DELIVERABLE

Project Acronym:	**OTN**
Grant Agreement number:	**620533**
Project Full Title:	OpenTransportNet – Spatially Referenced Data Hubs for Innovation in the Transport Section

D3.6 SEMANTIC REPORT

Version: 1.0

Authors:

Otakar Čerba (UWB)
Marek Šplíchal (HSRS)
Michal Šrédl (HSRS)
Přemysl Vohnout (HSRS)
Dmitrij Kozukh (HSRS)
Štěpán Kafka (HSRS)
Karel Charvát (HSRS)
Jan Bojko (HSRS)
Raitis Berzins (HSRS)

Reviewers:

KarelJedlicka (UWB)
Evangelos Argyzoudis (INTRA)
Irene Matzakou (INTRA)
Andrew Stott (CORVE)
Bart de Lathouwer (OGC)
Jan Martolos (EDIP)
Dzintars Vilnis (EX)
Pieter Colpaert (iMinds)
Phil Archer (W3C)

Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

This project has received funding from the European Union's Competitiveness and Innovation Framework Programme under grant agreement no. 620533.

Table of Contents

List of Figures

Revision History

Revision	Date	Author	Organization	Description
0.1	10/01/2015	Dmitrij Kozukh	HSRS	First draft
0.2	19/01/2015	Dmitrij Kozukh	HSRS	Input from partners Version for internal review
0.3	25/01/2015	Dmitrij Kozukh	HSRS	Internal review remarks integration
0.4	28/01/2014	Dmitrij Kozukh	HSRS	Final draft for consortium review
1.0	02/02/2014	Dmitrij Kozukh	HSRS	Update based on review comments Final version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

List of acronyms

API	Application Programming Interface
CKAN	Comprehensive Knowledge Archive Network
CSW	Catalog Service for the Web
D2RQ	Database to Relational Query
DCAT	Data Catalog Vocabulary
EU	European Union
FAO	Food and Agriculture Organization
GML	Geography Markup Language
GNNS	Global Navigation Satellite System
HSRS	Help Service - Remote Sensing
HTML	HyperText Markup Language
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
KML	Keyhole Markup Language
OSM	Open Street Map
OTN	OpenTransportNet
OWL	Web Ontology Language
PDF	Portable Document Format
RDBMS	Relational database management system
RDF	Resource Description Framework
REST	Representational State Transfer
RSS	Rich Site Summary
SOAP	Simple Object Access Protocol
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
TTL	Time To Live
URL	Uniform Resource Locator
URI	Uniform Resource Identifier
VGI	Volunteered geographic information
W3C	World Wide Web Consortium
XML	Extensible Markup Language

Executive Summary

Semantics is implemented as the fundamental property of spatial data and information which is supported and used by many tools such as ontological systems, thesauri etc. Semantic information (as part of spatial data or information) enables us to improve understanding and perception of particular objects of spatial data sets, because they are very often complicated due to big size (and number of records) and dependence (of their meaning) on context. The complex overview covering various semantic standards connected to spatial data contains:

- General standards
- Metadata standards
- Geographic Schemas
- Geo-domain vocabularies
- Ontologies

For the purpose of OTN Hubs, the following tools have been implemented for supporting semantic access to spatial data:

D2RQ[1]is a system for accessing relational databases as virtual, read-only RDF graphs. It offers RDF-based access to the content of relational databases without having to replicate it into an RDF store. Using D2RQ it is possible to query a non-RDF database using SPARQL, access the content of the database as Linked Data over the Web, create custom dumps of the database in RDF formats for loading into an RDF store, access information in a non-RDF database using the Apache Jena AP.

Virtuoso[2]is an innovative enterprise grade multi-model data server for agile enterprises and individuals. It delivers an unrivalled platform agnostic solution for data management, access, and integration.

MICKA[3] supports many output formats including HTML, PDF, JSON, GeoRSS, Atom, KML, OAI_PMH and OAI_MARC21. Internally, the metadata are stored in ISO 19139 format. The DCAT RDF-XML output was implemented with XSL template according to the defined rules. Keywords are mapped to original thesauri (e.g. GEMET) and INSPIRE registry (URI forms of INSPIRE data themes mapping).

CKAN[4] is an open source data management tool which helps to present data and make them discoverable.

All these components are currently installed as a part of OTN hub supporting semantic search and querying.

[1]For detailed description of D2RQ please see its official web-page http://d2rq.org/
[2]For detailed description of Virtuoso server see its official web-page http://virtuoso.openlinksw.com/
[3]For detailed description of MICKA in context of OTN project please see [1] and [2]
[4]For detailed description of CKAN please see its official web-page http://ckan.org/

1 Introduction

1.1 OpenTransportNet

OpenTransportNet's (OTN's) primary objective is to create an ambitious network of collaborative virtual service hubs that aggregate, harmonise and visualise open transport-related data to drive the rapid creation of innovative new applications and services. OTN specifically chose transport as a cross border focus-area because it touches upon almost every facet of 21st century living, making it an ideal target for the creation of solutions that can be enhanced by location based services such as GNNS and VGI resources such as OSM.

OTN service delivery hubs will address the following challenges:

1. need to better aggregate and harmonise data to improve accessibility and use,
2. need to link spatial and non-spatial data to extract value and increase accuracy,
3. need to provide innovators with easier APIs and GUIs to stimulate the creation of new services and commercial opportunities.

OTN innovation hubs will overcome the above challenges by using an automated flexible dataset aggregator to integrate and harmonise transport related data. OTN hubs will combine spatial (GI), dynamic data streams and non-spatial data and derive insights from the data through visualisation tools and pattern detection algorithms. OTN Hubs will go beyond the state-of-the-art by (a) improving the accuracy of data insights by enhancing knowledge with VGI and (b) deploying a sophisticated access control and identity management system that will manage privacy controls.

OTN will be validated in four pilot locations: the UK, Belgium, France and the Czech Republic. The OTN solution will likely use a social enterprise freemium business model. OTN Lite will be an open service that provides access to open data sets and an innovation environment. OTN Premium is considered to be available for a fee that enables users' access to business incubation.

1.2 The Aim of the Report

The aim of this deliverable is to describe the implementation of a semantic framework that will support access to geospatial metadata and data about transport in the form of Linked Open Data. The geospatial data and metadata will be transformed to the RDF format that supports express tree multihierarchy and other non-trivial structures better than traditional tools. The semantic framework will also provide a developed method for assessment of information which means a possibility of more qualified decision making. Moreover, the semantic structure will be clearer not only for experts or machines but also for common users.

The Semantic Architecture described in this document is based on transferring geospatial metadata stored in Micka into RDF form, usage of CKAN as metadata SparqlEnpint, use D2RQ as tool supporting querying of SQL data using SPARQL queries and Virtuoso as SPARQL endpoint for geospatial data in RDF form.

1.3 Structure of the Report

The first technical chapter of the report describes how semantics could be applied on spatial data. The second technical chapter is focused on description of particular tools (Micka, CKAN, D2RQ and Virtuoso). Next chapter describes the installation of these tools in OTN Hub and last chapter includes the conclusion.

2 Semantics and Spatial Data

Semantics (as a study of meaning) is the other keyword of contemporary research focused on spatial data and information (the document Data Semantics: what, where and how? by Sheth, 1996 is a very nice introductory document for a first meeting with data semantics; Figure 1).

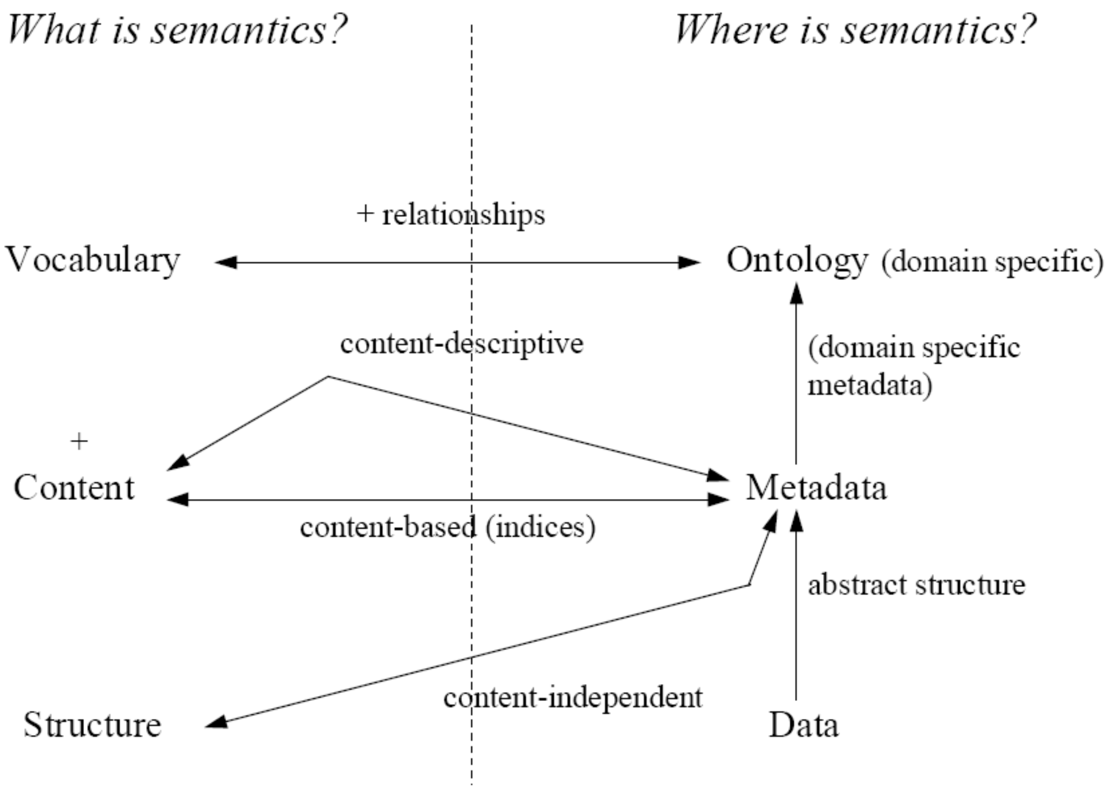

Figure 1.Data semantics (Sheth, 1996)

Semantics is implemented as the fundamental property of spatial data and information which is supported and used by many tools such as ontological systems, thesauri etc. Semantic information (as a part of spatial data or information) enables us to improve understanding and perception of particular objects of spatial data sets, because they are very often complicated due to big size (and number of records) and dependence (of their meaning) on context. Importance of semantics is shown in the Caracciolo's description of geographic concept (as representatives of geographical phenomena) – “a concept may be anything: an animal, a technique, and so on. Operationally, a concept is the set of all terms used in all languages to describe the same idea.” (Caracciolo, 2013). It is evident that geographical concept cannot be self-standing, because they are quite ambiguous and they could be unintelligible and incomprehensible.

It is evident from theoretical work dealing with geo-semantics (e.g. Bittner & Winter 2004 or Janowicz et al. 2010). Particular studies are focused on thedevelopment of semantics models (Huang et al., 2010, Zhang & Xu, 2011, Wang et al., 2011, Zhang et al., 2014), collecting of semantic data and information (Hazman et al., 2009, Pazienza et al., 2012), linking of semantic data resources (Severino, 2007, Lauser, 2008, Lopez-Pellicer et al., 2010), semantic similarity (Bae et al., 2014, Batet et al., 2014, Jiang et al., 2014) or development of semantic data resources in various domains (An & Zhao, 2007, Ping & Yong, 2009, Li et al., 2013, Gimenez et al., 2013). The majority of the above mentioned documents, which are connected to geographic domain, constitute the main theoretical background of this article as well.

The particular research of semantics and comprehensibility of geographic concepts isconcentrated on various topics and domains which are usually connected with one common approach – studying of similarity. Bennett (2001) studied above all a vagueness of geographic concepts, its types (sorites and conceptual) and nature. He also summarized a formalising of vagueness with use of multi-valued logic and fuzzy logic. The

paper Semantic Categories Underlying the Meaning of 'Place' (Bennett & Agarwal, 2007) follows previous above mentioned articles. The other study of semantics of geographic concepts (Kavouras et al., 2005) compared concepts connected to water (for example "lake" or "sea") in three well-known ontologies at that time (CORINE land cover, MEGRIN and WordNet). Very similar research on case of land cover concepts and four land cover nomenclatures were published in Deng (2008). Other similar research (Ballatore et al., 2013), deals with comparison and computation of similarity vectors and matrices of volunteered lexical definitions of geographic concepts. Fonseca et al. (2006) provide a more general study of interoperability of geographic concepts (from land cover domain), which results to relations (similarity and generalization) among particular concepts.
The complex overview covering various semantic standards connected to spatial data and geographical domain is provided by the project SmartOpenData (smartopendata.eu). This document is divided into several parts on the basis of type of semantic information:

- General standards (Resource Description Framework, modeling languages, Resource Description Framework Schema, Web Ontology Language, types of RDF serialization...)
- Metadata standards (metadata usage in semantic web, Dublin Core metadata /ISO 15836/, OWL metadata, semantic sitemaps and voiD, licenses, data norms, DCAT application profile for data portals in Europe, CKAN domain model, geospatial metadata standards – ISO TC 211 metadata standards, Open Geospatial Consortium standards, INSPIRE metadata profile)
- Geographic Schemas (Geospatial Resource Description Framework, Geographically Encoded Objects for RSS feeds, GeoRSS-Simple, GeoRSS GML, GeoSPARQL, NeoGeo Vocabulary, W3C Basic Geo, Location Core Vocabulary)
- Geo-domain vocabularies (AgroVoc, GEneral Multilingual Environmental Thesaurus, Agroontology, GeoSpecies Ontology, Landinndelingen i Norge Norway /Administrative division of Norway/, EuroVoc, STW Thesaurus for Economics)
- Ontologies / Linked data set containing spatial data (dbPedia, FAO Geopolitical Ontology, Ordnance Survey Products, U.S. National Map, Geonames.org, LinkedGeoData.org, Getty Thesaurus of Geographic Names®)
- Top-level ontologies (WordNet, Descriptive Ontology for Linguistic and Cognitive Engineering /DOLCE/, Cyc, Suggested Upper Merged Ontology /SUMO/)
- Existing technologies and software projects (for example data serialization libraries, noSQL databases, RDF databases, semantic frameworks, project GeoKnow)

The benefits of semantic technology for spatial metadata and data management are numerous and varied [18,20,31]. For example, ontologies have been used for formalization of taxonomies, classification schemes or nomenclatures on web portals (e.g. habitat or species taxonomies, categories of environmentally sensitive areas, or hierarchical land use classifications). The role of these ontologiesas self-standing products is however limited. They provide background knowledge, but only in some experimental prototypes they are used for constructing search requests or for grouping of search results into meaningful categories. Further, in experimental settings, there are examples of using OWL for bridging differences in conceptual schemas [16]. The role of ontologies and knowledge engineering in these prototypes is basically to provide methodologies for integration and querying [11,34]. Ontologies have played an important role in structuring data of geospatial domains [3]. However, semantic technology has not influenced spatial data management yet, and mainstream GIS tools are not yet extended with semantic integration functionality.
The use of the tools is not easy task for the data provider and it will probably not give him a lot of direct and immediate benefits. A data provider or a data processor should also use some third-party tool which could facilitate the transformation from a classical spatial data to a linked data or open linked data. However data which become linked have undoubtedly benefits in case of interconnections to other data structures and also some issues(for example redundancy of data) has to be resolved. The benefits and considerations raised during SmartOpenData project are listed below.

OpenDataSupport has identified the following benefits of using linked data[5]:

a. Allows for flexible integration of datasets from different sources, without needing the data to be moved (this principle is in harmony with INSPIRE requirements).
b. Fosters the reuse of information from reference/authoritative sources.
c. Caters for assigning common identifiers in the form of HTTP URIs to things (e.g. people, products, business, locations...).
d. Provides context to data – richer and more expressive data.
e. The use of standard Web interfaces (such as HTTP and SPARQL) can simplify the use of data for machines.

OpenDataSupport also identifies the following considerations for publishing Linked Data that has been raised during the SmartOpenData infrastructure building:

a. Linked Data is high-quality data. Considerable data cleansing and curation is required.
b. Managing the data lifecycle is a challenging task. Mechanisms for handling updates and deletions in the data should be devised.
c. The tools and software supporting linked data solutions are still not at production level/quality.
d. Various interpretations and views on semantics rules that have to be harmonized to provide clear relations between data.
e. A central authority should take the responsibility of publishing and maintaining persistent HTTP URIs for data resources. Existing identifiers should be reused to the extent possible, especially the ones coming from reference data sources, such as the INSPIRE Registry[6], EU Publications Office Metadata Registry (MDR)[7], and company registers[8].
f. Data is currently available under different licences and in most cases no licence actually exists. This hampers data reuse and integration. Possible licensing options for data and description metadata should be explored. The use of open licence, e.g. a public domain licence – CC0, is recommended, particularly for the metadata.
g. Alternative business model for publishing linked data should be further explored. The costs and benefits of the different alternatives need to identified, before governments can decide on the adoption of the linked data technological paradigm.

Open public data resources for re-use is one of the key priorities of the Digital Agenda for Europe. Data available in public European organisations have an enormous potential economic growth.

Producing and updating geospatial data is expensive and resource intensive. Hence, it becomes crucial to be able to integrate, repurpose and extract added value from geospatial data to support decision making and management of local, national and global resources. Spatial Data Infrastructures (SDIs) and the standardisation efforts from the Open Geospatial Consortium (OGC) serve this goal, enabling geospatial data sharing, integration and reuse among Geographic Information Systems (GIS). Geospatial data are now, more than ever, truly syntactically interoperable. However, they remain largely isolated in the GIS realm and thus absent from the Web of Data. Linked data technologies enabling semantic interoperability, interlinking, querying, reasoning, aggregation, fusion, and visualisation of geospatial data are only slowly emerging[9].

Spatial Data Infrastructures (SDIs) were created to promote the discovery, acquisition, exploitation and sharing of geographic information. They include technological and organisational structures, policies and standards that enable efficient discovery, transfer and use of geospatial data using the web[27]. R&D in this field is closely tied to standardisation activities led by international bodies, namely the ISO/TC 211[10], OGC[11] and W3C[12].

One of the purpose of choosing semantic technologies for the Open Transport data is to make them better available for citizens, but also for public and private organisations and SME developers. On one hand, Europe and the EU invest hundreds of millions of Euros in building the INSPIRE infrastructure. On the other hand, most European SMEs and citizens use Google maps for their applications. National and regional SDIs offer information which is not available on Google, but this potential is not used.

---

[5] See http://www.slideshare.net/OpenDataSupport/introduction-to-linked-data-23402165
[6] See https://inspire-registry.jrc.ec.europa.eu/
[7] See http://publications.europa.eu/mdr/
[8] See for instance http://opencorporates.com as an excellent open source of such information
[9] See Annex A for a discussion on Linked Data
[10] See ISO /TC 211 Geographic Information/Geomatics website: http://www.isotc211.org
[11] See Open Geospatial Consortium website: http://www.ogc.org
[12] See W3C website: http://www.w3.org

3 Semantics and Spatial Metadata

DCAT is a metadata format proposed for European portals based on principles of the semantic web. To identify how to create a DCAT-AP-compliant representation of INSPIRE metadata in order to enable their sharing with other sectors outside the thematic and normative framework of INSPIRE. This analysis is not meant to provide a complete representation of all INSPIRE metadata elements, but only of those included in DCAT-AP. More detail description is in OTN report related to Metadata [2].

4 Used Tools

Within the OTN project it is not planned to test the most suitable tools for semantic approach, fortunately the project can build on existing work which has been done during the other projects (for example GeoKnow or SmartOpenData).

The GeoKnow[13] project documents geospatial support in several semantic repositories that are widely known in the Semantic Web and Linked Data communities. It also covers topics related to RDF, OWL, SPARQL and GeoSPARQL languages, as well as research trends. It designed and conducted an evaluation study for some well known RDF stores (Virtuoso[14], OWLIM[15], Parliament[16], uSeekM[17], Strabon[18]) that provide geospatial support, and compared their performance against widely used geospatial databases (PostGIS[19] and Oracle Spatial[20]) in terms of bulk loading, spatial indexing and query execution time. It utilized properly converted Open Street Map data as a very large data source, and typical geospatial queries for evaluating performance and scaling attributes of current approaches, and provide quantitative metrics for evaluating the performance of prototypes that will be produced in the GeoKnow project.

The SmartOpenData[21] project has focused on more pragmatic approach for the tools and has tested the tools within pilot tasks. During the testing the D2RQ has shown as tool for data providers with less need for data harvesting and linking inside the infrastructure of the concrete provider. Virtuoso is on the other hand more suitable for complex tasks, it is very sophisticated tool for all the tasks with semantic search, linking, harvesting, transformation etc.

Based on the results of the mentioned projects the following tools were selected.

4.1 D2RQ

The D2RQ Platform is a system for accessing relational databases as virtual, read-only RDF graphs. It offers RDF-based access to the content of relational databases without having to replicate it into an RDF store. Using D2RQ you can:

- query a non-RDF database usingSPARQL
- access the content of the database as Linked Data over the Web
- create custom dumps of the database in RDF formats for loading into an RDF store
- access information in a non-RDF database using the Apache Jena API[22]

The D2RQ Platform uses the D2RQ Mapping Language to map the content of a relational database to RDF. The D2RQ also provides routines to generate the map represented by a mapping file according to a

[13] See at: http://svn.aksw.org/projects/GeoKnow/Public/D2.1.1_Market_and_Research_Overview.pdf
[14] See at: http://virtuoso.openlinksw.com/rdf-quad-store/
[15] See at: www.ontotext.com/owlim
[16] See at: http://parliament.semwebcentral.org/
[17] See at: https://dev.opensahara.com/projects/useekm
[18] See at: www.strabon.di.uoa.gr/
[19] See at: http://postgis.net/
[20] See at: www.oracle.com/us/products/database/options/spatial/overview/index.html
[21] See at: http://www.smartopendata.eu/
[22] See at: http://d2rq.org/

database model. The D2RQ can also follow references among database tables, therefore it is recommended first generate the mapping file using the D2RQ script and then edit it to user needs.

During other EU projects (e.g. SmartOpenData) it was tested that D2RQ is suitable for a data providers. Potential users can choose data source either directly from a TTL and a HTML format or JSON and XML through a SPARQL.

Many potential use cases with the D2RQ exists, one with data from the OSM would be how to create API for apps, which are working with a bus schedule. Let's imagine we have roads and bus stations from the OSM, at first we need calculate routing information for the network (route obstacles, types, weights etc.), afterwards adjust bus stations points to the network and finally create and fill suitable database model. After this we will have all necessary to deploy a D2RQ, which provides access to the routing data. One similar implementation, which uses another software, is Swiss Open Transport API[23].

Advantage of this approach is that the routing data can be extended with other linked information. If a provider has data about roads and another has data about bus tracks, bus departures or bus positions, D2RQ mapping file can be updated with this linkage. Applications or users can use a SPARQL endpoint and see bus positions, bus delays etc. Step by step tutorial for creation of a D2RQ SPARQL endpoint on the tomcat java container is below.

4.2 Virtuoso

Virtuoso is an innovative enterprise grade multi-model data server for agile enterprises and individuals. It delivers an unrivalled platform agnostic solution for data management, access, and integration.

The unique hybrid server architecture of Virtuoso enables it to offer traditionally distinct server functionality within a single product offering that covers the following areas:

- SQL Relational Tables Data Management (SQL RDBMS)
- RDF Relational Property Graphs Data Management (RDF Triple or Quad Store)
- Content Management (HTML, TEXT, TURTLE, RDF/XML, JSON, JSON-LD, XML)
- Web and other Document File Services (Web Document or File Server)
- Five-Star Linked Open Data Deployment(RDF-based Linked Data Server)
- Web Application Server (SOAP or RESTful interaction modes)[24].

4.3 Micka

MICKA is a complex metadata system developed by HSRS (license holder). It includes spatial metadata editor, metadata catalogue and metadata harvester. It supports the CSW 2.0.2 / INSPIRE discovery service. In addition to the CSW 2.0.2 / ISO-AP 1.0 profile it supports many output formats including HTML, PDF, JSON, GeoRSS, Atom, KML, OAI_PMH and OAI_MARC21. It is a system which is currently used for commercial and scientific purposes.

Apart fromthe CSW 2.0.2 / ISO-AP 1.0 profile, MICKA supports many output formats including HTML, PDF, JSON, GeoRSS, Atom, KML, OAI_PMH and OAI_MARC21. Internally, the metadata are stored in ISO 19139 format. The DCAT RDF-XML output was implemented with XSL template according to the rules defined in. Keywords are mapped to original thesauri (e.g. GEMET) and INSPIRE registry (URI forms of INSPIRE data themes mapping) More in [2].

4.4 CKAN

CKAN[25] is an open source data management tool which helps you to present your data and make them discoverable.

[23] See at: http://datahub.io/dataset/transport-opendata-ch
[24] http://virtuoso.openlinksw.com/

CKAN covers many possibilities:

Publish and manage data

a) Harvesting (importing) data from different sources, including:

- Geospatial CSW servers
- Web catalogues
- Simple HTML index pages or Web Accessible Folders
- ArcGIS, Geoportal Servers and Z39.50 databases
- Other CKAN instances

b) Entering data – editing and adding in many ways, including:

- Directly via the web interface
- Using CKAN rich JSON API
- Via custom spreadsheet importers

c) Publisher tools

- Publisher (Organization) admin dashboard: manage members, datasets, approve datasets to be public, manage harvest sources all from each organization admin page.
- Forms: Create portal or publisher specific forms that pre-fill certain fields or have additional required fields to fit individual requirements.

d) Workflow

Datasets can be public or private. Admins can manage access rights for datasets.

Search and discovery

You can use 'Google-style' quick keyword search, faceting tags system, browsing related datasets, search on datasets attributes, full-text search.

Metadata

Each dataset have rich set of metadata – Title, Unique identifier, Groups, Data preview, Revision history, License, Tags, Formats, Extra fields.

Geospatial

Advanced features covering data preview, search and discovery.

Community

Several key features that allow users of a CKAN portal to communicate and collaborate on data.

Visualization

Visualisation of stored data – Table view, Graph view, Mapping data, Image data, Custom view.

Customisation

- Themes - You can customise the appearance of your CKAN portal.
- Extensions - Can add new features.

[25]http://ckan.org

Data store

CKAN provide secure storage for data in any format.

Data history

Using Open Knowledge Foundation's Versioned Domain Model to keep a complete history of all edits and versions of dataset metadata.

Federation

Can create federated network of CKAN nodes which share data between each other.

API

Rich RESTful JSON API for querying and accessing dataset information.

5 Installation

In this section, we describe how to install D2RQ and Virtuoso. For installing Micka and CKAN, please refer to D3.2 Geospatial Metadata Catalogue [2].

5.1 D2RQ

To install D2RQ on Tomcat, perform the following steps:

1. Generate ttl

Unix:

bash generate-mapping -u username -p password jdbc:postgresql://localhost/d2rq >mapping.ttl

Windows:

generate-mapping.bat …

2. Add server configuration into the mapping.ttl

```
@prefix d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#> .
...
<> a d2r:Server;
 d2r:limitPerClassMap false;
 d2r:limitPerPropertyBridge false;
 d2r:vocabularyIncludeInstances true;
 d2r:port 8080;
 d2r:baseURI <http://localhost:8080/d2rq/>;
 d2rq:useAllOptimizations true;

 d2r:sparqlTimeout 300;
 d2r:pageTimeout 5;
 .
```

Tomcat by default usesby default port 8080, but D2RQ uses 2020, therefore the server configuration has to be set explicitly in the ttl mapping file.

3. Change configFile param in /webapp/WEB-INF/web.xml

```
<context-param>
<param-name>configFile</param-name>
<param-value>mapping.ttl</param-value>
</context-param>
```

4. Deploy to a Tomcat

- The mapping file comes to the /webapp/WEB-INF/ directory
- Run “ant war” in the root D2RQ directory. It will generate d2rq.war file
- Copy/Move d2rq.war to the tomcat server apps dir. E.g. /var/lib/tomcat7/webapps/
- Start/Restart tomcat (not mandatory)
- Go to http://localhost:8080/r2dq

5.2 Virtuoso

Virtuoso is available for all major OS (Windows, Linux and MacOS). The commonly available version is version 6 which unfortunately doesn’t contain all required functions. For that reason, virtuoso need to be compiled from source.

Source code can be obtained from github at https://github.com/openlink/virtuoso-opensource.git

Installation manual on Linux OS:

- After obtaining source code run **autogen.sh** script which is available in root directory of source code. This will generate configure file which is needed to setup compile process. If there will be some errors it will be probably because of missing dependency library.

- When this process finish successfully you will see configure file in root directory. Using this script you can enable (or disable) extra functionality. For our purposes is needed to run configure script with this switches. **./configure --enable-python --enable-php5 --enable-bpel-vad --enable-rdb2rdf-vad --enable-isparql-vad**

- If there is no error you can just compile and install virtuoso using **make && make install**

- If you are running virtuoso on Debian Linux you can use advantages of init.d script which makes life easier with maintaining and restarting applications. This init.d script is available in source code in debian directory.

Virtuoso is started by init.d script **/etc/init.d/virtuoso start** or by using command line **bin/virtuoso-t**

6 Conclusions

As it was already mentioned, semantic tools will become important part of Spatial Data Infrastructure. Selection of tools was done mainly on the basis of previous experiences coming from SmartOpenData.

All selected tools were implemented as part of OTN Hub, integrated with other parts of HUB and became part of Hub infrastructure.

7 References

[1] OTN D2.5 OTN PROJECT ARCHITECTURE BLUEPRINT AND HUB TECHNICAL SPECIFICATION

[2] OTN D3.2 GEOSPATIAL METADATA CATALOGUE

[3] Albrecht, Jochen, Derman, Brandon, Ramasubramanian, and Laxmi. Geoontology tools: The missing link. Transactions in GIS, 12(4):409{424, 2008, and Eva Klien and Florian Probst. Requirements for geospatial ontology engineering. In 8th Conference on Geographic Information Science (AGILE 2005), Citeseer, 2005 (pp.251-260).

[4] An, Y., & Zhao, B. Geo ontology design and comparison in geographic information integration. In Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on (Vol. 4, pp. 608-612). IEEE. ISBN: 978-0-7695-2874-8.

[5] Bae, M., Kang, S., & Oh, S. Semantic similarity method for keyword query system on RDF. Neurocomputing, 2014, 146, 264-275. ISSN: 0925-2312.

[6] Ballatore, A., Wilson, D. C., & Bertolotto, M. (2013). Computing the semantic similarity of geographic terms using volunteered lexical definitions. International Journal of Geographical Information Science, 27(10), 2099-2118.

[7] Batet, M., Harispe, S., Ranwez, S., Sánchez, D., & Ranwez, V. An information theoretic approach to improve semantic similarity assessments across multiple ontologies. Information Sciences, 2014, 283, 197-210. ISSN: 0020-0255.

[8] Bennett, B. (2001). What is a forest? On the vagueness of certain geographic concepts. Topoi, 20(2), 189-201.

[9] Bennett, B. and Agarwal, P. (2007) Semantic categories underlying the meaning of ‘place’. In: Winter, S., Duckham, M., Kulik, L. and Kuipers, B., (eds.) Spatial Information Theory : 8th International Conference, COSIT 2007. Proceedings. COSIT 2007, September 19-23, 2007, Melbourne, Australia. Lecture Notes in Computer Science (4736). Springer , Berlin / Heidelberg, pp. 78-95.

[10] Bittner, T., & Winter, S. (2004). Geo-semantics and Ontology Extended abstract. In Proceedings of the Bentley Empowered Conference, Orlando, Florida.

[11] Buccella, A., Cechich, A., Fillottrani, P. Ontology-driven geographic information integration: A survey of current approaches. Computers and Geosciences, 35(4), 2009(pp.710-723).

[12] Caracciolo, C. (2013). AGROVOC model description and analysis. With suggestion for improvements. (FAO internal document).

[13] Deng, D. (2008). Measurement of semantic similarity for land use and land cover classification systems. In International Conference on Earth Observation Data Processing and Analysis (pp. 72850J-72850J). International Society for Optics and Photonics.

[14] Fonseca, F., Câmara, G., & Miguel Monteiro, A. (2006). A framework for measuring the interoperability of geo-ontologies. Spatial Cognition and Computation, 6(4), 309-331.

[15] Gimenez, P. J., Tanaka, A. K., & Baião, F. A. A geo-ontology to support the semantic integration of geoinformation from the National Spatial Data Infrastructure. In GeoInfo, 2013 (pp. 103-114). ISSN: 2179-4820.

[16] Hart, G., Dolbear, C. Ontological bridge building - using ontologies to merge spatial datasets. In AAAI Spring Symposium: Semantic Scientfic Knowledge Integration, 2008 (pp. 15-20).

[17] Hazman, M., El-Beltagy, S. R., & Rafea, A. Ontology learning from domain specific web documents. International Journal of Metadata, Semantics and Ontologies, 2009, 4(1), 24-33. ISSN: 1744-263X.

[18] Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology, 1(1), 1-136.

[19] Huang, Y., Deng, G., Wu, X., & Zhao, Z. Research on Representation of Geographic Feature Based on Geo-Ontology. In Intelligent Systems and Applications (ISA), 2010, 2nd International Workshop on (pp. 1-5). IEEE. ISBN: 978-1-4244-5874-5.

[20] Jain, P., Hitzler, P., Yeh, P. Z., Verma, K., & Sheth, A. P. (2010, March). Linked Data Is Merely More Data. In AAAI Spring Symposium: linked data meets artificial intelligence (Vol. 11).

[21] Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maué, P., & Stasch, C. (2010). Semantic enablement for spatial data infrastructures. Transactions in GIS, 14(2), 111-129.

[22] Jiang, Y., Wang, X., & Zheng, H. T. A semantic similarity measure based on information distance for ontology alignment. Information Sciences, 2014, 278, 76-87. ISSN: 0020-0255.

[23] Kavouras, M., Kokla, M., & Tomai, E. (2005). Comparing categories among geographic ontologies. Computers & Geosciences, 31(2), 145-154.

[24] Lauser, B., Johannsen, G., Caracciolo, C., Keizer, J., van Hage, W. R., & Mayr, P. Comparing human and automatic thesaurus mapping approaches in the agricultural domain. 2008.

[25] Li, J., Liang, Y., & Wan, J. Geo-Ontology-Based object-oriented spatiotemporal data modeling. In Pervasive Computing and the Networked World, 2013 (pp. 302-317). Springer Berlin Heidelberg. ISBN: 978-3-642-37015-1.

[26] Lopez-Pellicer, F. J., Silva, M. J., & Chaves, M. Linkable geographic ontologies. In Proceedings of the 6th Workshop on Geographic Information Retrieval, 2010 (p. 1). ACM. ISBN: 978-1-60558-826-1.

[27] Nebert, D. Developing spatial data infrastructures: The sdi cookbook. Technical report, global spatial data infrastructure, 2004.

[28] Pazienza, M. T., Stellato, A., Tudorache, A. G., Turbati, A., & Vagnoni, F. An Architecture for Data and Knowledge Acquisition for the Semantic Web: The AGROVOC Use Case. In On the Move to Meaningful Internet Systems: OTM 2012 Workshops, 2012 (pp. 426-433). Springer Berlin Heidelberg. ISBN: 978-3-642-33618-8.

[29] Ping, D., & Yong, L. Building place name ontology to assist in geographic information retrieval. In Computer Science-Technology and Applications, 2009. IFCSTA'09. International Forum on (Vol. 1, pp. 306-309). IEEE. ISBN: 978-1-4244-5423-5.

[30] Severino, F. The term development in the thesauri of international organisations. The European Journal of Development Research, 2007, 19(2), 327-351. ISSN: 1743-9728.

[31] Shadbolt, N., & O'Hara, K. (2013). Linked data in government. IEEE Internet Computing, (4), 72-77.

[32] Sheth, A. (1996). Data Semantics: what, where and how. In 6th IFIP Working Conference on Data Semantics (DS-6), Atlanta, GA.

[33] Wang, Z., Li, M., & Li, F. Geo-ontology model based on description logic. In Geoinformatics, 2011 19th International Conference on (pp. 1-5). IEEE. ISBN: 978-1-61284-849-5.

[34] Zhao, T., Zhang, C., Wei, M., Peng, Z. Ontologybased geospatial data query and integration. In GIScience, volume 5266 of Lecture Notes in Computer Science. Springer, 2008 (pp. 370-392).

[35] Zhang, C., Cheng, J., Liu, J., Pang, J., Liang, C., Huang, Q., & Tian, Q. Object categorization in sub-semantic space. Neurocomputing, 2014. ISSN: 0925-2312.

[36] Zhang, X., & Xu, J. Construction of geo-ontology knowledge base about spatial relations. In Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on (pp. 234-237). IEEE. ISBN: 978-1-4244-8352-5.