Integrating Linked Open Data into Open Source Web Mapping

Wiafe Owusu-Banahene[1,2], Serena Coetzee[1]

[1]Centre for Geoinformation Science, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa.
[2]Computer Engineering Department, University of Ghana, Legon, Accra, Ghana.

Abstract. The application of the linked data principles to promote the availability of open interlinked datasets has resulted in the creation of the so-called Linked Open Data (LOD) cloud, consisting of 338 datasets, of which 51 are geographic data whilst the rest are data from non geographic domain. However, presently, integrating linked open data directly into open source web mapping is a challenging task since linked data sources are outside of open source web mapping environment. Moreover, web map servers cannot consume RDF data directly. Our current research is aimed at finding novel ways of visualising linked data in the form of thematic maps over the internet on-the-fly. In this paper, we present the results of experiments with integrating non-spatial linked open data into an open source web mapping environment and visualising them as thematic maps. We show experiments of web thematic maps created with choropleth and proportional symbol techniques using our geospatial thematic web service based on open source web mapping technology. Our results show that it is possible to integrate non-spatial linked open data with traditional geospatial data using open source web mapping technology. However, access, data conversion and data integration are some of the main challenges in creating web thematic maps with existing web mapping tools from traditional geospatial data integrated with non-spatial linked open data on-the-fly.

Keywords: Linked Open Data, Geographic Data, Thematic Maps, Web Mapping

1. Introduction

Linked data enables links to be set between items in different data sources and therefore connect these sources into a single global data space (Heath &

Bizer 2010). Linked data is open, accessible and its data representations are standardized (Heath & Bizer 2010). The creation of a web of Linked Open Data (LOD) aimed at promoting the availability of open interlinked datasets has resulted in the creation of the so-called LOD cloud. This presents an advantage to data consumers since data and information schema becomes available and accessible on the Web (Fensel 2011). Presently the *LOD Cloud*[1] is made up of 338 datasets out of which 51 datasets are from the geographic domain whilst the rest are data from non geographic domain.

There is a growing demand for thematic information for a multitude of applications but thematic data sets are highly heterogeneous in syntax, structure and semantics (Durbha et al. 2009). Producing thematic maps over the Internet and the WWW could therefore take advantage of the availability of linked open data in the LOD cloud. However, presently, integrating non-spatial linked data into open source web mapping is a challenging task. Our current research is aimed at finding novel ways of dynamically visualising linked open data in the form of thematic maps. One of our goals was to develop a geospatial web service dedicated to producing thematic maps with non-spatial linked data. We called it a geospatial thematic web service. In this paper, our objective is to present results of experiments with the implementation of this geospatial thematic web service which uses open source software and tools to integrate non-spatial linked open data with geospatial data on a web server to produce web thematic maps.

We show experiments with maps created using choropleth and proportional symbol mapping techniques using our GeoServer powered WMS based geospatial thematic web service which creates thematic maps by integrating non-spatial linked open data. The latter were accessed via DBpedia's SPARQL end point and integrated into a PostGIS spatial database via SQL script. Our results show that access, data conversion and data integration are some of the main challenges in creating web thematic maps with existing web mapping tools from traditional geospatial data integrated with non-spatial linked open data. We aim to automate this process in the next stage of our research so that from a single client request thematic maps are created on-the-fly, consuming linked data, applying styling, publishing data to the web service and displaying the thematic map to the client. We review concepts and related work in section 2 and section 3 respectively. We discuss our research approach in section 4 with design and implementation of our geospatial thematic web service. We present our results in section 5.

[1] Linking Open Data Cloud: http://datahub.io/group/lodcloud?q=. Accessed 22 March 2013

Discussions, Conclusions and Future Work are presented under section 6, section 7 and section 8 respectively.

2. Review of Concepts

In this section we briefly introduce the main concepts related to our work, namely linked data, thematic maps, open source web mapping and geospatial web services.

2.1. Linked Data

Berners-Lee (2006)outlines architectural principles of linked data which have been adopted by an increasing number of data providers, leading to the creation of a global data space containing billions of assertions - the Web of Data (Bizer et al. 2009(a)).

The creation of a web of linked open data is promoted by the Linking Open Data community project[2] aimed at promoting the availability of open interlinked datasets resulting in the creation of the so called LOD cloud. The LOD cloud diagram by Richard & Jentzsch (2011) is shown in *Figure 1*. The nodes are the linked datasets and the arrows show interlinks to other datasets in the cloud.

Information about resources on the Web is represented in RDF (Breitman et al. 2010, Fensel 2011). RDF makes it possible to write statements about resources with each statement consisting of a subject, predicate and object forming a triple. Several triples form a graph.

Prud'hommeaux & Seaborne (2008) define SPARQL as the query language for RDF. Some linked data providers provide RDF dump or SPARQL endpoint for their linked datasets. An RDF dump, usually a large RDF document, contains the RDF graph which makes up the entire linked dataset but a SPARQL endpoint is an HTTP-based query service that executes SPARQL queries over the linked dataset (Hartig & Langegger 2010).

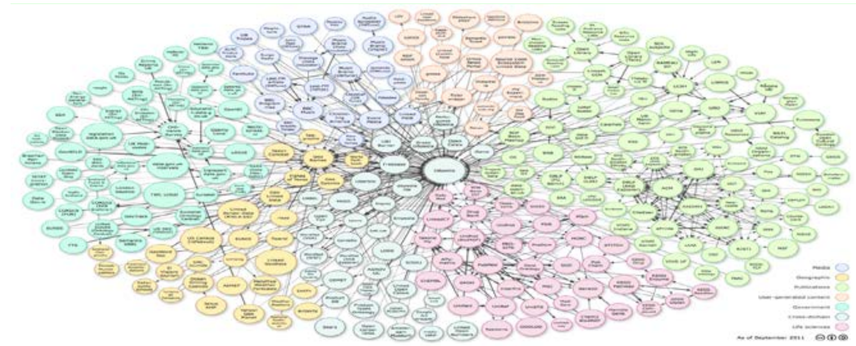
The geospatial thematic web service that we present in this paper consumes non-spatial linked data in RDF via a SPARQL end point.

[2] http:esw.w3.org/topicSweoIG/TaskForce/CommunityProjects/LinkingOpenData.

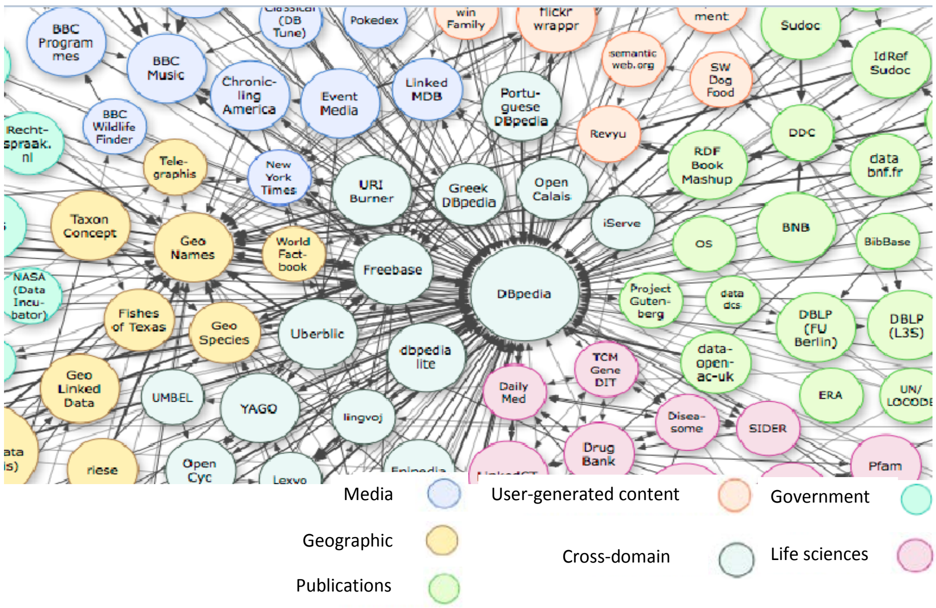
Accessed 22 March 2013

2.2. Thematic Maps

Cartography is the application of art, science and technology to make maps (Cartwright et al. 2009).Thematic cartography is a branch of cartography that deals with the production of thematic maps (Slocum et al. 201o). Thematic maps normally feature only a single distribution or relationship over a spatial background to help locate the distribution being mapped (Tyner 2010).



A) LOD Cloud diagram



B) An extract of the LOD Cloud diagram

Figure 1. LOD Cloud diagram and extract of the LOD Cloud diagram (Source: Richard & Jentzsch (2011))

Herzog (2003) observed that thematic cartography could more fully utilize the potential of the Web to communicate spatial information to larger audience and concluded that there are many possible applications of thematic mapping to the context of Web.

Choropleth and proportional symbol mapping are two of several cartographic techniques used in creating thematic maps (Slocum et al. 2010). Choropleth maps are constructed by grouping data for enumeration units (e.g. countries, states, districts) into classes and assigning either a colour or a gray tone to each class. Proportional symbol maps are created with symbols scaled in proportion to the magnitude of data arising from a particular point. In this paper we present results of experiments with web maps created using choropleth and proportional symbol techniques.

2.3. Open Source Web Mapping and OGC Web Services

A geospatial web service is a specialised type of web service that processes geospatial data. Geospatial web services are a convenient means of providing access to the large volume of geospatial data over the web (Breitman et al. 2010). In this paper, we refer to open source web mapping in the context of using open source software and tools to create geospatial web services.

The International Organisation for Standardisation's (ISO) technical committee, ISO/TC211 Geographic Information/Geomatics[3], develops standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth (Kresse & Fadaie 2010, Coetzee 2011). This includes geospatial web services. WMS is jointly published by ISO and OGC as ISO 19128:2005, Web Map Server interface and OpenGIS Web Map Service (WMS) Implementation Specification (ISO 2005). It standardises the way maps are requested over the Internet and how servers describe their data holdings (Kresse & Fadaie 2010). A WMS is defined by De la Beaujardierre (2006), as a service that produces maps of spatially referenced data dynamically from geographic information. Another OGC standard, the Styled Layer Descriptor (SLD) specifies how a WMS can be extended to allow user-defined styling (Lupp 2007).

Our geospatial thematic web service is a WMS that uses SLD to create and present thematic maps.

[3] ISO/TC 211 : http://www.isotc211.org . Accessed 22 March 2013

3. Related Work

Our geospatial thematic web service combines geographic information on a database server with non-spatial linked data and styling (SLD) to produce thematic web maps as shown in *Figure 2*. In its current implementation the GeoServer powered geospatial thematic web service combines geospatial data in PostGIS with non-spatial linked data from DBpedia. Our geospatial thematic web service is needed in cases where: 1) An existing WMS has to be migrated to using linked data, 2) Statistical agencies publish statistical data as linked data to the LOD cloud and 3) Geospatial data is too big to be accessed over the web and or it is already available locally. Moreover it is nice to create thematic maps from all the attribute data available on the web. In this section we present examples of efforts in representing and querying linked data in the geospatial context and explain how they are related to our work. We also highlight DBpedia as a source of non spatial linked data source for our geospatial thematic web service.

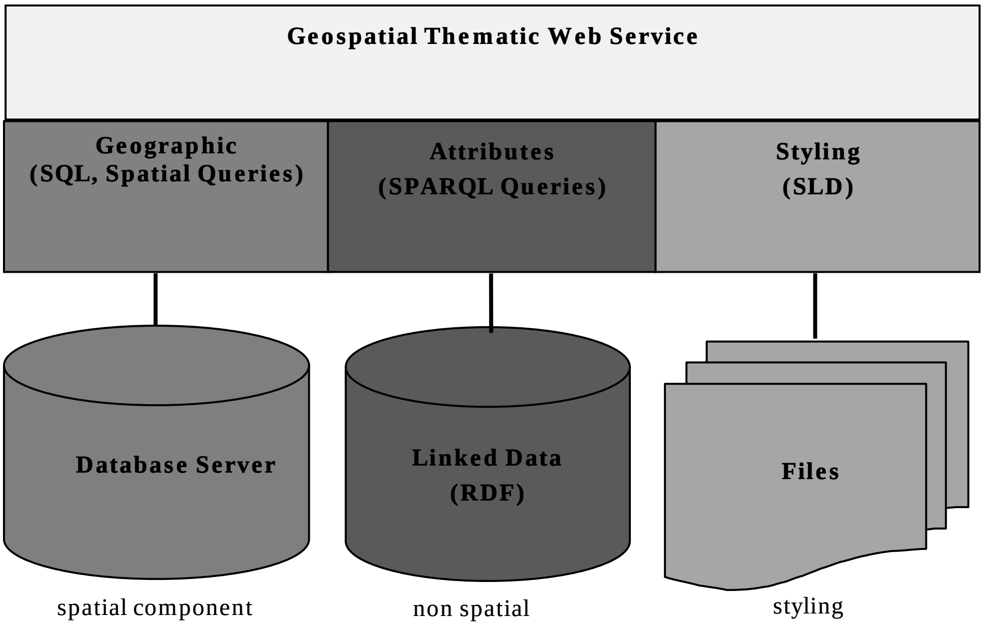


Figure 2. Spatial, non spatial (attributes) and styling components of the geospatial thematic web service.

3.1. Representing Geospatial Linked Data

W3C Geospatial Incubator Group (GeoXG)[4] published: Geospatial Vocabulary defines a basic ontology and OWL vocabulary for representing geospatial properties for Web resources (Lieberman et al. 2007(a)). This report presents a model for basic feature properties of Web resources and a realization of these feature property elements as XML and OWL/RDF vocabularies. Geospatial Ontologies, another report by Lieberman et al. (2007(b)), provide description of geospatial foundation ontologies that can be used to represent geospatial concepts and properties on the worldwide web. GeoRDF[5] is an RDF compatible profile for geospatial information. It defines profiles for points, lines and polygons. Our geospatial thematic web service in its current implementation consumes non-spatial linked data (attribute data) from the LOD cloud. In the next phase of our research we plan to integrate geospatial linked data modelled according to the Geospatial Vocabulary, Geospatial Ontologies and GeoRDF into our geospatial thematic web service.

3.2. Querying Geospatial Linked Data

Recently, OGC has published a standard, OGC GeoSPARQL which supports representing and querying geospatial data on the Semantic Web (Perry 2012). GeoSPARQL defines a vocabulary for representing geospatial data in RDF. It also defines an extension to the SPARQL query language for processing geospatial data. We do not need GeoSPARQL for the geospatial thematic web service in its current implementation because we are using only non-spatial attribute data from the LOD cloud.

3.3. DBpedia

DBpedia[6]: The DBpedia is a crowd-sourced community effort aimed at extracting structured data from Wikipedia. The DBpedia knowledge base consists of 1.89 billion pieces of information (RDF triples) out of which 400 million were extracted from the English edition of Wikipedia and 1.46billion extracted from other language editions. An increasing number of linked data providers have set data-level links to DBpedia resources, making DBpedia a central interlinking nucleus for the LOD cloud (Bitzer et al. 2009 (b)). Our geospatial thematic web service is a web service that con-

---

[4] www.w3.org/2005/Incubator/geo/. Accessed 22 March 2013

[5] www.w3.org/wiki/GeoRDF#Implementation. Accessed 22 March 2013

[6] http://wiki.dbpedia.org/DBpediaMobile. Accessed 22 March 2013

sumes non spatial linked data (RDF) from DBpedia and creates web thematic maps.

3.4. map4rdf

The GeoLinkedData Initiative developed map4rdf[7], a mapping and faceted browsing tool which can be configured with a SPARQL endpoint to provide exploration and visualisation of RDF resources enhanced with geometric information. map4rdf provides geospatial and geometrical visualisation using Google Maps and Open Street Maps. map4rdf provides visualization tools for a single source of geospatial RDF data, while our geospatial thematic web service links non-spatial linked data with geospatial data to produce thematic maps.

3.5. Summary

Integrating non-spatial linked data with geospatial data for web based thematic mapping has not received much attention to date. We follow this novel approach of integrating non-spatial linked data in an open source web environment to produce thematic maps.

4. Creating a Geospatial Thematic Web Service

In this section we show design and implementation of our geospatial web service that consumes non-spatial linked data from the LOD cloud.

4.1. Selecting State-of-the-art Open Source Tools

The following software packages were selected based on our previous research work (Owusu-Banahene & Coetzee 2012) and review of literature: GeoServer, PostgreSQL and PostGIS. GeoServer was required to create a WMS, to provide support for styling through SLD and SLD extensions, and to create a direct connection to the PostGIS database. PostGIS is a spatial extension to the PostgreSQL database management system.

4.2. Design and Implementation of the Geospatial Thematic Web Service

Our design of the geospatial thematic web service consisted of two main components: the web mapping environment and the linked open data access-and-integration mechanism. *Figure 3* is a high level architecture showing the main components of the geospatial thematic web service. The linked

[7] http://oegdev.dia.fi.upm.es/projects/map4rdf/. Accessed 22 March 2013

open data access-and- integration mechanism is shown in red. GeoServer with PostgreSQL and PostGIS spatial database provided the open source web mapping environment and support for WMS and SLD. Requests for thematic maps were made via a client such as a web browser. A map request was handled by the WMS which returned a thematic map based on the data published to it from the spatial database and styles (classification and symbolisation via SLD) associated with the data. The SPARQL end point exposed linked open data which were queried and stored temporarily as CSV files. SQL scripts were the mechanism through which linked data were fed into the web service environment.

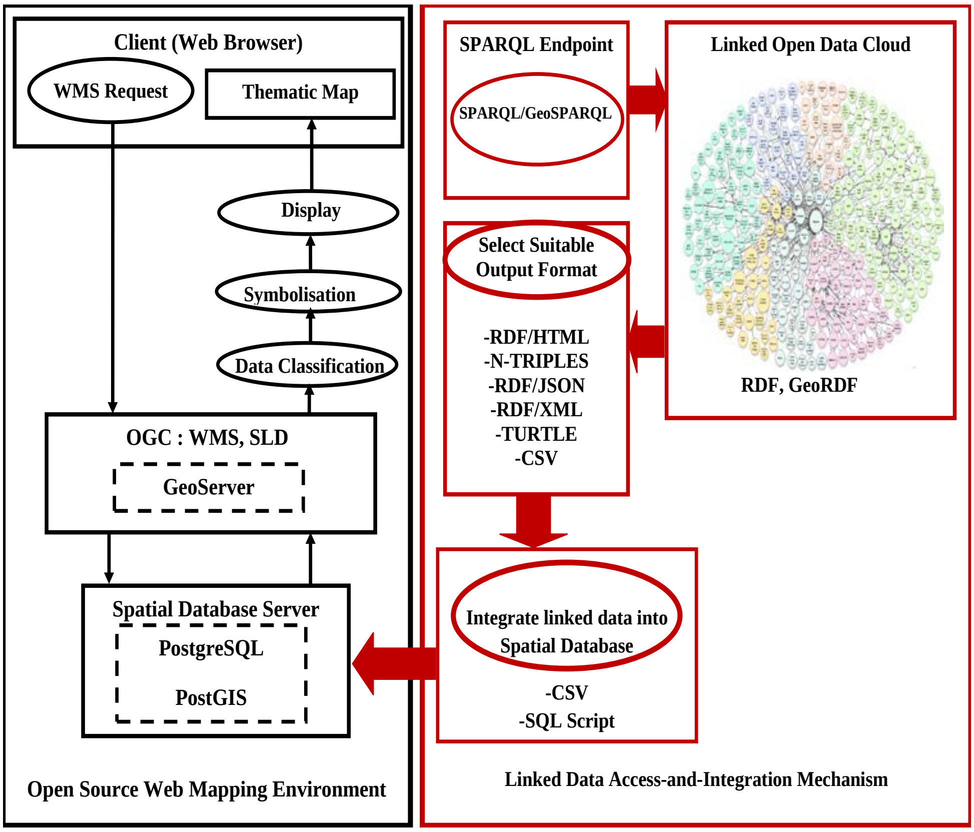


Figure 3. High level architecture of the geospatial thematic web service.

5. Producing Thematic Maps from Linked Data

In this section we present results of experiments with thematic maps created using non-spatial linked data from DBpedia[8] which is the nucleus of the LOD cloud. We show resulting thematic maps with two different thematic mapping techniques - choropleth and proportional symbols.

5.1. Accessing Linked Open Data

SPARQL queries were executed against DBpedia's SPARQL endpoint to retrieve names and population density per square kilometer of all land-locked countries. The results of the SPARQL query were stored in CSV format to allow for integration into the PostgreSQL database via SQL script.

5.2. Integrating Linked Data into PostGIS and Publishing to WMS

A spatial database in PostGIS called LOD with a table (World_Countries) containing the names of countries and their geometric data was created. *Figure 4* is a GetMap response from the WMS showing the WorldCountries_LandLocked layer. An SQL script (.sql file) was written to import the attribute data and to join based on country name to the geospatial data. The PostGIS spatial database (LOD) was connected to GeoServer. The table WorldCountries_LandLocked was then published as a new layer in WMS.

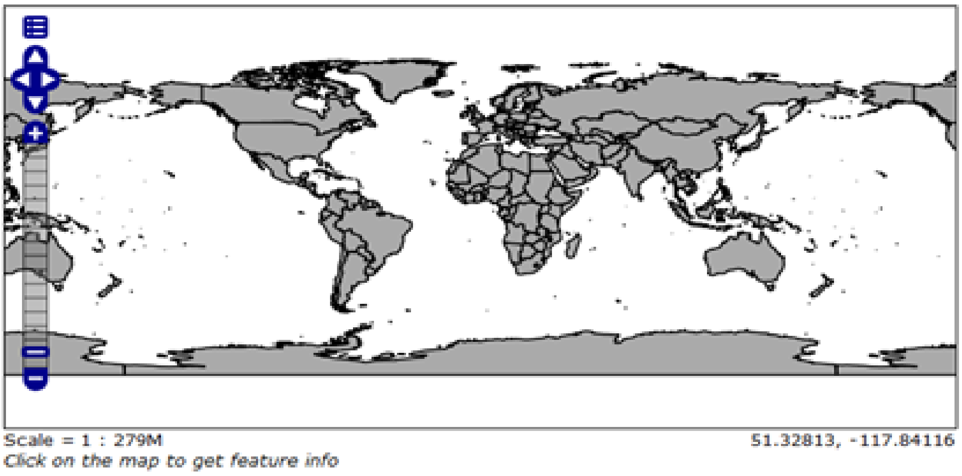


Figure 4. A GetMap response showing the WorldCountries_LandLocked layer.

[8] http://wiki.dbpedia.org/Interlinking. Accessed 22 March 2013

5.3. Creating Choropleth Maps

A style (SLD file) was created to produce a choropleth map showing landlocked and non-landlocked countries. Different colours were assigned to polygons based on attribute data classification. *Figure 5* is the resulting choropleth map showing landlocked and non-landlocked countries after applying the style to the WorldCountries_LandLocked layer. Another SLD file was created and applied to the same WorldCountries_LandLocked layer. A new choropleth map showing the population density (per square kilometre) of landlocked countries resulted (as shown in *Figure 6*).

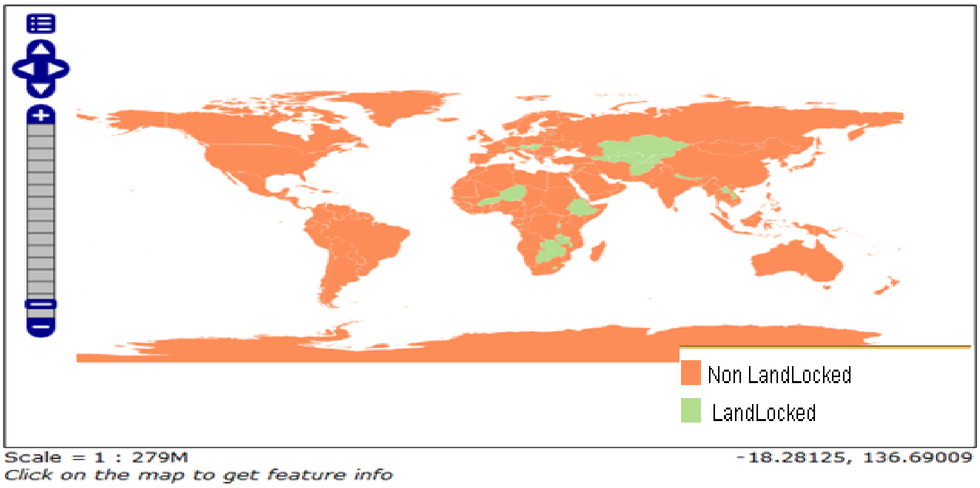


Figure 5. Choropleth map showing landlocked and non-landlocked countries

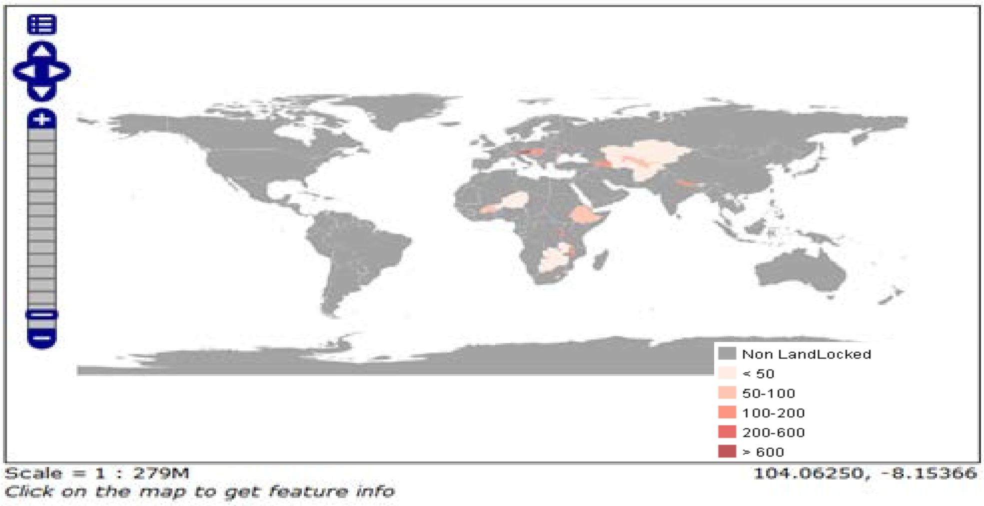


Figure 6. Choropleth map showing population density of landlocked countries.

5.4. Creating Proportional Symbols Map

Another SPARQL query was formulated to retrieve the name of a country and its nominal GDP per capita from DBpedia. In this example we applied different sizes of point based symbol (circle) to distinguish between different GDP per capita ranges: the size of a circle for a particular class is proportional to the GDP per capita. *Figure 7* shows the proportional symbols map showing nominal GDP per capita of countries in the world.

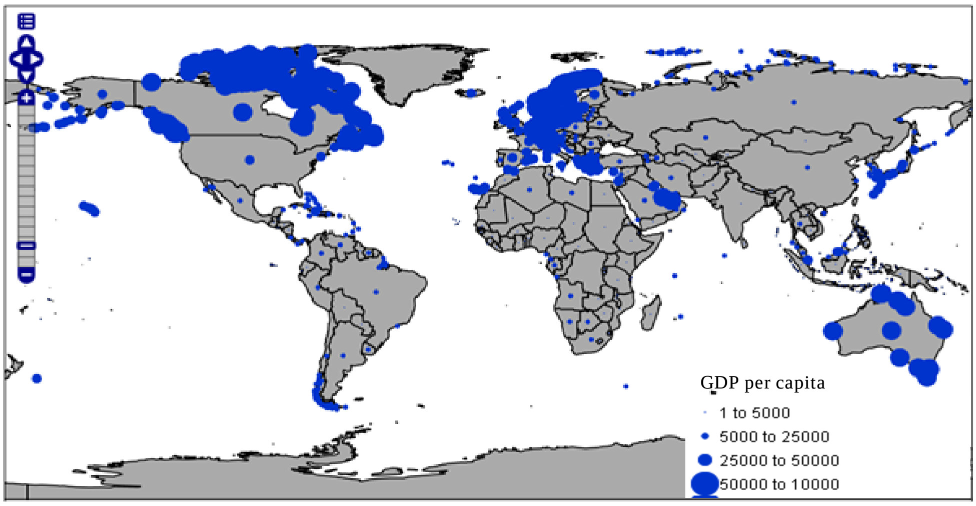


Figure 7. Proportional symbols map showing nominal GDP per capita of countries in the world.

6. Discussion

Our research brings to the fore some of the challenges involved in integrating non-spatial linked open data into existing web mapping tools. Linked open data can be accessed through a SPARQL end point or RDF dump. These sources are not part of the typical open source web mapping environment. Access, data conversion and data integration are some of the main challenges in creating thematic maps with linked data on-the-fly since SPARQL end point cannot be accessed directly from the web mapping environment. Moreover, web map servers cannot consume RDF data directly.

In order to overcome these challenges some middleware is required between the web map server that hosts the geospatial data and the non-spatial linked open data. Other challenges have to do with standardising data and creating classes on-the-fly. With the current implementation of our geospatial thematic web service, each thematic map requires a SPARQL query to be formulated. From a user's perspective, formulating SPARQL queries could be a daunting task. The next phase of our research is expected to provide solutions so that users can access linked data without having to formulate SPARQL queries themselves.

7. Conclusion

The availability of large volumes of non-spatial linked data over the web presents an opportunity for geospatial web services. The integration of non-spatial linked data into an open source geospatial thematic web service to create thematic maps, as presented by this research, demonstrates a novel approach to take advantage of the enormous opportunities that the Web of data presents to the geospatial community. The results of our experiments show that it is possible to create thematic maps from linked open data in the linked open data cloud but that it is a cumbersome process. Access, data conversion and data integration are some of the main challenges in creating thematic maps with linked data on-the-fly from the web mapping environment. In order to overcome these challenges there should be a bridge between object-relational database and linked open data.

8. Future Work

We aim to automate the process so that from a single client request, thematic maps can be created on-the-fly; consuming linked data (RDF and GeoRDF), applying styling, publishing data to the web service and displaying the thematic map to the client. For example, a client's request could be converted directly to SPARQL and GeoSPARQL queries and the results of those queries processed by the geospatial thematic web service at the backend with styles and presented back to the client as a thematic map.

Acknowledgements

The research in this paper is supported by a German Academic Exchange Service (DAAD)/African Network of Scientific and Technological Institutions (ANSTI) *In-Region* Scholarship Programme for PhD studies.

References

Berners-LeeT (2006) Linked Data-Design Issues. W3C,2009 (first published in 2006). http://www.w3.org/DesignIssues/LinkedData.html. Accessed 22 March 2013

Bizer C, Heath T, Berners-Lee T (2009(a)) Linked Data - the Story so far. International Journal on Semantic Web and Information Systems (IJSWIS). 5(3):1-22

Bizer C, Lehmann J, Kobilarova G, Auerb S, Becker C, Cyganiak R, Hellmann S (2009 (b)) DBpedia - A crystallization point for theWeb of Data. Web Semantics: Science, Services and Agents on the WorldWideWeb. 7 (2009): 154–165

Breitman KK, Casanova AM, Truszkowski W (2010) Semantic Web: Concepts, Technologies and Applications. Springer, London

Cartwright W, Gartner G, Meng L,Peterson PM(Eds) (2009) Cartography and Art. Lecture Notes in Geoinformation and Cartography. Springer, Berlin Heidelberg

Coetzee S (2011) Results from a Normative Dependency Analysis of Geographic Information standards.Computer, Standards & Interfaces 33 (2011):485–493.

Cyganiak R, Jentzsch A (2011) Linked Open Data Cloud Diagram. http://lod-cloud.net. Accessed 22 March,2013

De la Beaujardiere J (2006) OpenGIS® Web Map Service Implementation Specification. OGC. OGC® 06-042,Version 1.3.0.

Durbha SS, King RL, Shah VP, Youman NH (2009) A Framework for Semantic Reconciliation of Disparate Earth Observation Thematic Data. Computers & Geosciences 35 (2009): 761–773

Fensel D(Ed) (2011) Foundations for the Web of Information and Services. Springer-Verlag, Berlin Heidelberg

Hartig O,Langegger A (2010) Database Perspective on Consuming Linked Data on the Web. Datenbank Spektrum (2010) 10: 57–66. Springer-Verlag DOI 10.1007/s13222-010-0021-7

Heath T, Bizer C (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool

Herzog A (2003) Developing cartographic applets for the Internet. In Peterson MP (Ed)(2003) Maps and the Internet. Elsevier, Oxford

International Organization for Standardization (ISO) (2005) ISO 19128:2005 Geographic information -- Web Map Server Interface. ISO. Geneva, Switzerland

Kresse W, Fadaie K (2010) ISO Standards for Geographic Information. Springer-Verlag, Berlin Heidelberg

Lieberman J, Singh R, Goad C (2007(a)) W3C Geospatial Vocabulary. W3C. http://www.w3.org/2005/Incubator/geo/XGR-geo. Accessed 22 March,2013

Lieberman J, Singh R, Goad C (2007(b)) W3C Geospatial Ontologies. W3C. http://www.w3.org/2005/Incubator/geo/XGR-geo-ont.Accessed 22 March,2013

Lupp M(Ed) (2007) Styled Layer Descriptor Profile of the Web Map Service Implementation Specification. OGC. OGC® 05-078r4, Version1.1.0

Owusu-Banahene W, Coetzee S (2012) An Evaluation of the Suitability of Three Open Source Map Servers for their Suitability in Setting up a Geospatial Thematic Web Service. GISSA Ukubuzana 2012 Conference, Ekurhuleni, South Africa, 2-4 October, 2012

Perry M, Herring J (Eds) (2012) OGC GeoSPARQL - A Geographic Query Language for RDF Data.OGC® Implementation Standard.OGC. OGC 11-052r4, Version: 1.0

Prud'hommeaux E, Seaborne A (2008) SPARQL Query Language for RDF. W3C Recommendation, Jan. 2008. http://www.w3.org/TR/rdf-sparql-query/. Accessed 22 March,2013

Slocum TA, McMaster RB, Kessler FC, Howard HH (2010) Thematic Cartography and Geovisualisation. Pearson Prentice Hall,Upper Saddle, NJ

Tyner JA (2010) Principles of Map Design. The Guilford Press, New York, London