

# Servicios de generalización para el procesamiento de imágenes satélites

Díaz, L. (1), Granell, C. (2), Gould, M. (3)

(1) Centro de Visualización Interactiva  
Departamento de Lenguajes y Sistemas Informáticos  
Universitat Jaume I  
Avda. Vicent Sos Baynat, s/n.  
Castellón (España)  
+34 964 387043, +34 964 728435, laura.diaz@uji.es

(2) Centro de Visualización Interactiva  
Departamento de Lenguajes y Sistemas Informáticos  
Universitat Jaume I  
Avda. Vicent Sos Baynat, s/n.  
Castellón (España)  
+34 964 387043, +34 964 728435, carlos.granell@uji.es

(3) Centro de Visualización Interactiva  
Departamento de Lenguajes y Sistemas Informáticos  
Universitat Jaume I  
Avda. Vicent Sos Baynat, s/n.  
Castellón (España)  
+34 964 728317, +34 964 728435, gould@uji.es

## Resumen

Cada vez más las aplicaciones requieren de mayor funcionalidad y complejidad, y hacen uso de grandes cantidades de datos. Las aplicaciones que manejan imágenes satélites y datos raster representan en este sentido un gran reto en el contexto de las Infraestructuras de Datos Espaciales a la hora de mejorar el rendimiento del procesamiento distribuidos y disminuir el intercambio de datos entre clientes y servidores. Para intentar paliar estas limitaciones, en este artículo proponemos el uso de técnicas de generalización para mejorar la eficiencia de los servicios avanzados de geoprocésamiento para el tratamiento de imágenes satélites y minimizar de esta forma la cantidad de los datos transmitidos a las aplicaciones cliente. Además, dichas técnicas de generalización son implementadas como servicios de geoprocésamiento que pueden ser compuestos a su vez con otros servicios ofrecidos por las IDEs para formar aplicaciones a medida para el usuario final.

**Palabras clave:** generalización de datos espaciales, servicios de geoprocésamiento, generalización de imágenes satélites, simplificación, transferencia de datos

# 1 Introducción

Ya no sorprende encontrarse actualmente con aplicaciones que requieren cada vez de mayor funcionalidad y complejidad, y además hacen uso de grandes cantidades de datos. Estas características son habituales y a veces distintivas en las aplicaciones para Sistemas de Información Geográficas (GIS). Afortunadamente, las infraestructuras de datos espaciales (IDE) [1] minimizan las dificultades anteriores al proporcionar los medios tecnológicos necesarios (interfaces, estándares, protocolos, servicios, etc.) que facilitan la extracción y consulta de grandes volúmenes de datos espaciales que serán posteriormente compartidos por usuarios finales y aplicaciones cliente.

Sin embargo, los datos espaciales almacenados en los repositorios representan un nivel de detalle excesivo que a menudo no resulta adecuado para las necesidades de una aplicación en un escenario concreto. Un ejemplo típico de lo anterior sucede cuando un conjunto de objetos espaciales son seleccionados para ser representados en un mapa. Tal vez parte o algunos de estos objetos no pueden ser visualizados correctamente en el mapa debido por ejemplo a un problema de limitación de espacio del área de visualización o por cambios de escala de la información representada. Esta problemática es común en la generación y visualización de mapas cartográficos a diferentes escala. Para intentar solucionar esta limitación apareció el concepto de generalización aplicado a datos espaciales, que se define como una derivación de una representación abstracta [2]. Los mismos conceptos de generalización son aplicables a la reducción de volúmenes de datos intercambiados entre los proveedores de servicios y los consumidores de servicios o clientes. Por ejemplo esto facilitaría el acceso y consulta de datos espaciales en entornos donde las condiciones de conexión son muy limitadas o deficientes, como en el caso de algunas países en desarrollo, o en entornos donde los dispositivos clientes son limitados, como por ejemplo en escenarios donde predominan dispositivos móviles [3].

En este artículo estamos interesados pues en la aplicación de técnicas de generalización en contextos donde el intercambio de datos entre proveedores y consumidores puede ser voluminoso. Un ejemplo concreto es el procesamiento de imágenes satélites ya que éstas suelen ser a menudo de gran tamaño (del orden de varios megabytes). Evidentemente, la encadenación de varios servicios que involucran tareas relacionadas con el tratamiento de imágenes satélites acarrea problemas de rendimiento en la aplicación resultante, y más aún si estos servicios se ejecutan de forma distribuida y remota. La aplicación de las técnicas de generalización nos permitirán paliar en cierta medida estos problemas de rendimiento y trasiego de datos espaciales en la red.

A diferencia de otras aproximaciones de generalización basadas en aplicaciones o clientes GIS de escritorio, en este trabajo nos centraremos en el desarrollo de servicios de geoprocésamiento distribuidos para la aplicación de técnicas de generalización a imágenes satélites. Además de aplicar generalización para aumentar el rendimiento y minimizar el flujo de datos entre proveedores y clientes, dichas técnicas de generalización son en sí mismas implementadas como servicios de geoprocésamiento que pueden ser compuestos con otros servicios ofrecidos por las IDEs para formar aplicaciones más complejas y a medida [4].

En la siguiente sección introducimos brevemente las técnicas de generalización y servicios de geoprocésamiento, conceptos clave para ofrecer servicios de generalización para el procesamiento de imágenes satélites. La sección 3 describe un caso de uso para el estudio de cuencas hidrográficas donde el uso de imágenes satélites es necesario. Cuestiones de diseño para la exposición de técnicas de generalización como servicios de geoprocésamiento se proponen en la sección 4. Luego, la sección 5 resume la arquitectura y algunos detalles de la implementación de nuestra solución. Para terminar, se establecen las conclusiones y algunas líneas futuras de investigación en este campo.

## 2 Conceptos básicos

Como se ha mencionado en la introducción, en este trabajo nos centramos en el uso de las técnicas de generalización para mejorar la eficiencia de los servicios avanzados de geoprocésamiento y minimizar de esta forma la cantidad de los datos transmitidos a las aplicaciones cliente. Antes de describir el diseño e implementación en las secciones posteriores, es necesario introducir algunos conceptos generales sobre generalización de datos espaciales y servicios de geoprocésamiento.

### 2.1 Generalización de datos espaciales

El concepto de generalización ha sido ampliamente utilizado en cartografía como el proceso de reducir la escala cartográfica en mapas. La idea es muy simple, ya que básicamente se intenta reducir la complejidad de un mapa manteniendo los aspectos clave o interesantes para el usuario (observador) y eliminar aquellas otras características que no son interesantes o relevantes para el mapa en cuestión. Algunos autores [2, 5] denominan a este proceso *generalización cartográfica*, ya que el objetivo de aplicar las técnicas de generalización se debe únicamente a razones de mejorar la interpretación visual de los datos espaciales (mapas). En el campo de la cartografía es especialmente útil para la generación de mapas con variaciones de escala. Sin embargo, aparte de la generalización cartográfica aplicada para la visualización de datos espaciales, Weibel considera otros dos tipos de generalización: la *generalización de objetos* y la *generalización de modelos*. La primera tiene lugar en el momento de definir los repositorios y bases de datos originales. Aunque técnicas de generalización de objetos son necesarias para abstraer aspectos del mundo real en objetos concretos de una base de datos, nuestro interés en este artículo radica en la generalización de modelos que consiste en reducir la complejidad de los objetos espaciales ya definidos inicialmente en los repositorios de datos para otros propósitos específicos.

La generalización de modelos se define como un conjunto de operaciones de modelado sobre los datos espaciales sin tener en cuenta todavía aspectos de visualización. Podríamos decir que las técnicas de generalización de modelos permiten procesar los datos espaciales que van a ser visualizados posteriormente mediante técnicas de generalización cartográfica. Así pues, en este artículo únicamente estamos interesados en la generalización de modelos (datos) y no en la cartográfica (mapas). La generalización de modelos comprende técnicas y algoritmos para producir un conjunto de datos simplificado, a partir de un conjunto de datos fuente, idóneos para una aplicación dada, obteniéndose por ejemplo datos de menor complejidad, a una escala reducida, etc. Una práctica habitual para obtener un conjunto de datos simplificado consiste en eliminar de los datos fuentes detalles excesivos e irrelevantes para una aplicación dada. Sin embargo, aunque se reduzca la complejidad de los datos espaciales, éstos deben mantener la misma estructura que proporcionan los datos originales, es decir, la aplicación final debe ser capaz de extraer la misma información tanto a partir de los datos fuentes como de los simplificados. Evidentemente, esto permitirá reducir considerablemente el volumen de datos transmitido entre proveedores y consumidores de datos si se aplican servicios de generalización que procesan datos espaciales originales en datos espaciales mucho más simples y de menor volumen.

En definitiva, la generalización de modelos se basa en una reducción de datos controlada para mejorar la eficiencia en un proceso de análisis, disminuir el volumen de datos transmitidos por la red, o para reducir la precisión o resolución de los datos para ser integrados con otros datos espaciales.

## 2.2 Servicios de geoprociamiento

Open Geospatial Consortium (OGC) define una arquitectura abierta e interoperable permitiendo la creaci3n de nuevas aplicaciones de forma flexible y escalable, facilitando adem1s el cambio de componentes por otros similares, ya sea por motivos econ3micos, de funcionalidad o de rendimiento. Estas aplicaciones ofrecen al usuario la funcionalidad para solucionar un problema en concreto a partir de la uni3n y/o concatenaci3n de operaciones at3micas *ad hoc*, en contraste con las aplicaciones SIG monol3ticas, donde toda la funcionalidad est1 implementada en una caja negra a la que no tiene acceso el usuario.

Hasta el momento, se ha estado haciendo hincapi3 mayoritariamente en la implementaci3n de servicios OGC de acceso y visualizaci3n de informaci3n geogr1fica (WMS, WFS, WCS, servicio de cat1logo). Sin embargo, la nueva especificaci3n de OGC Web Processing Service (WPS) [6] permite ofrecer servicios de geoprociamiento para realizar operaciones m1s complejas de an1lisis y de tratamiento de informaci3n espacial (por ejemplo, an1lisis espacial, camino m3nimo en red, buffer, etc.). Mediante el uso de estos servicios de geoprociamiento distribuidos, los usuarios pueden acceder y compartir rutinas de procesamiento remoto, compartiendo no solo datos sino tambi3n los algoritmos y rutinas que los procesan. El acceso a servicios de geoprociamiento distribuidos incrementa la reutilizaci3n, disminuyendo el coste de mantenimiento y haciendo m1s eficiente el trabajo de los expertos en el contexto de las IDEs. La comunicaci3n con estos servicios se realiza f1cilmente mediante peticiones HTTP con XML [7] garantizando la interoperabilidad y la independencia entre sistemas. La especificaci3n WPS (en el borrador de la versi3n 1.0) no limita el tipo de procesos 3nicamente a datos espaciales, sino que es posible la ejecuci3n de rutinas sobre datos vectoriales, r1ster, literales, u otros tipos de datos.

La ventaja del procesamiento remoto de compartir rutinas y reutilizar m3todos sin la necesidad de mantenimiento de aplicaciones de forma local compite con la desventaja del rendimiento. En general cuando trabajamos con datos espaciales, 3stos suelen ser datos complejos y de gran tama1o [8], pudiendo hacer inviable el procesamiento remoto de este tipo de datos entre proveedores y consumidores de servicios. Por lo tanto, las t3cnicas de generalizaci3n complementan perfectamente a los servicios de geoprociamiento ya que permiten la extracci3n de la informaci3n relevante que evite el tr1fico de grandes vol3menes de datos.

## 3 Caso de estudio: Aplicaci3n de servicios de generalizaci3n al estudio de cuencas hidrogr1ficas

Actualmente la monitorizaci3n de nuestro entorno es una cuesti3n crucial, principalmente para la detecci3n de inundaciones, sequ3as, u otros posibles desastres naturales. Gracias al avance de las im1genes sat3lites y de la tecnolog3a de *remote-sensing* es posible realizar tareas de an1lisis sobre im1genes sat3lite de forma automatizada. El proyecto AWARE (<http://www.aware-eu.info>) se engloba dentro de esta pol3tica, y tiene como objetivo el desarrollo de servicios de geoprociamiento para la automatizaci3n de las tareas de monitorizaci3n y predicci3n de la disponibilidad de agua, y su distribuci3n por aquellas cuencas hidrogr1ficas donde el deshielo es el factor m1s importante del balance de agua anual. Nuestros servicios que dan soporte a las necesidades del proyecto AWARE se han implementado siguiendo la interfaz OGC WPS. A grandes rasgos, el usuario utiliza nuestros servicios WPS junto con otros servicios habituales en una IDE para realizar las siguientes tareas: la b3squeda de im1genes mediante cat1logos, el procesamiento de dichas im1genes y la codificaci3n de los resultados, la transferencia de los

resultados por la red para ser finalmente visualizados en la aplicación cliente. Básicamente, las técnicas de generalización pueden aplicarse a nuestro escenario para mejorar la eficiencia del procesamiento de imágenes satélites y minimizar al mismo tiempo la transferencia de datos por la red. Será necesario pues reducir el modelo de datos original de forma que sea más simple para que pueda ser procesado mediante WPS remotos, pero manteniendo al mismo tiempo el adecuado nivel de abstracción para que el modelo preserve sus características originales.

## 4 Diseño de la solución

En esta sección abordamos el diseño de la solución adoptada para el uso de las técnicas de generalización para mejorar la eficiencia de los servicios avanzados de geoprocamiento y minimizar de esta forma la cantidad de los datos transmitidos a las aplicaciones cliente. Para el primer objetivo se han implementado algunos servicios genéricos de geoprocamiento como por ejemplo un servicio de vectorización (conversión de datos raster a vector) con el fin de reducir la complejidad del procesamiento de datos satélite, eliminando, durante la fase de vectorización, atributos innecesarios para el escenario en cuestión, tratando la generalización como un servicio más de geoprocamiento disponible en una IDE. El procesamiento vectorial necesita y genera, gracias al servicio de vectorización previo, menor volumen de datos que el procesamiento directo de imágenes satélites. Además, este hecho repercute directamente sobre la transmisión de datos al cliente, minimizando la cantidad de datos enviado en el caso de datos vectoriales. Por la tanto, para determinados contextos (por ejemplo aplicaciones móviles) es absolutamente crucial encontrar un método adecuado de generalización que se ajuste a las necesidades de la aplicación específica, reduciendo la cantidad de datos pero manteniendo al mismo tiempo la estructura original de éstos. Todo ello, claro, sin que provoque un coste de procesamiento excesivo que ralentice la respuesta del servicio.

Para proporcionar servicios de geoprocamiento útiles y que cumplan las necesidades para los modelos sobre recursos hídricos, hemos identificado una serie de funcionalidades básicas que puedan ser reutilizadas en la cadena de tareas de análisis y procesamiento. El objetivo final es crear una librería de servicios de procesamiento cuyas tareas complejas y específicas del dominio deleguen en procesos simples, atómicos y robustos. De este modo la reutilización de servicios de procesamiento facilita la creación de nuevos servicios personalizados con el uso de esta librería compartida.

Nuestra estrategia de diseño de servicios ha consistido en identificar una serie de funciones atómicas requeridas para nuestros casos de uso. Inicialmente estudiamos el modelo de requisitos desde una perspectiva *top-bottom* para hacer un análisis que permita encontrar el conjunto adecuado de servicios básicos de procesamiento. En nuestro contexto consideramos como servicio adecuado uno que realice una función básica fácilmente comprobable y que puede ser independiente de forma que pueda reutilizarse en otros contextos. Una vez identificados este conjunto de procesos básicos, se agrupan en módulos acorde a su funcionalidad. Cada servicio de procesamiento implementa la norma OGC WPS, como un módulo básico de procesamiento.

## 5 Implementación

La Figura 1 muestra la arquitectura propuesta compuesta de tres capas donde el punto de entrada del usuario es un geportal. El geportal realiza el flujo de control para la ejecución del modelo, y cuando es necesario invoca a los WPS para realizar el procesamiento de los datos requerido en cada

paso. La conexión del geportal con los WPS se realiza a través del componente WPS API (representado en la Figura 1 como tres engranajes), que implementa la interfaz OGC WPS. Este componente ha sido desarrollado en Java basado en software libre.

Para la implementación de las rutinas de procesamiento hemos escogido la implementación de la especificación OGC WPS desarrollada por *52North Incubator* (<http://incubator.52north.org>), un componente específico dentro del amplio proyecto de software libre 52North que incorpora la implementación de otras especificaciones OGC. La implementación de 52North WPS ha sido desarrollada en Java al igual que los procesos que hemos desarrollado para la creación de los servicios de procesamiento de imágenes satélites.

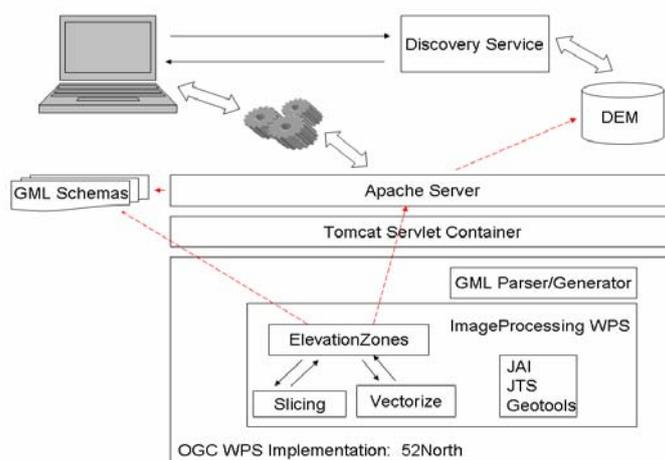


Figura 1. Arquitectura que soporta servicios de geoprocesamiento involucrados en el escenario de cuencas hidrográficas

La Figura 1 también muestra el servicio de geoprocesamiento *ImageProcessingWPS* que contiene algoritmos y procesamientos propios de imágenes raster, como puede ser *slicing*, donde cada celda de la imagen raster se clasifica en categorías de acuerdo con el valor de la celda de la banda indicada, y *vectorizing*, donde una vez clasificada la imagen raster se procede a su vectorización, obteniendo por ejemplo un fichero GML [9] con los polígonos de las celdas de un valor en la banda requerida. Este WPS permite a los usuarios finales extraer y procesar información necesaria a lo largo de la ejecución del modelo en nuestro caso de uso sin la necesidad de estar continuamente trabajando con el MDT (modelo digital de terreno) cuyo tamaño excede con creces un tamaño óptimo para trabajar con procesamiento remoto.

Consideremos el caso de cálculo de porcentajes de nieve en las zonas de elevación de la cuenca de estudio (concepto similar al cálculo de las curvas de nivel). Para evitar tratar con la imagen satélite o con el Modelo digital del terreno para las zonas de elevación usaremos el servicio *ImageProcessingWPS* que clasifica, vectoriza y extrae los polígonos en GML de cada una de las zonas de elevación que representan las zonas cubiertas de nieve de cada zona de elevación. El particular, el usuario busca las imágenes satélite y los MDT obtenidos a partir de los servicios de catálogo disponibles en la IDE. Una vez obtenidas las referencias, el geportal invoca mediante el WPS API al servicio *ImageProcessingWPS*, implementado con la librería JAI (*Java Advanced Imaging API*) para el procesamiento de imágenes, JTS (*Java Topology Suite*), GeoTools para el tratamiento de geometrías, y el generador de GML integrado en 52North para la conversión de estos objetos a formato GML. Como se puede apreciar todos los componentes son software libre.

## 6 Conclusiones y futuro trabajo

Utilizando la nueva norma de OGC Web Processing Services, hemos envuelto rutinas de procesamiento de datos espaciales creando servicios de geoprocesamiento web interoperables. La disponibilidad de rutinas de procesamiento de datos geoespaciales distribuidas mejora el intercambio de conocimiento a través de la red y la posibilidad de compartir no solo datos sino también los algoritmos dentro de una comunidad e incluso entre diferentes dominios del conocimiento. El geoprocesamiento es fácilmente adaptable proporcionando una gran escalabilidad de uso y desarrollo. Trasladar el procesamiento desde un plano local a distribuido hace que sea necesario el uso de rutinas que procesen datos para disminuir la cantidad de tráfico de datos para mejorar el rendimiento. Las técnicas de generalización de modelos permiten la simplificación de datos de forma que mejoramos la eficiencia en aplicaciones que requieren procesamiento distribuido. En nuestro caso de uso, tratar con imágenes raster de gran tamaño afectaba al rendimiento global de la aplicación. El desarrollo de WPS de tratamiento de datos raster para extraer en formato vectorial los datos relevantes en cada paso de la cadena de ejecución ha resuelto en parte los problemas de rendimiento, con el valor añadido de haber implementado una librería de software libre de procesos remotos de tratamiento de datos espaciales.

## Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto AWARE (SST4-2004-012257) de la Unión Europea, dentro del programa GMES.

## Referencias

- [1] C. Granell, M. Gould, M.A. Manso, y M.A. Bernabé, "Spatial Data Infrastructures". En: H. Karimi (ed): Handbook of Research on Geoinformatics. Information Science Reference (en prensa).
- [2] R. Weibel, "Generalization of spatial data: principles and selected algorithms". En: Algorithmic Foundations of Geographic Information Systems. Springer, 1997, pp. 99-152.
- [3] T. Foerster, J. Stoter, B. Köbben, y P. van Oosterom, "A Generis Approach to Simplification of Geodata for Mobile Applications". En: Proceedings of 10th AGILE International Conference on Geographic Information Science. AGILE, 2007.
- [4] C. Granell, M. Gould, y M.A. Esbrí, "Geospatial Web Service Chaining". En: H. Karimi (ed): Handbook of Research on Geoinformatics. Information Science Reference (en prensa).
- [5] R. Weibel, y G. Dutton, "Generalising spatial data and dealing with multiple representations". En: P. Longley, M. Goodchild, D. Maguire, and D. Rhind (Eds.): Geographic Information Systems – Principles and Technical Issues volume 1. John Willey & Sons, 1999, pp. 125-155.
- [6] Open Geospatial Consortium, Inc. Web Processing Service (WPS) Specification. Accedido el 10 de julio de 2007 en [https://portal.opengeospatial.org/files/?artifact\\_id=21670](https://portal.opengeospatial.org/files/?artifact_id=21670)
- [7] C. Kiehle, "Business logic for geoprocessing of distributed data". *Computers & Geosciences* 32 (10): 1746-1757, 2006
- [8] R. Lemmens, A. Wytzisk, R. de By, C. Granell, M. Gould, y P. van Oosterom, "Integrating Semantic and Syntactic Descriptions to Chain Geographic Services". *IEEE Internet Computing* 10 (5): 42-52, 2006.
- [9] Open Geospatial Consortium, Inc. Geography Markup Language (2.1.2). Accedido el 10 de julio de 2007 [http://portal.opengeospatial.org/files/?artifact\\_id=11339](http://portal.opengeospatial.org/files/?artifact_id=11339).