(Draft) Lecture Notes in Computer Science (LNCS). 2010, vol. 6261, p. 495-502. ISSN 0302-9743.

Geo Linked Data*

Francisco J. Lopez-Pellicer[1], Mário J. Silva[2], Marcirio Chaves[3], F. Javier Zarazaga-Soria[1], and Pedro R. Muro-Medrano[1]

[1] IAAA, Universidad de Zaragoza, Spain {fjlopez,javy,prmuro}@unizar.es
[2] LaSIGE, Universidade de Lisboa, Portugal mjs@di.fc.ul.pt
[3] Universidade Atlântica, Portugal marcirioc@uatlantica.pt

Abstract. Semantic Web applications that include map visualization clients are becoming common. When the description of an entity contains coordinate pairs, semantic applications often lay them as pins on maps provided by Web mapping service applications, such as Google Maps. Nowadays, semantic applications cannot guarantee that those maps provide spatial information related to the entities pinned to them. To address this issue, this paper proposes a refinement of Linked Data practices, named Geo Linked Data, which defines a lightweight semantic infrastructure to relate URIs that identify real world entities with geospatial Web resources, such as maps.

1 Introduction

Location pins on maps are a powerful way to convey spatial information, but they are just presentation tools in Semantic Web applications. Today, we can dereference a URI, such as `http://dbpedia.org/resource/Lisbon`, to obtain machine-processable data about the city of Lisbon, and use a location pin to show the user where Lisbon is. However, the use of such maps in Semantic Web applications can produce undesired situations, such as representing data about Lisbon over a map provided by Google Maps that shows the exception message: "*We are sorry, but we don't have imagery at this zoom level for this region*". We believe that the juxtaposition of data and maps in Semantic Web applications should not be justified by the location. The juxtaposition of data and maps could only be justified when the location and the semantics are compatible.

The research question behind this issue is *how to relate semantic and geospatial depictions of a real world entity in the Semantic Web*. In other words, we are looking for a feasible approach for using maps as data in Semantic Web applications. This paper proposes a solution that intertwines formal and geospatial depictions based on the Linked Data principles. Linked Data is an initiative for interconnecting data on the Web using the Semantic Web standards, so that users and machines can explore data on the Web. Linked Data [15] promotes (1)

* Short paper accepted in *21st International Conference o n Database and Expert Sytesm Applications DEXA 2010* `http://www.dexa.org/` August 30 – September 3, 2010, Bilbao, Spain

(Draft) Lecture Notes in Computer Science (LNCS). 2010, vol. 6261, p. 495-502. ISSN 0302-9743.

the use of HTTP URIs for identifying concepts and real world entities, and (2) a sound use of the HTTP protocol to assert that a given entity identified by a URI has a Web resource as description.

The proposed solution refines Linked Data practices to use descriptions available as geospatial Web resources, such as a map from a Web mapping service. Web mapping services are part of the Geospatial Web or Geoweb [16], a set of specifications and applications which adopts the Internet and Web services to publish, access and transform geospatial content related to real world entities. Our contribution involves (1) the *characterization of geospatial proxies*, that is, geospatial Web resources that could complement Semantic Web descriptions about entities in some scenarios, (2) the identification of the *roles* of these proxies in semantic applications, (3) the *recipe* for extending Linked Data with geospatial proxies, and (4) the *advertisement of presence, role and location* of geospatial proxies using RDF graphs. We use the term *Geo Linked Data*, as it specializes the practice followed by the Linked Data community.

This paper is organized as follows: Section 2 describes related work, Section 3 details Geo Linked Data concepts, Section 4 illustrates the role of Geo Linked Data in a scenario, and, finally, we present conclusions and future work.

2 Related Work

Tabulator [3] and DBPedia Mobile [2] represent the mainstream approach for the visualization of geographic properties in Semantic Web applications: the extraction of geographic points from a description in RDF, and then, its conversion into a marker added to a map viewer client that shows the location of the entity. That is, the geospatial meaning of an entity as complex as Lisbon is often (1) *simplified* to (x, y) pairs, and then (2) *overlaid* on a geospatial description generated in a remote Geoweb server.

The Geoweb emerged in 2000 with the service specification Web Map Server 1.0 [5] published by Open Geospatial Consortium (OGC). OGC has also published service specifications for data access and catalogues [12], and geographic markup languages, such as GML [13]. Nevertheless, the release in 2005 of Google Maps, which enabled an easy integration of maps to existing Web applications, made the Geoweb part of the mainstream Web.

The Geoweb is perceived from the Semantic Web community as a provider of datasets rich in geographic descriptions that need to be extracted from their silos. This is the approach of the publication as RDF of OpenStreetMap [1]. On the other side, the Semantic Web is perceived from the Geoweb community as the provider of formal infrastructure (i.e. Semantic Web standards, such as OWL). The application of Semantic Web technologies includes from enabling meaningful geospatial information retrieval using geospatial ontologies [6] to the development of profiles of Geoweb services with RDF and OWL support [7]. With Geo Linked Data, we propose an alternative approach: the use of Semantic Web best practices to publish Geoweb resources alongside their metadata in RDF.

(Draft) Lecture Notes in Computer Science (LNCS). 2010, vol. 6261, p. 495-502. ISSN 0302-9743.

3 Geo Linked Data

In the Linked Data approach, when a URI acts as identifier for an entity, which may exist outside the Web, the URI can be dereferenced to a Web resource that provides useful information about the entity [15]. The user agent works under the following assumption: when the URI gives access to a Web resource with a different URI at a given time, the Web resource could be interpreted as a *proxy for* an entity, at least in that given time. The concept *proxy for* is part of the *identity of resources and entities on the Web* (IRE) model proposed by Presutti and Gangemi [14], a framework for reasoning when a Web URI can be associated to an entity. The *proxy for* association between a Web resource (e.g. a semantic description) and an entity (e.g. Lisbon) means that the representation of the Web resource (e.g. a RDF document) materializes information (e.g. a pair of geographic coordinates) about the entity.

The IRE model classifies the *proxy for* relations as *exact* or *approximate*, and as *formal* or *informal*. An *exact proxy for* relation means that the Web resource only describes one entity, and otherwise it is *approximate*. Nevertheless, an *exact proxy* may contain references to related entities. For example, satellite images about Portugal may contain parts of the Atlantic Ocean and Spain. However, satellite images are meant to be *exact proxies for* Portugal. A *formal proxy for* relation means that the Web resource is represented in a formal language. If not, it is an *informal proxy for* relation.

The *proxy for* concept is independent of the technology and the information about the entity. Hence, the definition of *proxy for* is applicable when the Web resource is a Geoweb resource conveying spatial characteristics of other entities. We designate as a *geospatial proxy* any Web resource conveying spatial information about other entities using Geoweb standards (e.g. a GML document describing the location of Lisbon, a satellite image in JPEG describing the Earth). A *geospatial proxy* is an *exact geospatial proxy* if it only describes one entity.

Linked Data principles require that the information must be available in RDF and, for humans, should be available in HTML. We can characterize RDF and HTML representations as having *exact formal proxy for* and *exact informal proxy for* relations with an entity, respectively. For humans, an *exact geospatial proxy* and a HTML representation describing the same entity could contain equivalent spatial information. From a machine-processable point of view, an *exact geospatial proxy* could be considered as an alternative representation of the spatial information of an RDF representation when it is possible to map the content of the geospatial proxy to a formal model. The mapping is possible because the information in Geoweb representations is specified by standardized data models [9], and we can find in the literature mappings to formal models, such as the proposed in the Geospatial Semantic Web Interoperability Experiment [10]. We can conclude that *exact geospatial proxies* conform to Linked Data rules. That is, when a semantic application requires spatial information, they could be an alternative for HTML and RDF representations.

(Draft) Lecture Notes in Computer Science (LNCS). 2010, vol. 6261, p. 495-502. ISSN 0302-9743.

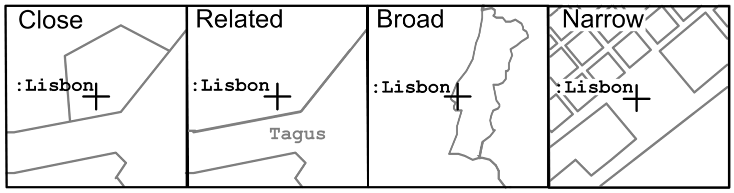


Fig. 1: Different examples of maps that may act as *geospatial proxy for* the official boundary of Lisbon, capital of Portugal: *close* is a sketch of the boundary, *related* shows the area where the boundary of Lisbon is, *broad* shows the boundary of Portugal, and *narrow* shows the *Praça do Comércio* (Commerce Square), a landmark of Lisbon.

3.1 Roles of a Geoweb description

We identify four roles that could be useful to understand how to use Geoweb descriptions in Semantic Web applications (Figure 1):

close A *close proxy* is a proxy that some information systems can use as source for an alternative identifier of the entity. For example, a marker that identifies an entity in a map provided by a GML document is a *close proxy*.

related A *related proxy* provides an indirect description of the resource through the spatial characteristics of the proxy. For example, a satellite image of an area is a *related proxy* of the entities of the area.

broad A *broad proxy* is a kind of related proxy that realizes essential characteristics of an entity that is a parent of the described entity. For example, a large regional map is a broad geospatial proxy for capital cities.

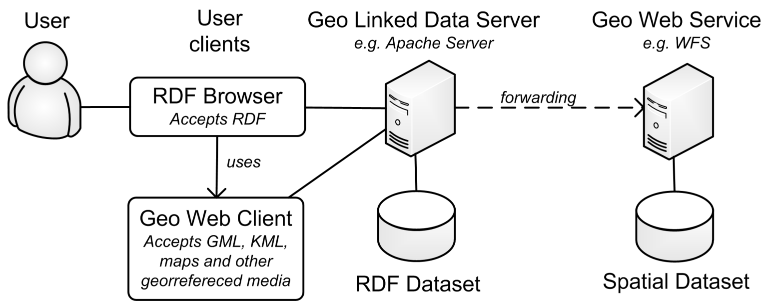
narrow A *narrow proxy* is similar to a broad proxy but realizes essential characteristics of an entity part of the described entity. An example is a map that shows a landmark of a city when the described entity is the city.

3.2 Geo Linked Data recipe

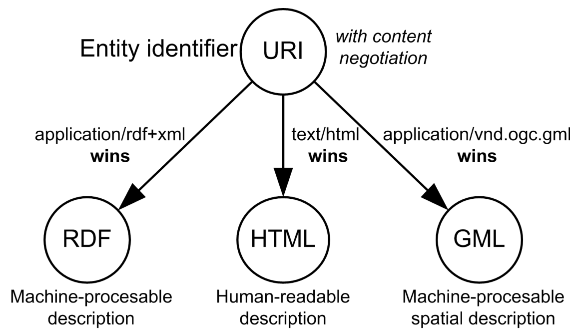
We can summarize the Geo Linked Data recipe as follows (see Figure 2b and Figure 2a):

1. Entity URIs are dereferenced following the Linked Data principles (see Bizer et al. [4]).
2. If such a URI is dereferenced accepting a Geoweb MIME-type (e.g. *application/vnd.ogc.gml*), the server must return a geospatial description of the entity that matches the MIME-type (e.g. GML) when available.
3. RDF descriptions may contain properties that advertise the presence and role of Geoweb descriptions; clients can use these clues to discover their MIME-type and the best use of these representations.

(Draft) Lecture Notes in Computer Science (LNCS). 2010, vol. 6261, p. 495-502. ISSN 0302-9743.



(a) Simple setup with a Geoweb data access service which serves geospatial resources in GML.



(b) The content negotiation model of Linked Data is extended with support to Geoweb MIME-types.

Fig. 2: Schematic representation of the relations in a simple setup. With a URI that identifies a real world resource (e.g. Lisbon), the user client can access a RDF or HTML representation through the Geo Linked Data Server in the same server, or a geospatial representation (GML) in a Geoweb service.

4. RDF descriptions may contain RDF links to navigate to geospatial descriptions provided by Geoweb services.

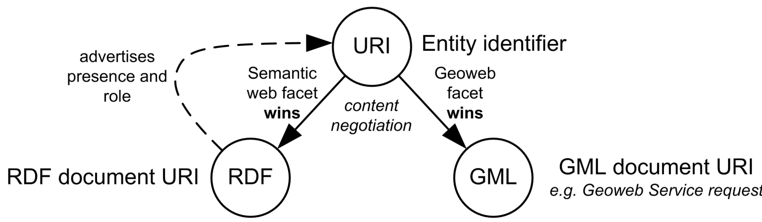
3.3 Making proxies visible

We can use RDF to advertise the *presence*, the *role* and the *location* of a geospatial proxy. We can relate a URI to a geospatial proxy adding the following statement to the RDF description of the entity:

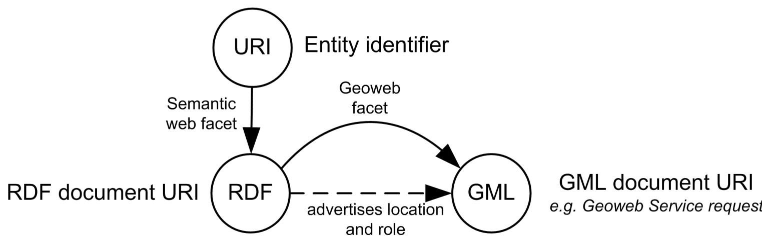
```
<entity URI> p <MIME-type> .
```

The semantics of the property `p` should entail that there is a dereferenceable geospatial proxy for the entity identified by the URI. The property `p` could also

(Draft) Lecture Notes in Computer Science (LNCS). 2010, vol. 6261, p. 495-502. ISSN 0302-9743.



(a) The document asserts presence and role of the proxy.



(b) The document asserts explicit location and role of the proxy.

Fig. 3: Advertisement of geospatial proxies in RDF documents.

describe the role of the proxy. Figure 3a shows how this assertion should be interpreted. The URI can be dereferenced again, but this time requesting the MIME-type asserted in the statement, to retrieve the geospatial representation. The server that owns the URI is responsible for redirecting the user to the effective location of the geospatial proxy. This way of advertising presence is limited to only one proxy for each content type. If the server cannot be properly configured, or the entity has several geospatial proxies with the same MIME-type, we could make an explicit advertisement of the location of each geospatial proxy. The advertisement requires the assertion of statements like:

```
<entity URI> q <Geoweb document URI> .
```

The semantics of the property `q` should assert that the Geoweb resource is a geospatial proxy, and, additionally, describe its role (see Figure 3b).

4 A Geo Linked Data scenario

Geo-Net-PT 02 [11] is an authoritative RDF dataset about named places of Portugal available in the XLDB Node of Linguateca[4]. Geo-Net-PT 02 defines 701,209 instances, most of them named places. Postal codes, streets and settlements are the most common types. Geo-Net-PT 02 has 21 different sources. The

[4] `http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT_02`

(Draft) Lecture Notes in Computer Science (LNCS). 2010, vol. 6261, p. 495-502. ISSN 0302-9743.

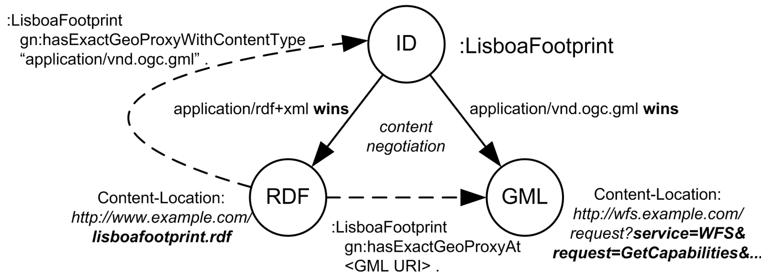


Fig. 4: Use case scenario: geospatial proxies for Geo-Net-PT 02; the prefix `gn:` identifies the terms added to the vocabulary for advertising Geo Linked Data.

Instituto Geográfico Português provides the Official Administrative Boundaries Map (CAOP) of Portugal. The CAOP dataset is available through a Geoweb data access service application [8], which makes accessible up-to-date GML documents that are *exact geospatial proxies* of the footprints of the administrative units. Figure 4 describes an example of the advertisement of presence, role and location of geospatial proxies in the Geo-Net-PT 02 dataset. The exact link is possible for administrative units. Each administrative unit in Portugal has a unique official identifier as attribute, which exists in the CAOP server and in Geo-Net-PT 02. The setup of the Geo Linked Data endpoint follows the recipe presented in Subsection 3.2. It allows clients to dereference a Geo-Net-PT 02 URI, such as `:LisboaFootprint`, and discover that a representation in GML is also available. Then, the client can dereference again but asking for the MIME-type of GML, and then, for example, display the result on a map.

5 Conclusions

This paper introduced Geo Linked Data, an approach for incorporating references to Geoweb data and services in the Semantic Web. The references, named geospatial proxies, were characterized, and the example based on Geo-Net-PT 02 showed how to use this approach in a real scenario. Geo Linked Data requires the update of RDF datasets with assertions of presence of spatial descriptions, which may include information about their role and location. The roles ease the use of Geo Linked Data by Semantic Web applications. The locations allow users to navigate between Linked Data and Geoweb datasets. Future work will include the formalization of the categorization of geospatial proxies, and the automatic generation and classification of bindings between RDF datasets and Geoweb resources.

Acknowledgments

Thanks to Aneta J. Florczyk and Walter Renteria Agualimpia for their comments and insights. This work was partially supported by FCT (Portuguese research funding agency) for its LaSIGE Multi-annual support, GREASE-II project (grant PTDC/EIA/73614/2006), by the Spanish Government (projects "España Virtual" CENIT 2008-1030 and TIN2009-10971), the Aragón Government (project PI075/08), and GeoSpatiumLab S.L.

References

1. Auer, S., Lehmann, J., Hellmann, S.: LinkedGeoData – Adding a spatial Dimension to the Web of Data. In: Proceedings of 8th ISWC (2009)
2. Becker, C., Bizer, C.: Exploring the Geospatial Semantic Web with DBpedia Mobile. Web Sem.: Science, Services and Agents on the WWW 7(4), 278–286 (2009)
3. Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., Sheets, D.: Tabulator: Exploring and Analyzing linked data on the Semantic Web. In: Proceedings of the 3rd International Semantic Web User Interaction Workshop (2006)
4. Bizer, C., Cyganiak, R., Heath, T.: How to Publish Linked Data on the Web [online] (2007), `http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/`
5. Doyle, A.: OpenGIS Web Map Server Interface Implementation Specification. OGC Standard OGC 00-028, Open Geospatial Consortium Inc. (2000), version 1.0.0
6. Egenhofer, M.J.: Toward the Semantic Geospatial Web. In: GIS '02: Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems. pp. 1–4. ACM, New York, NY, USA (2002)
7. Janowicz, K., Schade, S., Bröring, A., Keß ler, C., Maué, P., Stasch, C.: Semantic Enablement for SDIs. Transactions in GIS 14(2), 111–129 (April 2010)
8. Julião, R.P., Mas, S., Rodriguez, A., Furtado, D.: Portugal and Spain twin SDI's from national projects to an Iberian SDI. In: GSDI-11:Spatial Data Infrastructure Convergence, Rotterdam, The Netherlands 15-19 June 2009 (2009)
9. Kresse, W., Fadaie, K.: ISO Standards for Geographic Information. Springer, Berlin (2004), ISBN 978-3-54020-130-4
10. Lieberman, J.: Geospatial Semantic Web interoperability experiment report. OpenGIS® Discussion Paper OGC 06–002r1, OGC, Inc. (2006)
11. Lopez-Pellicer, F.J., Silva, M.J., Chaves, M.: Linkable geographic ontologies. In: GIR '10: Proceedings of the 6th Workshop on Geographic Information Retrieval. pp. 1–8. ACM, New York, NY, USA (2010)
12. Nogueras-Iso, J., Zarazaga-Soria, F.J., Lacasta, J., Béjar, R., Muro-Medrano, P.R.: Metadata standard interoperability: application in the geographic information domain. Computers, Environment and Urban Systems 28(6), 611– 634 (2004)
13. Portele, C.: OpenGIS Geography Markup Language (GML) Encoding Standard. OpenGIS Standard OGC 07-036, Open Geospatial Consortium Inc. (July 2007)
14. Presutti, V., Gangemi, A.: Identity of Resources and Entities on the Web. International Journal on Semantic Web and Information Systems 4(2), 49–72 (2008)
15. Sauermann, L., Cyganiak, R.: Cool URIs for the Semantic Web [online]. W3C note, W3C (Mar 2008), `http://www.w3.org/TR/2008/NOTE-cooluris-20080331/`
16. Scharl, A., Tochtermann, K.: The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)