

Cacheado de datos procedentes de servicios WFS en la aplicación web del proyecto EuroGeoSource

R. Béjar^{1a}, D. Gayán-Asensio¹, M. Á. Latre¹, R. Rioja², M. Usón²

¹ Universidad de Zaragoza, Zaragoza, España

² GeoSpatiumLab, Zaragoza, España

^a rbejar@unizar.es

Resumen

El proyecto EuroGeoSource tiene como uno de sus objetivos el desarrollar una aplicación web que de acceso a, y permita explotar, datos de recursos minerales y energéticos en Europa. El sistema está siendo desarrollado como una arquitectura de IDE y siguiendo normas de INSPIRE. Los servicios más importantes son los WFS puestos en marcha por los socios, que son principalmente institutos geológicos europeos. Este artículo describe como se está creando y manteniendo una base de datos centralizada a modo de caché de datos para estos WFS, de manera que la aplicación web ofrezca un rendimiento que no se vea penalizado por la distribución física de estos servicios WFS.

Palabras clave: IDE. Servicios Web. INSPIRE. Caché. WFS.

1 Introducción

EuroGeoSource ¹ (*EU Information and Policy Support System for Sustainable Supply of Europe with Energy and Mineral Resources*) es un proyecto financiado por la Comisión Europea (*7FP, ICT Policy Support Programme, Theme 6 – Public Sector Information CIP-ICT-PSP-2009-3*) desarrollado por un consorcio coordinado por el *Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Onderzoek* (TNO) y que integra varios institutos geológicos europeos, dos empresas tecnológicas (incluyendo la española GeoSpatiumLab S.L.) y una universidad (la Universidad de Zaragoza).

Uno de los objetivos del proyecto es el desarrollo de un sistema de información geográfica en web que incluya servicios basados en especificaciones OGC, estándares ISO y las normas de la directiva europea INSPIRE. Este sistema debe permitir a sus usuarios localizar, acceder a y explotar información geográfica agregada sobre recursos minerales y energía en los países cuyos institutos geológicos son miembros del consorcio, así como de otros países europeos que quieran incorporarse al sistema. Los servicios web fundamentales en este sistema son servicios de tipo WFS (*Web Feature Service*), que permiten acceder a información vectorial en formato GML (*Geography Markup Language*) y que por tanto pueden ser la fuente para soportar consultas elaboradas.

La arquitectura del sistema web en desarrollo tiene algunas de las características técnicas de las arquitectura de las IDE (Infraestructuras de Datos Espaciales): los servicios web WFS están descentralizados y bajo la responsabilidad de los distintos miembros del consorcio. Además hay un portal y una aplicación web centralizados para dar acceso y poder explotar

¹ <http://www.eurogeosource.eu/>

estos servicios. La Figura 1 muestra una vista arquitectural del sistema focalizada en estos elementos.

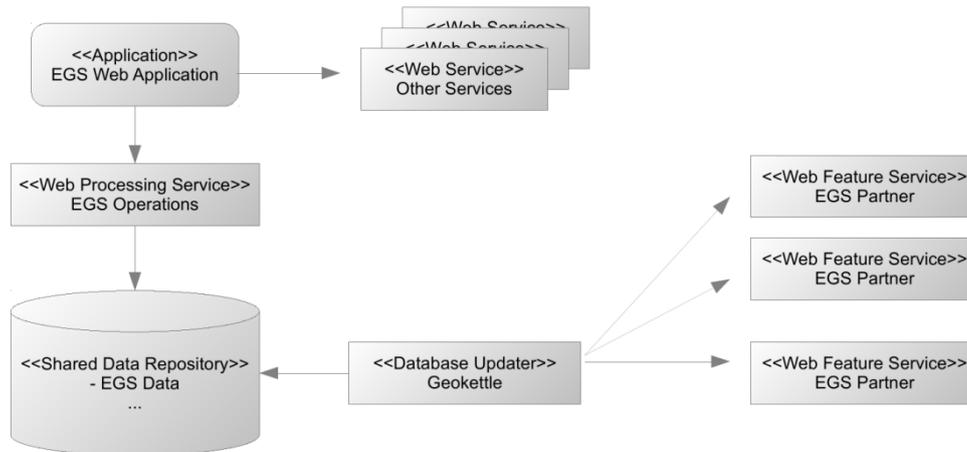


Figura 1: Vista arquitectural del sistema EuroGeoSource

Las prestaciones del sistema son una de las prioridades del consorcio. Para mejorar éstas se siguen diversas estrategias:

- Teselado de los servicios de mapas WMS (*Web Map Service*) que se incluyan.
- Recopilación (*harvesting*) de los metadatos de los socios del consorcio en un catálogo central.
- Caché de los datos de los servicios WFS de los socios en una base de datos centralizada que será a la que acceda la aplicación web.

El resto del artículo se centra en la tercera de las estrategias, por ser la menos habitual en este tipo de sistemas.

2 Servicios WFS distribuidos y base de datos centralizada

Como resultado de uno de los paquete de trabajo, el consorcio EuroGeoSource ha publicado un modelo de datos, compatible con INSPIRE en todo lo posible, de intercambio para recursos minerales y

energéticos [1]. Este modelo de datos se ha establecido como el modelo común en que los institutos geológicos miembros del consorcio deben publicar sus propios datos a través de servicios WFS, según el proceso descrito en [2].

Tal y como se ve en la Figura 1, el sistema EuroGeoSource tiene una base de datos centralizada que es actualizada por un proceso basado en la aplicación de ETL (*Extraction, Transformation, Load*) espacial Geokettle [3]. Esta base de datos comparte el modelo de datos con los WFS de los miembros del consorcio, lo que permite la carga de los datos que estos ponen a disposición del sistema a través de sus servicios WFS. Los servicios de procesamiento WPS (*Web Processing Service*) que soportan la funcionalidad de la aplicación web acceden a esta base de datos centralizada, lo que agiliza su funcionamiento. Sin esta base de datos, los servicios de procesamiento tendrían que interrogar a todos los WFS de los socios del consorcio y, en el mejor de los casos, el tiempo de esta consulta sería el del WFS que tardase más tiempo en contestar. Incluso en ese caso, este tiempo sería mayor que el tiempo en acceder a una base de datos local. A cambio, el sistema EuroGeoSource no accede siempre a los datos más recientes disponibles aunque, en el caso de los datos de recursos minerales y energéticos, los socios del consorcio consideran que esto es un problema menor, puesto que son datos que no se actualizan con frecuencia, y el proceso de actualización se puede ajustar para minimizar problemas.

3 Proceso de actualización de la base de datos central

El proceso de carga de los datos a partir de los WFS ha sido creado en GeoKettle utilizando su aplicación gráfica (llamada *Spoon*). Este componente se ejecuta como un trabajo a través de la herramienta *Kitchen*, otro componente de GeoKettle. A grandes rasgos el proceso es como sigue:

1. Vaciar los datos de la base de datos centralizada.
2. Obtener las direcciones de los WFS de los socios.
3. Consultar si hay permisos para copiar los datos de cada WFS.

4. Obtener los datos de estos WFS a través de las peticiones de tipo *GetFeature* correspondientes.
5. Cargar los datos en la base de datos centralizada.
6. Calcular varios datos agregados a partir de los datos cargados y poner estos resultados en tablas auxiliares. Estos agregados son usados típicamente en consultas desde los servicios de procesamiento y es conveniente tenerlos precalculados.

Para cada WFS el proceso se divide en tres, uno por cada tema de datos INSPIRE relacionado con el proyecto: uno para recursos energéticos (ER), otro para recursos minerales (MR) y otro para zonas de gestión, restricción y regulación y unidades de reporte (AM). Para cada uno de estos tres temas de INSPIRE hay un proceso ETL (**extraer** del WFS, **transformar** en el esquema de la base de datos y **cargar** en la base de datos) que se aplica a cada objeto geográfico. El proceso tiene un orden bien definido para evitar inconsistencias de datos debidas a las relaciones entre los objetos geográficos en el modelo. Para transformar los atributos espaciales se usan los pasos (*steps*) OGR de GeoKettle. Para los datos alfanuméricos y fechas se usan los pasos de procesamiento XML.

4 Conclusiones

Este artículo presenta un trabajo en curso, y por tanto hay cosas que todavía se están definiendo, concretando y probando. La estrategia de crear una caché centralizada de datos de los WFS de los socios del consorcio está funcionando bien para proporcionar un sistema con un buen rendimiento. El trabajo adicional de crear el proceso de carga y de actualización se compensa con creces con la estabilidad y las prestaciones que se consiguen. Además GML es principalmente un formato de intercambio, y por lo tanto es común cargarlo en una base de datos espacial para sacarle todo el rendimiento, así que en realidad esta estrategia no implica mucho trabajo adicional.

Es posible plantearse si podemos seguir hablando de un sistema distribuido. De hecho algunos miembros del consorcio mostraron esta duda al principio. Nosotros sostenemos que si. Simplemente hemos habilitado

un mecanismo de caché para que el sistema sea más rápido. Pero las responsabilidades de crear y mantener los servicios WFS, de actualizar los datos y mantenerlos en el modelo de datos común y la decisión de retirar un servicio WFS del sistema en cualquier momento siguen estando plenamente distribuidas y en manos de los socios². Y el sistema físicamente también, sólo que el acceso a los servicios WFS es periódico, y lo realiza el proceso de actualización de la base de datos, en lugar de ser bajo demanda, y realizado por los servicios de procesamiento cuando los usuarios de la aplicación solicitan cierta funcionalidad de la misma.

Agradecimientos



Este trabajo ha sido parcialmente financiado a través del proyecto EuroGeoSource (número de proyecto 250532), del ICT Policy Support Programme de la Unión Europea como parte del Competitiveness and Innovation Framework Programme. Este trabajo refleja sólo la visión de los autores y la Unión Europea no es responsable del uso que pueda hacerse de la información en él contenida. Este trabajo también ha sido parcialmente financiado por GeoSpatiumLab S. L.

Referencias

[1]

http://www.eurogeosource.eu/docs/D4_2_EGSource_Data_exchange_format.pdf

[2]

http://www.eurogeosource.eu/docs/D5_1_EGSource_Service_Implementation.pdf

[3] <http://www.spatialytics.org/projects/geokettle/>

2 Otra de las cosas en las que se está trabajando es en un protocolo para que los socios establezcan qué y hasta cuándo se pueden copiar los datos de sus WFS.