

Publicación de datos geográficos utilizando software libre en el ámbito público

Daniel Fernandez¹

¹ Instituto Geográfico Nacional, Dirección General de Servicios Geográficos,
Av. Cabildo 381, Capital Federal, Argentina
dfernandez@ign.gob.ar
<http://www.ign.gob.ar>

Resumen. Múltiples organismos públicos cuentan con bases de datos, tanto informáticas como de otra índole, con valiosa información geoespacial. Por otro lado, el movimiento de apertura de datos a nivel nacional, provincial y municipal o departamental en la República Argentina ya es un hecho. Se considera que la información representada, en muchos de los casos, tendría un valor superlativo si a la misma podría adicionarse datos que permitan georreferenciarla y hacer estos datos de acceso público. El proyecto espera contribuir en la implementación de un sistema de publicación de datos geográficos haciendo uso de software libre dentro del ámbito de una organización pública.

1 Introducción

Múltiples organismos públicos cuentan con bases de datos, tanto informáticas como de otra índole, con valiosa información geoespacial. Por otro lado, el movimiento de apertura de datos a nivel nacional, provincial y municipal o departamental en la República Argentina ya es un hecho, conociendo múltiples portales de acceso a dicha información, como ser, Datos Argentina de la Secretaría de Modernización [<https://datos.gob.ar/>], Portal de Información de Ciencia y Tecnología Argentino del Ministerio de Educación, Cultura, Ciencia y Tecnología [<http://datos.mincyt.gob.ar/>], Datos Abiertos del Municipio de Tigre [<http://www.tigre.gov.ar/datosabiertos/>], Datos Abiertos del Gobierno de la Provincia de Buenos Aires [<https://www.gba.gov.ar/datosabiertos/>].

Teniendo esto en cuenta, se considera que la información representada, en muchos de los casos, tendría un valor superlativo si a la misma podrían adicionarse datos que permitan georreferenciarla y hacer estos datos de acceso público.

Es por esta razón que el proyecto propone contribuir en la implementación de un sistema de publicación de datos geográficos haciendo uso de software libre dentro del ámbito de una organización pública.

La intención de este trabajo es actuar como punto de partida dentro de un proceso dinámico de construcción donde se irán añadiendo nuevas ideas, nuevos casos, distintas motivaciones y más problemáticas a lo propuesto.

2 Objetivo

Contribuir en la implementación de un sistema de publicación de datos geográficos haciendo uso de software libre dentro del ámbito de una organización pública.

3 Definiciones contextuales

Para lograr un marco contextual en el presente trabajo, es necesario mencionar dos temas fundamentales: software libre y sistemas de información geográfica.

Software libre

En la Sociedad de la Información, compartir conocimiento es lo mismo que redistribuir riqueza. Es base fundamental para el desarrollo de una Sociedad de la Información más democrática y con un poder menos concentrado. Por lo tanto, cuando hablamos de software abierto y no-propietario, de Software Libre, estamos señalando una nueva política tecnológica [1].

Las consecuencias de la implementación de este tipo de software componen un gran abanico, pero las más destacables podrían ser el coste y la combinación de dos poderosos mecanismos, la competencia (pudiendo usar el mismo programa base) y la colaboración (incluso voluntaria) [2].

En cuanto a los procesos de desarrollo en el software libre, se puede mencionar que no está necesariamente asociado con un proceso de desarrollo específico, sino que existe un amplio consenso sobre los procesos más comunes que se utilizan. Esto no quiere decir que no existan proyectos de software libre que hayan sido creados utilizando procesos clásicos como el modelo en cascada. Generalmente el modelo de desarrollo en proyectos de software libre suele ser más informal, debido a que gran parte del equipo de desarrollo realiza esas tareas de manera voluntaria y sin recompensa económica, al menos directa, a cambio [2]. Este proceso informal, está complementado, en el caso de software libre con un alcance moderado, a una comunidad que da soporte, planifica y desarrolla mejoras y, algo de lo más importante, realiza pruebas sobre su funcionamiento.

Sistemas de información geográfica

Hoy en día, es indiscutible que la información georreferenciada ocupa un lugar relevante tanto para entidades públicas como privadas (y hasta para muchas disciplinas e individuos), y no solo para el terreno científico, sino para la vida cotidiana [3]. El

término información georreferenciada se refiere a aquella información a la cual puede asignarse una posición geográfica, y es por tanto información que viene acompañada de otra información adicional relativa a su localización.

El avance de la tecnología ha impulsado aún más esta necesidad de utilizar técnicas de posicionamiento geoespacial de entidades en una localización geográfica única y bien definida en un sistema de coordenadas y datum específicos [4]. De hecho, existen estudios que indican que aproximadamente un 70% de la información que manejamos en cualquier tipo de disciplina está georreferenciada [5].

La implementación de estos tipos de sistemas ha acercado información cartográfica al público más diverso en el mundo [6], hasta el punto en que la elaboración de cartografía ha pasado de ser terreno exclusivo de un sector de profesionales, a ser de acceso público; hasta llegar al punto donde casi cualquier persona con acceso a un celular, puede generar cartografía.

Un SIG es un sistema compuesto por hardware, software y datos geográficos, diseñado para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada con el fin de resolver problemas complejos de planificación y gestión. El SIG funciona como una base de datos con información geográfica (datos alfanuméricos) que se encuentra asociada por un identificador común a los objetos gráficos de un mapa digital. Dada su versatilidad, son capaces de resolver problemas de localización, eficiencia, tendencias y modelos, siendo útiles en una gran rama de disciplinas.

La razón fundamental para utilizar un SIG es la gestión de información geoespacial. El sistema permite separar la información en diferentes capas temáticas y las almacena independientemente, permitiendo trabajar con ellas de manera rápida y sencilla, y facilitando al profesional la posibilidad de relacionar la información existente a través de la topología de los objetos, con el fin de generar otra nueva que no podríamos obtener de otra forma [7].

En un momento, se hizo fundamental que los SIG sean interoperables. Para que esto fuese posible, o sea, para que las funcionalidades de visualizar y manipular datos mediante un entorno de red informática fueran posible, surgió la necesidad de generar catálogos de datos (con sus respectivos metadatos) y distintos servicios de accesos estándar y para ello surgieron las Infraestructuras de Datos Espaciales (IDE) [8].

Las IDE permiten acceder a datos, productos y servicios geoespaciales, publicados en internet bajo estándares y normas definidos, asegurando su interoperabilidad y uso, como así también la propiedad sobre la información por parte de los organismos que la publican y su responsabilidad en la actualización [9].

3 Problemática

La problemática propuesta en el presente trabajo está centrada en un sector particular dentro de las Infraestructuras de Datos Espaciales (IDE) y se trata de la publicación de sus datos teniendo en cuenta las distintas formas con las que se encuentra una organización al momento de decidir cómo publicar sus datos geoespaciales y, a su vez, los elevados costos que esto conlleva y la complejidad del proceso en caso de no contar con una guía adecuada dado que el proceso incluye, por un lado un grupo interdisciplinario de recursos humanos, los cuales deben ser orquestados con cierta destreza para lograr el éxito; y por otro lado, una serie de activos tecnológicos y de software.

A nivel Nacional se encuentra la iniciativa Infraestructura de Datos Espaciales de la República Argentina (IDERA), la cual es una comunidad de información geoespacial que tiene como objetivo propiciar la publicación de datos, productos y servicios, de manera eficiente y oportuna como un aporte fundamental a la democratización del acceso de la información producida por el Estado y diversos actores, y al apoyo en la toma de decisiones en las diferentes actividades de los ámbitos público, privado, académico, no gubernamental y sociedad civil. A través de su representación, IDERA busca mantener un carácter nacional y federal [9].

Las recomendaciones aquí expuestas son compatibles con la filosofía de IDERA como así también con los estándares internacionales impulsados por Open Geospatial Consortium (OGC).

4 Propuesta

El presente trabajo propone una metodología para la implementación de un sistema de información adecuado para publicar datos geográficos dentro del ámbito de una organización pública estatal de modo de que estos datos sean abiertos, centrándose en la parte tecnológica de la solución, dejando fuera del alcance la otra parte igual de importante, los recursos humanos. Se pretende incluir esta última parte en trabajos futuros.

La primera problemática a sobrellevar es el elevado costo de los software propietarios que permiten realizar este tipo de trabajo y, más allá del costo, el impedimento que tienen los organismos estatales al momento de la adquisición de productos internacionales. Es por esta razón que se cree que el aprovechamiento del software libre será mucho más interesante para un organismo público. En la Figura 1 se pueden apreciar los diferentes componentes de un sistema de publicación de datos geoespaciales.

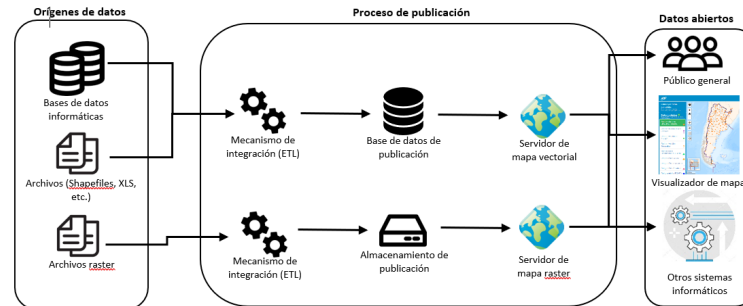


Fig. 1. Principales componentes de la implementación de un sistema de publicación de datos geoespaciales.

Organización y persistencia de los datos

En primera instancia, es importante definir en qué forma la organización tendrá organizados, disponibles y persistidos sus datos. Una propuesta interesante es partir de un catálogo de objetos geográficos predefinidos. Es recomendable pensar este catálogo antes de realizar cualquier otro paso ya que será el que defina como estarán organizados los datos.

Uno de los objetivos principales del establecimiento de una Infraestructura de Datos Espaciales (IDE) es facilitar la disponibilidad y acceso a la información geográfica. Este principio está basado en la interoperabilidad, para lo cual es necesaria la estandarización y la normalización de la información geográfica. Los estándares aplicados a este ámbito corresponden a normas que definen recomendaciones, terminologías, procedimientos y conceptos que aseguran el intercambio de información geográfica [10].

La definición de un Catálogo de Objetos asegura que los procesos de producción de información geográfica sean los ideales, apoya la evaluación de los elementos de calidad, permite bases de datos homogéneas, difundiendo y aportando al intercambio de información a cualquier escala. La catalogación en el marco de la estandarización de la IG tiene un rol fundamental para el desarrollo de una IDE.

Un Catálogo de Objetos Geográficos se basa en la abstracción de la realidad clasificada en conjuntos de datos geográficos, cuyo nivel básico es el tipo de objeto. En el Catálogo se organizan esos Objetos Geográficos, sus definiciones y características, es decir, sus atributos, valores de dominios, relaciones y operaciones.

A modo de ejemplo, en la Figura 2 puede observarse la organización del catálogo IDERA.

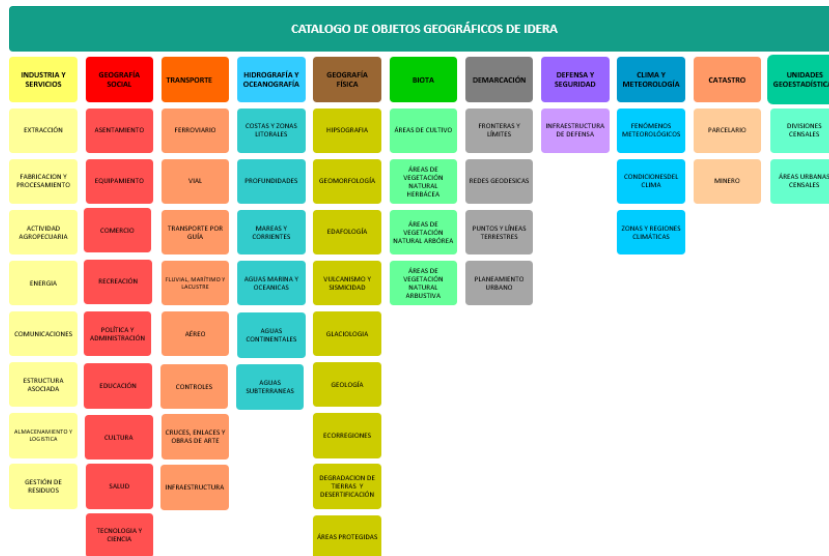


Fig. 2. Organización del catálogo de objetos geográficos propuesto por IDERA.

Algo importante dentro un catálogo es contar con agrupaciones lógicas de cada objeto geográfico para poder tener reunidos ciertos objetos similares o que tratan de la misma temática.

Desde el portal IDERA se puede tener la documentación completa para la confección de un catálogo de objetos geográficos.

Luego de tener los datos catalogados, se recomienda la utilización de una base de datos relacional para la brindar tanto la persistencia como la accesibilidad de los datos a los actores interesados. Dicha base de datos podría ser PostgreSQL dado su famoso complemento PostGIS que brinda todas las propiedades necesarias para convertirse en una base de datos geográfica.

Asimismo, queda a criterio de la organización, la forma en que dicha base de datos será mantenida y actualizada. En la generalidad de los casos, existen múltiples fuentes de información las cuales deberán ser concentradas en una única base de datos, con algún mecanismo de integración, para poder partir la publicación de esa información de una fuente de datos única. La forma en que se importarán los datos en dicha base será tan variada como orígenes de datos y organizaciones existan. Es por ello por lo que no puede recomendarse un único proceso válido para todos los casos.

En caso de que se desee profundizar en los mecanismos de integración, se pueden seguir las recomendaciones de procesos de extracción, transformación y limpieza (ETL) referidos a la minería de datos [11]. Dichos procesos son similares a los requeridos en esta etapa.

A modo de ejemplo, se podría mencionar que se obtienen buenos resultados al utilizar archivos con formato Shapefiles [12] como transportadores de los datos ya que tanto el complemento PostGIS mencionado como la mayoría de las herramientas GIS existentes en el mercado, tiene herramientas preparadas para manejar estos tipos de datos y permitir que se importen dentro de una base de datos, formando parte del mecanismo de integración necesario.

Por supuesto, en caso de que la organización cuente con sus datos inmersos en una base de datos informática, los procesos de migración hacia PostgreSQL podrán ser resueltos de una forma más automatizada y, a su vez, con más controles y más seguridad.

Todo lo expuesto hasta el momento, es útil para los datos vectoriales¹. En caso de que se tengan datos en formato raster² su persistencia será una forma distinta, dado que se trata de archivos en su mayoría binarios y que suelen tener un tamaño importante. En la Figura 3 pueden apreciarse las diferencias de representación de estos dos tipos de datos.

Los datos de este último tipo deberán ser almacenados en un espacio de disco que siga las normas básicas de la seguridad informática para poder tener cierto nivel de prestaciones. Queda fuera del alcance de este trabajo mencionar aspectos relativos a seguridad informática en cuanto al almacenamientos de archivos, aunque se recomendando consultas la norma del Banco Central de la República Argentina en lo que respecta al resguardo de la información [13].

En caso de contar con alta demanda de información o mucha cantidad de datos publicados, se recomienda utilizar un servidor de mapas para los datos vectoriales y otro para los datos de tipo raster. Esto es, por un lado, para dividir la carga de trabajo, y por otro para configurar los servidores de mapas de manera optima y especializada para un único tipo de dato, ya que, en ciertas ocasiones, la configuración óptima para datos vectoriales no es la misma que para datos raster y viceversa.

¹ Estructura de datos utilizada para almacenar datos geográficos. Los datos vectoriales utilizan geometrías (puntos, líneas y polígonos) para representar el mundo real.

² Los raster se componen de una matriz de píxeles (también llamadas celdas), cada uno con un valor que representa las condiciones de la zona cubierta por dicha celda.

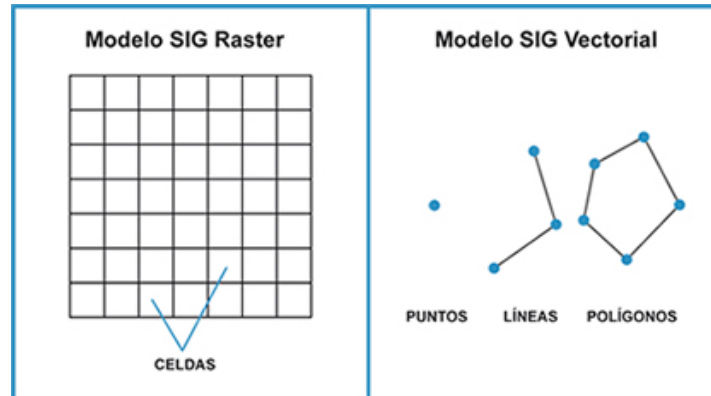


Fig. 3. Diferencias entre el modelo de representación de datos raster y vectorial.

En el caso de necesitar insertar, editar, mantener o revisar datos geoespaciales, se propone la utilización de la herramienta QGIS [14] como herramienta de soporte informático también dentro del software libre. La herramienta mencionada cuenta con todas las funcionalidades necesarias para generar, mantener, procesar y controlar información geoespacial, la misma permite: visualizar información espacial, crear y editar datos, realizar análisis espaciales y hasta crear cartografía. Asimismo, la herramienta cuenta con conexión a los orígenes de datos más utilizados, como ser, archivos Shapefiles, bases de datos, servicios OWS, entre otros [15].

APERTURA DE LOS DATOS

No es recomendable hacer públicos los datos directamente desde la base de datos, sino haciendo uso de un servidor de mapas. De esta manera, no solo permite agregar una capa de seguridad al acceso de la información geográfica, sino también hacer una distribución siguiendo varios de los estándares de publicación más importantes a nivel mundial.

Estos tipos de servidores, además de permitir la visualización de datos como mapas, permite compartir información en múltiples formatos. Los más reconocidos y utilizados de estos formatos son los compatibles con estándares propuestos por OGC [16], ellos son WMS, WMTS, WFS y TMS. Asimismo, también dan soporte a otros formatos que no están dentro de OGC, como ser, KML, Shapefiles, entre otros.

Un poderoso servidor de mapas y de software libre es Geoserver [17]. Entre sus características más destacadas, se puede mencionar: cuenta con una variedad de acceso a almacenamientos de datos tanto para datos vectoriales como para raster; las solicitudes pueden ser atendidas como imágenes rápidas y seguras mediante el protocolo WMTS; soporta el envío de datos vectoriales mediante el protocolo WFS; permite realizar proyecciones “al vuelo”; cuenta con el complemento GeoWebCache [18] integrado, permitiendo realizar caché de mosaicos; cuenta con una administración facilitada por su entorno gráfico.

Un aspecto muy importante dentro del servidor de mapas es que el mismo cuente con soporte para procesos de memoria intermedia (o caché) [19] de modo tal que, al contar con grandes volúmenes de datos y alta demanda de solicitudes por parte de los clientes, se pueda hacer uso de este mecanismo para mejorar los tiempos de respuesta del servidor ante las dichas solicitudes.

PUBLICACION DE LOS DATOS

La forma en que una organización desea publicar sus datos puede ser tan basta como organizaciones existen, pero, la recomendación básica es publicar los datos en formatos de archivos para su descarga y en formato visual para ser accedido desde herramientas gráficas. Ambas formas son soportadas por el servidor de mapas Geoserver recomendado.

Mediante solicitudes a dicho servidor, se pueden obtener distintos formatos de archivos listos para ser publicados, como ser KML, Shapefiles, GeoPackage, entre otros. Este mismo servidor, permite ser accedido mediante servicios como WMS, WMTS y WFS mediante herramientas SIG convirtiéndose en fuente de datos para dichas herramientas.

En cuanto a la publicación de datos de manera visual y sobre plataformas web, se pueden hacer uso de las múltiples bibliotecas de mapas existentes. Una de las más conocidas dentro del software libre se trata de Leaflet [20]. Una de las características importantes de esta biblioteca, es su tamaño. En su versión básica pesa alrededor de 150Kb y esto tiene relación con la segunda de sus características sobresalientes y es que está diseñado con una modalidad modular, o sea, la versión básica trae soporte solo para lo más básico de un visualizador de mapas (mostrar mapas base y capas mediante WMS). El resto de las funcionalidades que se desean agregar se deben utilizar como complementos, o sea, que deben ser adquiridas por separado del código básico. Esta modularidad permite tener un mayor control sobre los complementos que realmente se desean utilizar y no cargar complementos indeseados. Asimismo, desde el sitio web oficial de la biblioteca se puede acceder a un centenar de complementos listos para ser utilizados.

Como una alternativa interesante al uso de Leaflet de forma directa, el Instituto Geográfico Nacional (IGN) de la República Argentina ha desarrollado un visualizador de mapa que está pensado para ser utilizado por el público en general sin necesidad de tener conocimientos de desarrollo de software. El mismo se encuentra disponible en <https://github.com/ign-argentina/argenmap> para su utilización.

ARGENMAP

Argenmap es un proyecto financiado y desarrollado por el IGN [21]. Surge a raíz de la necesidad de publicación de datos geográficos de esta institución y se enmarca dentro de un proyecto superior cuyo objetivo es realizar una reingeniería del proceso de publicación de datos geoespaciales.

El objetivo de Argenmap es desarrollar un visualizador de mapas orientado al público general, compatible con estándares y reutilizable [22].

Sus características principales son:

- **Fácil implementación:** los pasos para su implementación básica son: descargar el código, configurar los orígenes de datos (desde el archivo de configuración) y seleccionar una plantilla (también desde el mismo archivo de configuración).
- **Multiplataforma:** está desarrollado con tecnologías web, en su mayoría con el lenguaje de programación Javascript. Funciona en navegadores de escritorio y móviles. Se implementa en la mayoría de los servidores web comerciales sin necesidad de componentes adicionales.
- **Arquitectura modular (Figura 4):** se basa en dos módulos principales, el primero es de presentación, y el segundo el listado de capas; de modo tal de facilitar el desarrollo, escalabilidad y mantenimiento. La plantilla personalizable permite cambiar desde el estilo del mapa hasta la biblioteca de visualización a utilizar (Leaflet, Openlayer, ArcGIS, etc.).

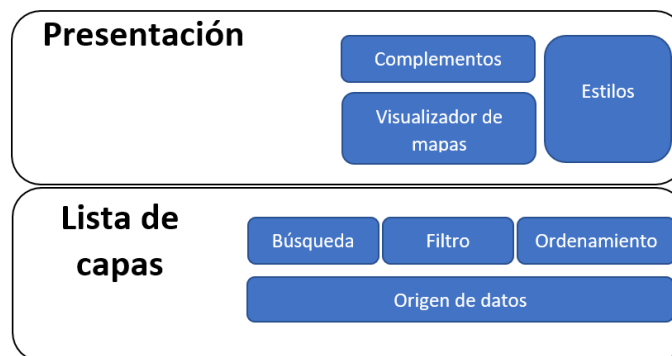


Fig. 4. Arquitectura del visualizador de mapas Argenmap.

- **Código abierto y libre:** basado en componentes de software libre y distribuido del mismo modo, asegurando que el conocimiento desarrollado sea público y reutilizable. Publicado en GitHub bajo licencia GPL v3.0.
- **Mapa base oficial del IGN:** incluye de forma predeterminada el mapa base oficial del Instituto Geográfico Nacional, poniendo a disposición de los usuarios un mapa oficial de la República Argentina. Además, se incluye un archivo de configuración de ejemplo con algunos WMS del IGN.
- **Compatible con datos heterogéneos:** El listado de capas admite actualmente como orígenes de datos a servicios WMS, WMTS, TMS. Se prevé la inclusión de más tipos de datos, tanto de servicios OGC como otros formatos de archivos.

- Herramientas básicas: la plantilla básica utiliza Leaflet como visualizador de mapa y viene con las herramientas zoom y paneo, posición actual, pantalla completa, cuadrícula, medición de distancias y herramientas de dibujo.

7 Resultados esperados

Mediante la implementación del sistema informático propuesto se espera contar con un proceso de publicación de datos geográficos con un alto nivel de servicio, haciendo uso de software libre y permitiendo abrir los datos al público general. Bajando los costos de implementación en cuanto a licencias y demás exigencias de software privativo y concentrando los recursos financieros en el desarrollo de los recursos humanos necesarios para la implementación.

Otros resultados esperados es el aporte que se realizará a la comunidad de cada uno de estos software utilizados, permitiendo mejorarlos.

Asimismo, se espera utilizar este material como punto de partida y, mediante sucesivas mejoras, ir profundizando tanto en aspectos técnicos como así también contando experiencias de implementaciones (ya sean exitosas o no) y generar una transferencia del conocimiento adquirido de manera libre, abierta y gratuita.

Referencias

1. Branco, M. D. Software Libre en la Administración Pública Brasileña. Barcelona, España: Universitat Oberta de Catalunya (2015).
2. González J., Seoane J., Robles G. Introducción al software libre. Barcelona, España: Universitat Oberta de Catalunya (2008).
3. Silva, A. El uso de información censal para análisis medioambiental. Santiago, Chile: Comisión Económica para América Latina y el Caribe (2010).
4. Nieto Masot, A., Cárdenas Alonso, G. Sistemas de información geográfica y teledetección: aplicaciones en el análisis territorial. Extremadura, España: Grupo de Investigación Geo-Ambiental de la Universidad de Extremadura (2018).
5. Olaya, V. Sistemas de Información Geográfica. CreateSpace Independent Publishing Platform (2014).
6. Sarría, F. Sistemas de Información Geográfica. Murcia, España: Universidad de Murcia (2006).
7. Buzai, G. Sistemas de Información Geográfica (SIG) y cartografía temática. Lugar Editorial (2008).

8. Valencia Martínez de Antuña, J. Pasado, presente y futuro de las infraestructuras de datos espaciales (2011).
9. IDERA. (s.f.). Infraestructura de Datos Espaciales de la República Argentina. Recuperado el 25 de 03 de 2019, de <https://www.idera.gob.ar/>
10. IDERA. (s.f.). Infraestructura de Datos Espaciales de la República Argentina. Recuperado el 25 de 03 de 2019, de https://www.idera.gob.ar/images/stories/downloads/catalogo/Descripcion_Catalogo_IDERA_V3.pdf
11. Hernández Orallo, J., Ramirez, M., et al. Introducción a la minería de datos. Madrid: Pearson Prentice Hall (2004).
12. Gómez, F. Compartiendo Información Publica Georreferenciada. VII Congreso de Exploracion y Desarrollo de Hidrocarburos. Mar del Plata, Argentina (2008).
13. Banco Central de la República Argentina, Comunicación “A” 4609: Requisitos mínimos de gestión, implementación y control de los riesgos relacionados con tecnología informática y sistemas de información (2006).
14. QGIS. (s.f.). Sistema de Información Geográfica libre y de Código Abierto . Recuperado el 18 de 03 de 2019, de <https://www.qgis.org>
15. Graser, A. Learning QGIS. Birmingham, UK: Packt Publishing (2016).
16. Goodchild, M., Egenhofer, M., & et al. Interoperating geographic information system. New York, USA: Springer Science+Business Media (1999).
17. Iacovella, S., Youngblood, B. GeoServer Beginner's Guide. Birmingham, UK: Packt Publishing (2013).
18. Huayi, X. G. A High-Concurrency Web Map Tile Service Built with Open-Source Software. WuhanP.R. China: Springer Science+Business Media (2013).
19. Martínez, N., Wladimir, A. Evaluación del rendimiento de los servicios WMS de Mapserver y Geoserver para la implementación IDE. Sangolquí, Ecuador: Universidad de las Fuerzas Armadas (2013).
20. Crickard, P. Leaflet.js Essentials. Birmingham, UK: Packt Publishing (2014).
21. IGN. (s.f.). Instituto Geográfico Nacional. Recuperado el 15 de 04 de 2019, de <https://github.com/ign-argentina/argenmap>
22. Perez P., Fernandez, D., et al. Argenmap: nueva versión del visualizador de mapas del IGN. FOSS4G-AR. Capital Federal, Argentina (2019).