

# Datos enlazados de información geográfica del Ecuador

Víctor Saquicela\*  
Fernando Baculima\*  
Lucia Lupercio\*\*

*Recibido 14 de octubre de 2018, aceptado 07 de noviembre de 2018*

## Resumen

Actualmente, la gran cantidad de información geográfica generada por diferentes instituciones exige que se propongan nuevas soluciones de gestión y publicación de estos datos. En Ecuador, la heterogeneidad de la información geográfica generada por diversos entes y publicada en la Web no es una excepción. Con el objetivo de dar solución a esta problemática, el proyecto Geo Linked Data Ecuador integra información geográfica proveniente de diversas fuentes y la publica en formato RDF bajo estándares y recomendaciones de la comunidad de la Web Semántica, creando así el repositorio de datos enlazados geográficos del Ecuador.

Palabras clave: *Datos enlazados geográficos, Web semántica, GeoRDF.*

## Resumo

Atualmente, a grande quantidade de informação geográfica gerada por diferentes instituições exige que se propague novas soluções de gestão e publicação desses dados. No Equador, a heterogeneidade da informação geográfica gerada por diversas entidades e publicada na Web não é uma exceção. Com o objetivo de solucionar essa problemática, o projeto *Geo Linked Data* Equador integra informação geográfica proveniente de diversas fontes e a publica em formato RDF por meio de padrões e recomendações da comunidade Web Semântica, criando assim o repositório de dados geográficos vinculados ao Equador.

Palavras chave: *Dados vinculados geográficos, Web semântica, GeoRDF.*

\* Departamento de Ciencias de la Computación, Universidad de Cuenca, Cuenca, Ecuador  
{victor.saquicela; fernando.baculima }@ucuenca.edu.ec

\*\* Facultad de Ciencias Agropecuarias, Universidad de Cuenca, Cuenca, Ecuador  
lucia.lupercio@ucuenca.edu.ec

## Abstract

Nowadays, different institutions generate a large amount of geographical information, which require some new management and publication solutions to be proposed. In Ecuador, the heterogeneity of geographic information generated by different entities and published on the Web is not the exception. Aiming to solve this problem, the Geo Linked Data Ecuador project integrates geographic information from different sources and publishes it in RDF format, under the standards and recommendations of the Semantic Web community, as result, the Ecuadorian Geographic Linked Data Repository has been created.

Key words: *Geo linked data, Semantic web, GeoRDF.*

## Introducción

En Ecuador existen varias instituciones del gobierno que generan y publican grandes cantidades de información geográfica (IG) referentes a una misma o diferentes temáticas (Lupercio *et al.*, 2015). A esto se le une el hecho de que las instituciones al ser autónomas utilizan sus propios modelos de datos en los procesos de generación y publicación, dando como resultado diferentes niveles de heterogeneidad en la información. Por lo tanto, se crea la necesidad de sistemas que faciliten el análisis, integración, interpretación, publicación y explotación de estos datos.

Debido a la gran cantidad de datos que se generan diariamente, varios trabajos relacionados con la gestión de los mismos se dirigen hacia la aplicación de Linked Data y la Web Semántica (Berners-Lee *et al.*, 2001). En este contexto, la utilización de RDF (*Framework* de Descripción de Recursos) permite el intercambio de datos en la Web Semántica, brindando facilidades para el establecimiento de las relaciones entre los datos provenientes de diferentes fuentes (RDF-WG, 2014). En el caso particular de la iniciativa Geo Linked Data, la adaptación de estas propuestas permite generar y publicar RDF de datos geográficos a través de tecnologías de la Web Semántica (Vilches-Blázquez *et al.*, 2010).

Durante los últimos años, en Ecuador surge la iniciativa denominada “Geo Linked Data Ecuador”,<sup>1</sup> vinculada a un proyecto que nació en el año 2015, cuyo objetivo es integrar información proveniente de diferentes fuentes de datos de IG del país para ser representados y publicados utilizando las tecnologías y recomendaciones de la Web Semántica. Además, esta iniciativa permitió el desarrollo de herramientas de software para llevar a cabo el proceso de generación, publicación y explotación de datos geográficos en el contexto de la mencionada Web.

<sup>1</sup> <<http://linkedata.ec/>>.

Adicionalmente, “Geo Linked Data Ecuador” basa su proceso de generación, publicación y explotación de RDF de información geográfica (GeoRDF) en los principios Linked Data (Bizer, 2011) y el ciclo de vida de Linked Data propuesto en (Villazón-Terrazas *et al.*, 2011) utilizando estándares y herramientas recomendadas por la comunidad.

### Proceso de generación, publicación y explotación de Geo Linked Data del Ecuador

El proceso de generación, publicación y explotación de Geo Linked Data del Ecuador se lleva a cabo a través de diversas actividades, reflejadas en la Figura 1, donde se utilizan fuentes de datos de diferentes instituciones ecuatorianas.

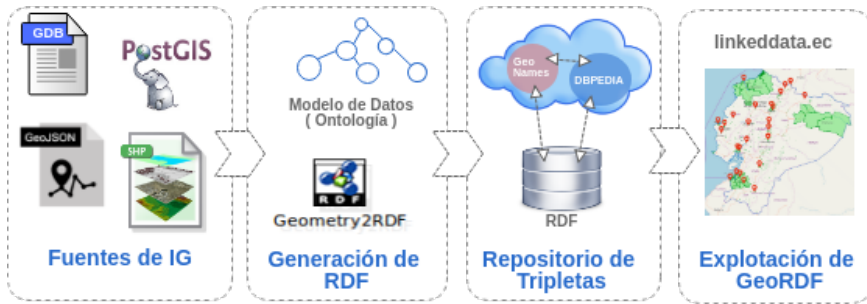


Figura 1. Proceso de generación, publicación y explotación de Geo Linked Data del Ecuador.

#### Fuentes de información geográfica

El análisis de las fuentes de datos permite obtener un conocimiento claro de la IG existente en diferentes instituciones en el Ecuador, así como de los medios para acceder a ellas y los formatos en los cuales se encuentran disponibles. A través de buscadores Web, se pueden encontrar servicios Web, archivos CSV y *shapefiles* alojados en diferentes repositorios de instituciones públicas del Ecuador con datos geográficos.

Los archivos de tipo *shapefile* resultan entre los formatos de datos con mayor presencia en la distribución de IG del Ecuador. Entre la información disponible se encuentran entidades espaciales con diversas geometrías (puntos, líneas y polígonos) referente a información de provincias, ciudades, instituciones educativas, ríos, lagos, entre otras. De esta manera, para la iniciativa Geo Linked Data Ecuador se abordan más de 80 entidades (*features*) con IG del país.

Algunas de las principales instituciones de las que proceden estos conjuntos de datos en el Ecuador son las siguientes:

- Instituto Geográfico Militar (IGM). La agencia que se encarga de planificar y controlar las actividades relacionadas con la cartografía ecuatoriana.
- Secretaría del Agua del Ecuador (SENAGUA). Genera y publica información geográfica de lugares de interés para la ejecución y desarrollo de proyectos relacionados con el manejo y control del agua en el país.
- PROMAS. Este centro de I+D+i de la Universidad de Cuenca, ejecuta proyectos de investigación relacionados con el manejo de los recursos hídricos del país.

Tras la obtención de los conjuntos de datos de los diversos productores mencionados, se concluye que, en Ecuador, no existe un repositorio centralizado de IG, debido a que cada institución genera su información y la almacena en sus repositorios de manera autónoma con el inconveniente de la heterogeneidad que se produce.

### **Generación de RDF**

Para la generación de GeoRDF resulta necesario definir modelos ontológicos, los cuales permiten representar el conocimiento de un dominio. Las ontologías definen conceptualizaciones (conceptos y relaciones) de un dominio determinado, de forma compartida y consensuada, donde su representación debe ser realizada de una manera formal, legible y utilizable por los computadores (Gruber, 1993). En este trabajo específicamente se optó por la utilización de modelos ontológicos que permitan la descripción formal de la IG para alcanzar la estandarización, reutilización e interoperabilidad de la IG del Ecuador.

En este proceso se utilizan herramientas como Protégé,<sup>2</sup> que permite conformar el modelo ontológico del dominio geográfico considerado en Geo Linked Data. Para ello, se reutilizan recursos ontológicos como GeoSPARQL (Perry y Herring, 2011), hydrOntology (Vilches-Blázquez *et al.*, 2009) y WGS84.<sup>3</sup> Estos modelos son poblados utilizando la herramienta GeoKettle,<sup>4</sup> que permite realizar procesos ETL (*Extract, Transform and Load*) desde fuentes de datos geográficas hacia repositorios RDF a través de componentes creados específicamente para cubrir los procesos de generación, publicación y explotación de GeoRDF, y son consultados utilizando el lenguaje de consulta *SPARQL Protocol and RDF Query Language* (SPARQL).

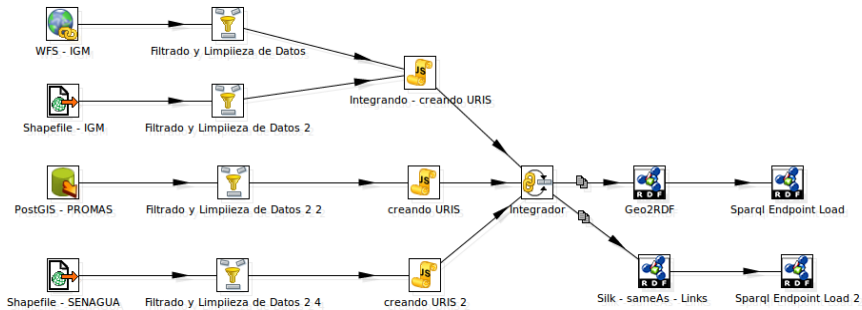
Con la definición del modelo ontológico se procede a generar el RDF a partir de las diferentes fuentes de información del Ecuador. El proceso de generación de RDF se ejecuta mediante la utilización del *framework* Geo-Kettle (software basado

<sup>2</sup> <<https://protege.stanford.edu/>>.

<sup>3</sup> <<https://www.w3.org/2003/01/geo/>>.

<sup>4</sup> <<http://www.spatialytics.org/projects/geokettle/>>.

en Pentaho Integrator). Como se mencionó anteriormente, Geo-Kettle permite realizar procesos ETL, de una manera amigable para el usuario, configurar procesos de extracción, preprocesamiento, integración y transformación de datos de diferentes fuentes. En el marco de este trabajo, se utilizaron componentes para extraer los datos de las fuentes de IG y otros que se desarrollaron específicamente para el tratamiento de IG y la creación y manejo de RDF.



**Figura 2.** Proceso de generación de RDF sobre la plataforma Geo-Kettle.

Así, en la Figura 2 se muestra la configuración del proceso de transformación sobre la plataforma Geo-Kettle para la generación de RDF a partir de diferentes fuentes de IG. El proceso inicia con la utilización de los componentes de extracción de datos que posee Geo-Kettle de las fuentes de IG. En este caso, un WFS (“WFS – IGM”), dos *shapefile* (“shapefile – IGM”), una base de datos PostGIS (“PostGIS – PROMAS”).

A continuación, los datos extraídos de cada fuente son procesados con el objetivo de poseer datos de calidad, mediante la aplicación de técnicas de preprocesamiento de datos, a través de los componentes de filtrado y limpieza proporcionados por Geo-Kettle. Siguiendo las recomendaciones de la propuesta metodológica adoptada en este trabajo (Villazón-Terrazas *et al.*, 2011), se procede a crear las URI (*Uniform Resource Identifiers*) para cada fuente de datos e integrar todas ellas.

Posteriormente, basado en las ontologías utilizadas se procede a realizar el emparejamiento correcto entre los datos integrados y las ontologías utilizadas en el proceso de generación. Este emparejamiento y la consiguiente generación de RDF son posibles gracias al desarrollo del componente “Geo2RDF”. A continuación, se procede a detectar posibles recursos similares, para ello se desarrolla el componente “Silk-sameAs-Links”, el cual permite buscar recursos RDF similares en la nube de datos enlazados.

Finalmente, se procede a publicar los datos generados en formato RDF en *triple store* (base de datos semántica) para posteriormente ser explotada. Para ello, se desarrolla el componente “*SPARQL Endpoint Load*” que permite cargar el RDF a un repositorio semántico.

### **Publicación de GeoRDF**

Los datos generados en formato GeoRDF son cargados a un repositorio semántico, que permite la consulta de los mismos a través de GeoSPARQL, donde se pueden utilizar operadores espaciales como: intersecciones, *buffers*, vecindades, entre otras.

El repositorio semántico en el que se encuentran almacenados los datos de la iniciativa “Geo Linked Data Ecuador” es Apache Marmotta.<sup>5</sup> La elección de este repositorio se debe a la implicación de esta iniciativa en el desarrollo de este proyecto de Apache, ya que a través de los esfuerzos de Geo Linked Data Ecuador se ha contribuido activamente al soporte de los datos geográficos en el contexto de la Web Semántica a través de GeoSPARQL.

La iniciativa “Geo Linked Data Ecuador” refleja los resultados en el portal <<http://linkeddata.ec>> donde se encuentra información referente al proceso de generación de Geo Linked Data, acceso al SPARQL Endpoint, visualizador de mapas, contactos para soportes, links de descargas para las herramientas desarrolladas y artículos publicados, entre otros aspectos.

### **Explotación de datos**

Los datos cargados al repositorio semántico son explotados utilizando un visualizador de mapas, el cual permite consultar toda la información generada (GeoRDF). La Figura 3 muestra un ejemplo de cómo el visualizador presenta la información publicada asociada con su representación geométrica.

El visualizador desarrollado dispone de un panel izquierdo que permite navegar sobre los recursos de los cuales se ha generado RDF (capas de información de los diversos productores mencionados con anterioridad) y el panel principal muestra los recursos graficados en un mapa a través de OpenStreetMap.

Cabe mencionar que el proceso de generación al estar basado en modelos únicos permite, de manera automática, integrar las diferentes fuentes de IG, es decir, si una fuente de datos de una institución “X” provee información de ciudades del Ecuador y una institución “Y” provee también información de ciudades del Ecuador, por el hecho de que las fuentes están modeladas de manera común, a través de una ontología, permitiría extraer todos los datos, de manera conjunta, de las dos instituciones para una ciudad específica. De esta manera, se logra un proceso de integración

<sup>5</sup> <<http://marmotta.apache.org/>>.

de los conjuntos de datos, se resuelven los problemas de heterogeneidad presente en los diferentes modelos de datos y se enriquecen los datos con la información presente en las distintas fuentes.

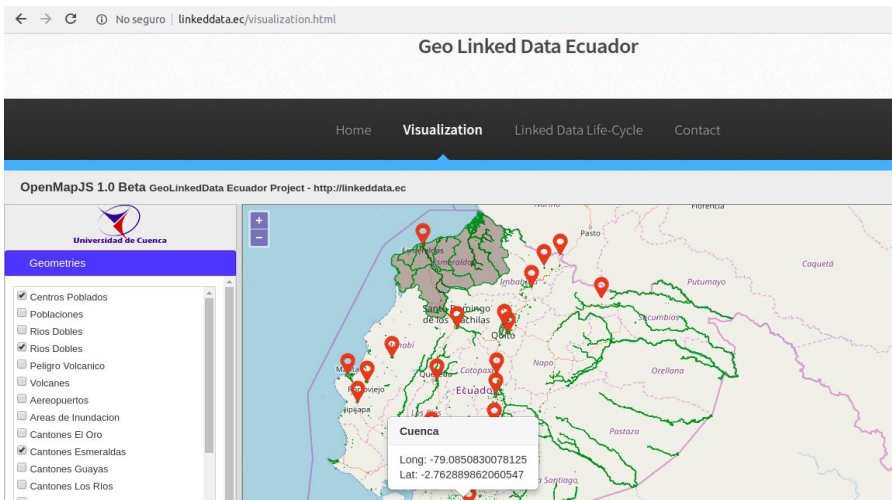


Figura 3. Visualizador de Geo Linked Data Ecuador.

### Conclusiones

En este trabajo se presenta un proceso para la generación de RDF de fuentes de información geográfica, enfocando la investigación en los esfuerzos científicos actuales y desafíos abiertos en este dominio. Además, se han creado componentes que permiten tener herramientas que faciliten la publicación y explotación de datos geográficos enlazados.

El proceso descrito permitió generar un repositorio semántico de información geográfica del Ecuador (*geo.linkeddata.ec*), el cual puede ser usado por instituciones interesadas en detectar fuentes de datos similares, así como por científicos interesados en proyectos de integración de datos, etc.

Es necesario continuar con la investigación, por lo que se pretende mejorar el proceso de generación a través de la automatización del proceso de emparejamiento entre la fuente de datos y la ontología utilizando recursos externos. Además, se pretende experimentar con fuentes de información geográfica en otros formatos, tales como *Geography Markup Language (GML)*, *Keyhole Markup Language (KML)*, etc.

## Bibliografía

- Berners-Lee, T.; Hendler, J. and Lassila, O. (2001). "The semantic web" *Sci. Am.*, 284(5), 28-37.
- Bizer, C.; Heath, T. and Berners-Lee, T. (2011). "Linked Data: The Story so Far", en *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. IGI Global, 205-227.
- Gruber, T.R. (1993). "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition*, 6(2), 199-221.
- Lupercio, L.; Baculima, F.; Espinoza, M. and Saquicela, V. (2015). "Explotación de información en el dominio geo-hídrico ecuatoriano utilizando tecnología semántica", en *Maskana*, 6 (Supl.), 69-77. Recuperado a partir de <<https://publicaciones.ucuenca.edu.ec/ojs/index.php/maskana/article/view/699>>.
- Perry, M. and Herring, J. (2011). "OGC GeoSPARQL - a geographic query language for RDF data". Recuperado a partir de: <<http://www.opengeospatial.org/standards/geosparql>>.
- RDF Working Group. (2014), "Resource Description Framework (RDF)". W3C Semantic Web. Disponible Online <<http://www.w3.org/RDF/>>.
- Vilches-Blázquez, L.M.; Ramos, J.A.; López-Pellicer, F.J.; Corcho, O. and Nogueras-Iso, J. (2009) "An approach to comparing different ontologies in the context of hydrographical information", in Popovich *et al.*, (eds.): IF&GIS'09. LNG&C Springer, St. Petersburg, Russia, 193-207.
- Vilches-Blázquez, L.M.; Villazón-Terrazas, B.; León, A.; Priyatna, F. and Corcho, O. (2010). *An Approach to Publish Spatial Data on the Web: The GeoLinked Data Case*, en "Workshop on Linked SpatioTemporal Data 2010 in conjunction with the 6th International Conference on Geographic Information Science , GIScience 2010", Zurich, Suiza.
- Villazón-Terrazas, B.; Vilches-Blázquez, L.M.; Corcho, O. and Gómez-Pérez, A. (2011). "Methodological Guidelines for Publishing Government Linked Data", en *Linking Government Data*, Springer New York, 27-49.