The standards bodies soup recipe: an experience of interoperability among ISO-OGC-W3C-IETF standards

Francisco J Lopez-Pellicer
Universidad Zaragoza
fjlopez@unizar.es

Javier Lacasta
Universidad Zaragoza
jlacasta@unizar.es

Borja A Espejo
Universidad Zaragoza
borjaeg@unizar.es

Jesus Barrera
GeoSpatiumLab
jesusb@geoslab.es

José M Agudo
GeoSpatiumLab
joseam@geoslab.es

Abstract

Preparing a yummy soup and implementing successfully a SDI have resemblances. Both require following recipes that contains two lists. The first is the list of ingredients (e.g. a suite of standards for data, metadata and services); the second is the directions to mix them (e.g. a best practice to interoperate). There are different kinds of soups depending on the culinary tradition (or influential standards body). Theydiffer mainlyin the list of ingredients because the tradition marks what is in the list. But, what about when we prepare fusion cuisine? That is, what happens when we develop a geospatial project based on standards elaborated by different standardisation bodies? The Linked Map project is an example.The project required the joint use of emerging and established standards at data, metadata and service level elaborated by different standardisation bodies. The project faced a thought-provoking problem: the lack of standard directions to mix them. This work describes this experience, the inherent problem of our solutions (the lack of standard directions) and how should cope with it in similar scenarios.

Introduction

In 2012, Wired magazine published a feature on the desire of Ferran Adrià, world-famous chef, to bring innovation to all with a partnership with the telecommunications company Telefónica [1]. In that feature, the R+D Director at Telefónica affirmed "*the worlds of science and gastronomy may share similar processes and methodologies (...)*". We can extend his comparison to processes and methodologies we can find in the worlds of engineering and gastronomy. For example, implementing successfully a complex infrastructure such as a Spatial Data Infrastructure (SDI) and preparing a dish such as a yummy soup have resemblances. Both require practitioners to follow interoperability recipes that contain a list of ingredients and a list of directions to mix them. The list of ingredients in a SDI includes a suite of standards for data, metadata and services provided by mainly by the Open Geospatial Consortium (OGC) and the International Organization for Standardization (ISO). The directions to mix them, not surprisingly, can be found in the SDI Cookbook [2] maintained by the GSDI.

There are different kinds of soups depending on the culinary tradition. In the SDI world, it is possible to identify different SDI approaches [3]. The soups differ mainly in the list of ingredients because the tradition determines what is in the list. But, what about when we prepare fusion cuisine? That is, what happens when we develop geospatial projectsthat mix OGC and ISO standards with standards endorsed by the World Wide Web Consortium (W3C) and the Internet Engineering Task Force (IEFT)?The Linked Map[1] project is an example.It was focused on the evaluation of the usefulness of Linked Data, a W3C set of best practices and standards, for developing an infrastructure for the quality assessment of geospatial data. During its development we faced the challenge of creating our own interoperability recipes because we wanted to put emerging standards (and not so emerging) from different standardization bodies to work together. We faced a thought-provokingproblem: the lack of standard directions to mix them.

Below we are going to present the Linked Map project with more detail. Afterward, the list of standards (our ingredients) that were chosen by their properties for solving a specific set of interoperability issues related to services, data and metadata. Next, we will present three fusion recipes developed during the project that deal with data transformation, provenance metadata, and semantically enhanced services. We end discussing the inherent problemof our recipes, that is, the lack of standard directions for interoperability among standards from different standardisation bodies, and how the geospatial community copes with it.

The Linked Map project

The Linked Map project (2014) was a research subproject developed within the European FP7 Network of Excellence PlanetData (2010-2014)[2]. The aim of this network was the creation of an interdisciplinary community of researchers in Europe focused on helping organizations to expose large and vertical data on the Web taking into account cross-cutting aspects such as quality, provenance, privacy, trust and access rights. Within this network were developed research projects that addressed combinations of these aspects, proposed new methods and techniques, and opened new datasets as Linked Data.

As it was aforementioned in the introduction, the focus of the Linked Mapproject was the evaluation of the usefulness of Linked Data in a complex challenge: the quality assessment of geospatial data based on crowdsourcing techniques.We envisioned that a reviewer would have at hand a map, a semantically rich description of the map with provenance data, and comments from other reviewers as data. Next, he can add his quality reviews, which in turn

[1]http://linkedmap.unizar.es/
[2]http://planet-data.eu/

are stored as data linked to the map data and previous reviews. The review activity would be supported by an infrastructure enablingworkflows such as the proposed in Figure 0. A community of volunteer reviewers could use the infrastructure, for example, to review the integration of a National Map and Volunteer Geographic Information (VGI) (this was the application case of the project). This vision required the development of a Web client that gives access to maps and liked data, and enables to edit quality data as linked data.On the server side, our technological vision was alsobold: a seamless interoperability bridge between the geospatial community and the semantic community. It consisted in a standard OGC Web map service (WMS) that, simultaneously, was a W3C linked data node offering not only access to the geographic data rendered by the WMS but also to quality data and provenance data.Wedubbed this service*Linked Map Service*.

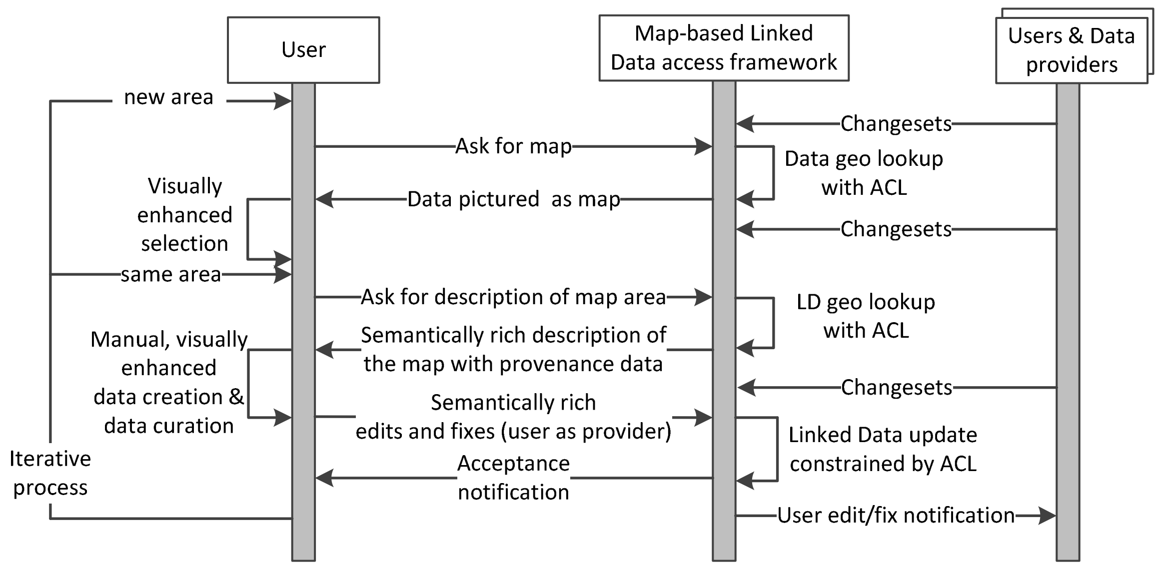

Figure 1. Quality Assessing workflow planned in Linked Map

The list of standards

The standards adopted (see Table 1) solve interoperability issues related to services, data and metadata. Data related-standards guaranteed that the infrastructure uses a general formal methodfor conceptual description (W3C RDF family), a formal ontology language (W3C OWL2), a formal vocabulary for representing geographic information (OGC GeoSPARQL 1.0) based in turn in standard definitions (OGC/ISO SFA), several serialization formats for formal languages (W3C RDF serializations), a SQL implementation of the representation of simple geographic features (OGC/ISO SFA SQL), and a well-know and ubiquitous text markup language for representing vector geometry objects (OGC/ISO WKT). Selected metadata related-standards allowed to describe formally and to serialize provenance data (W3C PROV family), to describe data transformations from the relational domain to the semantic domain (W3C R2RML), and to serve as reference model for checking the suitability of the lineagemodel implemented (ISO 19115 family). Finally, regarding to services, the chosen standards enabled the infrastructure to discover new resources by connecting semantically them (W3C Linked Data, IETF RFC 5988), view portrayals of geospatial data (OGC/ISO WMS 1.3.0), download machine-readable representation of resources (W3C Linked Data, W3C SPARQL 1.1), and invoke edit operations on data (W3C SPARQL 1.1).

Table 1 The list of standards, candidate standards and best practices adopted

Class	Kind	Bodies	Standard	Is standard?	Date
Data	Model	W3C	OWL2	Yes	2.012
		OGC/ISO	SFA	Yes	2.006
		OGC	GeoSPARQL 1.0	Yes	2.011
	Storage	W3C	RDF serializations	Yes	2.014
		OGC / ISO	SFA/SQL	Yes	2.006
			WKT	Yes	2.006
Metadata	Model	W3C	PROV-O	Yes	2.013
			R2RML	Yes	2.012
		ISO	ISO 19115 family	Yes	2.014
	Storage	W3C	PROV serializations	Yes	2.013
Service	Discovery	W3C	Linked Data best practices	Note	2.014
		IETF	RFC 5988 Web Linking	Candidate	2.010
	View	OGC / ISO	WMS 1.3.0	Yes	2.006
	Download	W3C	Linked Data best practices	Note	2.014
			SPARQL 1.1	Yes	2.013
	Invoke	W3C	SPARQL 1.1	Yes	2.013

Fusion recipes: interoperating ISO-OGC-W3C-IETF standards

In this section, we present three scenarios that involved the joint use of standards at data, metadata and service level elaborated by different standardisation bodies, namely ISO, OGC, W3C and IETF. Each scenario required the development of its own fusion recipe. These scenarios are the transformation of geospatial data from a non-standardized storage to an RDF storage,the capture with W3C PROV of the lineage of each feature instance during the conversion process, and the development of the semantically enhanced *Linked Map service*.

A data transformation recipe

Moving data from a data storage to another may involve the use of standards of different origin and purpose. For example, the source or the target storage may implement the SFA specification of ISO/OGC for two-dimensional simple features. If the project involves the use of RDF data, we would need tools able to deal with the different RDF serialization formats defined by the W3C. In many situations, we are going to extract, to transform and to load data into a new storage. Even in these scenarios we can find useful standards such as R2RML, a W3C language for expressing customized mappings from relational databases to RDF datasets.

The recipe of interoperability used in the Linked Map project used the aforementioned standards in the process of the conversion of National Map databases and VGI databases into RDF data. We can distinguish several steps in this process (Figure 2). Third party data providers publishgeospatial data relevant for the project using non-standard storage formats (e.g. ESRI Shapefiles, Microsoft Access). We extract data from these storagesand loadsextracted data in a staging relational database that implements the SFA specification. This process is described in a geospatial mapping language provided by the tool used (GeoKettle[4]). Next, we extracts data form this database and transform using a mapping tool (MorphRDB[5]) into RDF data stored ina semantic geospatial database (Strabon[6]) that supports the GeoSPARQL vocabulary. Finally, RDF data is served through a SPARQL endpoint.

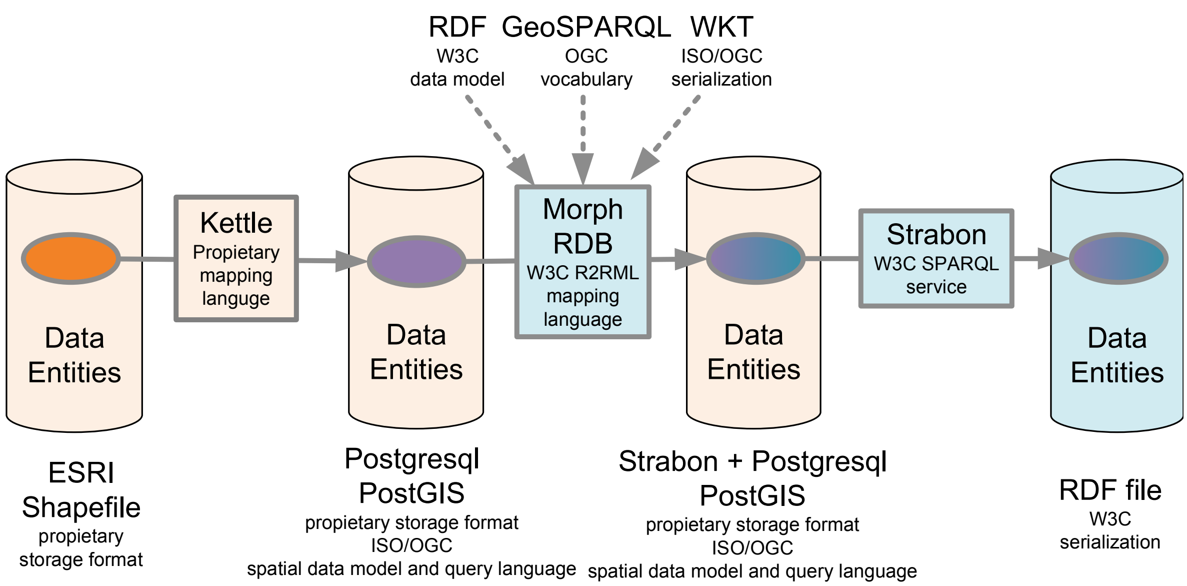

Figure 2. Our data transformation recipe

A lineage metadata recipe

The above process may generate a tremendous amount of information useful for quality control. We planned the capture of provenance metadata of each feature instance during this process. Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. Neither the ISO nor the OGC provided us a standalone and extensible model for capturing lineage. The W3C PROV data model offered us a domain-agnostic and extensible solution able to describe agents and activities involved in producing, influencing, or delivering data.

The development of a recipe of interoperability required first the development of guidelines for extending W3C PROV for including geographic information. The reference for developing them was the ISO 19115 family of standards. We also decided to keep provenance data serialized in a W3C PROV serialization format inspired by the previous decision of keep geospatial data serialized in WKT. This decision enabled us to transport provenance metadata through the different storages easily (Figure 3).

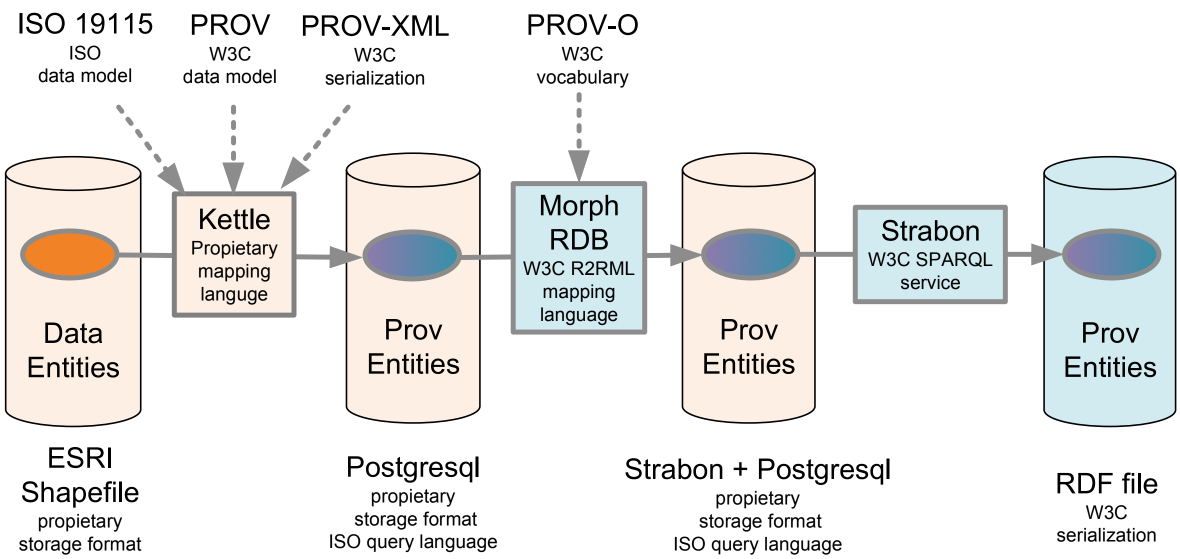

Figure 3. Our lineage metadata recipe

A semantically enhanced view service

Nowadays, geospatial processes are exposed through different kinds of Web services APIs (e.g. OGC Web Services, W3C Web Services, RESTful APIs). The interaction between user agents and Web services is ruled by the architecture of the Web that W3C defined[7]. This architecture defines howuser agents and Web services communicate over a network using the protocols of the Web by exchanging messages about resources identified by means of URIs. Each message may include data and, additionally, metadata about the resource, the data, and the message itself.In messages exchanged by the HTTP protocol, HTTP headers are the way to convey metadata. The proposed standard IETF RFC 5988defines a specific HTTP header (the *Link* header) that allows the declaration of typed links between resources on the Web. RFC 5988 is a useful tool for resource discovery. For example, Google promotes its use by webmasters to signal the preferred URI of a resource when it is served from multiple URIs or available in different formats[8].

RFC 5988 is a key ingredient in our recipe (Figure 4). Itenables to embedsemantic annotations about the output of a geospatial processes in the message metadata. In particular, it enables to add a semantic annotation to WMS responses that link them to Linked Data related to their content (e.g. the data portrayed in a map, quality comments made by reviewers, etc.) and vice versa.

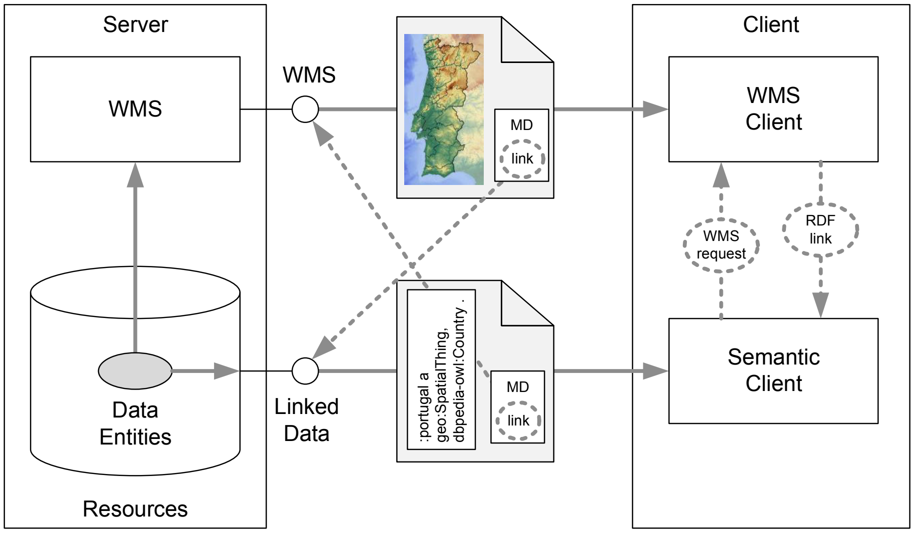

Figure 4. Our semantically enhanced view recipe (first step)

The architecture of the Web also defines how Web clients and servers use the HTTP protocol to exchange messages about resources. When a user agent makes a HTTP request it may send along some HTTP headers to indicate what data format it prefers. The server then should generate a response in the format that best matches and send it back to the client. This is known as content negotiationandmay be used for provide agnostic extension points to endpoints[9]. Content negotiation is part of the W3C Linked Data best practices.

Content negotiation is the second ingredient in thisrecipe (Figure 5). Itis used in the Linked Map project to extend WMS serversas Linked Data servers: each possible KVP-encoded request is extended to be used as the URIthat identifies the data used as input in such request and related data. It is implemented as follows. When a WMS client performs a KVP-encoded request, it sends standard HTTP headers (if any relevant) as OGC specification defines. Next, the server interprets the request as a WMS request and delegates the generation of the response to a WMS service. When a semantic client uses such KVP-encoded request as a resource identifier and dereferences it, the client adds HTTP headers that explicitly demand RDF data as response. Due to the presence of such headers, the server interprets the request as a Linked Data request and delegates the generation of the response to a Linked Data service. Extending WMS endpoints has allowed in the project the development of a client application that can get map tiles and their corresponding RDF data seamlessly.

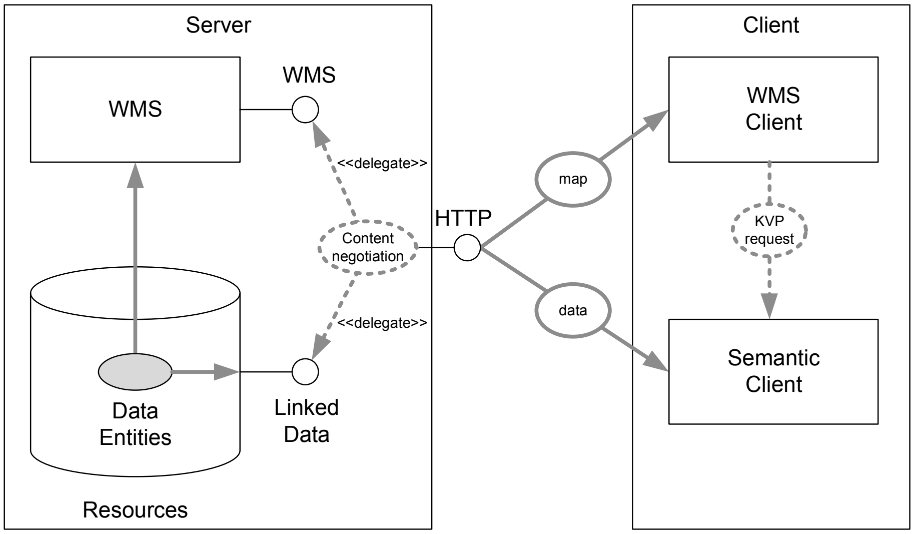

Figure 5. Our semantically enhanced view recipe (second step)

The need ofstandard interoperability directionsfor mixing standards

The above recipesinclude directions for using emerging standards or new approaches based on existing standards in scenarios dominated by geospatial standards produced by "traditional" geospatial standard bodies. The lack of standard directions has forced us to develop our own homemade recipes. It is possible to find similar proposals in other research projects (e.g. LEO [10], GeoViQua[11]).There is an inherent problem in this approach. When you find directions in a research project thatmay even seem smart only people prone to be early adopters of technologies can be tempted to reuse them.We should acknowledge that in the SDI world weare never enthusiastic about what appear to be good ideas unless standard bodies endorse their worth.

The recently formed OGC/W3C Spatial Data on the Web Working Group acknowledges the need of clarifying and formalizing the standards landscape around spatial data on the Web. It is a good step in fixing the consistency of the standard bodies soup by identifyingbest practices anddeveloping new standards (e.g. time ontology, semantic sensor network vocabulary, coverage in Linked Data). However, from our experience, a true clarification requires joint standards providing explicit directions to combine existing standards to get advantage in geospatial processes (e.g. W3C R2RML for producing OGC GeoSPARQL data, W3C PROV for producing ISO 19115 compliant lineage metadata, IETF HTTP headers and W3C Linked data best practices to semantically enhance OGC Web services).

This work began by quoting the R+D Director at Telefónica about the similitudes between two different communities: science and gastronomy. The quote was incomplete. It ends with a pessimistic "*but they rarely intersect*".I would like to end with an optimistic wish. We should create a stronger bond between the world of W3C/IETF and the world of

OGC/ISO to effectively integrate the geographic information in the Web and the Web in the geographic information.

Acknowledgement

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 257641. This work has been partially supported by the Spanish Government (project TIN2012-37826-C02-01), the National Geographic Institute (IGN) of Spain, GeoSpatiumLab S.L. and the Universidad Zaragoza (project JIUZ-2014-TEC-05).

References

[1] G. Williams, "After el Bulli: Ferran Adrià on his desire to bring innovation to all," *Wired*, no. October, 2012.

[2] "The SDI Cookbook," *GSDI Association Web Page*. [Online]. Available: http://www.gsdidocs.org/GSDIWiki/index.php/Main_Page. [Accessed: 25-Apr-2015].

[3] K. Tóth, C. Portele, A. Illert, M. Lutz, and M. N. de Lima, "A conceptual model for developing Interoperability Specifications in Spatial Data Infrastructures," European Commission Joint Research Centre, EUR 25280 EN, Apr. 2012.

[4] T. Badard, E. Dubé, B. Diallo, J. Mathieu, and M. Ouattara, *GeoKettle: a powerful open source spatial ETL tool*. FOSS4G 2009 Conference, 2009.

[5] F. Priyatna, O. Corcho, and J. F. Sequeda, "Formalisation and Experiences of R2RML-based SPARQL to SQL Query Translation using Morph ," presented at the WWW 2014, 2014.

[6] K. Kyzirakos, M. Karpathiotakis, and M. Koubarakis, "Strabon: A Semantic Geospatial DBMS," in *Lecture Notes in Computer Science*, vol. 7649, no. 19, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 295–311.

[7] I. Jacobs and N. Walsh, "Architecture of the World Wide Web, Volume One," W3C, 2004.

[8] "Supporting rel='canonical' HTTP Headers," *Google Webmaster Central Blog*, 2011. [Online]. Available: http://googlewebmastercentral.blogspot.com/2011/06/supporting-relcanonical-http-headers.html. [Accessed: 26-Apr-2015].

[9] T. V. Raman, "On Linking Alternative Representations To Enable Discovery And Publishing," *W3C*, 2006. [Online]. Available: http://www.w3.org/2001/tag/doc/alternatives-discovery-20061101.html. [Accessed: 26-Apr-2015].

[10] K. Kyzirakos, I. Vlachopoulos, D. Savva, S. Manegold, and M. Koubarakis,

"GeoTriples: a Tool for Publishing Geospatial Data as RDF Graphs Using R2RML Mappings.," *Posters & Demos*, pp. 393–396, 2014.

[11] J. Masó, G. Closa, and Y. Gil, "Applying W3C PROV to Express Geospatial Provenance at Feature and Attribute Level," in *Provenance and Annotation of Data and Processes*, vol. 8628, no. 31, S. Spaccapietra and E. Zimányi, Eds. Cham: Springer International Publishing, 2015, pp. 271–274.