

DATOSURBANOS.CL UNA INFRAESTRUCTURA DE DATOS ESPACIALES PARA INVESTIGACIONES

STEFAN STEINIGER¹, HELEN DE LA FUENTE²

Resumen

Las Infraestructuras de Datos Espaciales (IDE) permiten la difusión e intercambio de información geográfica. Por esto, la implementación de una IDE en una universidad o centro de investigación trae consigo variados beneficios para investigadores, instituciones y el público en general. Entre los beneficios se destacan, por ejemplo, el intercambio rápido de nueva información producto de una investigación y, a su vez, también se evita la doble adquisición de datos costosos. El centro de investigación CEDEUS, con su enfoque en una planificación urbana sustentable en Chile implementó un “Observatorio”, el cual es una IDE con herramientas para la comunicación entre usuarios y herramientas para análisis y procesamiento de datos geográficos. Este artículo presenta sus principales requisitos, funciones y datos base de esta IDE solicitados por los miembros del centro a través de una encuesta. Además, se discute el uso de software libre para la implementación de la plataforma web, y se presenta un análisis de fortalezas y debilidades después de dos años de operación.

Palabras clave: Infraestructura de Datos Espaciales (IDE), software libre, GeoNode, investigaciones urbanas, acceso y difusión de datos.

Abstract

Spatial Data Infrastructures (SDIs) allow the distribution and exchange of geographic information. Implementations of such an SDI in a university or research centre will bring several benefits to researchers, the institution itself, but also to the public. Among the benefits are an easy and fast way to interchange information and data generated during research, also costly data does only need to be acquired once and can be used by several courses and research projects. The Chilean research centre CEDEUS, with a focus on urban sustainable planning has created an “Observatory” platform, which consists of an SDI with a set of web tools that allow communication with users, and a set of tools for the analysis and processing of spatial data. In this article, we present an analysis of needs for this particular SDI, including needs for tools and data, obtained from a potential-user survey. We discuss the free and open source software used to implement the observatory website and present an analysis of strengths and weaknesses based on the experiences collected in the first two years of operation.

Keywords: Spatial Data Infrastructure (IDE), free software, GeoNode, urban research, access and distribution of data.

¹ Dr. Sc. Nat., CEDEUS & Pontificia Universidad Católica de Chile, Depto. Ing. Transporte y Logística, Avda. Vicuña Mackenna 4860, Macul - Santiago, Chile, Email: ssteiniger@uc.cl

² Geógrafa, CEDEUS & Universidad de Concepción, Edmundo Larenas 219, Concepción, Chile, Email: helendelafuente@udec.cl

1. Introducción

Hace aproximadamente diez años, con la llegada de las redes sociales y más recientemente las redes de sensores, la sociedad ha entrado de lleno en la era de la información. Por ende, utilizar esta información y las nuevas fuentes que los proveen proporciona una gran oportunidad para las ciencias de la Tierra y también para aquellas que poseen su enfoque en la descripción y análisis de procesos a nivel territorial, es decir, la interacción entre el hombre y la naturaleza. Algunos ejemplos de disciplinas con una alta demanda de datos son la ecología (McDermid et al., 2005), geografía (Haggett, 2001), ingeniería en transporte (Thill, 2000) o la planificación urbana (Drummond & French, 2008). A partir de esta última disciplina mencionada nace la Infraestructura de Datos Espaciales (IDE) del Centro de Desarrollo Urbano Sustentable (CEDEUS). El centro de investigación CEDEUS tiene cuatro clústeres o líneas de investigación: (i) recursos críticos, (ii) movilidad, (iii) planificación integrada y (iv) entorno construido. Los grupos de trabajo en estas líneas de investigación están compuestos por investigadores de múltiples disciplinas y distintos departamentos de universidades como: ingeniería hidráulica, salud pública, ecología, geografía, ingeniería en transporte, arquitectura, planificación urbana, sociología, entre otras.

Con el propósito de facilitar el acceso a datos geográficos básicos e información elaborada por los propios investigadores del centro se ha creado en el año 2013 el Observatorio CEDEUS (<http://observatorio.cedeus.cl>). Haciendo realidad un sistema web que permite difundir e intercambiar información entre investigadores. En la primera etapa de la IDE, el equipo del observatorio comenzó analizando el estado del arte en relación a datos existentes en los departamentos, estudiando la demanda de coberturas de los afiliados al centro, y conociendo los software que permiten implementar una infraestructura de datos espaciales. Luego de esta fase inicial se implementó la IDE CEDEUS junto con diversas herramientas para procesar y analizar datos y también para comunicación. En una tercera etapa, los trabajos del equipo se enfocaron en la adquisición de bases geográficas para las investigaciones. En las siguientes secciones se describirán los resultados de las tres etapas y luego se comentan las fortalezas y debilidades. Se concluye con recomendaciones.

2. Análisis de la situación inicial – Usuarios potenciales, Datos

Para llegar a conocer a los usuarios potenciales de la IDE CEDEUS y sus necesidades, se aplicó una encuesta a los investigadores en el año 2013. A partir de esto, se recibieron 36 respuestas de los cerca de 100 asociados al Centro. Con respecto a los temas de investigación se descubrió que los investigadores trabajan en temas muy diversos, por ejemplo: (i) redes ecológicas, (ii) accidentes de tráfico, (iii) calidad del agua, (iv) segregación demográfica espacial, (v) contaminación de aire, (vi) planificación y operación de corredores de buses, (vii) áreas verdes urbanas, entre otras. En consecuencia, los miembros de CEDEUS tienen diferentes necesidades de datos. Un resultado esperado es que sólo una o dos personas estén a cargo de la información, pero, contrario a esto, la mayoría de los

departamentos tienen información distribuida en distintos computadores de escritorio - no servidores o una base de datos con acceso común.

Un resultado que llamó la atención es que el 49% de los encuestados tiene muy poca o simplemente no tienen experiencia con coberturas espaciales – refiriéndose a “muy poco” como experiencias de los usuarios en Google Earth. También sorprende el alto porcentaje (41%) de los investigadores que utilizan AutoCAD y Excel, en lugar de un SIG para trabajar con datos geográficos (ArcGIS 77% de 36 personas). Esto significa que los formatos utilizados son muy variados, incluyendo, por ejemplo, documentos Word y PDF, documentos Excel y CSV, ESRI Shapefile, Google KMZ, Autocad DWG, y TIFF. Por lo tanto, el desafío de la plataforma a crear debía permitir a trabajar con muchos de estos formatos y fortalecer el uso de otros.

Sobre la base de estos resultados se hizo evidente la necesidad de ofrecer talleres introductorios de SIG para los investigadores y estudiantes, esenciales para que el observatorio tenga un impacto. Además, se consideró fundamental crear un equipo observatorio que funciona como un “Centro de consulta y ayuda SIG” permitiendo a los miembros de CEDEUS la oportunidad de preguntar sobre cualquier tema relacionado al mapeo y el análisis espacial. Teniendo en cuenta los distintos temas de investigación y sus requerimientos se decidió que el primer enfoque del Observatorio es la adquisición y generación de una base de datos geográficos básicos para crear mapas y, en segundo lugar, la obtención y búsqueda de cierta información específica requerida por unas pocas personas. Asimismo, se concluyó que la plataforma a implementar debe contar con tres niveles de acceso: (i) acceso sólo para investigadores CEDEUS, (ii) acceso para personas e instituciones colaboradoras de CEDEUS, y (iii) acceso público. El acceso público a los datos y resultados de investigaciones se creó con el objetivo de transparentar el proceso tanto al usuario común como a la institución que financia al centro y, además, fortalecer el impacto de resultados. El acceso público a datos también favorece nuevas iniciativas de ciencia abierta, permitiendo la replicación de experimentos y resultados (Morin et al., 2012; Rey, 2014; Singleton et al., 2016; Nature Editorial, 2016).

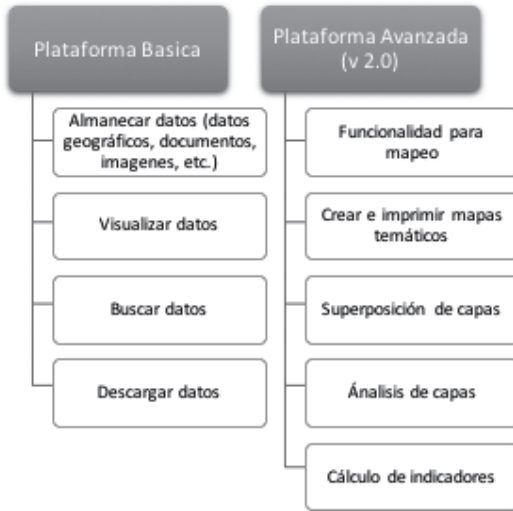
3. Implementación de la IDE y otras herramientas útiles

3.1 Análisis de necesidades

Basándose en el objetivo de crear una plataforma para la difusión y el intercambio de información se ha definido la funcionalidad de la plataforma técnica del observatorio CEDEUS (ver Figura 1). Según esto las prioridades de implementación y las capacidades del software se distinguen entre (i) una plataforma básica y (ii) una plataforma avanzada. En cuanto a las herramientas necesarias, es evidente la demanda de un software de infraestructura de datos espaciales que permita cargar, buscar, explorar y descargar información espacial y no espacial. Además, se requiere una plataforma para la documentación y comunicación (difusión de actividades y noticias). También es útil poseer una herramienta de geo-codificación, la cual permite obtener coordenadas a partir de una dirección

(por ejemplo, el Aguilucho #3236, Providencia => coordenada X, Y). Algunas personas incluso solicitaron formas que permitan la creación “fácil” de mapas temáticos, que no requieran conocimientos de SIG. En este sentido, se trata de imitar el modelo de la plataforma web Carto.com que permite crear mapas sin una educación profunda en cartografía, donde se puede elaborar un mapa después de cargar una hoja de cálculo (*.xls).

FIGURA 1.
 FUNCIONALIDAD DESEADA DE LA PLATAFORMA WEB DE LA IDE CEDEUS.



Fuente: elaboración propia.

3.2 Arquitectura

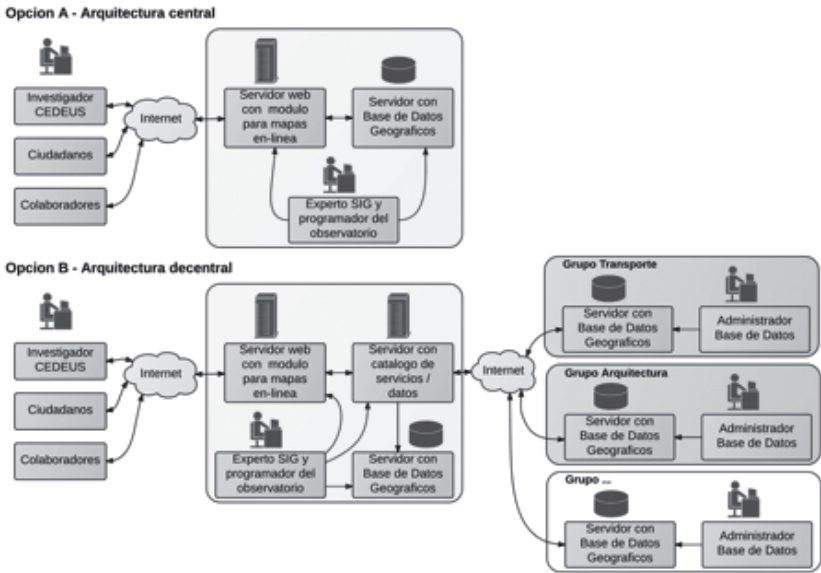
CEDEUS concentra investigadores en dos universidades, en múltiples facultades y a su vez diferentes departamentos. A partir de esto se puede responder con dos posibles arquitecturas técnicas (Figura 2):

- (A) una arquitectura centralizada con una sola base de datos, o
- (B) una arquitectura distribuida (“federada”) con base de datos independientes que pueden comunicarse entre ellas (Coetzee & Bishop, 2009).

La opción (A), con una base de datos central, tiene ventajas al momento de hacer copias de respaldo (más simple), requiere un equipo más pequeño y requiere menos trabajo en su mantención. Pero tiene también desventajas con respecto a la comunicación con los investigadores por la separación espacial del equipo. La opción (B), en donde hay al menos una persona responsable en cada departamento, es mejor con respecto a la comunicación e intercambio de información/datos, pero tiene sus desventajas con respecto a los costos de personal, equipamiento y mantención.

Revisando los fondos disponibles para personal y equipamiento y considerando el tamaño del equipo CEDEUS (aproximadamente 100 personas entre investigadores, personal técnico y alumnos) se opta por la opción (A), una base de datos central, trabajando con dos servidores.

FIGURA 2.
OPCIONES DE ARQUITECTURA COMPUTACIONAL PARA REALIZAR LA IDE CEDEUS.

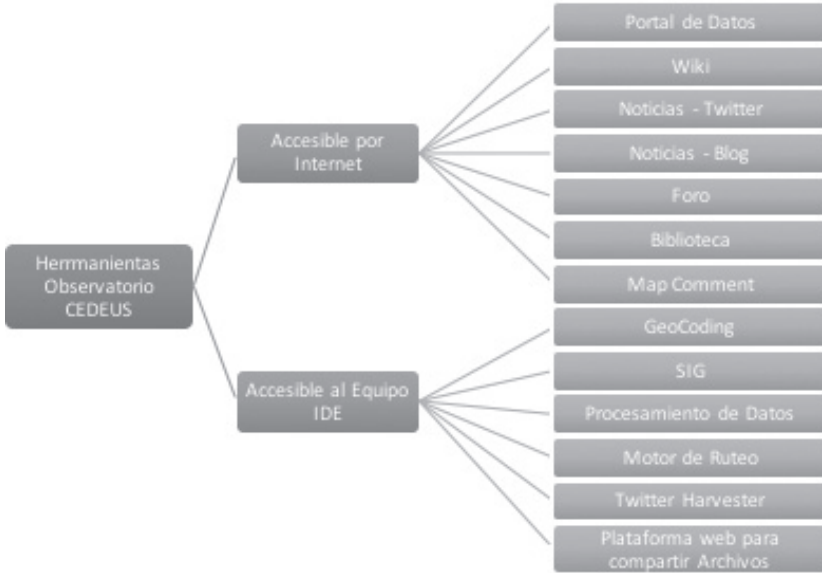


Fuente: elaboración propia.

3.3 Herramientas disponibles

Teniendo en cuenta la necesidad de herramientas de análisis se realizó una búsqueda de las más apropiadas para (a) distribuir datos, (b) documentar el uso de herramientas y (c) comunicar con los usuarios de la IDE. Dado que los fondos para software son limitados, la búsqueda se concentró en un software gratuito y de código abierto (Steiniger & Hunter, 2013). La página web principal del Observatorio www.datosurbanos.cl muestra algunas de sus funcionalidades accesibles para el usuario del observatorio e IDE CEDEUS (ver Figura 3).

FIGURA 3.
HERRAMIENTAS DEL “OBSERVATORIO CEDEUS”



Fuente: elaboración propia.

Ejemplos de soluciones completas e integradas para implementar una IDE son las proporcionadas por el software de ESRI “Portal for ArcGIS”, la cual posee costos prohibitivos y una licencia propietaria. Pero también hay plataformas libres y de código abierto como Geonodo (del SNIT Chile), GeoNode, y Deegree. El equipo de CEDEUS seleccionó GeoNode.org como plataforma base para difundir y compartir datos, por su uso fácil y amplias funcionalidades (ver Figura 4, captura de pantalla). Los componentes de GeoNode consisten en:

- (1) un interfaz web basado en Django y con GeoExplorer como visor de mapas;
- (2) la base de datos PostGIS;
- (3) el software GeoServer para generar servicios de mapas y servicios de procesamiento de datos; y,
- (4) el software pyCSW para consultas al catálogo de base de datos.

Es importante mencionar que GeoNode permite guardar datos geográficos y datos no-geográficos como documentos de texto, archivos ZIP, tablas de Excel y fotos (no geo-referenciados).

Para la comunicación con los usuarios se utiliza las plataformas de Twitter y Elgg. Elgg es un software de red social con funcionalidades de foro y publicación de noticias en forma de blog (Costello, 2012). Además, se ha desarrollado un plugin para Elgg que permite colocar comentarios de planificación urbana parti-

cipativa en un mapa web (ver: <http://miciudad.geosteiner.cl>). Para documentar avances de los diferentes grupos de trabajo, como el grupo de indicadores para ciudades sustentables, y proveer guías para el usuario se utiliza el software MediaWiki. Por otra parte, también se instaló el software Omeka, un programa de biblioteca web, para guardar y distribuir documentos y publicaciones generados por CEDEUS.

Existe un segundo conjunto de herramientas para procesar datos y que no son accesibles de forma directa por usuarios de IDE. Estas herramientas son:

- Nominatim un software para la geo-codificación de direcciones a partir de OpenStreetMap;
- QGIS como SIG para crear y procesar información, y generar mapas;
- Yupiter Notebook, un software que permite escribir scripts en el lenguaje Python para la exploración y el procesamiento de información;
- el software OwnCloud para compartir archivos entre los miembros de equipo, similar a Dropbox;
- el software OpenTripPlanner para el análisis de tiempos de viaje en redes de transporte; y,
- el CF Twitter Harvester, una herramienta de desarrollo propio que permite almacenar los comentarios de redes sociales de Twitter para su posterior análisis en el contexto de planificación urbana.

FIGURA 4.

EL GEOPORTAL DEL "OBSERVATORIO CEDEUS" BASADO EN GEONODE 2.0



Fuente: elaboración propia.

4. Datos

Como se ha señalado anteriormente las necesidades por datos son muy diversas (ver Sección 2). Según la encuesta inicial existen al menos dos grupos de datos clave en términos de demanda: en primer lugar, los datos demográficos, tales como los registros del último censo o aquellos obtenidos a través de la encuesta origen-destino desarrollada por el Ministerio de Transporte y Telecomunicaciones de Chile (MTT); en segundo lugar, los datos cartográficos básicos, como límites administrativos, información de uso de suelo o modelos de elevación digital. Otros solicitados de manera especial y relacionado a un tema de investigación son, por ejemplo, las redes de transporte público, áreas verdes, nivel socioeconómico.

Al momento (octubre de 2016) se cuenta con un conjunto de 390 coberturas geográficas, de las cuales el 95% (aproximadamente 360) están disponibles públicamente. Entre ellas se encuentran cerca de 50 imágenes aéreas de alta resolución (30cm/píxel) que muestran las zonas urbanas de Chile, así como predios obtenidos a través del proyecto CEDIZ por www.IDE.cl. Esta base de datos está disponible en <http://datos.cedeus.cl> y ofrece un conjunto de capas geográficas básicas para Chile, que son útiles para la preparación de mapas. La mayor parte de los datos geográficos de base se recibieron de www.Rulamahue.cl y www.NaturalEarthData.com.

Después de tener el portal de datos operativo durante casi dos años, los cinco datos geográficos más visitados por los investigadores y el público han sido los siguientes:

1. Modelo de Elevaciones – SRTM 250m - Chile (975 vistas);
2. Clasificación de Cubiertas vegetacionales, Luebert y Pliscoff (750 vistas);
3. Ríos y Esteros de Chile (556 vistas);
4. Zonas EOD Santiago 2012 (544 vistas); y,
5. Cobertura de Suelo - Región XV - 2012 - Arica y Parinacota (471 vistas).

Además de las coberturas geográficas, la IDE CEDEUS alberga alrededor de 160 documentos, incluyendo archivos de imágenes, hojas de cálculo (CSV, XLS), PDF y archivos ZIP. La mayoría de estos documentos son de tipo climáticos, de transporte público (como archivos GTFS), mapas en formato de imagen y documentos relacionados a la planificación territorial tanto a nivel comunal como regional, por ejemplo, los PRC – Plan Regional Comunal - y PLADECO – Plan de Desarrollo Comunal. Los documentos más visitados son los siguientes:

1. PLADECO Lo Espejo (1.745 vistas);
2. PLADECO La Florida (1.659 vistas);
3. Mapa de Población de Santiago 2002 según comuna (977 vistas);
4. PLADECO El Bosque (928 vistas); y,
5. PLADECO Independencia (772 vistas).

Es sorprendente ver la alta tasa de visitas para los documentos PLADECO en comparación con las visitas de los datos espaciales. La diferencia entre los números de visitas (para documentos con la misma fecha de subida) indica que

la demanda de archivos PLADECO es, con alta probabilidad, del público general y no de investigadores internos o externos. Esto se puede explicar en parte por el hecho de que estos documentos se encuentran fácilmente en una búsqueda de Google, y que en muchos casos esta información no está accesible directamente a través de las páginas web de las municipalidades correspondientes.

5. Análisis de fortalezas y debilidades

Analizando el estado actual del observatorio con respecto a su implementación y funcionamiento en los últimos dos años, surge el siguiente listado de fortalezas y debilidades. Entre las fortalezas se encuentran:

- La provisión de un repositorio de coberturas geográficas básicas de Chile junto con datos temáticos para investigaciones específicas;
- Un equipo pequeño de dos expertos en SIG que manejan bases de datos y la infraestructura (software y hardware) y que, además, están disponibles para consultas por distintos medios;
- Una plataforma web, basado en el software GeoNode, que permite una difusión fácil de capas geográficas y documentos no-geográficos;
- Un servidor propio de CEDEUS que permite controlar el acceso a los datos en un 100%, y, en consecuencia, permite guardar y procesar información confidencial; y,
- Una infraestructura de bajo costo respecto a licencias de software.

Entre las debilidades de la implementación del observatorio se encuentran los siguientes problemas:

- La decisión de una arquitectura central en el modelo de la IDE, con respecto al almacenamiento y difusión de datos, concluyó en una desconexión con algunos investigadores del centro, y en consecuencia la pérdida de ellos y sus alumnos como usuarios;
- Tener un equipo de sólo dos personas no es suficiente para todo el trabajo que requiere mantener la IDE, esto incluye trabajos de procesamiento y mantención de datos, creación de mapas, publicación de noticias, buenas prácticas y mantención del software (y hardware);
- El uso de diferentes software en un propio servidor del centro, en comparación con un servicio de la nube (cloud service), requiere de bastante tiempo de mantención. Además, con cada software añadido aumenta la probabilidad de conflicto entre los software ya instalados;
- A pesar de que GeoNode es relativamente fácil de usar, aún es necesario impartir más capacitación para aumentar su uso;
- Relacionado a este último punto es necesario el conocimiento básico de SIG, si una persona quiere (o debe) contribuir datos (geográficos) para difundirlos a otros usuarios de la IDE; y,
- El uso del software libre llega a sus límites cuando se reciben coberturas creadas con software pagados. En nuestro caso recibimos, por ejemplo, da-

tos y mapas creados con ESRI ArcGIS los cuales no son accesibles sin tener este software (costoso). Similar es el caso con archivos creados con Autocad en el formato DWG.

A estos problemas mencionados se suma la falta de funcionalidad, resaltada en la Figura 1 “Plataforma Avanzada”, como la creación y edición de datos, elaboración de mapas temáticos y funciones básicas de análisis de información geográfica. Los servicios de mapas en la nube (por ejemplo: ArcGIS Online, Carto.com, InstaGIS.com) ofrecen este tipo de funciones, pero los autores no conocen ningún software libre que ofrezca estas capacidades por defecto y, por lo tanto, su implementación requiere de programación propia.

6. Conclusiones y recomendaciones

La implementación de infraestructuras de datos para investigaciones está recibiendo más y más atención en universidades y centros de investigación para habilitar la difusión de datos geográficos y para permitir compartir información entre investigadores. Como consecuencia, investigadores pueden obtener acceso rápido a nuevos resultados, el público puede “ver que están haciendo los científicos”, y se evita la doble adquisición de datos, ahorrando tiempo y fondos. El centro de investigación CEDEUS implementó un observatorio el cual consiste (i) en una IDE, para dar acceso a datos, (ii) un conjunto de herramientas para la comunicación con usuarios, y (iii) herramientas para análisis y procesamiento de datos.

En los dos años desde su operación pública, el número de usuarios crece lento, pero continuamente. La clave para el éxito consiste en atraer nuevos usuarios, sabiendo quienes son los usuarios potenciales y, segundo: desarrollar estrategias de cómo llegar a estas personas. En el caso del Observatorio CEDEUS estos usuarios son los investigadores de CEDEUS, sus alumnos, investigadores e instituciones afiliados, y el público general. Para atraer a los investigadores y (unos) ciudadanos se consideró importante el uso de Twitter para difundir noticias y mapas. Para ganar usuarios como alumnos la clave no es la comunicación, pero sí la capacitación en forma de talleres, dado que muchos de los alumnos no tienen la oportunidad de tomar un curso en SIG, pero tienen alta motivación de conocer nuevas herramientas. Además, pensando y actuando como unidad de servicio visitamos y atendemos las necesidades de los investigadores - para proveer datos atractivos. Finalmente, para convertir al público general en usuario de la IDE es evidente la necesidad de tener los datos indexados en Google.

Agradecimientos

Los autores agradecen soporte y financiamiento por CEDEUS y Conicyt (Conicyt/Fondap/15110020, Fondecyt/1150239).

Bibliografía

- Coetzee, S. & Bishop, J. 2009. Address Databases for National SDI: Comparing the Novel Data Grid Approach to Data Harvesting and Federated Databases. *International Journal of Geographical Information Science*. 23 (9): 1179–1209.
- Costello, C. 2012. *Elgg 1.8 Social Networking: Create, Customize, and Deploy Your Very Own Social Networking Site with Elgg*. Birmingham: Packt Publishing.
- Drummond, W.J., & French, S.P. 2008. The future of GIS in planning: converging technologies and diverging interests. *Journal of the American Planning Association*. 74 (2): 161–74.
- Editorial. 2016. Announcement: Where are the data? *Nature*. 537 (7619): 138–138. doi:10.1038/537138a.
- Haggett, P. 2001. *Geography: A Global Synthesis*. 4th ed. Englewood Cliffs, NJ: Prentice Hall.
- McDermid, G.J., Franklin, S. E., LeDrew, E. 2005. Remote sensing for large-area habitat mapping. *Progress in Physical Geography*. 29 (4): 449–74.
- Morin, A., Urban, J., Adams, P. D., Foster, I., Sali, A., Baker, D., Sliz, P. 2012. Shining light into black boxes. *Science*. 336 (6078): 159–60.
- Rey, S.J. 2014. Open Regional Science. *The Annals of Regional Science*. 52 (3): 825–37.
- Singleton, A.D., Spielman, S., Brunsdon, C. 2016. Establishing a framework for open geographic information science. *International Journal of Geographical Information Science*. 30 (8): 1507–21.
- Steiniger, S., Hunter, A.J.S. 2013. The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption. *Computers, Environment and Urban Systems*. 39(1): 136-150.
- Thill, J.-C. 2000. Geographic information systems for transportation in perspective. *Transportation Research Part C: Emerging Technologies*. 8 (1–6): 3–12.