

Evaluación de la reusabilidad de los conjuntos de datos abiertos de un portal de Infraestructura de Datos Espaciales. Caso de Estudio: IDERA

María A. Barrera¹, Carlos Salgado²

¹Departamento de Informática
Facultad de Tecnología y Ciencias Aplicadas
Universidad Nacional de Catamarca
Maximio Victoria N° 55 – (4700) Catamarca - Argentina
Tel.: +54-383- 4435112 – Int. 168
marita.latingeo@gmail.com

²Departamento de Informática
Universidad Nacional de San Luis
Ejército de los Andes 950 – (5700) San Luis – Argentina
Tel.: +54-266-4520300– Int. 2101
csalgado@unsl.edu.ar

Resumen

Cada vez más organismos públicos ponen a disposición de los ciudadanos datos geoespaciales a través de sus Infraestructuras de Datos Espaciales (IDE), denominados también geoportales. Estos, son aplicaciones web que se utilizan para buscar y acceder a información geográfica (IG) y servicios geográficos asociados a través de Internet.

Sin embargo, estudios a nivel mundial sobre la utilización de las aplicaciones, muestran que aquellas creadas con datos abiertos (DA) públicos no se encuentran ni siquiera entre las 100 aplicaciones más descargadas por los usuarios y no se cuenta con estándares que evalúen los recursos que componen los conjuntos de datos de una IDE, por lo que pueden existir recursos difíciles de reutilizar debido a diversas características de los geoportales, y los consumidores de los mismos tienen que realizar un gran esfuerzo para encontrar datos de calidad y que se ajusten a sus necesidades.

Esta investigación realiza un análisis de la información publicada en los geoportales, con el objeto de medir el grado de reutilización de sus conjuntos de datos, para esto se propone una adaptación de la métrica MELODA,

teniendo en cuenta la naturaleza específica de los geodatos y tomando como caso de prueba la IDE de la República Argentina.

Palabras Claves: Datos Abiertos, Datos Geoespaciales, Información Pública, Infraestructura de Datos Espaciales, Métricas e Indicadores.

Contexto

Se presenta aquí la investigación que se viene realizando en el marco de una de las líneas de investigación correspondiente al proyecto "Las TIC al Servicio del Dato Abierto: Situación actual, conceptualización e iniciativas de apertura de información Pública", aprobado y subsidiado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca y que se desarrolla en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa.

La labor se realiza en forma colaborativa con el proyecto de investigación: Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N° P-031516. Además, ambos proyectos se encuentran reconocidos por el

programa de Incentivos.

Introducción

“Todos los datos producidos por las administraciones públicas, son datos públicos” (CEPAL, 2012). Muchas ciudades publican sus datos en formato abierto, pero el gran desafío en la actualidad es demostrar la capacidad y calidad de la información publicada, la reutilización que hace el ciudadano con esos datos y la forma en la que encuentran la información para que los datos puedan ser consultados, descargados o incluidos en aplicaciones externas.

Por otra parte, la disponibilidad, accesibilidad y el intercambio de información georeferenciada son requerimientos que cobran paulatinamente mayor importancia para una sociedad interesada en incorporar este tipo de información como apoyo fundamental en los procesos de planificación y toma de decisiones, tanto a nivel local como regional e inclusive global (Rix et al., 2011). Esta utilización de conjuntos de datos espaciales conlleva múltiples abordajes para el diseño y construcción de herramientas tecnológicas dirigidas hacia este fin (Rajabifard & Williamson, 2001). Sin embargo, cualquiera que sea el punto de partida, es primordial que los datos georeferenciados se brinden de forma precisa, sean mantenidos adecuadamente y garanticen su veracidad (Williamson, Rajabifard, & Binns, 2006).

Ante dicha coyuntura surge la noción de Infraestructura de Datos Espaciales (IDE) como alternativa para suplir información georeferenciada y, a la vez, cumplir con las demandas mencionadas anteriormente en relación a su integridad y validez. Una IDE consiste en una plataforma tecnológica distribuida que permite a los proveedores publicar sus datos geoespaciales y a los usuarios acceder a dicha información. Cabe resaltar que las IDE incluyen un conjunto de políticas, estándares y buenas prácticas que guían su uso (Rix et. al., 2011).

Estas infraestructuras de los países, se hacen

visibles a través de los denominados geoportales. Estos son aplicaciones web que actúan como un punto de entrada a la información geográfica distribuida y disponible para ser compartida (European Commission, 2005).

Cada vez hay más ciudades que cuentan con portales de IDE que ponen a disposición de los ciudadanos datos, productos y servicios geográficos, publicados bajo estándares y normas, asegurando su interoperabilidad y uso, la propiedad sobre la información por parte de los organismos que la publican y su responsabilidad de actualización, que han sido publicados para que los ciudadanos puedan consultarlos, descargarlos o incluirlos en aplicaciones externas abriendo nuevas posibilidades para generar valor, tanto en el ámbito social como en el económico. Ahora, el desafío es demostrar las capacidades de esa información y la reutilización que se hace de esos datos. Sin embargo, los estudios a nivel mundial sobre la usabilidad de las aplicaciones, muestran que aquellas creadas con datos abiertos públicos no se encuentran ni siquiera entre las 100 aplicaciones más descargadas por los usuarios. Igualmente, el Barómetro de Datos Abiertos (Open Data Barometer en inglés, es un proyecto de la ONG World Wide Web Foundation que consiste en una revisión del estado de la apertura de los datos públicos en múltiples países del mundo. <https://opendatabarometer.org/>), indica que menos del 7% de los datos publicados son reutilizables adecuadamente. Esta es una de las mayores preocupaciones que se presentan en las discusiones de DA y servicios geográficos, probablemente debido a la complejidad de los metadatos y de los estándares, la tecnología usada para la publicación, la confianza de los usuarios, la baja integración y la actualización de los datos disponibles, entre otros.

Comienzan a observarse iniciativas que han tomado este nuevo reto y han planteado algunas estrategias para mejorar el nivel de reuso de la información geográfica, sin embargo hay interrogantes que siguen en la mira: ¿Cuál es el

nivel de utilidad de la información geográfica publicada en los portales de DA de las ciudades?, ¿los ciudadanos están aprovechando efectivamente la publicación de estos servicios para promover la transparencia, la colaboración y el emprendimiento?, ¿o son solo iniciativas que corresponden a intereses de las administraciones actuales en cuanto a los mandatos de transparencia?. Es por esto que la aplicación de métricas se presenta como una necesidad estratégica para calificar la información disponible y evaluar su grado de reutilización.

En esta línea de investigación se está realizando un análisis general de la información publicada en los portales de IDE, con el objeto de medir el grado de reutilización efectiva de sus conjuntos de datos, teniendo en cuenta estándares técnicos, acceso legal, formatos de datos, visualización, su utilización en la recuperación de información, y las exigencias que las mismas tienen por parte de organismos internacionales de normalización para asegurar su calidad e interoperabilidad, para que las IDE sean herramientas ciertas que mejoren el acceso y reutilización de los datos abiertos geoespaciales a toda la ciudadanía.

En este sentido, según las encuestas de la Cámara de la Industria Argentina de Software, <http://www.cessi.org.ar/opssi> (2016) las métricas son un buen medio para entender, monitorizar, controlar, predecir y probar el desarrollo de software y los proyectos de mantenimiento, y para poder asegurar que un proceso o sus productos resultantes son de calidad, o poder compararlos, es necesario asignar valores, descriptores, indicadores o algún otro mecanismo mediante el cual se pueda llevar a cabo dicha comparación. Para ello, y según lo establecido en la nueva Ley de Promoción de la Industria del Software, (<http://servicios.infoleg.gob.ar/infolegInternet/a-nexos/185000-189999/185701/norma.htm> - 2011) es necesario llevar a cabo un proceso de medición del software que, en general, persigue tres objetivos fundamentales: (i) ayudarnos a entender qué ocurre durante el desarrollo y el

mantenimiento, (ii) permitirnos controlar qué es lo que ocurre en los proyectos y (iii) poder mejorar los procesos y productos.

Abella, Ortiz-de-Urbina-Criado y De Pablos-Herederó (2017) han elaborado la métrica MELODA que permite evaluar el grado de reutilización de los datos abiertos. MELODA analiza seis dimensiones: Marco Legal, Estándares Técnicos, Acceso a la Información, Modelo de Datos, Geolocalización de la Información y Actualización en Tiempo Real. Cada una de estas dimensiones consta de cinco niveles mediante los cuales se obtiene un porcentaje de calidad de las mismas, dicho porcentaje se utiliza posteriormente en una fórmula que proporcionará el nivel de reusabilidad de los conjuntos de datos disponibles en los portales abiertos.

Esta última métrica se ha utilizado por otros autores como Vicente-Paños y Jordán-Alfonso (2017) para estudiar el acceso a la información del sector público y su reutilización en las comunidades autónomas de España. Estos autores observaron que cuando se calificaban a través de MELODA los portales de datos abiertos de esas comunidades, las puntuaciones más bajas se obtenían en las dimensiones de información geolocalizada e información en tiempo real.

Es por esto que se propone una adaptación de la métrica MELODA, con el propósito de evaluar el grado de reutilización de los conjuntos de datos publicados en una IDE, teniendo en cuenta la naturaleza específica de esos datos y tomando como caso de prueba la Infraestructura de Datos Espaciales de la República Argentina (IDERA).

Líneas de investigación y desarrollo

La presente es una investigación que se desprende del Proyecto: "Las TIC al Servicio del Dato Abierto: Situación actual, conceptualización e iniciativas de apertura de información Pública".

Para la metodología de trabajo se establecieron métodos comparativos y descriptivos y análisis documental y de contenidos, que abordan el estudio de los conceptos fundamentales sobre DA e IDE, su importancia y las diversas metodologías, técnicas, herramientas y propuestas que se encuentran disponibles para evaluar y diagnosticar su implementación. La revisión bibliográfica y el análisis del estado del arte en DA son la base para determinar las capacidades de los mismos en cualquier organismo y en especial en la Infraestructura de Datos Espaciales de la República Argentina (IDERA: <http://www.idera.gob.ar/>). Con respecto a la obtención de información se prevé recolección de muestras mediante examen de características de las administraciones públicas que ofrecen a la ciudadanía sus DA. La fuente de información se centra en la observación indirecta, a través de documentación de las tecnologías informáticas utilizadas en la administración pública nacional y provincial. En cuanto a la métrica utilizada para el análisis de los datos se ha seleccionado a MELODA (MEtric for reLeasing Open DATA) para el análisis exploratorio de los principios de DA, y sobre la situación actual de iniciativas de DA, con el fin de que el proyecto sea de máxima aplicabilidad, se realizan pruebas de la métrica MELODA sobre conjuntos de datos de IDERA y de los portales de otras IDE regionales.

Resultados y Objetivos

Esta línea de I-D-I se encuentra en una etapa inicial. El objetivo principal planteado para la ejecución del trabajo fue: Evaluar la reusabilidad de los conjuntos de datos disponibles en las Infraestructura de Datos Espaciales, ejemplificando los pasos seguidos por medio del geoportal de la Infraestructura de Datos Espaciales de la República Argentina (IDERA).

En este sentido, se está realizando una adaptación de la métrica MELODA para la evaluación de la reusabilidad de los conjuntos

de datos de los geoportales que requieren un análisis especial dado sus características y requerimientos exigidos por organismos internacionales de estandarización e interoperabilidad.

Formación de Recursos Humanos

El primer autor se desempeña como docente de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca. Mientras que el segundo autor es docente de la Facultad de Ciencias Físico-Matemáticas y Naturales de la Universidad Nacional de San Luis. La temática abordada en este proyecto de investigación en desarrollo, fue concebida a priori en el marco de investigación de la tesis para optar al título de Magister en Ingeniería de Software. El mismo se desarrolla en la Facultad de Ciencias Físico-Matemáticas y Naturales de la Universidad Nacional de San Luis, en donde el Mgter. Carlos Salgado, se desempeña como director de dicha Tesis.

También se desarrollan dentro de los trabajos de investigación becas CIN y trabajos finales de carreras de grado con relación a la temática por alumnos de la carrera de Ingeniería Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca y de la carrera Licenciatura en Computación, y trabajos finales de carrera de la Ingeniería en Informática e Ingeniería en Computación de la Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis.

Bibliografía

- Barómetro de los Datos Abiertos, 4ª Edición. Informe Global, 2017 - World Wide Web Foundation.

<http://opendatabarometer.org/doc/4thEdition/O DB-4thEdition-GlobalReport-ES.pdf>

-Conferencia Regional de Datos Abiertos para América Latina y el Caribe. *Datos* <https://www.cepal.org/es/discursos/conferencia-regional-datos-abiertos-para-el-desarrollo-de-america-latina>

- Albella A., Ortiz-De-Urbina-Criado, M Y De-Pablos-Heredero C. "*Meloda, métrica para evaluar la reutilización de datos abiertos*". El profesional de la información, 2014, noviembre-diciembre, v. 23, n. 6. ISSN: 1386-6710

- "meloda 40v1 – MELODA," 2016. [Online]. Available:

<http://www.meloda.org/meloda40v1/>.

- Iniesto M., Núñez et all. (2014). "*Introducción a las Infraestructuras de Datos Espaciales*". Edición digital. Catálogo general de publicaciones oficiales <http://publicacionesoficiales.boe.es>. NIPO: 162-14-022-0 - DOI: 10.7419/162.12.

- Sommerville, I. (2011) *Ingeniería de Software*. PEARSON EDUCACIÓN, México, ISBN2011:978-607-32-0 603-7 – 9na Edición

gubernamentales en formato abierto?, 2014.

- Infraestructura de Datos Espaciales de la República Argentina (IDERA) (2020) <https://www.idera.gob.ar/>

- Vicente-Paños, A.; Jordán-Alfonso, A. (2017). "*Acceso a la información pública y su reutilización en las comunidades autónomas: evaluación de la reutilización de datos abiertos*". El profesional de la información, v. 26, n. 3, pp. 381-391. <https://doi.org/10.3145/epi.2017.may.04>

- Nueva Ley de Promoción de la Industria del Software, (2011) <http://servicios.infoleg.gob.ar/infolegInternet/anexos/185000-189999/185701/norma.htm>.

- CESSI, "Encuestas. Cámara de la Industria Argentina de Software" (2016) <http://www.cessi.org.ar/opssi>.