

Geoservicios web SIOSE: un sistema de información como servicio público.

X. Fernández Villarino⁽¹⁾, J. Delgado Hernández⁽¹⁾, N. Valcárcel Sanz⁽¹⁾, M. E. Caballero García⁽¹⁾, M. A. Benito Saz⁽¹⁾ y A. Porcuna Fernández-Monasterio⁽²⁾

⁽¹⁾ Servicio de Ocupación del Suelo, Subdirección General de Geodesia y Cartografía, Instituto Geográfico Nacional, calle Ibáñez de Ibero, 3, 28003 Madrid, {xfernandez, jdherandez, nvalcarcel, mecaballero, mabenito}@fomento.es.

⁽²⁾ Tragsatec, aporcuna@fomento.es.

RESUMEN

Tras finalizar la producción del Sistema de Información de Ocupación de Suelo en España (SIOSE) afrontamos su publicación en la red. A la hora de decidir la estrategia de difusión de la información debemos tener en cuenta 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. SIOSE tiene, para realizar esta tarea de forma adecuada, dos características que nos condicionan: su gran volumen de información y la complejidad semántica de su modelo de datos orientado a objetos.

A partir de una arquitectura construida con componentes de software libre, desde el sistema operativo hasta las aplicaciones servidoras; los geoservicios web estándar son la primera opción para comunicar información geográfica a través de la web, pero las posibles soluciones son múltiples y muy variadas. Además los servicios deben combinarse y coordinarse con otras herramientas de publicación, como la libre descarga de los datos en bruto, para conseguir que SIOSE cumpla con su función de infraestructura de información territorial sobre la que generar valor añadido y para que sea una herramienta útil de análisis del territorio.

Describiremos como hemos 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 y como hemos intentado optimizar el rendimiento de estos servicios para ofrecer tiempos de respuesta adecuados ante cualquier posible petición.

Palabras clave: SIOSE, ocupación de suelo, geoservicios web, software libre, modelo orientado a objetos.

ABSTRACT

Once SIOSE's production is finished (SIOSE is the acronym for the Spanish land cover and land use information system), is time to publishing it on the Internet. At the time of deciding the information's spreading strategy, user's needs should be taken into account: access to visualizing, querying and analysing original data and its different thematic views must be given. To correctly accomplish these tasks, SIOSE has two determining features: its large volume of data and the semantic complexity of its object oriented data model.

Based on a systems architecture build with free software components, from the operating system to server applications, standard geo-webservices are the first option to communicate geographic information on the web, but possible solutions are multiple. Web services should be combined and coordinated with other publishing tools, such as downloading raw data, to get to the final goal of SIOSE being a real land information infrastructure on which to add value and being an useful land analysis tool.

It will be described how view and query services have been designed in order to show on the Internet all the information of SIOSE and the different points of view from which SIOSE can be analysed. Also, it will be discussed how performance has been tried to be improved to give reasonable response times against any possible request.

Key words: *SIOSE, land cover, geo web services, free software, object oriented model.*

INTRODUCCIÓN

En esta introducción realizaremos una breve descripción del proyecto SIOSE y de las razones, debidas a sus propias características, por las que no es sencillo publicar este sistema de información en la web.

¿Qué es SIOSE?

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 en lo que respecta a coberturas y usos de suelo. 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 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 al analizar el territorio, como se necesiten según los distintos requerimientos de los usuarios.

Características 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 gran 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 de suelo existentes en España.

Todo este volumen de datos plantea un problema de transmisión a una velocidad razonable a la hora de publicarlo 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 a través de una respuesta suficientemente ágil como para no dar sensación de mal funcionamiento.

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) o desde un punto de vista más centrado en una temática de análisis territorial determinada: 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 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.

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.

La otra premisa de partida es el tratar de maximizar la eficiencia en el uso de recursos, económicos y materiales, en un contexto de limitación de estos recursos y en un entorno de producción de una infraestructura de servicio público por parte de la Administración.

Arquitectura de software libre

A la hora de elegir los componentes de software del sistema la mejor opción para aumentar la eficiencia en el uso de los recursos disponibles era la que proporcionaban los desarrollos de software libre. Por una parte el ahorro en los costes de implementación frente a otras soluciones era significativo y, por otro lado, creemos que maximiza el beneficio socioeconómico para el conjunto de los ciudadanos y las empresas el hecho de que la Administración pública revierta, sin costes añadidos, en la sociedad los frutos de la inversión presupuestaria, en este caso un sistema de información territorial.

Además, se consideró que las aplicaciones geográficas de software libre están hoy en día en un punto de madurez tecnológica y estabilidad más que suficiente para considerarlas la solución más adecuada para unos servicios que deben ser lo bastante sólidos como para sostenerse en el tiempo indefinidamente y, a la vez, lo bastante flexibles como para poder aumentar fácilmente sus funcionalidades de forma progresiva y tener siempre abierta la posibilidad de evolucionar y adaptarse a los cambios tecnológicos que puedan surgir.

El sistema consiste, básicamente, en un servidor de bases de datos y un servidor de aplicaciones. Ambos tienen instalada, como sistema operativo, la distribución GNU/Linux CentOS 5.5.

Como base de datos se utiliza PostgreSQL 9.1 con su extensión espacial PostGIS 1.5.3. Como aplicación servidora de datos geográficos se ha elegido Geoserver (2.1), ejecutándose en una instancia del contenedor de aplicaciones Apache Tomcat 6. La configuración de software del servidor de aplicaciones se completa con un servidor web Apache, conectado con Tomcat, que es el que recibe y distribuye las peticiones. También se utiliza, para generar y gestionar los servicios web de visualización de información cacheada (WMTS), la aplicación GeoWebCache, embebida en la aplicación servidora Geoserver.

Tanto para la base de datos como para el conjunto de aplicaciones servidoras instaladas en el servidor de aplicaciones se realizaron diferentes pruebas de «sintonización» de la configuración con el fin de aprovechar al máximo las prestaciones del hardware en el que están instaladas. El rendimiento de estas aplicaciones, y por lo tanto del conjunto del sistema, varía sustancialmente si la configuración es la adecuada y está adaptada a los medios físicos con los que se cuenta (hardware, comunicaciones). En este sentido, dedicar el tiempo necesario a corregir esta configuración, evitando la configuración con la que las aplicaciones se instalan por defecto, es un paso imprescindible, previo a la puesta en carga, dado que hace que nuestro sistema de publicación sea capaz de «mover» un mucho mayor volumen de datos sin necesidad de aumentar los recursos de hardware destinados al sistema, con el consiguiente ahorro en medios materiales o la consiguiente mejora en rendimiento de respuesta.

Base de datos

Una vez diseñada la estructura e implementado el software del sistema, comenzamos a enfrentarnos a los retos que supone publicar el SIOSE. Para optimizar la respuesta, además de configurar adecuadamente las aplicaciones que forman parte del sistema, es necesario comenzar por actuar sobre la estructura de los datos.

Dado el volumen de datos manejado en SIOSE, más allá de las aplicaciones utilizadas, es crucial optimizar la implementación del modelo de datos en la base de datos relacional. Por una parte es necesario para mejorar el rendimiento del sistema y, por otra, evitar redundancias innecesarias en la información mantiene controlado el espacio de almacenamiento requerido.

En el caso de la base de datos del SIOSE para su publicación en web se ha mantenido en lo fundamental el modelo de datos del sistema de información pero adaptándolo a las necesidades de los servicios. De esta forma, se ha procurado separar, en la medida de lo posible, la información geométrica de los atributos alfanuméricos para tratar de independizar los procesos que sólo necesitan de estos últimos (más rápidos) de la geometría y utilizar ésta únicamente en el momento de la representación espacial. Así podemos ir generando todas las vistas y clasificaciones temáticas que sean necesarias, a partir de consultas a los atributos de la base de datos, e ir asociándolas a una única repetición de la geometría. En sentido inverso, esta separación nos permite no repetir conjuntos de atributos cuando es necesario repetir geometrías dado que cada polígono puede tener diferentes representaciones espaciales (representación en coordenadas geográficas y proyectadas, generalizaciones de la geometría, etc.). Esta estructura de datos unida a una correcta gestión de los tipos de dato, índices, índices espaciales y metadatos espaciales permite una mejor respuesta de la base de datos ante las peticiones de las aplicaciones servidoras.

Servicios de visualización

Para diseñar los 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 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 se han considerado las dos posibilidades y el objetivo es llegar a una solución final que combine ambas. En todo caso se van a utilizar servicios WMTS, replicando al servicio WMS implementado

inicialmente, para mejorar los rendimientos de la visualización a pequeñas escalas. Además se va a proceder a un proceso de generalización de la información de SIOSE para escalas 1:50.000 y 1:100.000. Pero este proceso de generalización geométrica, que implica una generalización semántica asociada (con el consiguiente riesgo de la aparición de sesgos por pérdida de información temática en el resultado final), no es trivial en el caso de SIOSE (cobertura continua de 2,5 millones de polígonos y 10,5 millones de coberturas) y se está realizando en paralelo a la implementación de los servicios. Por ello, en una primera fase y hasta que se complete el proceso de generalización de la información, los servicios de visualización de SIOSE presentarán para escalas 1:150.000 y menores la información de la base de datos Corine Land Cover 2006 (de resolución equivalente a escala 1:100.000) y la capa de polígonos de SIOSE aparecerá a partir de esa escala y para escalas mayores. A todas las escalas de visualización la información será consultable.

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.

Servicios de 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 por el elevado volumen de datos a transmitir que generan las consultas a estos servicios.

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

deseo 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 puntual 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. Pensando en este caso de uso, el modelo de datos de SIOSE cuenta con un campo, asociado a cada polígono, cuyo valor es un texto XML que describe la cobertura completa de ese polígono, su «etiqueta». Haciendo que la respuesta a la petición *GetFeatureInfo* consista en el contenido de este campo adecuadamente formateado en forma de ficha obtenemos una descripción rápida y detallada del polígono que nos interesa.

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 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 cerrado y predefinido, 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. Los ejemplos más claros los tenemos en la utilización de servicios WFS como parte de los procesos de un servicio WPS que encadena varias funciones (petición de información, cálculo y respuesta). De hecho, el IGN-CNIG tiene implementado un servicio WPS de análisis territorial que utiliza la información proveniente de un servicio WFS de ocupación de suelo (Corine Land Cover) como parte de un proceso, pero ese servicio WFS no se publica como tal. El de SIOSE sería un caso similar.

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.

Servicios de 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, el poco

desarrollo y falta de homogeneidad que ha tenido hasta ahora este tipo de servicios y la falta de recursos humanos para afrontar esta tarea; no se ha considerado por el momento la posibilidad de desarrollar servicios WPS a partir de la información del SIOSE. El desarrollo de estos 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. También cabe la posibilidad de que estos servicios sean desarrollados por el sector privado y la comunidad de usuarios, dado que son servicios que aportan mayor valor añadido a la información y que, por lo tanto, su desarrollo, para explotar todo el potencial de análisis territorial del SIOSE, puede ser económicamente rentable.

ESTRATEGIA GLOBAL DE PUBLICACIÓN DE LA INFORMACIÓN GEOGRÁFICA DE SIOSE

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 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>).

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.