caspian Sea

44 http://aims.fao.org/standards/agrovoc/functionalities/hierarchy

45 http://aims.fao.org/standards/agrovoc/functionalities/search

46 http://aims.fao.org/standards/agrovoc/webservices

47 http://www.eionet.europa.eu/gemet/about

48 http://www.eionet.europa.eu/gemet/theme_concepts?th=16&langcode=en

or South America). GEMET also includes INSPIRE Spatial Data Themes[49]. GEMET can be used online (searching and browsing in web pages), as downloaded data and by web services.

4.3 Agroontology

URL	http://aims.fao.org/aos/agrontology
Provider	Food and Agriculture Organization
Resource	http://aims.fao.org/aos/agrontology

Agrontology is the OWL vocabulary providing a set of domain properties to Agrovoc. The ontology contains above all object and other relations (including a few spatial relations such as isSpatiallyIncludedIn) that can be not described in traditional is-a relation in AGROVOC vocabulary.

4.4 GeoSpecies Ontology

URL	http://lod.geospecies.org/
Provider	Peter J. DeVries
Resource	http://rdf.geospecies.org/ont/geospecies
License	http://creativecommons.org/licenses/by-sa/3.0/us/

This ontology was designed to help integrate species concepts with species occurrences, gene sequences, images, references and geographical information (see also Taxonconcept.org). The ontology is stored in OWL format. The ontology contains 107 classes and 705 individuals. The most detailed information on this ontology is given in LOV[50] and datahub[51].

4.5 Landinndelingen i Norge Norway (Administrative division of Norway)

URL	http://vocab.lenka.no/geo-deling#
Provider	Kjetil Kjernsmo
Resource	http://vocab.lenka.no/geo-deling.rdf
License	Creative Commons Attribution

This is a vocabulary of administrative divisions in Norway, the land register, places and other geographical things.

49 http://www.eionet.europa.eu/gemet/inspire_themes?langcode=en

50 http://lov.okfn.org/dataset/lov/details/vocabulary_geosp.html

51 http://datahub.io/cs_CZ/dataset/geospecies

4.6 EuroVOC

URL	http://eurovoc.europa.eu/
Provider	European Union
Resource	http://eurovoc.europa.eu/drupal/?q=legalnoticeredirect&cl=en
License	Reuse of the EuroVoc Thesaurus for commercial or non-commercial use is authorised provided appropriate acknowledgement is given as follows: "© European Union, 2014, http://eurovoc.europa.eu/"

The multilingual thesaurus of the European Union is not only focused on geographical domain, but some sub-domains (e.g. Geography or Environment) contain terms related to geography and spatial data. The data can be downloaded in SKOS/XML format. EuroVoc provides information according to traditional vocabulary structure (narrow terms, related terms, broader terms, language equivalents, definitions - scope notes). The version 4.4 includes 6883 thesaurus concepts of which 85 concepts are new, 142 have been updated and 28 have been classified as obsolete concepts.

4.7 STW Thesaurus for Economics

URL	http://zbw.eu/stw/versions/latest/about.en.html
Provider	ZBW - Leibniz Information Centre for Economics
Resource	http://zbw.eu/stw/versions/latest/download/about.en.html
License	Creative Commons Attribution-Noncommercial-Share Alike 3.0 Germany License

The STW Thesaurus for Economics contains the subthesauri G Geographic names. But also other subthesauri are connected to geographical domain. According to the authors the thesaurus provides a vocabulary on any economic subject: more than 6,000 standardized subject headings and about 19,000 entry terms to support individual keywords. You can also find technical terms used in law, sociology, or politics, and geographic names.

5 LOD data set containing geo-data

This chapter contains examples of spatial data resources or data resources containing spatial data coded in RDF formats or ontological systems. Each resource is described by basic identification information, standards used (e.g. data formats), contained information and data, data harvesting and transformation (methodologies, workflows) and links to other data.

5.1 dbPedia

URL	http://dbpedia.org/
Provider	University of Leipzig ,Freie Universität Berlin , OpenLink Software
Resource	Latest versions of the datasets (3.9): http://wiki.dbpedia.org/Downloads39
License	From the version 3.4 is DBpedia data licensed under Creative Commons Attribution-ShareAlike 3.0 license and the GNU Free Documentation License.

The aim of a DBpedia project is to extract structured information from Wikipedia and making it available on the WWW. It enables users to query relationships and properties connected with Wikipedia sources, inclusive of links to external datasets (such as GeoNames, CIA World Fact Book, Eurostat or US Census). The first public DBpedia dataset was published in 2007. The current version is 3.9, which was released in September 2013. The dataset describes 4 million entities, out of which 3.22 million are classified in a consistent ontology. The project uses RDF to represent the information and consists of almost 2.5 billion RDF triples. But more than 45 million of them is the DBpedia dataset interlinked with various other Open Data datasets on the Web.

5.2 FAO Geopolitical Ontology

URL	http://www.fao.org/countryprofiles/geoinfo/en/
Provider	Food and Agriculture Organization of the United Nations
Resource	http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource/
License	FAO copyright

The geopolitical ontology ensures that FAO and associated partners can rely on a master reference for geopolitical information, as it manages names in multiple languages (English, French, Spanish, Arabic, Chinese, Russian and Italian); maps standard coding systems (UN, ISO, FAOSTAT, AGROVOC, etc); provides relationships among territories (land borders, group membership, etc); and tracks historical changes. The ontology is connected to dbPedia.

5.3 Ordnance Survey Products

URL	http://data.ordnancesurvey.co.uk/
Provider	Ordnance Survey , Great Britain's national mapping authority
Resource	http://data.ordnancesurvey.co.uk/datasets/os-linked-data/downloads
License	Ordnance Survey OpenData License

The Ordnance Survey is the British executive agency providing geographic data. As part of the opening up of this data Ordnance Survey has published three datasets as Linked Data: the 1:50 000 Scale Gazetteer (containing over 250 000 place names), Code-Point Open (containing approximately 1.7 million postcode units) and Boundary-Line (containing all levels of administrative boundaries). Each of the datasets is accessible via SPARQL endpoint or can be downloaded as raw data, which is packaged as zipped RDF N-Triples files. These datasets are described in a semantically meaningful way via ontologies.

5.4 U.S. National Map

URL	http://nationalmap.gov/
Provider	U. S. Geological Survey
Resource	Ontologies and Vocabularies, Sample data (9 test areas) is available via a SPARQL endpoint , currently out of order.

License	Information have not been found

The National Map is a project of the U. S. Geological Survey (USGS). The aim of this effort is to become a source for trusted, integrated and current topographic information available online for a broad range of uses. The National Map consists of eight thematic layers (such as Land Cover, Hydrography or Geographic Names) which are integrated from various datasets. One of the National Map's project goals is to build an ontology to specify geospatial feature semantics for richer data models. The data is in this case archived as RDF triplestores and can be accessed by SPARQL endpoint, downloaded or accessed by URI for mashups with other data. The data is linked across triplestores by URIs. The USGS dataset is connected with other datasets from the Open Linked Data cloud.

5.5 Geonames.org

URL	http://www.geonames.org/
Provider	GeoNames Team
Resource	GeoNames Data
License	Creative Commons Attribution 3.0 License

The GeoNames geographical database is available for download free of charge under a creative commons attribution license. It contains over 10 million geographical names and consists of over 8 million unique features which include 2.8 million populated places and 5.5 million alternate names. All features are categorized into one out of nine feature classes and further subcategorized into one out of 645 feature codes. The data is accessible free of charge through a number of web-services and a daily database export. GeoNames is already serving up to over 30 million web service requests per day. GeoNames is integrating geographical data such as names of places in various languages, elevation, population and others from various sources. All lat/long coordinates are in WGS84 (World Geodetic System 1984). Users may manually edit, correct and add new names using a user friendly wiki interface.[52]

5.6 LinkedGeoData.org

URL	http://linkedgeodata.org/
Provider	Agile Knowledge Engineering and Semantic Web
Resource	http://downloads.linkedgeodata.org/releases/
License	Open Database License

According to the WEB page LinkedGeoData.org LinkedGeoData is an effort to add a spatial dimension to the Web of Data / Semantic Web. LinkedGeoData uses the information collected by the OpenStreetMap project and makes it available as an RDF knowledge base according to the Linked Data principles. It interlinks this data with other knowledge bases in the Linking Open Data initiative.

52 http://www.geonames.org/about.html

The resource contains data divided into groups based on OpenStreetMap classification such as emergency, historic, tourism. The geometries (nodes and ways) are also separated.

5.7 Getty Thesaurus of Geographic Names®

URL	http://www.getty.edu/research/tools/vocabularies/tgn/index.html
Provider	The J. Paul Getty Trust
Resource	http://www.getty.edu/research/tools/vocabularies/lod/
License	ODC-By 1.0 license (Getty Vocabularies as Linked Open Data)

Getty Thesaurus of Geographic Names® (TGN) is together with other Getty Vocabularies provided in the form of LOD. All information on LOD version (including documentation) is in the web page Getty Vocabularies as Linked Open Data. The structure of information follow traditional geopolitical division of the world (continents - nation - region - place). The thesaurus contains definitions, descriptions and links to other resources such as Times Atlas of the World, Encyclopaedia Britannica or NIMA, GEOnet Names Server (see an example). The Getty Vocabularies as Linked Open Data are published as AAT Semantic Representation.

6 Top-level ontologies

The upper ontologies (or top-level or foundations ontologies) represent resources to define and re-use general terms such as people, events, organizations, time, units, general relations; sensors, observations etc. These terms are not crucial in the terms of spatial data and information, but they can be used to construct more expressive and broad linked data structures, because spatial data and information are not self-standing objects and have overlaps with other common structures.

Top-level ontologies describe very general concepts that are the same across all knowledge domains. An important function of an upper ontology is to support very broad semantic interoperability between a large number of ontologies which are accessible ranking "under" this upper ontology (Wikipedia)[53]. The portfolio of different top-level ontologies is very large, but the following list contains just several important examples (in the list ontologies such as Semantic Web Technology Evaluation Ontology[54], General Formal Ontology, Basic Formal Ontology[55], IDEAS Group Ontology[56], Upper Mapping and Binding Exchange Layer[57] or KR Ontology[58] are not mentioned, because they are not key

53 http://en.wikipedia.org/wiki/Upper_ontology

54 http://lsdis.cs.uga.edu/projects/semdis/sweto/

55 http://www.ifomis.org/bfo

56 http://www.ideasgroup.org/

57 http://umbel.org/

58 http://www.jfsowa.com/ontology/

items of this report). Documents A Comparison of Upper Ontologies[59] and Toward the use of an upper ontology for US government and US military domains: An evaluation[60] provide a useful description, evaluation and comparison of top-level ontologies.

6.1 WordNet

URL	http://wordnet.princeton.edu/
Provider	WordNet team
Resource	http://wordnet.princeton.edu/wordnet/download/
License	http://wordnet.princeton.edu/wordnet/license/

According to the WordNet web page the WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with a browser. WordNet is also freely and publicly available for download. WordNet's structure makes it a useful tool for computational linguistics and natural language processing. WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.

6.2 Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)

URL	http://www.loa.istc.cnr.it/old/DOLCE.html
Provider	Laboratory for Applied Ontology
Resource	http://www.loa.istc.cnr.it/old/ontologies/DLP3971.zip
License	Free available

The DOLCE is targeted to the sphere of common information science and its interconnection to philosophy and natural languages. Wikipedia (and also DOLCE web page) mentions that as implied by its acronym, DOLCE has a clear cognitive bias, in that it aims at capturing the ontological categories underlying natural language and human common sense. DOLCE, however, does not commit to a strictly referentialist metaphysics related to the intrinsic nature of the world. Rather, the categories it introduces are thought of as cognitive artefacts, which are ultimately depending on human perception, cultural imprints and social conventions. In this sense, they intend to be just descriptive (vs prescriptive) notions, that assist in making already formed conceptualizations explicit. DOLCE is an ontology of particulars, in the sense that its domain of discourse is restricted

59 http://www.disi.unige.it/person/MascardiV/Download/DISI-TR-06-21.pdf

60 www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA459575

to them. Of course, universals are used to organize and characterize the particulars, but they are not themselves subject to being organized and characterized (e.g., by means of metaproperties).

6.3 Cyc

URL	http://www.cyc.com/ (there is also open platform OpenCyc)
Provider	Cycorp, Inc.
Resource	https://sourceforge.net/projects/opencyc/files/ (OpenCyc)
License	Commercial (OpenCyc is licenced under as OpenCyc License)

Cyc is a commercial product the following information (proceeding from OpenCyc web page) describes OpenCyc platform as the open subset of the original Cyc. OpenCyc contains about 239.000 terms and 2.093.000 triples. From the viewpoint of spatial data the existence of about 19.000 information on places is very important. The OpenCyc is interconnected (through owl:sameAs property) to dbPedia (about 47.000 links), WordNet, CIA World Factbook or RDFAbout states and counties. OpenCyc also includes specification of CycL (language which Cyc and OpenCyc is written), Cyc API, java-based Cyc Inference Engine and the Cyc Knowledge Base Browser.

6.4 Suggested Upper Merged Ontology (SUMO)

URL	http://www.ontologyportal.org/
Provider	Institute of Electrical and Electronics Engineers; Technical Editor - Adam Pease
Resource	http://sigmakee.cvs.sourceforge.net/sigmakee/KBs/
License	GNU General Public License

SUMO is owned and managed by the IEEE (Institute of Electrical and Electronics Engineers). It is coded in SUO-KIF language, but there is a translation to OWL. SUMO is free and licensed by GNU GPL licence. SUMO contains about 25.000 terms and about 80.000 logical axioms. The set of terms and axioms covers not only SUMO itself, but also the MId-Level Ontology (MILO), and ontologies of communications, countries and regions, distributed computing and user interfaces, economy, finance, auto mobiles and engineering components, Food, Dining, Sports, Shopping catalogues and Hotels, geography, government and Justice, language taxonomy, media and Music, Military (general, devices, processes, people), North American Industrial Classification System, people and their Emotions, physical elements, transnational issues, transportation and its Details, viruses, world airports A-K, world airports L-Z, weapons of mass destruction. SUMO is interconnected to dbPedia, WordNet, YAGO and Open Biomedical Ontologies.

7 Overview of existing technologies

SmOD is not the only project working with geospatial data and the semantic web. There already exists EC and other projects.[61] In this chapter we will review for SmOD project relevant projects and their results, and their potential for reuse of work already done.

7.1 Initial Results from the Linking Geospatial Data Workshop

At the beginning of March, the Smart Open Data Project put on a workshop with W3C, OGC, Google Maps, the Ordnance Survey and the UK Government called Linking Geospatial Data. At the time of writing the full report on the workshop is not complete but it is possible to summarise the event.

72 papers were received, authors of most of which were invited to attend and participate in some way. During the two days, over 100 participants discussed common problems with using and linking geospatial data. Participants included representatives from the national mapping, environment, oceanographic, geological and meteorological agencies from many European countries as well as Australia, Japan, Ecuador, Canada and the USA. The diversity of participants was reflected in the diversity of use cases and projects presented. Some need very general data about locations including the concepts of a 'fuzzy URI' that may indicate a general area without formal boundaries through to cases where individual coordinates in a polygon need their own metadata to describe how that specific point was determined. Other EC-funded projects participating included MELODIES, DaPaaS and GeoKnow.

As this report shows, there is no shortage of available standards, vocabularies and software libraries. Nevertheless, the status of some of the existing work is uncertain and the plethora of available choices is itself a problem. Therefore the hope and expectation is that new work will be undertaken jointly by W3C and OGC working together to address:

- a formal standard for GeoJSON as a profile of JSON-LD;
- best practice guidance on spatial ontology use (i.e. use of GeoSPARQL, NeoGeo and the ISA Core Location Vocabulary);
- a geospatial cookbook/best practice guide;
- formal standardisation of existing ontologies for time and semantic sensor networks.

7.2 Data serialization libraries

Data serialization formats and processing available in a number of popular programming environments.

	C#	JavaScript	Java	PHP	Python	Ruby
RDF/XML	dotNetRDF http://www.dotnetr	rdflib.js https://githu	Apache Jena https://jena.a	EasyRDF http://www.	RDFLib https://githu	Rdf-rdfxml* http://ruby-

61 See also section 5 of D2.1 “Requirements of the SmartOpenData Infrastructure”, January 2014

	df.org/, SemWebhttp://razor.occams.info/code/semweb/	b.com/linkeddata/rdflib.js	pache.org/documentation/rdf/	easyrdf.org/	b.com/RDFLib/rdflib	rdf.github.io/rdf-rdfxml/
N-Triples	-	Rdflib.js	Apache Jena		RDFLib	rdf-turtle* https://github.com/ruby-rdf/rdf-turtle
TriG	-	-	Apache Jena		RDFLib	-
RDFa	-	-	Apache Jena		RDFLib	RDF-rdfa* http://ruby-rdf.github.io/rdf-rdfa/
Notation3	SemWeb	Rdflib.js	Apache Jena		RDFLib	Rdf-n3* http://ruby-rdf.github.io/
Turtle	-	Rdflib.js	Apache Jena	EasyRDF	RDFLib	Rdf-turtle
JSON/JSON-LD	json-ld.net https://github.com/NuGet/json-ld.net	https://github.com/digitalbazaar/jsonld.js	Apache Jena, JSONLD-JAVA https://github.com/jsonld-java/jsonld-java	EasyRDF php-json-ld https://github.com/digitalbazaar/php-json-ld and JsonLD https://github.com/lanthaler/JsonLD	PyLD https://github.com/digitalbazaar/pyld, Fiona https://pypi.python.org/pypi/Fiona able to add JSON-LD context to GeoJSON data	JSON-LD reader/writer* https://github.com/ruby-rdf/json-ld/
rdf/JSON	-	-	Apache Jena	EasyRDF	RDFLib	-
Microdata	-	-	-	-	RDFLib	rdf-microdata* https://github.com/ruby-rdf/rdf-microdata

*project is a part of Ruby RDF/Linked Data for Ruby

8 Conclusion

In this deliverable we collected information about existing metadata and data ontologies, sources and data serialization with the scope on geospatial information and datasets. As was highlighted in the W3C/OGC Linking Geospatial Data workshop (5th - 6th March 2014, Campus London, Shoreditch)[62] vocabulary reuse is important, but the adoption of existing vocabularies needs care.

The same topic may cover several ontologies, but we need to be careful choosing the right one, because some of them might be outdated, with different interpretation of the same thing. It is not an easy selection, for example about “Geometry” we are speaking several ontologies – GeoSPARQL , NeoGeo , Location Core , W3C Geo. There are too many options to choose from.

Meaningful notes were made during Linking Geospatial Data (5th - 6th March 2014, Campus London, Shoreditch) barcamp session: “Determining which RDF ontologies provide best practices in areas of overlap”[63].

62 http://www.w3.org/2014/03/lgd/

63 http://www.w3.org/2014/03/06-lgd-minutes.html#tandyBarCamp

