

Estrategia de evaluación de la reusabilidad de los conjuntos de datos de los portales de Infraestructuras de Datos Espaciales.

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
Facultad de Ciencias Físico-Matemáticas y Naturales
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

La información geoespacial se ha transformado en un bien de consumo primario en nuestra sociedad. A nivel mundial, las organizaciones gubernamentales han contribuido a esta realidad a través de las Infraestructuras de Datos Espaciales (IDE), promoviendo el desarrollo económico, estimulando una mejor gobernanza y fomentando la sostenibilidad ambiental. (Vilches-Blázquez, Ballari, 2020)

Las IDE, son aplicaciones web que se utilizan para buscar y acceder a información geográfica (IG) y servicios geográficos asociados a través de Internet (Inierto M., Núñez et al, 2014). Están definidas por un conjunto de políticas, tecnologías y acuerdos institucionales que aprovechan la provisión y uso de datos y servicios espaciales estandarizados.

Sin embargo, se ha prestado poca atención al dominio de la información geoespacial, que está llamando cada vez más la atención de los tomadores de decisiones que buscan mejorar los

modelos de decisión considerando factores espacio-temporales. En este contexto, esta investigación realiza un análisis de la información publicada en los geoportales IDE, 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.

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

Contexto

La presente investigación está contextualizada en el trabajo colaborativo entre dos grupos de investigación: Los integrantes del Proyecto de Investigación: Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis.

Proyecto N° P-03-2020; y, por otra parte, el grupo de investigación interviniente, del proyecto de investigación: “Las TIC al Servicio del Dato Abierto: Situación Actual, Conceptualización e Iniciativas de Apertura de Información Pública”, del departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

El trabajo colaborativo entre las universidades aportará dimensiones analíticas prácticas, condición indispensable para integrar y comparar la información obtenida a partir de los proyectos que genere cada grupo. En este contexto, el grupo de la Universidad Nacional de Catamarca aporta su conocimiento y experticia en lo referente al análisis de los Datos Abiertos Gubernamentales y desde la Universidad de San Luis, el grupo de investigación proveerá el conocimiento y la experiencia necesaria para definir y adaptar conceptos y modelos de calidad a este nuevo escenario de las ciudades modernas. Su experiencia en definir modelos, métodos, métricas e indicadores.

Introducción

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).

Algunos autores resumen la idea anterior diciendo que “la información geoespacial se ha transformado en un bien de consumo primario en nuestra sociedad. A nivel mundial, las organizaciones gubernamentales han contribuido a esta realidad a través de Infraestructuras de Datos Espaciales (IDE), promoviendo el desarrollo económico, estimulando una mejor gobernanza y fomentando la sostenibilidad ambiental”. (Vilches-Blázquez, Ballari, 2020).

Las IDE son aplicaciones web que actúan como un punto de entrada a la información Geográfica (IG) distribuida y disponible para ser compartida. Esta IG proveniente de instituciones oficiales, es publicada, con frecuencia, bajo estándares y normas, asegurando su interoperabilidad y uso, ofreciendo como mínimo algunas funcionalidades básicas, tales como: visualización de mapas, localización y descarga de datos y servicios, para el procesamiento de datos. La idea de publicar estos datos para los ciudadanos, es que éstos puedan consultarlos, descargarlos o incluirlos en aplicaciones externas que modelan, analizan, muestran y/o generan datos, información y productos geoespaciales de valor agregado en múltiples formatos de archivo para algunos propósitos de información (por ejemplo, visualización) o toma de decisiones (es decir, a través del análisis), lo que abre nuevas posibilidades para generar valor, tanto en el ámbito social como en el económico.

El costo de producir esa información es bastante elevado, por lo que ahora el desafío es demostrar sus capacidades y la reutilización que se hace de esos datos. En este sentido, y en lo que respecta al diseño que presentan los portales IDE, se puede observar que no responde a las necesidades de los distintos perfiles de usuarios que necesitan acceder en busca de información, lo que determina un uso y

reuso bastante limitado de los mismos y, en algunos casos, que solo sean utilizados por profesionales muy especializados de IG. Además, aunque estos datos se encuentran potencialmente abiertos, muchos de ellos resultan difíciles de encontrar, desorganizados y en un formato que no es el adecuado para análisis espacial, es decir, estos datos no son “técnicamente” abiertos, por lo que resulta necesario hacer un análisis de la efectiva utilidad de los datos ofrecidos por una IDE.

Por otra parte, no se cuenta con estándares específicos de análisis de calidad que describan las características que deben tener los recursos que componen los conjuntos de datos geográficos disponibles en una IDE para ser efectivamente reutilizados, ya que hay que tener en cuenta que no es igual el acceso a los mismos y su difusión, que su reutilización.

La presente investigación realiza el 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 (<http://www.meloda.org>), teniendo en cuenta la naturaleza específica de los geodatos, su topología y su localización espacial y tomando como caso de prueba la IDE de la República Argentina (<https://www.idera.gob.ar>).

MELODA (Metric for releasing open data) es una métrica de evaluación que permite calificar la información y evaluar el grado de reutilización de los datos públicos y el valor de la información publicada en los portales de DA. Ayuda a los organismos productores a hacer que todos los datos que publiquen sean buenos para la reutilización. También ayuda a los consumidores de open data a encontrar los mejores datos a reutilizar con el objetivo de crear nuevos productos

y servicios de calidad. (Albella A. et al, 2014).

Si bien es cierto que ya comienzan a observarse iniciativas y guías de buenas prácticas que han tomado este nuevo reto y han planteado algunas estrategias para mejorar el nivel de reuso de la información geográfica abierta, aún hay muchos interrogantes que siguen sin resolverse.

En el presente trabajo se pretende realizar 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.

Abella, Ortiz-de-Urbina-Criado y De Pablos-Heredero (2014; 2017a) han elaborado la métrica MELODA que permite evaluar el grado de reutilización de los datos abiertos. MELODA analiza seis dimensiones: Estándares Técnicos, Acceso a la Información, Marco Legal, 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, por lo que estimamos que esta métrica es la que más se adecúa como base de análisis para la evaluación de los datos disponibles en una IDE.

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".

La naturaleza de este estudio es la investigación descriptiva con dos propósitos principales. En primer lugar, se ocupa de la explicación de las motivaciones, procesos y dificultades asociados a la adopción de métricas de evaluación de reusabilidad a nivel general para cualquier portal de DA. El segundo propósito apunta a describir los procedimientos que una IDE realiza para la apertura de los conjuntos de datos que pone a disposición en su geoportal. Por lo tanto, el método cualitativo resulta ser el enfoque más apropiado para esta investigación. Este enfoque ayuda a comprender mejor los fenómenos y obtener un conocimiento exhaustivo sobre el estudio de los procesos y los participantes. 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 geoespaciales a

través de las Infraestructuras de Datos Espaciales.

En cuanto a la métrica utilizada para el análisis de los datos, se ha seleccionado a MELODA. Esta métrica se utiliza realizando un análisis de los conjuntos de datos, en seis dimensiones: Estándares Técnicos, Acceso a la Información, Marco Legal, 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.

Resultados y Objetivos

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. En este sentido, nos planteamos redefinir las dimensiones e indicadores de la métrica MELODA, y de su fórmula aplicada para el cálculo, presentando una propuesta cuantitativa para evaluar la reusabilidad de los geodatos disponibles en una IDE, teniendo en cuenta la naturaleza específica de la IG.

Entre las nuevas dimensiones propuestas, podemos destacar aquellas que tienen que ver con el análisis de los conjuntos de datos, teniendo en cuenta aspectos tales como: estándares técnicos, acceso legal, formatos de datos y metadatos, sistemas de referencia espacial, visualización en servicios de mapas, su utilización en la recuperación de información, además de las exigencias que las mismas tienen por parte de organismos internacionales de normalización, y que tienen relación

directa con los servicios web que ofrecen las IDE a través de sus geoportales, tales como el visualizador o servicio de mapas, servicio de fenómenos, servicios de coberturas, servicios de catálogos, etc., ya que estos son los componentes más utilizados para la realización de los análisis de datos espaciales en estas infraestructuras y los que son requeridos por profesionales y desarrolladores de nuevas aplicaciones.

El análisis de los conjuntos de datos, teniendo en cuenta las dimensiones propuestas nos dará un porcentaje de calidad de las mismas, dicho porcentaje se utiliza posteriormente en una nueva fórmula de cálculo que proporcionará el nivel de reusabilidad de los conjuntos de datos disponibles en los portales abiertos de una IDE.

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, dio lugar al desarrollo de una tesis para optar al título de Magister en Ingeniería en Software. El mismo se desarrolla en la Facultad de Ciencias Físico-Matemáticas y Naturales de la Universidad Nacional de San Luis, en donde el Mg. Ing. Carlos Salgado, se desempeña como director de dicha Tesis. Además, se trabaja en el desarrollo de tesinas de grado y trabajos finales de carrera en el Marco de la Ingeniería informática y la Licenciatura en Ciencias de la Computación de la FCFMyN de la UNSL y la carrera de

Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca

Bibliografía

- Vilches-Blázquez Luis M. y Ballari Daniela (2020): Unveiling the diversity of spatial data infrastructures in Latin America: evidence from an exploratory inquiry, *Cartography and Geographic Information Science*, DOI: 10.1080/15230406.2020.1772113; <https://doi.org/10.1080/15230406.2020.1772113>
- Zarazaga-Soria, F.J. et al. (2012) “Hacia las Infraestructuras de Datos Abiertos Espaciales” - Actas de las III Jornadas Ibéricas de Infraestructuras de Datos Espaciales (JIIDE'2012).
- Iniesto M., Núñez et all. (2014). “Introducción a las Infraestructuras de Datos Espaciales”. <http://publicacionesoficiales.boe.es>. NIPO: 162-14-022-0 –
- Metric for releasing open data - MELODA “<http://www.meloda.org>
- Infraestructura de Datos Espaciales de la República Argentina - IDERA (2020) <https://www.idera.gob.ar/>
- Abella, Aberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen (2019). “Meloda 5: A metric to assess open data reusability”. *El profesional de la información*, v. 28, n. 6, e280620. <https://doi.org/10.3145/epi.2019>.
- 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.
- Infraestructura de Datos Espaciales de la República Argentina - IDERA (2020) <https://www.idera.gob.ar/>