

Metodología de análisis y estudio comparativo de Infraestructuras de Datos Espaciales en el territorio español

Juan Ignacio Varela García¹, Daniel Díaz Grandío¹, Fco. Alberto Varela García¹,
Manuel Gallego Priego² e Inmaculada Serantes Durán².

¹CartoLab
Universidade da Coruña
C/ Campus de Elviña, A Coruña
nachouve@gmail.com
cartolab@udc.es

²SITGA
C/ Estrada de Noia Km 1, Santiago de Compostela
manuel.gallego.priego@xunta.es
inmaserantes@hotmail.com

Resumen

OGCAnalyzer es una herramienta en desarrollo que nace con el fin de ayudar a conseguir una metodología objetiva y eficaz para valorar y comparar IDEs. Permite extraer información de los servicios OGC de forma masiva y puede resultar de gran utilidad para determinar aspectos a mejorar en este tipo de infraestructuras. En el presente artículo se hace una introducción a la problemática del análisis y control externo de IDEs y se presenta además un extracto de algunos resultados interesantes obtenidos con la metodología desarrollada.

Palabras clave: IDE, análisis, metodología, OGCAnalyzer

1 Introducción

Este estudio pretende ser una aproximación a unos criterios objetivos para el análisis y valoración de Infraestructuras de Datos Espaciales. Para ello, se han estudiado una serie de IDEs, en este caso las IDEs regionales, abarcando todo el territorio español. La metodología de análisis definida nos permite evaluar ciertos aspectos interesantes de estos servicios, establecer unos valores numéricos y poder compararlos entre sí. A pesar de estar en una fase inicial, las métricas utilizadas, así como los desarrollos informáticos creados, dejan la puerta abierta para nuevas propuestas que amplíen o proporcionen una mejor evaluación del funcionamiento de estas plataformas.

1.1 Las IDEs: importancia y marco europeo

Las Infraestructuras de Datos Espaciales (IDE) juegan un importante rol en la difusión, utilización y mejora de la Información Geográfica (IG). La IG es un recurso muy valioso para la sociedad y disponer de una IDE eficaz permite optimizar su uso a la vez que se reducen costes. En los últimos años son muchos los organismos que han puesto en marcha plataformas de este tipo, sobre todo a raíz del impulso en este campo que ha significado la Directiva INSPIRE[1].

La directiva INSPIRE delega gran responsabilidad en las Infraestructuras Nacionales de Datos Espaciales y éstas a su vez en las IDEs autonómicas, provinciales y locales. Esta estructura jerárquica basada en niveles provoca la aparición de diferencias entre nodos IDE que afectan a su organización, implementación, disponibilidad y, evidentemente, a la información ofertada. Se debe resaltar el papel del Comité Técnico ISO/TC 211 para la Información Geográfica así como del Open Geospatial Consortium (OGC)[2] en su labor de definición de estándares que garanticen la interoperabilidad. Los estándares son un punto común para todas las IDEs a la hora de afrontar su creación a nivel técnico. Sin embargo, los perfiles de los usuarios y los usos que realizan de la información suelen ser muy dispares. Además, es difícil ofertar un servicio homogéneo en una red tan amplia y compleja como la propuesta en la directiva INSPIRE.

Se necesitan guías detalladas sobre cómo ha de desplegarse una IDE más allá de las propias especificaciones técnicas y los estándares. En este sentido, la directiva INSPIRE, no solamente pone las bases para que las IDEs proliferen, sino que propone una serie de recomendaciones sobre aspectos más concretos de cómo deben ser esas IDEs. A su vez, el Consejo Superior Geográfico también enfatiza a

este respecto con unas recomendaciones basadas en las de la propia directiva INSPIRE y desarrolladas por el "Grupo de Trabajo de la IDEE". Sin embargo, el carácter opcional de estas recomendaciones hace que su seguimiento no sea generalizado y muchas IDEs no las cumplan.

1.2 Por qué estudiar las IDEs

El auge en los últimos años de las IDEs, y la rápida aparición de nuevas plataformas en España, nos invita a realizar una reflexión acerca de la oferta actual de estos espacios de información. Una IDE es una tecnología compleja y los datos que ésta alberga exigen un tratamiento, gestión y organización no trivial. Poner en marcha y mantener una Infraestructuras de Datos Espaciales requiere muchos recursos y una notable inversión.

Dado el alto interés de la sociedad y el creciente respaldo político al acceso público de la información, parece claro que se necesitan estudios que permitan valorar estas plataformas, hacer comparativas e incluso determinar si su funcionamiento y servicios son adecuados a las expectativas que podría esperar un usuario. Además, no existen criterios estandarizados para determinar la calidad de una IDE en cuanto a la información alojada, jerarquía de datos, tiempo de respuesta, etc.

La Infraestructura de Datos Espaciales de España (IDEE)[3] hace una labor centralizadora de todas estas iniciativas en el territorio nacional, siendo un lugar de referencia para la consulta y búsqueda de servicios IDE. Sin embargo, parece que los nodos IDEs son muy heterogéneos. Las diferencias pueden afectar al nivel de calidad y/o cantidad de datos y servicios ofrecidos en distintas regiones. Esto supone un agravio comparativo para los ciudadanos, instituciones, empresas o investigadores de esos lugares con IDEs menos eficientes. Por ello, el cuantificar estas diferencias puede suponer un gran impulso para que se tomen medidas y que situaciones de este tipo se equilibren.

2 Metodología

En este estudio se han desarrollado dos metodologías para el estudio de las IDEs: por un lado, una metodología basada en procesos manuales de supervisión y diagnóstico de portales web con geoservicios, y por otro, una metodología orientada a automatizar, empleando una herramienta informática propia (todavía en

sus primeras fases de desarrollo) análisis de servicios del OGC. Estas dos metodologías son complementarias e independientes.

En este punto, se debe decir que hemos empleado el documento de Recomendaciones elaboradas por el GT-IDEA [4] como base de partida para determinar los parámetros a tener en cuenta a la hora de analizar los elementos que constituyen las IDEs.

2.1 Estudio asistido de geoportales

En esta primera fase del estudio, se ha procedido a analizar únicamente las IDEs regionales¹. De esta manera quedan fuera del estudio el resto de IDEs nacionales y locales, ya que excede el marco de este artículo. Se han estudiado en total 18 IDEs de 16 CC.AA. que a día de hoy disfrutan de este servicio.

Se ha intentado evitar la subjetividad de los parámetros a estudiar escogiendo sólo aquellos que pudieran ser medibles y no sujetos a opiniones personales. El proceso de estudio se lleva a cabo diagnosticando y revisando los siguientes atributos:

- **Geoportal:** Se ha recogido la existencia o no de Geoportal, considerando este como un sitio de internet que ofrece al usuario, de forma fácil e integrada, una serie de recursos y servicios espaciales.
- **Información General:** se ha recogido la existencia o no de referencias a la directiva INSPIRE.
- **Información Genérica:** se ha recogido si en la Geoportal se nos ofrece documentación sobre la propia IDE objeto de estudio.
- **Datos:** se considera si en la IDE es posible consultar o descargar información tabulada en cualquier formato.
- **Metadatos:** se considera si en la IDE es posible consultar o descargar información tabulada de los metadatos en cualquier formato.
- **Visor Cartográfico:** el objeto es comprobar si la IDE posee o no un visor cartográfico capaz de visualizar información geográfica y realizar operaciones básicas de navegación.
- **Facilidad de localización de la IDE:** para completar este campo hemos empleado la herramienta de búsqueda de Google. Se ha procedido a

¹ Las URLs son las listadas en el directorio de la web del IDEA. Acceso en Octubre del 2008.

realizar una consulta con el siguiente formato <IDE+NOMBRE IDE>. Se dió el valor F=1 (fácil de encontrar) a aquellas IDEs que aparecían entre los 10 primeros resultados, dando en caso contrario el valor D=0 (difícil de encontrar)

- **WMS, WFS y WCS:** se ha realizado una revisión de servicios ofertados por cada uno de los geoportales estudiados.

Como resultado de este estudio, se ha definido un Índice de Calidad de Servicio Ofertado (ICSO) ponderando, según su importancia, atributos como el acceso a datos, metadatos o tipo de servicios ofertados (WMS, WFS o WCS) frente a otros como persona de contacto o información general. El rango de valores del ICSO varía entre 0 y 60.

2.2 Análisis automatizados de servicios

El uso de programas que automatizan la recogida de datos y análisis de manera mecánica permiten reducir los tiempos dedicados a este tipo de tareas normalmente repetitivas. Además aumentan la capacidad de volumen de información que se puede tratar pudiendo ampliar el ámbito de estudio sin grandes penalizaciones temporales.

La idea de extraer de forma masiva datos accesibles públicamente surge del conocimiento y uso de las libressoft-tools[5]. Estas herramientas desarrolladas por el grupo GSyC/LibreSoft de la Universidad Rey Juan Carlos I (URJC) de Madrid [6] son capaces de obtener gran información de recursos como repositorios de código fuente, listas de correo, etc.

Para este estudio, se abordó el diseño e implementación de una herramienta propia para llevar a cabo las tareas de análisis llamada OGCAalyzer. Esta aplicación, pese a encontrarse en fase de gestación, ya cuenta con numerosas funcionalidades. Como su propio nombre indica, se trata de un programa informático pensado para realizar análisis de servicios definidos por el OGC, fundamentalmente WMS y WFS aunque su estructura arquitectónica permite su ampliación y el aprovechamiento de muchas de sus funcionalidades.

Como en el estudio de geoportales explicado en la sección 2.1, el origen de servicios a estudiar fue creado a partir de la lista de Servicios IDE regionales facilitada en la página web del IDEE. Por otro lado, se definió un paquete de análisis a realizar sobre los servicios seleccionados. Entre ellos, está el conteo del

número de capas de información, número de ortofotografías, comprobación de formatos de salida, etc. Entre los análisis que consideramos más interesantes está el grado de metadatos ofrecidos. Para ello, se han seleccionado algunos de los metadatos recogidos como mínimos en el Núcleo Español de Metadatos (NEM) conformando un subconjunto de los mismos. Entre ellos se contemplan, por ejemplo, datos tan básicos como información sobre la persona de contacto de ese servicio, los títulos de las capas, los sistemas de referencia de coordenadas soportados, etc.

En esta fase, estos metadatos se consideran válidos si existe una cadena de texto no nula en ese campo. Los datos relativos a la presencia o no de esos metadatos "básicos" se sintetizan con un tanto por ciento de presencia, correspondiendo el valor máximo 100 con un IDE que cuente con el total de los "tags" cumplimentados.

Los análisis con el programa fueron programados para ejecutarse de madrugada evitando conflictos en horas de oficina donde existen más usuarios potenciales.

3 OGCAalyzer

Como ya se ha comentado en la metodología, OGCAalyzer es un sistema informático creado *ad-hoc* para el estudio de Infraestructuras de Datos Espaciales. En concreto, OGCAalyzer nace con el fin de automatizar ciertas comprobaciones en servicios de OGC para realizar estadísticas, comparativas, diagnósticos, etc.

En las siguientes secciones veremos más detalles sobre OGCAalyzer, sus características y las posibilidades que ofrece a día de hoy.

3.1 Descripción técnica

OGCAalyzer se compone de un conjunto de funcionalidades informáticas que facilitan el chequeo y extracción de datos sobre servicios de OGC. Esto es posible debido a que las especificaciones de estos estándares son públicas existiendo varios clientes/librerías para acceder a servicios OGC. Hemos escogido la librería *owslib-0.3* [7] escrita en python y que cuenta con una licencia de software libre que permite libertad para su uso, acceso al código fuente, modificación y distribución.

Owslib-0.3, desarrollado por el "Ancient World Mapping Center" de la Universidad de "North Carolina"[8], tiene soporte para WMS, WFS y WCS y proporciona una API para el acceso a metadatos y wrappers para realizar peticiones "GetCapabilities", "GetMap", y "GetFeature".

En la figura 1, se puede ver la estructura modular de OGCAalyzer. Utiliza como núcleo owslib-0.3 recubriéndola para facilitar la ejecución de análisis más complejos a partir de las funciones básicas.

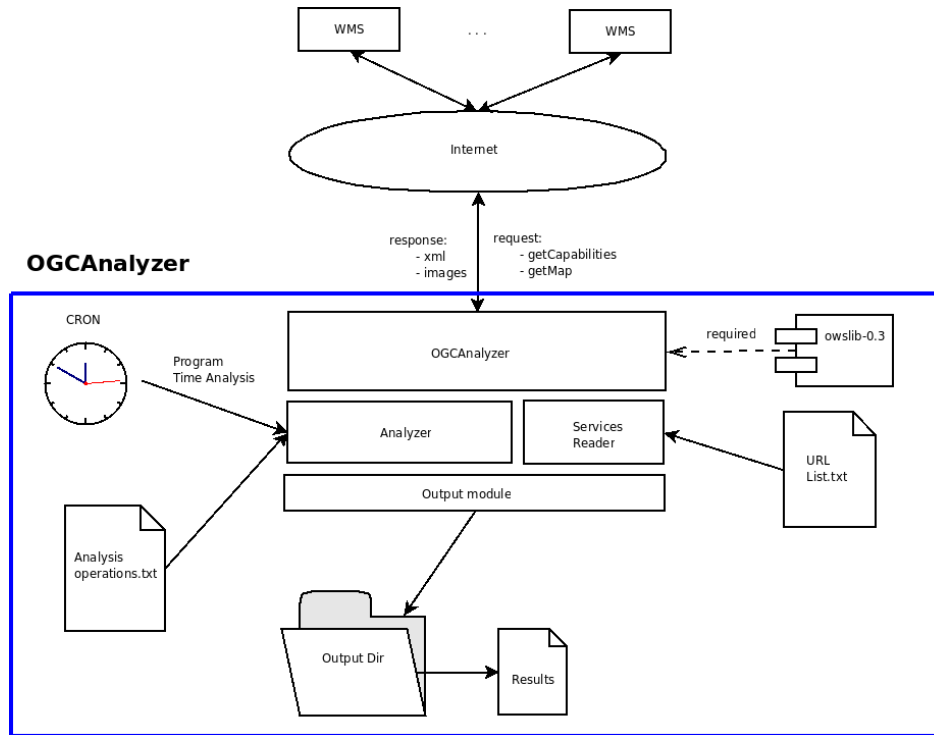


Figura 1: Esquema del OGCAalyzer

OGCAalyzer se divide en módulos que se encargan de tareas muy específicas. Por ejemplo, existen módulos que se encargan de la configuración de parámetros generales del sistema, de la carga de listados de direcciones de servicios a estudiar, de la comunicación con owslib, etc.

Debido a que está pensada para fines de investigación, actualmente no cuenta con interfaz gráfica, estando orientada a la línea de comandos o mediante la programación de scripts. Para obtener la totalidad de sus propiedades debe ser usado en sistemas Linux.

La configuración básica de la herramienta se realiza mediante la edición de ficheros de texto plano. Los resultados de los análisis se pueden visualizar a través de la salida estándar o con ficheros de texto o CSV.

3.2 Características

OGCAnalyzer cuenta con una serie de funcionalidades que lo dotan de mucha potencia a la hora de afrontar el diseño de análisis para estudiar IDEs. Se pasa a listar algunos de ellos:

- Tests de funcionamiento de operaciones básicas.
- Comprueba si en caso de error las excepciones son normativas.
- Examina los formatos de imagen soportados.
- Verifica que sistema de referencia de coordenadas esté en EPSG.
- Metadatos.
- Disponibilidad y Tiempos de respuesta.

4 Resultados

4.1 Extracto del estudio asistido de Geoportales

El resultado de esta parte del estudio queda reflejado en la siguiente tabla. En ella podemos apreciar los diferentes índices obtenidos por cada una de las IDEs.

La media de las calificaciones del ICSO es 32.07, mostrando en color verde aquellas IDEs que están por encima de la media, y en rojo quedan el resto.

NOMBRE	ICSO	NOMBRE	ICSO
IDE Andalucía (IdeAndalucia)	43,6	IDE OTALEX	27,3
IDE Aragón (SITAR)	43,6	IDE Galicia (IDEG)	32,7
IDE Canarias (IDECanarias)	43,6	IDE Galicia-N. Portugal	38,2
IDE Castilla la Mancha (IDEclm)	38,2	IDE Illes Balears	27,3
IDE Castilla y León	10,9	IDERioja	49,1
IDE Cataluña (IDEC)	43,6	GeoEuskadi	43,6
IDE Costas Catalunya	32,7	SITPA-IDEASturias	27,3
IDE Navarra (IDENA)	38,2	IDEPAsturias	38,2
IDE Comunidad Valenciana	16,4	IDRMurcia	27,3

Tabla 1: Índices de Calidad del Servicio Ofertado por CC.AA.

Citar algunos resultados genéricos negativos, como el hecho de que únicamente el 28% de las IDEs sometidas a estudio permiten la consulta o descarga de metadatos, o que prácticamente el 30% no son fácilmente accesibles con un buscador. En cuanto a los servicios ofertados cabe reseñar el alto índice de IDEs que ofertan el servicio WMS, un 94%, por el contrario, únicamente el 22%, dan la posibilidad de trabajar con WFS.

GEOPORTAL	CONTACTO	INF GRAL	INF.SPEC.	DATOS	METADATOS	VISOR CART.	LOCALIZ. IDE	WMS	WFS	WCS
89%	72%	61%	61%	28%	28%	100%	72%	94%	22%	6%

Tabla 2: Porcentajes generales por atributo analizado

4.2 Extracto resultados con OGCAalyzer

Como prueba de la utilidad de la herramienta OGCAalyzer para el estudio masivo de IDEs, se muestran un par de mapas temáticos donde se plasma el resultado del análisis de aprox. 180 URLs de servicios WMS regionales.

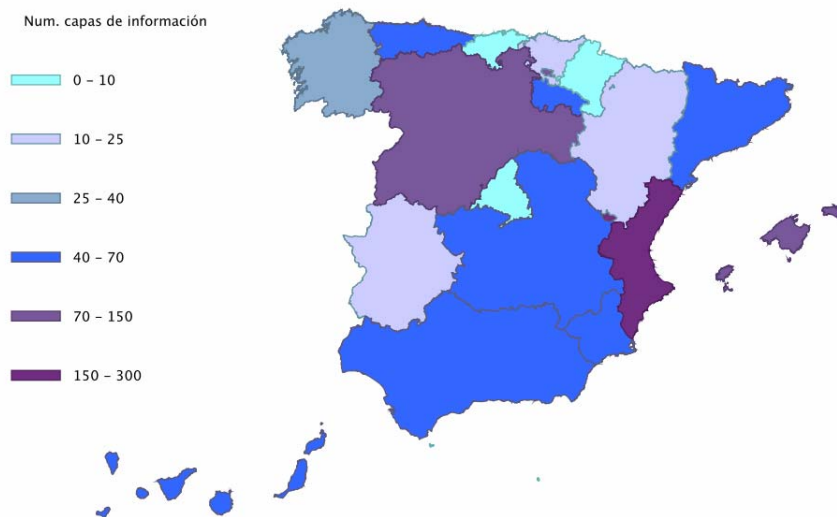


Figura 2: Mapa de número de capas ofrecidas en WMS regionales

En el mapa de la figura 2, se puede apreciar las diferencias en cuanto a número de capas de información ofertadas según la Comunidad Autónoma². La media de capas por cada región es de 68. Como dato interesante podemos decir que actualmente los WMS de este nivel administrativo en España suman un total de más de 3800 capas de información.

En el mapa de la figura 3, se muestra un temático que refleja el grado de existencia de campos de metadatos de información considerada como básicos³. En este caso, podemos apreciar que los metadatos no presentan un alto índice de presencia no llegando la media de este valor al 70%.

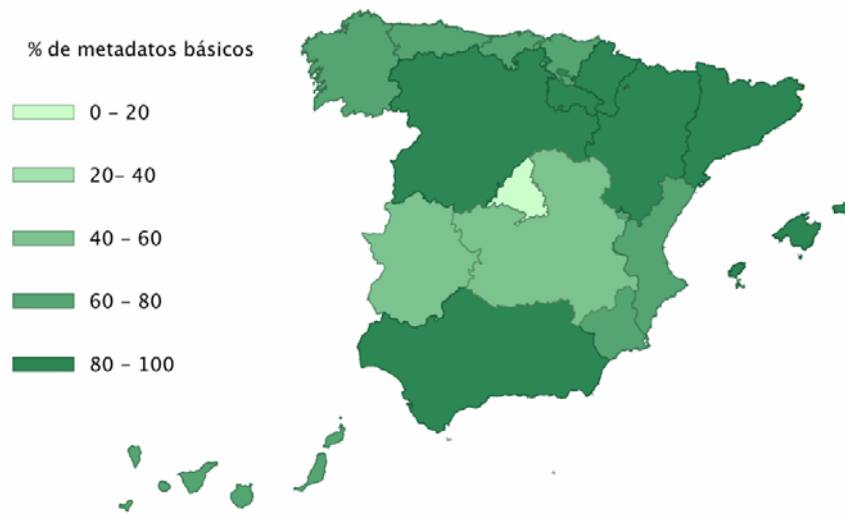


Figura 3: Porcentaje de metadatos "básicos" cubiertos en WMS regionales

5 Conclusiones

² Algunos servicios no se encontraban accesibles en el momento de la realización de las pruebas.

³ La Comunidad de Madrid no contaba con WMS en el momento de realizar este estudio.

El estudio y análisis de las IDEs es una tarea laboriosa y compleja, pero muy necesaria para el buen funcionamiento global de la red de datos espaciales. En este proyecto se ha avanzado en la metodología y en el desarrollo de herramientas que permitan de manera objetiva y semiautomática extraer y valorar gran cantidad de datos acerca de estos servicios.

A pesar de que se trata de un proyecto en fase inicial, los datos obtenidos revelan que el panorama de este tipo de tecnologías en España tiene aún mucho margen para progresar. Además se pueden imaginar las posibilidades que este tipo de análisis pueden ofrecer a las administraciones, instituciones y proveedores de servicios OGC para el chequeo, control y mejora de sus productos.

La automatización de análisis mediante OGCAalyzer parece ser vital para poder abordar el volumen de operaciones involucradas en este tipo de tareas. En el futuro, OGCAalyzer tiene previsiones de incrementar sus funcionalidades y se tiene intención de acometer estudios periódicos más ambiciosos acerca del estado de las IDEs.

Referencias

- [1] Portal de la directiva INSPIRE. <http://www.inspire-geoportal.eu>
- [2] Portal del OGC. <http://www.opengeospatial.org>
- [3] Portal de la IDEE. <http://www.idee.es>
- [4] Recomendaciones para servicios de mapas del GT-IDE.
<http://www.idee.es/resources/recomendacionesCSG/RecomendacionServicioMapas.pdf>
- [5] LibreSoft-Tools. <http://forge.morfeo-project.org/projects/libresoft-tools>
- [6] Página del grupo GSyC/LibreSoft. <http://libresoft.es>
- [7] Proyecto owslib. <http://pypi.python.org/pypi/OWSLib/0.3>
- [8] Grupo AWMC. <http://www.unc.edu/awmc>

