Huerta, Schade, Granell (Eds): Connecting a Digital Europe through Location and Place. Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, June, 3-6, 2014. ISBN: 978-90-816960-4-3

Little Steps Towards Big Goals. Using Linked Data to Develop Next Generation Spatial Data Infrastructures (aka SDI 3.0)

Francis Harvey
University of Minnesota
Minneapolis, MN, USA
fharvey@umn.edu

Jim Jones
Westfälische Wilhelms Universität, Münster, Germany
jim.jones@uni-muenster.de

Simon Scheider
Westfälische Wilhelms Universität Münster, Germany
simon.scheider@uni-muenster.de

Adam Iwaniak
Wrocław University of Environmental and Life Sciences
Wrocław, Poland
adam.iwaniak@up.wroc.pl

IwonaKaczmarek
Wrocław University of Environmental and Life Sciences
Wrocław, Poland
iwona.kaczmarek@up.wroc.pl

JaromarŁukowicz
Wrocław University of Environmental and Life Sciences
Wrocław, Poland
jaromar.lukowicz@struktura.eu

Marek Strzelecki
Wrocław University of Environmental and Life Sciences
Wrocław, Poland
marek.strzelecki@up.wroc.pl

Abstract

Society is moving at an increasing pace toward the next stage of the information society through linked data. Among the relevant developments in geographic information science, linked data approaches offer potential for improving SDI functionality [12]. Linked data uses Semantic Web technologies and makes it possible to link at a very granular level data resources of the web for a multitude of purposes. While the technological implementation in many ways is still in a phase of adolescence, vast amounts of data, including geographic information (GI) have been prepared, for example by the UK Ordinance Survey [8] and other governmental and non-governmental bodies. The overwhelming focus has been on producing RDF formatted data for linked data applications--the foundation for applications. In this short paper, we provide an overview of potentials of linked open data for SDI 3.0 developments. Through two exemplary use cases we illustrate specifically some first steps towards a more web-oriented and distributed approach to creating SDI architectures. The cases demonstrate applications based on the LOD4WFS Adapter, which opens the way for multi-perspective GI applications, created on-demand from multiple GI data resources. These applications automate geometry-based selections of data using spatial queries with the use of RCC8 and OGC Simple Features topological functions. Future work in this area includes adding semantic operators to refine GI processing with multiple ontologies.

1 Introduction

The information age offers a promise of improved access, efficiency, and new discoveries through information. Partly realized in Web 1.0 and 2.0 services and applications, Semantic Web technologies, linked data concepts, fundamental to Web 3.0 implementation, define the next big goals for the information society. Geospatial information occupies an important component of developments involving Linked Data and the Semantic Web. Many initiatives, both governmental and non-governmental continue to develop linked data resources and applications [28]. A number of commercial applications have also embraced these Semantic Web technologies. This article considers these developments points to the potentials of Linked Data for future SDI architectures and broader support.

These developments continue efforts to enhance online information access and use of geographic information, opening the doors to application potentials that only a few decades ago would have seemed to come directly from science fiction. We now have the capability to access GI from anywhere on the globe. Realistically, this potential faces many technological and organizational challenges. Achieving improved access through Semantic Web applications that can handle the semantic issues [9, 15, 16, 10] offers interesting means to support uses hitherto constrained by web 1.0 and 2.0 technologies [2]

Semantic Web technologies, understood as an important part of Web 3.0, in summary, serve as integrators across different content, information applications and systems. Already the applications are diverse with implementations in government [24, 8], commercial applications, entertainment, education [14] and other domains [18]. Linked open data (LOD) in particular refers to the practices for publishing and connecting data on the Web [3]. In other words, linked data offers the technologies for creating dynamic integration on the Web.

If we take a step back, we can see that many Semantic Web technologies offer information integration capabilities envisioned in SDI concepts [22, 20, 19, 31] but challenging to realize. The SDI implementations could only support links at the file level: each ftp:, http:, https: etc access retrieves an entire file. The operating system-based distribution of files requires additional steps to process data from the Web. A skilled and usually also knowledgeable user needs to locate, download, then prepare all required resources in order to answer a query in the common Web architecture used by SDI. This is clearly complex, time-consuming and inefficient for information integration [23]. The intent of SDI was to support feature level queries and processing, however, querying across repositories distributed on the Web is not feasible. This becomes possible using Semantic Web technologies, such as RDF, and opens possibilities to integrate or aggregate subsets of datasets based on logical criteria, e.g, topological operators determining if features are within a specified extent.

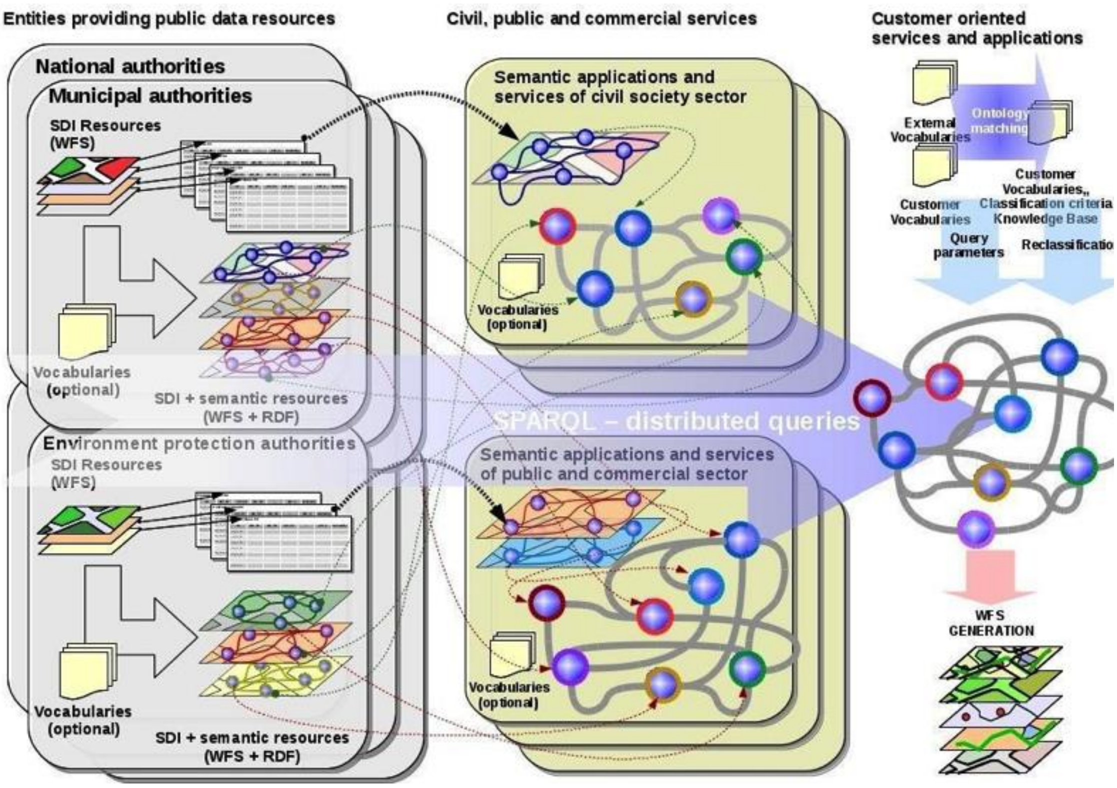
This paper describes a potential bridge between Web 3.0 based-architecture and SDI-type repositories to help make SDI-type applications more efficient and flexible. The proposed connection closely corresponds to OGC goals and

2 Technologies, Applications, and Institutions

Geographic applications and information needs range greatly [31] but every GI-Application needs data; and the data is needed by groups. Often portions of single or multiple files are required. Presently, most GI processing requires use of a GIS and/or SDI and proficiency in several areas beyond geoinformatics. The predominant implementation using Web 1.0 and 2.0 approaches only allows for the access of entire files. We need to consider technologies, applications, and institutions holistically to connect uses to data.

The main objective of SDI is to facilitate access to spatial data services through search and preview functionalities provided by portals. Public administration is the main actor as a provider of SDI [11] which, according to the INSPIRE directive, for example, is responsible for the publication of

Figure 1: Resources, services, and applications illustrating the capability to dynamically construct geographic information graphs to support multiple uses



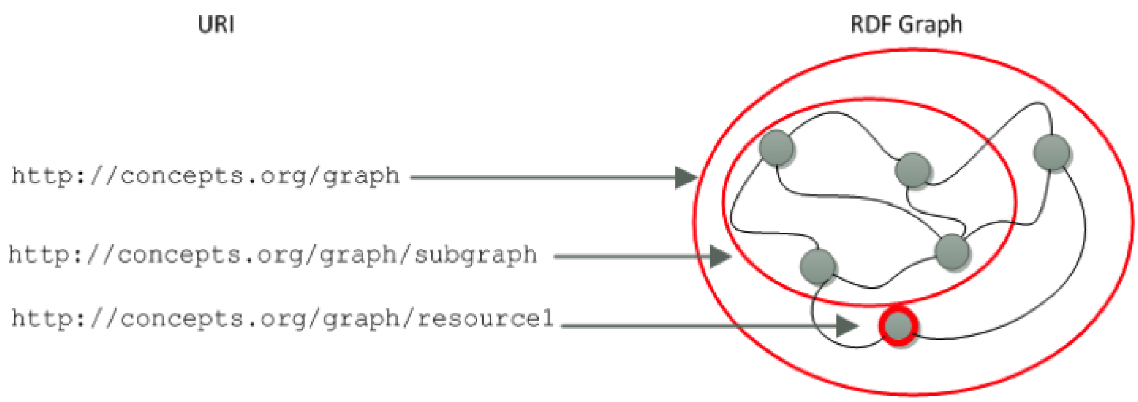
standards' development that implement Web 2.0 technologies, e.g., WMS, WFS [30]. LOD additionally provides the foundation for incorporating semantic considerations in GI processing.

In this paper we consider some LOD related technologies and how we can utilize common technologies to improve feature-level queries over distributed datasets. We focus particularly on developments using the LOD4WFS Adapter, which enables access to LOD data sources via OGC Web Services [13].

Web services according to OGC standards. The task of public administration is to publish structural data including spatial data, but in a LOD approach this occurs in a way that supports the ability to integrate data from multiple heterogeneous sources using Semantic Web technology, particularly RDF. Consuming RDF data is currently more complex for users compared to the well-known and widely implemented WMS and WFS services. Using the LOD4WFS Adapter we can connect the LOD, SDI repositories, and support multiple GI-Applications (see Figure 1).

Retrieving data from SDI resources into RDF triples can be done in two ways. First is a simple transformation. This transformation resembles a W3C Direct Mapping for the transformation from relational model into RDF graph. We use a similar mapping which binds WFS/GML objects to RDF nodes, WFS/GML attributes to RDF graph edges (properties) pointing to literals (OWL datatype properties), references between WFS/GML objects to RDF graph edges pointing to other resources (OWL object properties). The second possibility is associated with the use of ontology languages (e.g. RDF Schema or OWL Tbox) and a mapping ontology. It allows developing more sophisticated transformations, which create a "view" on original SDI resources in terms of an RDF representation. This is compatible to the W3C R2RML adapter (http://www.w3.org/TR/r2rml/). The result of a SPARQL query in this environment is a new graph that aggregates elements from multiple RDF data sources and constructs new connections between them (see Figure 1).

Figure 2: Dereferencing URIs of resources stored in graph for multiple purposes



Ontologies occupy a very important role in the nascent web of data. They provide information about properties which relate resources (nodes) in a graph. We can use the NeoGeo vocabulary (NeoGeo Vocabulary, 2014) to supplement repository and data set level ontologies. It makes it possible to formulate SPARQL queries to discover required data from multiple RDF data sources.

A distributed SPARQL query results in an RDF graph consisting of objects retrieved from various sources provided by separated institutions and organizations (see Figures 2 and 4). Using spatial information in the form of semantic representations we can discover spatial relations between heterogeneous data and enrich result graphs with new relations (see Figure 3). New graphs could include new literal values too, which were, e.g., derived from information transformations [4]

One of the main problems connected with using Semantic Web technologies with GI is the heterogeneity of the semantic and spatial data [9, 6, 7, 26, 29, 25]. To properly operate between these two different approaches it is needed to establish bridges. The first need is to be able to serve spatial data as semantic data with the use of RDF data model and appropriate vocabularies or ontologies. There are applications capable of exposing spatial data (geometry and properties) as RDF triples, which can work as online triple store server, e.g. Parliament [1], SemGeo, Strabon [17], and offline converter of spatial data TripleGeo (TripleGeo). Online triple stores also give user the possibility of querying graphs with the use of SPARQL queries with extensions like GeoSPARQL, which provide spatial analysis functions. The second need is to use semantic data within GI systems, which comes down to converting spatial data representation from RDF triples back to GIS-compliant representations like GML or Shapefiles. This part can be done with the LOD4WFS Adapter, which is capable of explore existing triple stores with spatial features and expose them as OGC WFS service, which can be opened and displayed directly in any GIS that implements the OGC WFS standard (Standard Data Access). The LOD4WFS Adapter also provides the possibility of creating on-demand WFS layers from SPARQL queries and executing them on distributed remote triple stores (Federated Data Access). If the query result contains spatial data, it can be converted on the fly into GML and served through a WFS service.

Figure 3: Example of a distributed query integrating data from two different sources.

```
PREFIX : <http://www.example.com/>
SELECT  ?o1 ?o2
WHERE
{
SERVICE <http://example1.com/sparql_endpoint1>
{
  ?s1 ?p ?o1 .
{SERVICE <http://example2.com/sparql_endpoint2>
{
    ?s2 ?p ?o2 .

}}
```

The result of the query is the list of East England administrative districts, which is published through a WFS service using LOD4WFS Adapter (Figure 5).

Figure 4: Example of federated query with DBpedia and Ordnance Survey Linked Data.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?abstract ?resource ?name ?entry (concat("POINT(", xsd:string(?long), " ", xsd:string(?lat), ")")
AS ?wkt) ?gss ?unitid
WHERE {
SERVICE <http://data.ordnancesurvey.co.uk/datasets/os-linked-data/apis/sparql>
 { ?x <http://www.w3.org/2000/01/rdf-schema#label> ?name .
  ?x <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat .
  ?x <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long .
  ?x <http://data.ordnancesurvey.co.uk/ontology/admingeo/gssCode> ?gss .
  ?x <http://data.ordnancesurvey.co.uk/ontology/admingeo/hasUnitID> ?unitid .
  ?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://data.ordnancesurvey.co.uk/ontology/admingeo/District>
  SERVICE <http://dbpedia.org/sparql/>
   { ?entry <http://www.w3.org/2000/01/rdf-schema#label> ?place .
    ?entry <http://dbpedia.org/ontology/abstract> ?abstract .
    ?entry <http://dbpedia.org/ontology/isPartOf> <http://dbpedia.org/resource/East_of_England>
    FILTER langMatches(lang(?place), "EN")
    FILTER langMatches(lang(?abstract), "EN")
    FILTER ( str(?place) = ?name )
  }}
```

3. Use cases

Semantic web technologies can use linked data for a large range of geographic information application. We consider two use cases in some detail to illustrate the potential of LOD and semantic web technology for extending and improving SDI functionality in both the administrative and civil society domains.

3.1 Supporting administration procedures

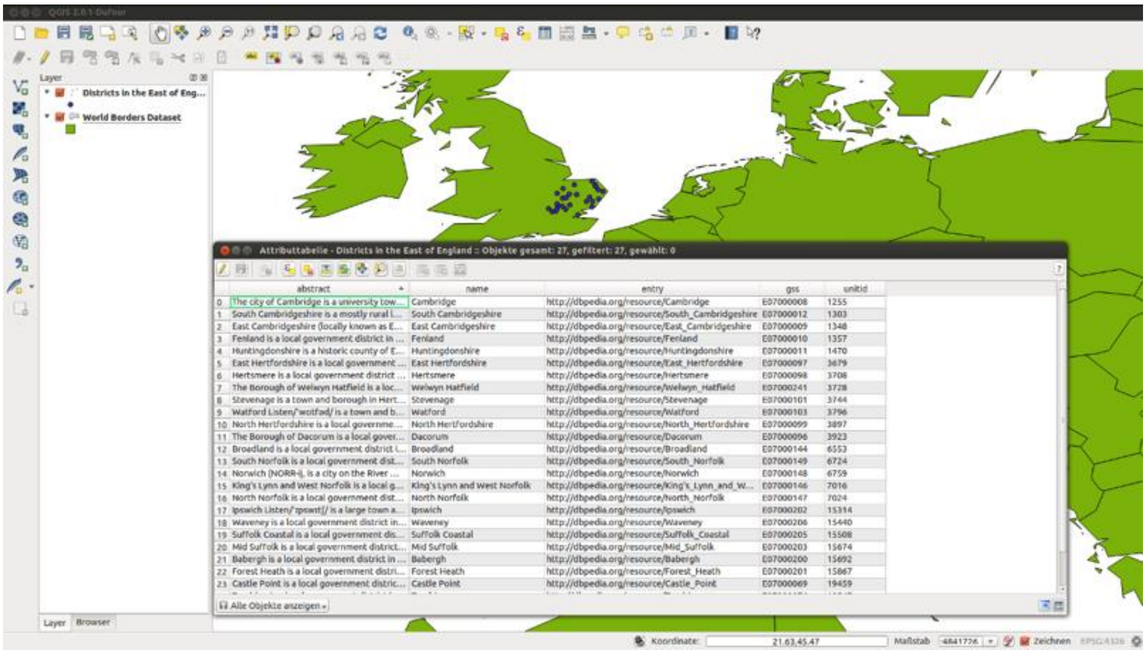
The first SDI-related case to consider is building permit application processing. The case focuses on providing simple access to relevant data by theme and by extent and legal requirements. In the prototype version of the application, administrative staff has to manually determine which data sets to include; the linked data implementation can automatically determine the extent and only access and download the corresponding data from multiple data sources. Later versions of the application can add richer semantic functionality to automate the selection of additional data sets based on criteria and relevant rules and laws.

Most important in the first version of the application is the support of the administrative officer, who must prepare a complete set of required information for review. The decision maker should be able to analyze and check all circumstances relating to the decision making procedure. This person can obtain information to determine,

- if a parcel in question is covered by local development plan,
- if a parcel has appropriate access to local transportation infrastructure,
- if it is possible to access all other required infrastructure systems (water supply, sewage system, electric power supply)
- if the proposed building includes measures to assure public safety

Important in the review and decision process is finding the data sets relevant to the planned uses and long-term

Figure 5: GML data converted by LOD4WFS adapter and published in WFS



developments of a given project. In particular, it is crucial to determine what rules and requirements apply and consider existing easements and liens. When the proposed building is multi-functional, multiple inter-related rules and regulations often apply.

Using a semantic web approach, the data needed for the review and decision come from heterogeneous sources: data repositories, specifications, guidelines, etc. They include public registers, cartographic and cadastral data, information about infrastructure facilities, planning regulations, past building activities and proposals. The application supports review through a browser-based user interface that allows the administrative staff to define geographic location and manually select relevant layers; the decision making process uses the same application architecture enhanced with annotations from the reviewing staff members. The application retrieves RDF resources using the *GeoSPARQL vocabulary*[1] and RCC8 operators [5] through SPARQL queries and assembles them into the linked data resource for supporting the query. Using the LOD4WFS Adapter, linked data can be transformed into GML and distributed through WFS distribution to GIS and other WFS capable browsers.

3.2 Supporting tourism and local tourist industry

SDI data is an important resource for many decisions. Topographic data constitutes an information resource familiar to a broad range of people and therefore extremely valuable in most cases in providing the necessary 'base data' for a literally unlimited number of applications. In the second use case, we explain how existing SDI topographic data encoded in an RDF store and made accessible using LOD can be aggregated by value-added-resellers in the tourism industry to produce tailored geographic information products for local area visitors.

Solutions based on LOD could remedy the drawbacks of data-centric applications and support a broader range of queries. We propose tourism application built-up in the way that makes it possible retrieving RDF geographic information from various sources, literally a LOD data cornucopia provided by governmental entities, tourism sector entities or social groups as well as volunteers and customers. Such data can be accessed through SPARQL Queries using GeoSPARQL vocabulary and RCC8 operators. A federated query should enable access to other RDF data sources that provide URI information. Each query assembles the retrieved data into a linked data resource for supporting the query. The query result can be then displayed in a regular GIS using the LOD4WFS Adapter. Using applications based on SPARQL, a user could compose a request that accesses multiple data sources. The query results can be linked to services enabling purchase of airlines tickets, maybe municipal public transport tickets (zone or time period tickets), hotel accommodation booking, museum tickets, entertainment sites entrance permissions, restaurant reservations, excursion vouchers and so on.

This LOD approach can support a broad range of queries. In particular it includes spatial location factors, spatial relations between subject of interest and methods of profiling attractions and services. Spatial queries (mutual location of objects, transportation bindings, accessibility) can be resolved by GeoSPARQLtriple stores, embracing finding locations and recognizing of topological relations (RCC8). It is also possible to build customized desktop GIS applications, based on QGIS (http://qgis.org), Kosmo (http://www.opensig.es/) or OpenJump (http://www.openjump.org/) open source tools. This will be possible thanks to the dynamic transformation of LOD resources into standard WFS documents. Network applications could be based on the available tools, such as OpenLayers (http://openlayers.org/) libraries or frameworks such as GeoMajas (http://www.geomajas.org/).

4. Outlook

This paper illustrates the potential of linked open data approaches for SDI-type applications based on semantic web technologies. Beyond this rather brief and conceptual overview, a number of issues remain for future work to consider: using SPARQL queries to create new objects; adding metadata generated automatically during queries;creating a reference implementation;assessing the use of ontologies to enrich semantic web application functionality;considering processes for integrating open data that address accuracy and quality concerns;exploring extensions to support real-time sensor data integration;supporting web-based data analytics operations.

Extending the proposed approach to new SDI organizations, the use of dereferable URIs supports more flexible queries. Every spatial data published as LOD (e.g. layers and spatial features) should have unique identifiers, which would allow users to separately acquire and use it in their applications. To emphasize the data hierarchy, the construction of URIs for spatial feature should also contain information about layer to which feature belongs. After URI dereferencing, the results should be available as a RDF document and, for using it in GIS systems, as an on-demand WFS layer created via the LOD4WFS Adapter. In this case the URI of a resources serves also as WFS service interface address.

While this paper only provides some small and future steps, we believe that they contribute to reaching larger goals of the Information Society that Geoinformatics can contribute to.

Acknowledgement

The research was partially supported by a project funded by the National Science Center granted on the basis of the decision DEC-2012/05/B/H/HS4/04197.

References

[1] Battle, Robert, and Kolas, Dave. 2011. Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL Semantic Web Journal.

[2] Berners-Lee, Tim, James Hendler, and OraLassila. 2001. The Semantic Web. *Scientific American* 501.

[3] Bizer, Christian, Heath, Tom and Berners-Lee, Tim. 2009. Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS). January 2009.

[4] Chrisman, N. R. 1999. What does 'GIS' mean? Transactions in GIS, 3(2), 175-186.

[5] Cohn Anthony G., Bennett, Brandon, Gooday John, GottsMicholas Mark 1997. Qualitative Spatial

Representation and Reasoning with the Region Connection Calculus. *GeoInformatica*, 1, 275–316.

[6] de Man, W. H. Erik. 2006. Understanding SDI: Complexity and Institionalization. *International Journal of Geographical Information Science* 20 (3) : 329-43.

[7] Georgiadou, Yola, S. K. Puri, and S. Sahay. 2005. Towards a Potential Research Agenda to Guide the Implementation of Spatial Data Infrastructures: A Case Study From India. *International Journal of Geographical Information Science* 19 (10) : 1113-30.

[8] Goodwin, J., Dolbear, C. and Hart, G. 2008. Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. Transactions in GIS, 12: 19-30.doi: 10.1111/j.1467-9671.2008.01133.x

[9] Harvey, F., Kuhn, W., Bishr, Y., Pundt, H., &Riedemann, C. 1999. Semantic Interoperability: A Central Issue for Sharing Geographic Information. *Annals of Regional Science*, 33(2), 213-232.

[10] Harvey, Francis, Adam Iwaniak, Serena Coetzee, and Antony K. Cooper. 2012. *Sdi Past, Present and Future: a Review and Status Assessment*. In *Spatial Enabling Government, Industry and Citizens*, edited by Abbas Rajabifard, and David Coleman. Needham, MA: GSDI Association Press.

[11] Hjelmager J., Moellering H., Cooper A., Delgado T., Rajabifard A., Rapant P., Danko D., Huet M., Laurent D., Aalders H., Iwaniak A., Abad P., Daren U., Martynenko A. 2008. An initial formal model for spatial data infrastructures International Journal of Geographical Information Science, Vol. 22 No. 11-12 (Nov. 2008), pp. 1295-1309

[12] 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.

[13] Jones, Jim, Kuhn, Werner, Keßler, Carsten, Scheider, Simon. Making the Web of Data Available via Web Feature Services. AGILE 2014, Castellón, Spain.17th AGILE Conference on Geographic Information Science, 2014.

[14] Kessler Carsten, KauppinenTomi, Linked Open Data University of Münster. Infrastructure and Applications, "9th Extended Semantic Web Conference (ESWC2012)", 2012, http://data.uni-muenster.de/context/cris/publication/75876

[15] Kuhn, W. (2001). Ontologies in support of activities in geographical space.*International Journal of Geographic Information Science*, 15, 613-631.

[16] Kuhn, W. (2003). Semantic reference systems.*International Journal of Geographic Information Science*, 17(5), 404-409.

[17] Kyzirakos, K., Karpathiotakis M., and Koubarakis, M. 2012.Strabon: A Semantic Geospatial DBMS. In the 11th International Semantic Web Conference (ISWC 2012), Boston, USA, 11-15 November 2012

[18] Marshall, M. Scott and Boyce, Richard and Deus, Helena F. and Zhao, Jun and Willighagen, Egon L. and Samwald, Matthias and Pichler, Elgar and Hajagos, Janos and Prud'Hommeaux, Eric and Stephens, Susie, 2012. Emerging practices for mapping and linking life sciences data using RDF: A case series, Web Semantics: Science, Services and Agents on the World Wide Web, Volume 14, July 2012, Pages 2-13, ISSN 1570-8268, http://dx.doi.org/10.1016/j.websem.2012.02.003.

[19] Masser, I. 1999. All shapes and sizes: the first generation of national spatial data infrastructures. IJGIS, 13(1), 67-84.

[20] Nebert, D. (ed.). 2001. The SdiCookbook, Version 1.1. 2003 :

[21] NeoGeo Vocabulary. Retrieved from: http://geovocab.org/doc/neogeo.html. (Accessed 2014, February, 21th)

[22] Onsrud, H., & Rushton, G. 1995. Sharing Geographic Information: An Introduction. In H. Onsrud& G. Rushton (Eds.), Sharing Geographic Information (pp. xiii-xviii). New Brunswick, NJ: Center for Urban Policy Research.

[23] Rautenbach V., Coetzee S., Iwaniak A. 2013. Orchestrating OGC web services to produce thematic maps in a spatial information infrastructure.Computers Environment and Urban Systems, Vol. 37, pp. 107-120

[24] Shaon, A, Woolf, A, Boczek, R, Rogers, W & Jackson, M. 2011. 'An Open Source Linked Data Framework for Publishing Environmental Data under the UK Location Strategy'. in R Grutter, D Kolas, M Koubarakis& D Pfoser (eds), Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web. vol. 798, CEUR Workshop Proceedings, Terra Cognita 2011 Workshop on Foundations, Technologies and Applications of the Geospatial Web, Bonn, Germany, 23-23 October.

[25] Schade S. and Smits, P. 2012. Why linked data should not lead to next generation SDI. IGARSS 2012: 2894-2897.
http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=06350721. Accessed 6Mar2014

[26] Tosta, Nancy. 1999. *NSDI Was Supposed to be a Verb*. In *Innovations in GIS 6*, edited by B. Gittings. London: Taylor and Francis.

[27] TripleGeo: An open-source tool for extracting geospatial features into RDF triples. Institute for the Management of Information Systems at Athena Research Center. (Accessed 2014, February, 2nd). Retrieved from: https://web.imis.athena-innovation.gr/redmine/projects/geoknow_public/wiki/TripleGeo

[28] Usery, E Lynn, and Dalia Varanka. 2012. Design and Development of Linked Data From the National Map. *Semantic Web* 3 (4) : 371-84.

[29] van Loenen, Bastiaan, Jaap Besemer, and Jaap Zevenbergen. 2009. *Spatial Data Infrastructure Convergence*.In *Sdi Convergence. Research, Emerging Trends, and Critical Assessment*, editedby Bastiaan van Loenen, Jaap Besemer, and Jaap Zevenbergen. Delft, The Netherlands: Nederlandse Commissie voor Geodesie.

[30] Vretanos, Panagiotis A. 2003. Web Feature Service Implementation Specification, Version 1.0.0. Open Geospatial Consortium.

[31] Williamson, Ian, Abbas Rajabifard, and Andrew Binns. 2006. Challenges and Issues for SDI Development. *International Journal of Spatial Data Infrastructures Research*,1 : 24-35.