Ecuadorian Geospatial Linked Data

Víctor Saquicela*[1], Mauricio Espinoza†[1], Nelson Piedra‡[2] and Boris Villazón Terrazas§[3]

[1]Department of Computer Science, Universidad de Cuenca
[2]Department of Computer Science, Universidad Técnica Particular de Loja
[3]iSOCO, Intelligent Software Components

January 19, 2014

Abstract

Much of the attention lately has been on geospatial data and their combination with Linked Data. In this paper we present our approach for creating the Ecuadorian Geospatial Linked Data repository, pointing out some existing efforts, and open issues we have to cope with in order to make it easier the publication and exploitation of Geospatial Linked Data.

1 Introduction

Geographical Information Systems (GIS) help geographers, topology experts and map-makers to create specific maps for special needs. Moreover, nowadays geospatial information is present in most of the aspects of our lives.

Publishing Government data on the Web as Linked Open Data is one important goal of Governments around the world. In a nutshell, these initiatives can be grouped into two categories : (1) creating government data catalogs; and (2) applications that publish government as Linked Data that provide RESTful APIs and interfaces such as SPARQL query endpoints.

In the context of geospatial domain exists diverse types of data sources (shapefiles, geo-databases, CSVs, excel spreadsheets, GML, XML, KML, etc.), which are represented in a structured manner. Geospatial data spatially describe geometries, points, lines, polygons and multipolygons. Several studies have addressed positively the problem of RDF generation from geospatial information[1]. Once this information is available in Semantic Web formats, it can be easily shared and exchanged.

With respect to geospatial linked data initiatives, several studies have addressed positively the problem of RDF generation from geospatial information

*victor.saquicela@ucuenca.edu.ec
†mauricio.espinoza@ucuenca.edu.ec
‡nopiedra@utpl.edu.ec
§bvillazon@isoco.com

[1]Locations and Address Community Group `http://www.w3.org/community/locadd/`

[5, 2, 4]. GeoLinkedData[2] is an initiative to add spatial dimension to Web of Data. LinkedGeoData[3] uses the information collected by the OpenStreetMap[4] project and makes it available as RDF. Ordnance Survey Linked Data [5] is a Great Britain's national mapping agency. Ordnance Survey publish a number of different kind of information from its products, products that can be access through a SPARQL service. National Information Institute of France[6] is working as well on publishing geospatial data on linked data format. Trentino Government Linked Open GeoData has published its data following open data principles.

Having this in mind, our aim is to create an Ecuadorian Geospatial Linked Data Repository[7], and represent the geospatial data sources in RDF. To this end, we carry out transformation of those resources by including valuable metadata information. We combine the ideas of several initiatives, and propose a new framework to deal with generation and integration of geospatial data. Moreover, we plan to use, reuse and extend the capabilities of ETL tools for generating Linked Data based on available RDF transformation libraries but adapted to geospatial domain.

The paper is organised as follows: Section 2 we present the Geospatial Linked Data Life Cycle, including some open discussions and issues. In Section 3, we present some conclusions and outline some future work.

2 Geospatial Linked Data Life Cycle

In this section we describe our proposal for generating, publishing and consuming Geospatial Linked Data. The proposed life cycle is based on the upcoming W3C Working Group Note [3], and guidelines proposed in [6]. We have extended and adapted those proposals to cover geospatial data features and needs (see Figure1).

Selection of the data sources. The initial step is to identify and select the geospatial data sources that are available in this particular context, i.e., Ecuadorian Government agencies. We have collected and created a draft version of a data catalog of geospatial data sources from Ecuador. This catalog is available at `http://geo.linkeddata.ec/sources/sources.xls`. The data sources consists of geo-databases, shapefiles, spreadsheets, and KML.

URIs for geospatial resources. Within the Web of Data, URI is a single global identification system. In [3] there are some guidelines for defining URIs for RDF resources. However in the context of geospatial data we have to consider some specific characteristic. For instance, RFC5870[8] defines an URI for Geospatial locations. There are some discussions about this within the INSPIRE Forum [9]. However, as a community we need to still reach a consensus on how to define URIs for geospatial resources, specially to points, linestrings, polygons, and multipolygons.

[2] `http://geo.linkeddata.es`
[3] `http://linkedgeodata.org/`
[4] `http://www.openstreetmap.org`
[5] `http://data.ordnacesurvey.co.uk`
[6] `http://data.ign.fr`
[7] `http://geo.linkeddata.ec`
[8] `http://tools.ietf.org/search/rfc5870`
[9] `http://inspire-forum.jrc.ec.europa.eu/pg/forum/topic/29922/cool-uris-for-linked-geospatial-data/`

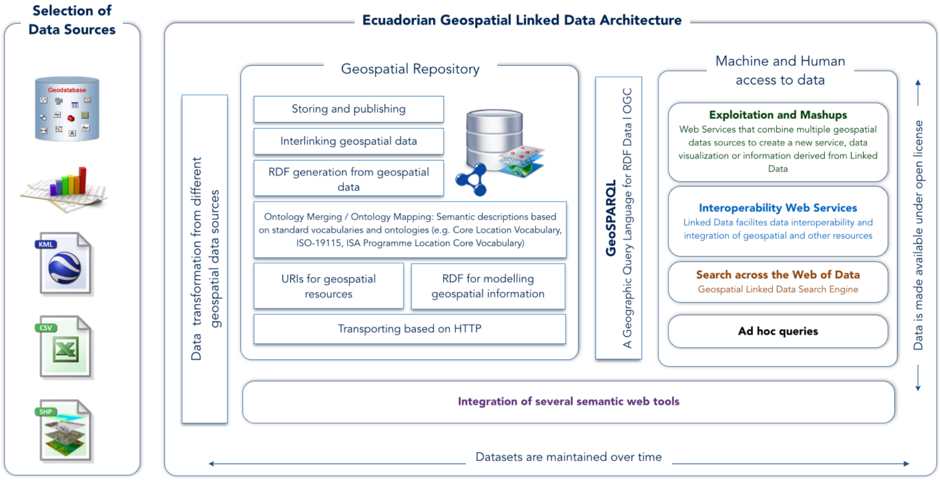

Figure 1: Architecture for the Ecuadorian Geospatial Linked Data Repository

Geospatial vocabularies. According to [3], standardised vocabularies should be reused as much as possible to facilitate inclusion and expansion of the Web of data. Concerning geospatial vocabularies most authors describe their models and benefits.

- One research work using numerous VoCamps related to geospatial data developed[10] a RDF vocabulary for modelling geospatial information.
- GeoSPARQL, a Geographic Query Language for RDF Data (GeoSPARQL) is an emerging standard from the Open Geospatial consortium (OGC). It provides a common representation of geospatial RDF data and the ability to query and filter on the relationships between geospatial entities[11].
- Core Location Vocabulary[12] is a simplified, reusable and extensible data model that capture the fundamental characteristics of a location.
- ISA Programme Location Core Vocabulary[13], provides a minimum set of classes and properties for describing any place in terms of its name, address or geometry.
- Within the `http://schema.org` there are two schemas for geospatial data, GeoShape[14], and GeoCoordinates[15].

In the context of the Ecuadorian Geospatial data we have selected the GeoSPARQL ontology because this is the standard defined in the geospatial world. However, we do not discard the reuse and extension of other geospatial vocabularies. Moreover, we claim that geospatial metadata is also important, and because

[10] http://geovocab.org/doc/neogeo/
[11] htpp://opengeospatial.org/standards/geosparql
[12] https://joinup.ec.europa.eu/asset/core_location/description
[13] http://www.w3.org/ns/locn
[14] `http://schema.org/GeoShape`
[15] `http://schema.org/GeoCoordinates`

of this, we have selected ISO-19115 ontology[16]. ISO-19115 defines the schema required for describing geographic information and services. It provides information about the identification, extent, quality, spatial and temporal schema, spatial reference, and distribution of digital geographic data.

RDF generation from geospatial data. Some of the difficulties that arise in the RDF generation of geospatial information are briefly described in [1]. There are several tools for generating RDF, however, all these tools are not integrated into a single framework, and not able to generate geo-RDF. Moreover, tools are limited to specific formats. For example, we have a relational database and want generate RDF, in this case, we can use D2RQ tool[17]; but if we have an excel file, we can use NOR2O[18].

In order to cope with some of those issues, we rely on techniques and process that allows (1) RDF generation using different tools, and (2) GeoKettle[19] plugin implementation. GeoKettle is a powerful, metadata-driven spatial ETL tool dedicates to the integration of different geospatial data sources. Using GeoKettle, we can extract data from different data sources, transform the data, perform some data cleansing, and change their structure. Our final goal is to integrate several RDF tools into a framework.

Linking of geospatial data. Interlinking in the Semantic Web is closely related with data integration in database. The aim is to bring together data from different and heterogeneous data sources, and detect whether descriptions of resources refer to the same entity in the real world. We conducted a study to generate links between RDF data sources. Linking is obtained using Silk [7], which is a tool for discovering and maintaining data links between RDF sources in the Web of Data.

Interlinking geospatial data is a process based on an algorithm for matching of similarity string between URIs and algorithms that process geospatial operations (near, distance, etc).

Publication and exploitation. The RDF generated can be loaded into a triplestore. Some triplestores support geospatial data, such as Virtuoso, and Parliament. The generated RDF need to be exploited, to make sure that the RDF are useful. We can execute several queries to validate our approach. Furthermore, on top of this repository, we intend develop a web based applications to enhance the visualisation of geospatial linked data generated.

3 Conclusions and Future Work

In this paper, we have presented our ideas for generating geospatial linked data. Pointing out some existing efforts, and open issues we have to cope with in order to make it easier the publication and exploitation of geospatial Linked Data.

In the future we plan to participate actively in the geospatial linked data community, for reaching consensus on the open questions we still have in this domain. Moreover, we plan to introduce Latin American organisations as key stakeholders in the community.

[16] `https://www.seegrid.csiro.au/subversion/xmml/metadata/ISO19115/iso-19115.owl`
[17] `http://d2rq.org/`
[18] `http://oeg-dev.dia.fi.upm.es/nor2o/`
[19] `http://www.spatialytics.org/projects/geokettle/`

References

[1] G. Atemezing and R. Troncy. Comparing vocabularies for representing geographical features and their geometry. In *ISWC 2012, 11th International Semantic Web Conference, Terra Cognita Workshop, November 11-15, 2012, Boston, USA*, Boston, UNITED STATES, 11 2012.

[2] S. Auer, J. Lehmann, and S. Hellmann. Linkedgeodata: Adding a spatial dimension to the web of data. In *Proceedings of the 8th International Semantic Web Conference*, ISWC '09, pages 731–746, Berlin, Heidelberg, 2009. Springer-Verlag.

[3] B. Hyland, G. Atemezing, and B. Villazon-Terrazas. Best Practices for Publishing Linked Data. `http://www.w3.org/TR/ld-bp/`, 2014.

[4] P. Shvaiko, F. Farazi, V. Maltese, A. Ivanyukovich, V. Rizzi, D. Ferrari, and G. Ucelli. Trentino government linked open geo-data: A case study. In P. Cudré-Mauroux, J. Heflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, J. Parreira, J. Hendler, G. Schreiber, A. Bernstein, and E. Blomqvist, editors, *The Semantic Web – ISWC 2012*, volume 7650 of *Lecture Notes in Computer Science*, pages 196–211. Springer Berlin Heidelberg, 2012.

[5] L. M. Vilches-Blázquez, B. Villazón-Terrazas, V. Saquicela, A. de León, O. Corcho, and A. Gómez-Pérez. Geolinked data and inspire through an application case. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, pages 446–449, New York, NY, USA, 2010. ACM.

[6] B. Villazón-Terrazas, L. Vilches, O. Corcho, and A. Gómez-Pérez. Methodological Guidelines for Publishing Government Linked Data. In D. Wood, editor, *Linking Government Data*, chapter 2. Springer, 2011.

[7] J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov. Discovering and maintaining links on the web of data. In *Proceedings of the 8th International Semantic Web Conference*, ISWC '09, pages 650–665, Berlin, Heidelberg, 2009. Springer-Verlag.