



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

SIOSE EN LA WEB: CÓMO PUBLICAR UN SISTEMA DE INFORMACIÓN COMPLEJO EN LA RED

Fernández Villarino, Xalo ⁽¹⁾; Delgado Hernández, Julián ⁽¹⁾; Valcárcel Sanz, Nuria ⁽¹⁾; Caballero García, M^a Elena ⁽¹⁾; Benito Saz, M^a Ángeles ⁽¹⁾; Porcuna Fernández-Monasterio, Ana ⁽²⁾; Villa Alcázar, Guillermo ⁽¹⁾; Arozarena Villar, Antonio ⁽¹⁾.

⁽¹⁾ Instituto Geográfico Nacional; ⁽²⁾ Tragsatec.

Tras finalizar la producción del Sistema de Información de Ocupación del Suelo en España (SIOSE) se inicia la fase de su publicación y difusión. Esta fase del proyecto debe satisfacer las necesidades de los usuarios dando acceso a la visualización, consulta y análisis de los datos originales del sistema y de sus diferentes vistas temáticas. La base de datos de SIOSE está compuesta por 2.5 millones de polígonos y cerca de 10.5 millones de registros de coberturas de suelo asociados a estos polígonos. Su modelo de datos se completa con 85 clases de entidades simples y compuestas predefinidas que se combinan en un total de 820632 casos de coberturas de suelo diferentes en el conjunto de polígonos. Además, durante la producción, se ha conformado un banco de fotografías georreferenciadas de paisajes de toda España con más de 250000 imágenes que también se quiere publicar como parte del SIOSE.

¿Cómo publicar todo este volumen de información, en toda su complejidad, en la web, mediante tecnologías IDE para que pueda visualizarse, consultarse y explotarse mediante análisis? ¿son los geoservicios web suficientes o es necesario complementarlos con otras soluciones? ¿podemos dar una buena solución a estas cuestiones en el actual entorno de crisis en el que disponemos de recursos económicos, materiales y humanos limitados?

Esta presentación explica el sistema de difusión adoptado para dar respuesta a estos retos y llegar, siendo útiles, a nuestra comunidad de usuarios. Describiremos como hemos organizado y configurado nuestros servicios de visualización y consulta de los datos para mostrar en la red toda la jerarquía de información y los diferentes enfoques de análisis con los que se puede abordar SIOSE, como hemos intentado optimizar el rendimiento de estos servicios para ofrecer tiempos de respuesta adecuados para cualquier posible petición sin dejar de mostrar toda la información.

Mostraremos el encuadre de nuestros componentes IDE dentro de un conjunto más amplio de herramientas de comunicación en la red; que integra el portal web del proyecto, medios de descarga de datos y una herramienta cliente de explotación; que se complementan entre ellas para mejorar la forma de revertir a la sociedad una base de datos de información geográfica y ambiental que se concibe, desde el inicio del proyecto, como un servicio público, como una infraestructura de información territorial



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

básica sobre la que construir estudios, proyectos y servicios de valor añadido, ya sea en forma de utilidad social o económica.

SIOSE en la web: cómo publicar un sistema de información complejo en la red

Xalo Fernández Villarino¹, Julián Delgado Hernández¹, Nuria Valcárcel Sanz¹, M^a Elena Caballero García¹, M^a Ángeles Benito Saz¹, Ana Porcuna Fernández-Monasterio², Guillermo Villa Alcázar¹, Antonio Arozarena Villar¹

¹Subdirección General de Cartografía

Instituto Geográfico Nacional

{xfernandez, jdhernandez, nvalcarcel, mecaballero, mabenito, gmvilla, arozarena}@fomento.es

²Tragsatec

aporcuna@fomento.es

1 Introducción: características de SIOSE. Dificultades para su publicación en web

El Sistema de Información de Ocupación del Suelo de España (SIOSE) es un proyecto del Plan Nacional de Observación del Territorio (PNOT) que recoge, a partir de fotointerpretación de imágenes, la información de coberturas y usos del suelo de nuestro país a escala 1:25.000.

Se concreta, fundamentalmente, en una base de datos geográfica formada por una cobertura completa de España de polígonos de ocupación de suelo y por todos sus atributos semánticos. Además de la base de datos hay otros productos complementarios, como el banco de fotografías de campo o la documentación técnica, y a partir de ella se generan productos derivados en diferentes formatos de archivo y servicios web de visualización y consulta de la información.

Tanto el Instituto Geográfico Nacional (IGN-CNIG) como el resto de administraciones participantes en el proyecto sirven estos datos al público, a través de diferentes canales y diversos formatos, en un marco general de gratuidad y libre distribución. Se entiende la información que proporciona SIOSE como un servicio público y una infraestructura básica de información territorial que posibilita la realización de estudios y la generación de actividad económica con valor añadido.

El marco legal que determina la publicación y el uso de los productos SIOSE está fijado por sendas leyes en los dos campos a los que afecta este sistema de información: el de la información geográfica y el de la información ambiental. Como sistema de información geográfica SIOSE está sujeto a la Ley 14/2010, de 5 de julio, sobre las infraestructuras y los servicios de información geográfica en España (LISIGE) (resultado



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

de la transposición al ordenamiento jurídico español de la Directiva europea INSPIRE, 2007/2/CE) y como información ambiental a la Ley 27/2006, de 18 de julio, por la que se regulan los derechos de acceso a la información, de participación pública y de acceso a

la justicia en materia de medio ambiente (la cual incorpora la Directiva europea 2003/4/CE relativa a esta materia).

El SIOSE se estructura según un modelo de datos orientado a objetos que permite describir los objetos, atributos, relaciones y reglas de consistencia que lo forman. Tiene dos entidades fundamentales: el polígono (geometría) y la cobertura (semántica, no geométrica) asociada al polígono. Este modelo de datos flexibiliza la asignación de información temática a cada polígono ya que posibilita la asociación de una o varias coberturas de suelo diferentes para un único polígono mediante porcentajes de ocupación y atributos. Esto aumenta la riqueza semántica de la información contenida en el sistema respecto a las tradicionales bases de datos jerárquicas de esta temática que se limitan a la clasificación de polígonos en función de una nomenclatura dada. La de SIOSE es una información de ocupación de suelo más compleja pero mucho más versátil en su explotación: con ella se pueden generar tantas consultas o vistas temáticas, partiendo de diferentes puntos de vista, como se necesiten según las distintas necesidades de los usuarios.

1.1 Cifras de SIOSE que dificultan su difusión en la red

Las dificultades que SIOSE plantea para su publicación en web vienen dadas por sus propias características, en concreto dos de ellas: el volumen de datos geográficos que contiene y la complejidad de su modelo de datos.

La base de datos completa de SIOSE está formada por unos 2,5 millones de entidades geométricas poligonales con resolución equivalente a la escala 1:25.000 (tamaño mínimo de polígonos de 0,5 – 2 ha.). Esto supone muchos cientos de millones de pares de coordenadas a transmitir por red y dibujar en pantalla a la hora de representarla. Además, estos polígonos tienen asociadas cerca de 10,5 millones de coberturas de suelo mediante relaciones 1 a n; aunque, ciertamente, estos 10,5 millones de registros alfanuméricos suponen en realidad un menor volumen de datos que los 2,5 millones de registros geométricos. También se quiere publicar, como un producto más del sistema de información, el banco de fotografías de campo georreferenciadas recogidas durante las tareas de control de calidad, que son unas 250.000 fotos de los diferentes paisajes y ocupaciones del suelo existentes en España.

Todos este tamaño de datos plantea un problema de transmisión a una velocidad razonable a la hora de publicarlos en la web. Nuestro condicionante en este caso son las telecomunicaciones y debemos afinar en este sentido la eficiencia de nuestros sistemas para completar la comunicación de la información con un nivel de velocidad tal que se consiga que la percepción del usuario sea de buen funcionamiento, no es suficiente con que las peticiones se respondan correctamente, además debe parecerlo.



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

La segunda cuestión que se nos plantea como dificultad es la complejidad del modelo de datos de SIOSE y su versatilidad semántica. Este modelo de datos se compone de 85 clases de coberturas simples y compuestas predefinidas que pueden combinarse libremente y en cualquier proporción en los polígonos. En la base de datos completa de SIOSE se dan 820.632 casos de coberturas de suelo diferentes y todas ellas pueden analizarse desde un punto de vista global y generalista (por ejemplo analizando la cobertura mayoritaria), tratando de analizar la parte humanizada del territorio (ej.: porcentaje de edificación, porcentaje de suelo sellado), queriendo estudiar las cubiertas de vegetación natural (ej.: fracción de cabida cubierta), extrayendo la información agrícola (ej.: tipología de cultivos), etc.

Esta profundidad semántica nos plantea un problema serio de representación gráfica y cartográfica de los datos, los límites de percepción y los de asimilación de un conjunto extenso de variables comunicadas a la vez juegan en este caso en nuestra contra. Debemos buscar soluciones de representación complementarias que nos ayuden a transmitir el conjunto completo de información que contiene la base de datos y, al mismo tiempo, publicar extracciones temáticas, generalizaciones y clasificaciones agrupadas que ayuden a la mejor comprensión de los datos.

2 Arquitectura de geoservicios

El objetivo inicial con el que planteamos el diseño de los geoservicios web de SIOSE es el de dotar a los usuarios de herramientas que permitan la visualización, consulta y análisis de los datos. Para ello consideramos la utilización de toda la familia de geoservicios web estándar que existen y que, según sus especificaciones técnicas, permiten completar estas tres funcionalidades.

2.1 Visualización

Para diseñar nuestros servicios de visualización debemos tener en cuenta nuestros dos problemas principales: la velocidad de respuesta y la forma de representación gráfica. En cuanto a las alternativas de geoservicio con que contamos, tenemos dos especificaciones estándar: WMS y WMTS.

Teniendo en cuenta que uno de los problemas principales es la velocidad de renderizado de la información (debido a la gran cantidad de geometrías a dibujar) parece claro que los servicios cacheados deberían proporcionar una mejor respuesta en este sentido. Tras diferentes pruebas comprobamos que, en este caso, la teoría se cumple y que, no solo el WMTS es más rápido sino que, para el caso del SIOSE, únicamente los servicios cacheados proporcionan una respuesta de visualización adecuada en un tiempo razonable. Esto es así hasta escalas de visualización del orden



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

de 1:50.000 – 1:25.000, a partir de estas, y para escalas mayores, la velocidad de respuesta de servicios WMS y WMTS similares se iguala, con lo cual pasan a ser más

determinantes las mayores capacidades del servicio WMS (posibilidad de consulta de la información mediante la petición *GetFeatureInfo*).

Una solución alternativa a la utilización de servicios WMTS; que tienen las desventajas de necesitar en todo caso un servicio WMS que cachear, no permitir ninguna consulta de la información y las rigideces derivadas de ser una «foto fija»; es actuar directamente sobre los datos generalizando la información y generando una capa de datos adecuada a cada escala o rango de escalas de visualización. La generalización de las geometrías hace que se mantenga la relación *pares de coordenadas a renderizar:rango de la ventana geográfica a visualizar* con lo que se pueden mantener tiempos de respuesta razonables.

En el caso de los servicios web de visualización de SIOSE consideramos las dos posibilidades, ambas tienen sus pros y sus contras, pero hemos detectado que la generalización de los datos, en este caso, exige una fuerte inversión de potencia de cálculo de máquina y un control muy estricto y preciso del proceso para evitar sesgos indeseados en la información temática del resultado final. En todo caso, la generalización de las geometrías para su visualización en red a todas las escalas exige también una generalización semántica en paralelo y conlleva una pérdida de información.

Para afrontar la segunda de las cuestiones, la forma de representación gráfica para conseguir una transmisión clara de la información, consideramos la posibilidad de publicar diferentes visualizaciones temáticas de SIOSE como distintas capas de un mismo servicio de visualización. De esta manera, partiendo de una visualización genérica básica, el geoservicio de visualización contendrá varias representaciones gráficas predefinidas (las de uso más común y mayor demanda entre nuestros usuarios) de la base de datos de SIOSE.

Esta solución plantea dificultades al combinarla con cualquiera de las dos soluciones contempladas para resolver adecuadamente la velocidad de respuesta de los servicios (uso de servicios cacheados y generalización de las geometrías) dado que multiplica el tamaño de los datos a almacenar (y publicar) en un factor igual al número de capas predefinidas que se quieran generar.

Una solución alternativa, y arquitecturalmente más «elegante» (dado que la información servida es siempre la misma y sólo cambiamos el punto de vista según el que la representamos), sería publicar estas diferentes visualizaciones temáticas de SIOSE como distintos estilos de una misma capa del servicio de visualización. Pero, en la práctica, esta opción se ha desechado dado que numerosas aplicaciones cliente de



visualización no contemplan la posibilidad de cambiar el estilo de visualización de una capa de información, lo que limita las posibilidades de comunicación del servicio.

2.2 Consulta

Para generar servicios web de consulta de geoinformación vectorial la primera opción parece la utilización del estándar de servicio WFS. Sin embargo, tras diversas pruebas con los datos de SIOSE, hemos comprobado que con nuestros condicionantes no es viable la utilización de este estándar para responder a las necesidades de consulta de la información de nuestros usuarios. Esto es debido a las propias características de este estándar, que lo hacen ineficiente, sobre todo en términos de velocidad de respuesta, para algunas de las necesidades a las que, supuestamente, debería responder. En el caso de información geográfica de entidades poligonales, como en SIOSE, la capacidad de respuesta de estos servicios es muy limitada.

Hemos analizado tres posibles utilidades de consulta/descarga de los datos para las cuales la especificación WFS podría parecer la más adecuada: consulta individual de atributos de las entidades, descarga de información y consulta de entidades geográficas como parte de un proceso.

Para resolver consultas puntuales o individuales de geoinformación visualizada en red los servicios WMS con la petición *GetFeatureInfo* adecuadamente configurada pueden ser tan útiles como los servicios WFS, con la ventaja de que son servicios más «ligeros» (y por lo tanto más rápidos) y que, en general, habrán de implementarse como complemento de visualización para cualquier WFS que se desee que cumpla con esta función. La única diferencia es que la entidad geométrica no es llamada propiamente a nuestro ordenador, en forma de XML, pero éste no parece un requerimiento real del usuario que sólo está haciendo una consulta de información. La clave de la cuestión está en explotar al máximo las posibilidades de los servicios WMS y enriquecerlos con respuestas a la petición *GetFeatureInfo* lo más completas posible, de forma que satisfagan las necesidades de información del usuario.

Como utilidad de descarga más o menos masiva de geoinformación a local el servicio WFS se revela claramente más ineficaz que los otros métodos utilizados comúnmente para estos fines (centros de descarga, directorios de ficheros, servidores ftp), que siempre resultan ser más rápidos dado que no tienen que generar el fichero de descarga en tiempo real. La ventaja que puede ofrecer el WFS en cuanto a la mayor flexibilidad y precisión con la que podemos definir la ventana geográfica de datos que deseamos, se ve en la práctica compensada con creces con la desventaja en tiempos de respuesta de la que adolece frente a otros medios de descarga.

Por último, la utilidad de los servicios WFS como parte de procesos más complejos de tratamiento de la geoinformación parece ser la única a la que estos servicios responden adecuadamente. Podríamos decir que cuando el usuario directo del WFS es una



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

persona (que puede ordenar consultas a gran número de entidades y que exige una inmediatez subjetiva a sus peticiones) este servicio no responde adecuadamente; mientras que si el usuario de este servicio es otro servicio o un proceso cerrados y

predefinidos, en los que se ha delimitado adecuadamente el número y tipo de entidades a las que llamar en cada petición, el servicio WFS es el interfaz adecuado para realizar estas tareas, por ejemplo como parte del proceso de un servicio WPS que encadena varias funciones (petición de información, calculo y respuesta).

Por todo lo expuesto anteriormente, para cumplir con las necesidades de consulta de la información de los usuarios de SIOSE hemos optado por la opción de enriquecer al máximo las posibilidades de nuestros servicios WMS. De esta forma daremos respuesta a la primera de las necesidades citadas. Por un lado la capa de polígonos del geoservicio de SIOSE responde a la petición *GetFeatureInfo* con la información de coberturas completa del polígono estructurada en un esquema que la hace más legible y, por otro lado, la capa puntual de fotografías de campo del proyecto devuelve a esta petición la propia imagen.

Para proporcionar la utilidad de descarga de información se ha decidido que la mejor opción pasa por utilizar centros de descarga y servidores ftp, fuera de la estricta arquitectura de servicios IDE, que son más eficientes para el usuario final y que además, en nuestro caso, son medios de publicación de datos que ya estaban operativos en el IGN-CNIG.

2.3 Análisis

En cuanto al análisis y proceso de los datos, dentro de la arquitectura de geoservicios estándar, es la interfaz WPS la que nos ofrece este tipo de funcionalidades. Debido a la complejidad de la información y a la gran variedad de consultas complejas y geoprocursos que se podrían aplicar a SIOSE, no se ha considerado la posibilidad de desarrollar múltiples servicios WPS sino que se ha optado por el desarrollo de una aplicación finalista, del lado cliente, que explote la información mediante acceso directo a la base de datos y que cuente con un abanico de operaciones de consulta espacial y alfanumérica configurables por el usuario. El desarrollo de servicios WPS se plantea como una posibilidad a considerar en un futuro, cuando este estándar sea más estable, más utilizado y cuando el conjunto de servicios y utilidades de visualización y consulta de SIOSE esté más asentado.

3 Estrategia global de publicación de la información geográfica de SIOSE. Encaje de las tecnologías IDE en el conjunto

El conjunto de los geoservicios web de SIOSE son una parte de la estrategia de difusión y puesta a disposición pública de este sistema de información territorial. Por una parte los componentes IDE necesitan ser complementados por otros medios de publicación de datos para completar sus carencias y, por otra parte, es necesario estar presente en



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

todos los canales de comunicación a través de los que nuestros usuarios nos reclaman la información. Esta estrategia de difusión de la información geográfica del proyecto se resume en estar presentes en todos los canales de publicación de datos del IGN-CNIG.

Para ello el SIOSE está presente en el portal web del IGN-CNIG (www.ign.es/siose) y tiene también su propio dominio: www.siose.es. En ambas páginas se puede acceder a toda la documentación técnica del proyecto y a los datos estadísticos. Como ya se ha dicho, la base de datos completa del SIOSE es accesible vía web a través del Centro de Descargas del CNIG (<http://centrodedescargas.cnig.es/CentroDescargas/index.jsp>) en formatos shp (entidades geométricas) y mdb (datos alfanuméricos) y dividida en bloques autonómicos. También se puede acceder a los datos a través del servidor ftp del proyecto (<ftp://ftp.siose.ign.es>). Los geoservicios web de SIOSE pueden utilizarse con cualquier herramienta de visualización y consulta geográfica que admita servicios estándar y, además, el Instituto Geográfico Nacional también publica el SIOSE a través de su visualizador de imágenes Iberpix, que permite ver y consultar la información asociada a los polígonos y las fotografías de campo (<http://www.ign.es/iberpix/visoriberpix/visorign.html>). Como también se ha citado anteriormente, está en desarrollo una herramienta web cliente de consulta y análisis de la información de SIOSE, que permitirá además el análisis combinado con otros conjuntos de datos geográficos. Esta aplicación estará accesible desde el portal web del proyecto.

Por último, cualquier consulta, sugerencia o duda es atendida por el equipo técnico de SIOSE en la dirección de correo de contacto del proyecto: siose@fomento.es.