Implantación de un servidor SOS para la publicación de datos de sensores medioambientales en la IDE OTALEX-C

Nacho Brodin [(1)], Carlos Sánchez [(1)] , Jorge Gaspar Sanz [(1)] y Pedro Vivas [(2)]

(1) PRODEVELOP, S.L Pza. D. Juan de Villarrasa, 14, entlo pta. 5 46001 Valencia, ibrodin@prodevelop.es, csanchez@prodevelop.es, jsanz@prodevelop.es.

(2) Centro Nacional de Información Geográfica, Instituto Geográfico Nacional General Ibáñez de Íbero, 3, 28003 Madrid, pedro.vivas@cnig.es

RESUMEN

El proyecto IDE OTALEX está financiado por el programa europeo INTERREG III A y su objetivo es estudiar y mostrar la realidad del territorio, compuesto por las regiones del Alentejo en Portugal y Extremadura en España, separado convencionalmente por una frontera administrativa pero unido por sus características físicas, ambientales, sociales y económicas.

La fase actual del proyecto denominada OTALEX C, tiene como objetivo que la IDE OTALEX actual contemple los resultados de estudios medioambientales, por lo que es necesario la catalogación, estandarización, geoprocesamiento, la publicación de datos procedentes de sensores medioambientales, así como la publicación de mapas temáticos de isolíneas y mapas continuos procedentes de la interpolación de los datos de sensores medioambientales.

En la actualidad se está trabajando en esta fase. La primera parte del proyecto ha consistido en la transformación y carga de datos en la base de datos central (PostGIS) mediante procesos ETL con la herramienta GeoKettle.

Por otra parte, para hacer posible la estandarización e integración de los datos procedentes de los sensores y teniendo en cuenta la iniciativa SWE (Sensor Web Enablement), uno de los objetivos del proyecto será la implantación de un servidor SOS integrado en la arquitectura actual del proyecto IDE OTALEX, así como la implantación de un cliente SOS embebido en el Geoportal que permitirá la visualización y consulta de los datos de los sensores.

También se generarán mapas temáticos puntuales y de isolíneas por cada uno de los observables medidos en los sensores para ser publicados posteriormente a través de la especificación WMS y visualizados en el Visor de Mapas de la IDE OTALEX.

En la presentación se presentarán los trabajos realizados hasta la fecha, compartiendo los resultados obtenidos, los problemas detectados y las soluciones adoptadas.

ABSTRACT

OTALEX SDI project is funded by the INTERREG III European programe and its main objective is to study and expose the reality of the land composed by Alentejo region in Portugal and Extremadura in Spain. Both regions, administrative are separated by country frontiers but share physical, environmental, social and economic characteristics.

Current project phase is called OTALEX C and its focused on giving to OTALEX SDI the ability to show the results of environmental studies. It's necessary to catalogue, standardize, geoprocess and publish data from environmental sensors, as well as publishing thematic contour and continuous maps derived from the interpolation of these sensors data.

The first part of the project was the development of processes to load an heterogeneous group of data from different types of sensors into a central repository using Open Source and self developed ETL(Extract Transform and Load) tools.

A Sensor Observation Service server was deployed and integrated into the OTALEX infrastructure. At this time, the technical team is working on the development of a SOS client to let users query and visualize SOS servers besides the OTALEX repository of environmental data.

On the other hand, besides the punctual SOS data, the geoportal will share thematic maps (coropleth and contours) for all observed properties gathered and at any time (aggregated per day). The maps are published using WMS standard and accessed through the OTALEX geoportal as new layers per observed property.

This talk will present an updated status of all the work done, results and issues.

PALABRAS CLAVE

Jornadas, IDE, Portugal, España, IDE-OTALEX, *sensores medioambientales, procesos ETL, SWE, Servidor SOS, WMS*

INTRODUCCIÓN

El proyecto IDE OTALEX está financiado por el programa europeo INTERREG III A y su objetivo es estudiar y mostrar la realidad del territorio, compuesto por las regiones del Alentejo en Portugal y Extremadura en España, separado convencionalmente por una frontera administrativa pero unido por sus características físicas, ambientales, sociales y económicas. La web del proyecto es: www.ideotalex.eu.

La fase actual del proyecto denominada OTALEX C, tiene como objetivo que la IDE OTALEX contemple los resultados de estudios medioambientales, por lo que es necesario la catalogación, estandarización, geoprocesamiento, la publicación de datos procedentes de sensores medioambientales, así como la publicación de mapas temáticos de isolíneas y mapas continuos procedentes de la interpolación de los datos de sensores.

Figura 1: Acceso actual a la IDE-OTALEX.

En la actualidad se está trabajando en esta fase y se espera que los trabajos estén finalizados e integrados a finales de 2013.

TECNOLOGÍAS UTILIZADAS

Para llevar a cabo el proyecto se ha evolucionado el código existente del Geoportal y las tecnologías existentes, utilizando otras tecnologías 100% open source. Así las tecnologías utilizadas para el desarrollo del presente proyecto han sido:

- **Subversion**[1]: Sistema de control de código fuente para el versionado, branching, edición concurrente, parches, etc.
- **Apache HTTP**[2] como servidor web
- **Apache Tomcat[**3] como Contenedor de Servlets
- **PostgreSQL/PostGIS**[4][5] como servidor de bases de datos para alojamiento del repositorio de información geoespacial
- **GeoServer**[6] como servidor de mapas para publicación de servicios WFS y WMS
- **OpenLayers**[7] como cliente Web JavaScript de mapas y aplicaciones geográficas
- **52th North**[8] como servidor SWE para publicación de servicios SOS, SOS-T
- **GeoKettle**[10] como herramienta para la carga y transformación de datos geoespaciales Procesos ETL y control de calidad de los datos.
- **Java** + **Jsoup**[17] Lenguaje de programación y librería de "web scrapping" (en español "Rascado de la red") que se utiliza para obtener datos ya publicados en internet. Se há elaborado un programa que trata estos datos y los inserta en la base de datos SOS que há sido creada en la base de datos PostgreSQL.

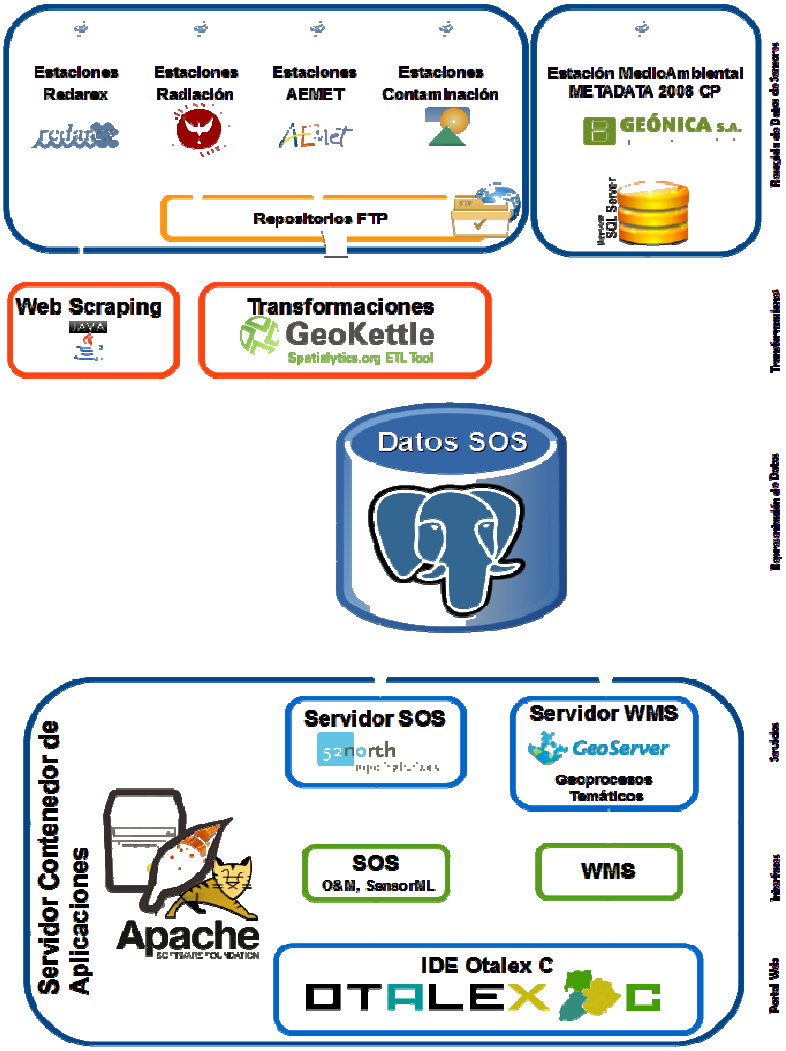

Figura 2: Tecnologías utilizadas para añadir SOS a OTALEX C.

DESCRIPCIÓN DEL TRABAJO REALIZADO

Se ha creado un sistema que permite la automatización de la catalogación, estandarización, geoprocesamiento y publicación en la IDE OTALEX-C de los datos heterogéneos procedentes de los sensores con información ambiental, medioambiental y radiológica en el ámbito del territorio OTALEX-C: Alentejo, Extremadura y región centro del proyecto.Este sistema permite de una manera automática el almacenamiento de los datos procedentes de las observaciones de los sensores medioambientales en la base de datos central de la IDE de OTALEX-C (PostGIS), la posibilidad del acceso al histórico de datos así como la actualización periódica de éstos y su publicación en tiempo real en la IDE OTALEX-C.

A continuación se describen las fases del proyecto:

- Captura de datos generados por fuentes de datos propias del proyecto
- Captura de datos de fuentes externas
- Configuración del servidor SOS
- Transformación y carga de datos al repositorio central
- Publicación de mapas temáticos
- Implementación de un cliente SOS
- Actualización del geoportal

Para la captura de datos internos del proyecto, se ha adquirido una estación medioambiental portátil que mide los siguientes parámetros en tiempo real:

- Dirección y velocidad del viento
- Presión atmosférica,
- Humedad relativa del aire
- Temperatura
- Radiación IR, UV, Beta y Gamma.

Para el acceso a datos de fuentes internas procedentes de la estación medioambiental portátil también se utilizará procesos ETL que permitan integrar las observaciones almacenadas en la base de datos ACCESS/SQL Server por el programa GEONICA SUITE en el repositorio centraCaptura de datos internosl de datos del proyecto OTALEX-C.
Estos procesos ETL desarrollados en GeoKettle son capaces accediendo a la base de datos Access /SQL Server de descubrir los sensores pertenecientes a la estación para registrarlos en el servidor SOS del proyecto. El proceso ETL para incorporar las estaciones a registrar se muestra en la imagen siguiente.

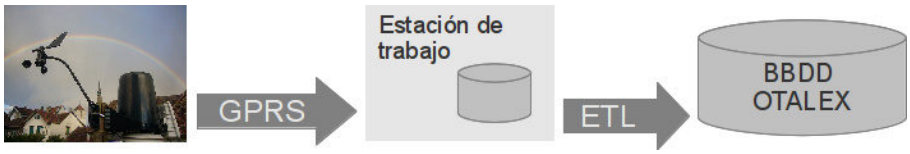

Figura 3: Transferencia de información de la estación meteorológica.

La estación medioambiental transmite la información mediante GPRS a una estación de trabajo que almacena la información en una base de datos propia. Esta base de datos será consultada mediante los procesos de transformación que se detallarán más adelante.

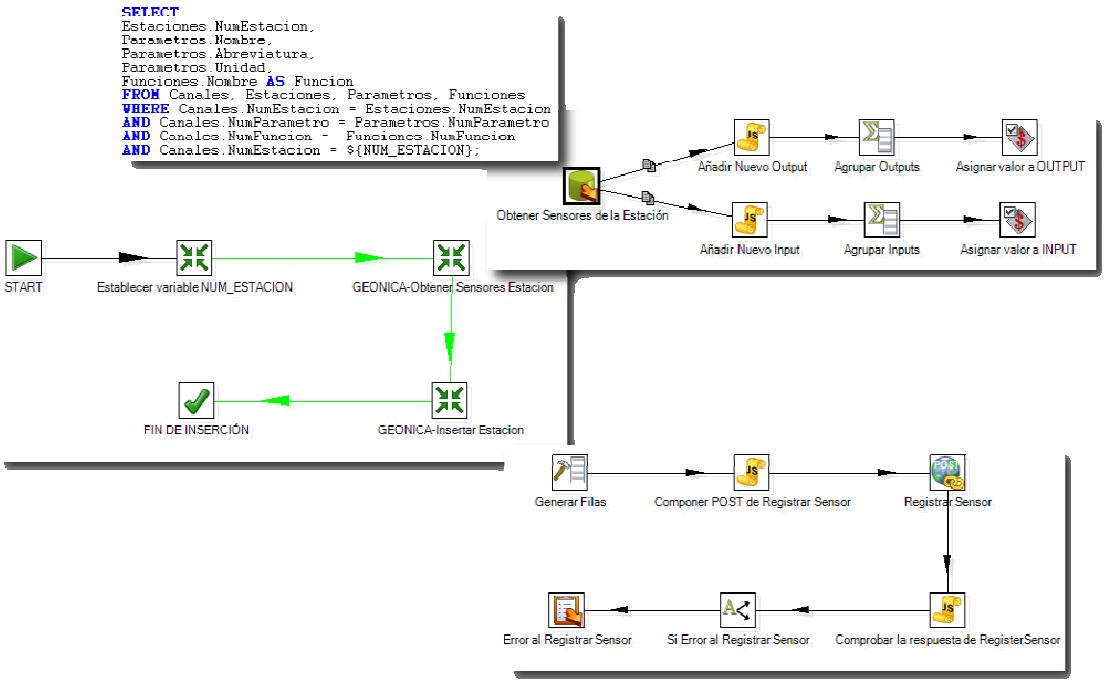

Figura 4: Carga y Transformación de de datos con GeoKettle.

Fuentes de datos externas

Se ha llevado a cabo el estudio, análisis y reutilización de los datos de las fuentes ambientales de los siguiente datos externos:

- **A**gencia **E**spañola de **Met**eorología: **AEMET**[11] dispone de un servidor de ficheros ftp con el propósito de que la información meteorológica y climatológica generada en el uso de sus funciones oficiales sea accesible como un servicio de pago. En este servidor dispone de acceso a un amplio conjunto de datos de observación, boletines del Sistema Mundial de Telecomunicaciones de la OMM, datos de radiación solar, ozono, contaminación de fondo, radar, rayos, modelos numéricos de predicción y series climatológicas de la Agencia Estatal de Meteorología. En el ámbito del proyecto de OTALEX se encuentran 51 estaciones dentro de comunidad de Extremadura que proporcionan información al proyecto. Los datos que van a ser procesados de los que ofrece AEMET para publicarlos en el servicio SOS de OTALEX C son:

  1. Temperatura
  2. Velocidad Viento
  3. Dirección Viento
  4. Presión Atmosférica
  5. Precipitaciones
  6. Insolación
  7. Dióxido de Azufre SO2
  8. Monóxido de Nitrógeno NO
  9. Dióxido de Nitrógeno NO2
  10. Partículas PM10
  11. Radiación IR
  12. Radiación UV-B
  13. Radiación Difusa
  14. Radiación Directa
  15. Radiación Global

  Estos datos no van a ser actualizados al cerrarse el ftp de Aemet, por tanto quedarán solo como histórico en la base de datos.

- **Red** de **A**sesoramiento al **R**egante de **Ex**tremadura: **REDAREX**[12]. En Extremadura se han instalado 22 estaciones agrometeorológicas completas, que se suman a las 10 que ya poseía la Consejería de Agricultura a través del Servicio de Investigación y Desarrollo Tecnológico. Por tanto, la Región cuenta con un total de más de 32 estaciones, que cubren la totalidad de las Zonas Regables actuales y futuras. Se ha realizado un proceso de reciclado "web scrapping" para obtener los datos publicados en la web http://sw-aperos.juntaex.es/redarex/ e insertarlos en la base de datos SOS del proyecto Otalex, ya que no se encuentran publicados en un formato estándar. De las estaciones correspondientes a esta red se capturan los datos correspondientes a:

  16. Temperatura
  17. Humedad Relativa
  18. Pluviómetro
  19. Velocidad del Viento
  20. Dirección Viento
  21. Radiación Solar

  Estos datos al estar ya publicados en la página web de Redarex necesitan del proceso previo de web scrapping para poder obtener los datos, luego tratarlos y finalmente introducirlos dentro de la base de datos a través del servicio SOS de 52North.

- Datos recogidos por la **Universidad de Évora**. Para incorporar estos datos se realizarán procesos ETL a partir de un fichero que se publica vía FTP. Esta estación ubicada en Portugal proporcionará los datos siguientes:

- Temperatura
- Humedad
- Viento
- Radiación Solar
- Radiación Difusa
- Radiación Atmosférica
- Radiación Ultravioleta UVA y UVB
- Radiación Infraroja IR
- Radiación Solar
- Precipitación Acumulada

- **R**ed **E**xtremeña de **P**rotección e **I**nvestigación de **C**alidad del **A**ire. **REPICA**[13] es una red para la vigilancia e investigación de la calidad del aire en el entorno regional, diseñada y gestionada por la Junta de Extremadura (Consejería de Agricultura, Desarrollo Rural, Medio Ambiente y Energía) con la colaboración de la Universidad de Extremadura (grupo de investigación AQUIMA, Análisis Químico del Medio Ambiente). La incorporación de estos datos al servicio SOS se realizará vía FTP descargando los ficheros que contienen cada uno las medidas tomadas por una estación concreta durante una semana. Estos ficheros se actualizan regularmente e introducen en la base de datos mediante un proceso ETL. Dispone de 6 estaciones de las cuales se va a obtener las observaciones de las siguientes propiedades observadas:

- Dióxido de Azufre SO2
- Monóxido de Carbono CO
- Monóxido de Nitrógeno NO
- Dióxido de Nitrógeno NO2
- Concentración Total de Nitrógeno NOX
- Ozono O3
- Benzeno BENZ
- Velocidad Viento
- Dirección Viento
- Temperatura
- Humedad Relativa

Repositorio central de la base de datos

Sobre una base de datos PostgreSQL 9.1 se ha creado un esquema para albergar toda la información proporcionada por las redes de sensores. Además se ha creado un esquema en el que se han definido vistas que ofrecen una simplificación de la información apta para ser consumida por el servidor de mapas temáticos como más adelante se detallará.

Instalación de un servidor SOS

Una vez identificadas las fuentes de datos y previo a la carga de los mismos, fue necesario la instalación y configuración del servidor SOS de 52 north. Esto es así porque se ha utilizado el estándar SOS-T para la inserción de la información en la base de datos. Es decir, en lugar de insertar directamente la información en la base de datos, haciendo por tanto dependiente el proceso del esquema de ésta, se ha utilizado un estándar que abstrae el proceso del esquema de la base de datos, permitiendo mantener el proceso, aún variando el software SOS o la base de datos que soporta el sistema.
En concreto se ha instalado la última de las releases de la serie 3.x del servidor, la cual soporta tanto el estándar SOS 2.0 como el SOS 1.0. No se ha utilizado la versión 4.0 porque solo soporta la versión 2.0 del estándar e impone limitaciones técnicas, más allá de que se trata de una versión no estable todavía.

Transformación de datos procedentes de sensores heterogéneos externos

Para la extracción, transformación y carga de datos en el repositorio central PostgreSQL/PostGIS del proyecto OTALEX-C, tanto de los procedentes de la estación medioambiental portátil, así como de fuentes de datos externas, se han diseñado procesos ETL (Extraction, Transformation &Load) utilizando el proyecto open source GeoKettle así como procesos desarrollados ad-hoc para poder hacer la extracción (scrapping) de orígenes de datos en formato HTML.
Tanto los procesos en GeoKettle como estos últimos y tal como se ha comentado en el punto anterior se insertan en la base de datos mediante llamadas HTTP al servidor SOS.
Un ejemplo de la inserción de observaciones en el servicio SOS consiste en una llamada al servidor SOS de tipo HTTP POST proporcionando el XML siguiente:

```
<InsertObservation xmlns="http://www.opengis.net/sos/1.0"
        xmlns:ows="http://www.opengis.net/ows/1.1"
        xmlns:ogc="http://www.opengis.net/ogc"
        xmlns:om="http://www.opengis.net/om/1.0"
        xmlns:sos="http://www.opengis.net/sos/1.0"
        xmlns:sa="http://www.opengis.net/sampling/1.0"
        xmlns:gml="http://www.opengis.net/gml"
        xmlns:swe="http://www.opengis.net/swe/1.0.1"
        xmlns:xlink="http://www.w3.org/1999/xlink"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.opengis.net/sos/1.0
        http://schemas.opengis.net/sos/1.0.0/sosInsert.xsd
        http://www.opengis.net/sampling/1.0
        http://schemas.opengis.net/sampling/1.0.0/sampling.xsd
        http://www.opengis.net/om/1.0
        http://schemas.opengis.net/om/1.0.0/extensions/observationSpecialization_override.xsd"

        service="SOS" version="1.0.0">

        <AssignedSensorId>urn:ogc:object:feature:Sensor:aemet-badajoz-3410</AssignedSensorId>

        <om:Measurement>
                <om:samplingTime>
                        <gml:TimeInstant>
                                <gml:timePosition>2012-06-14'T'00:00:00.000Z</gml:timePosition>

                        </gml:TimeInstant>
                </om:samplingTime>

                <om:procedure xlink:href="urn:ogc:object:feature:Sensor:aemet-badajoz-3410" />

                <om:observedProperty xlink:href="urn:ogc:def:phenomenon:OGC:1.0.30:DirViento" />


                <om:featureOfInterest>
                        <sa:SamplingPoint gml:id="foi-aemet-badajoz-3410">
                                <gml:name>Estación Badajoz</gml:name>
                                <sa:sampledFeature xlink:href="Estación Badajoz"/>
                                        <sa:position>
                                        <gml:Point>
                                                <gml:pos srsName="urn:ogc:def:crs:EPSG::4326">
                                                        38.5243 -6.5815
                                                </gml:pos>
                                        </gml:Point>
                                </sa:position>
                        </sa:SamplingPoint>
                </om:featureOfInterest>
                <om:result uom="°">30</om:result>
        </om:Measurement>
</InsertObservation>
```

Publicación de mapas temáticos

En OTALEX se está utilizando GeoServer como servidor de mapas WMS/WFS. Este servidor ha introducido en sus últimas versiones dos funcionalidades que permiten agilizar y simplificar de forma significativa este proceso. Las dos características son las capas SQL[14] y las transformaciones de renderizado[15].

Capas SQL

Estas capas se definen no como una tabla o vista de una base de datos, sino directamente como una consulta SQL que además admite parámetros. Esto es, en la sentencia SQL de acceso a los datos se introducen variables a las cuales se les da un valor por defecto pero que pueden ser definidas como un parámetro en la petición GetMap (ver figura). De este modo se pueden definir capas que obtengan datos en función de un parámetro fecha que traerá de la base de datos la información adecuada para un momento solicitado.Además, a cada parámetro se le puede asignar una expresión

regular de validación, con el objetivo de evitar ataques de inyección SQL.

Figura 5: Capa SQL con parámetro fecha.

Transformaciones de renderizado

GeoServer ha agregado capacidades de geoprocesamiento mediante la adición al proyecto del plugin de Web Processing Service. Este plugin permite por tanto publicar geoprocesos mediante estándares y compartir en definitiva funcionalidades de análisis geoespacial en la web. Aprovechando que se dispone de esta funcionalidad, los desarrolladores de este servidor de mapas han agregado la capacidad de definir geoprocesos a la definición de estilos SLD. De este modo, se ejecutan geoprocesos durante el renderizado de los mapas en peticiones GetMap del estándar WMS.

Es decir, se permite generar visualizaciones alternativas directamente en la definición del estilo de la capa, sin necesidad de alterar el origen de datos. Actualmente se soportan tres procesos:

- Creación de mapas de calor
- Interpolación de Barnes[16]
- Creación de curvas de nivel a partir de una imagen raster

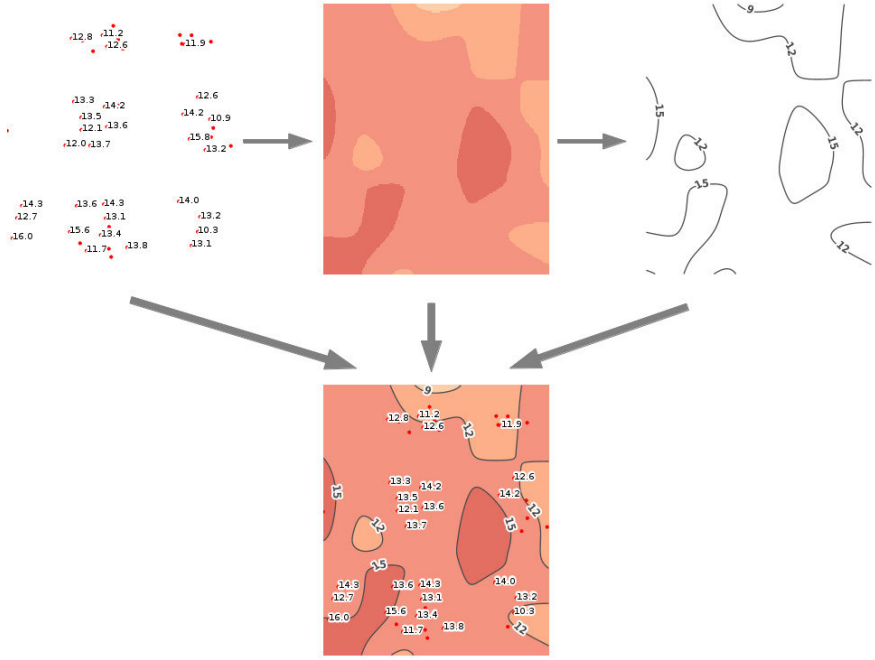

Figura 6: Composición de estilos con transformaciones de renderizado.

En nuestro caso esto permite generar mapas de interpolación a partir de una capa de puntos con la medida de una variable medioambiental. Un segundo estilo permite concatenar el proceso de interpolación al de generación de curvas de nivel con lo que tenemos tres estilos para la capa de puntos que podemos unir en el visor:

- Estilo de punto para mostrar el valor de la medida
- Estilo de interpolación usando tintas graduadas
- Estilo de curvas de nivel

Estos mapas se integrarán en el actual visor de OTALEX para dar al usuario del geoportal esta información medioambiental de forma sencilla y usable.

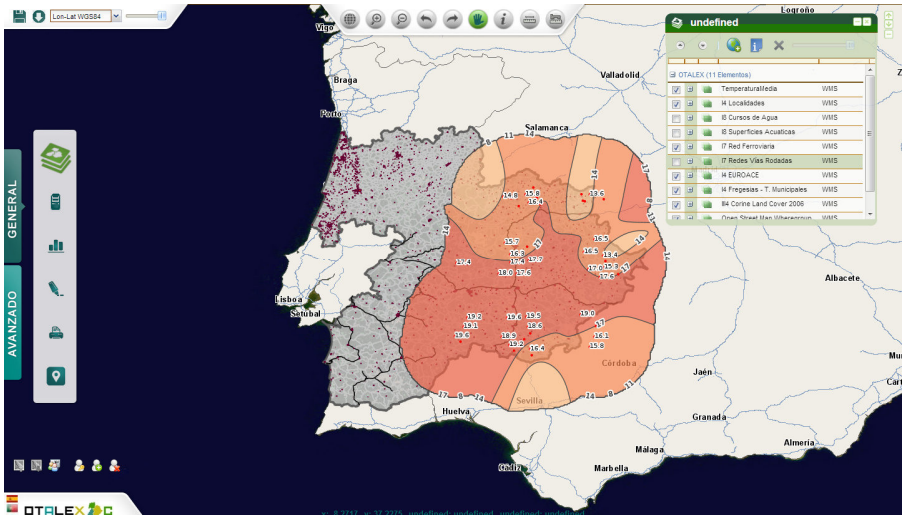

Figura 7: Mapa temático en el geoportal de OTALEX-C.

Implementación de un cliente SOS

Para la implementación del cliente, éste permitirá añadir servidores SOS y cargar la capa de puntos sobre el mapa. Se há tomado de base para su implementación el actual prototipo de cliente de SOS implementado en OpenLayers.

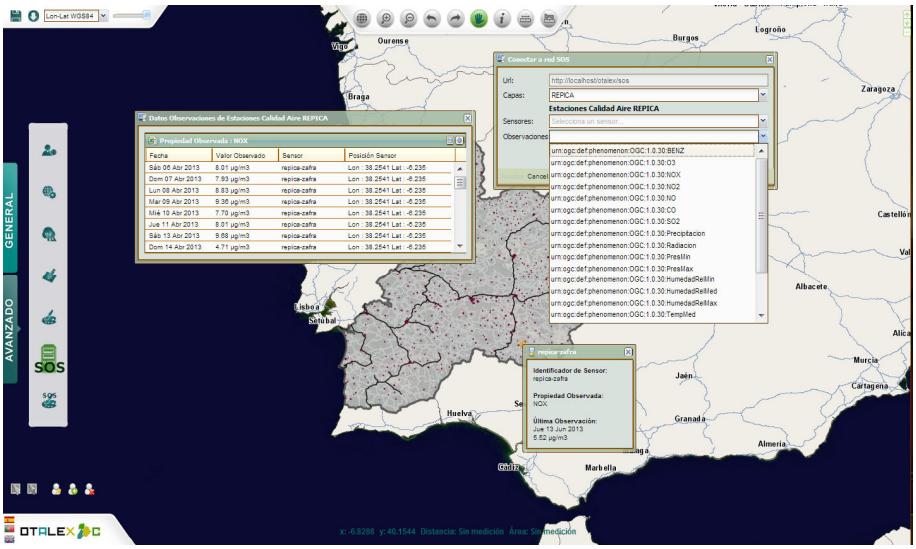

Figura 8: Cliente SOS en el geoportal de OTALEX-C.

Por defecto el cliente tiene seleccionado el própio servidor de Otalex para consultar los datos SOS que desee el usuario, muchos nos visualizables en los temáticos al ser datos que no se miden en suficientes estaciones como para crear un nuevo temático.
El cliente permite:

1. Conectarse a un servidor SOS (por defecto el servidor SOS de Otalex)
2. Seleccionar un “Offering” que equivale a una capa de puntos con datos relacionados.
3. Seleccionar un Fenómeno Observado, Observaciones (requerido para poder visualizar la capa)
4. Seleccionar un Sensor o “procedure” (si no se selecciona ninguno se mostrarán todos los que dispongan de datos para el fenomeno observado)

Una vez cargada la capa se visualiza una tabla con los datos obtenidos a partir de la selección hecha por el usuario. Estos datos són descargables.
Aparte de mostrar los datos en la tabla, la nueva capa de puntos se visualiza en el mapa, permitiendo seleccionar uno de los puntos, que representa un sensor, para pedirle información de la última observación.

CONCLUSIONES

Actualmente el proyecto está en fase de implantación. A partir del proyecto se pueden sacar algunas conclusiones de utilidad:

- Es necesario que las administraciones publiquen la información recogida por sus redes de sensores.
- Es necesario que lo hagan en formatos estructurados y consistentes para poder integrarlos en servicios que añadan valor.
- La utilización de protocolos y estándares SWE permiten no solo el consumo, sino también la alimentación de información medioambiental de forma interoperable e independiente del sistema de almacenamiento de la misma.
- GeoServer dispone de la funcionalidad necesaria para publicar información medioambiental de forma dinámica y accesible mediante estándares, evitando procesos intermedios que

incrementarían la complejidad del sistema sensiblemente.

REFERENCIAS

[1] Subversion: http://subversion.tigris.org/

[2] Apache HTTP: http://httpd.apache.org/

[3] Apache Tomcat: http://tomcat.apache.org/

[4] PostgreSQL: http://www.postgresql.org/

[5] PostGIS: http://postgis.org/

[6] GeoServer: http://geoserver.org

[7] OpenLayers: http://openlayers.org

[8] 52 north SOS: http://52north.org/communities/sensorweb/sos/index.html

[9] Deegree catalog: http://deegree.org/

[10] GeoKettle: http://www.spatialytics.org/projects/geokettle/

[11] AEMET: http://www.aemet.es

[12] REDAREX: http://aym.juntaex.es/servicios/redarex/

[13] REPICA: http://xtr.extremambiente.es/repica/index.html

[14] Capas SQL: http://docs.geoserver.org/stable/en/user/data/database/sqlview.html

[15] Transformaciones de renderizado: http://docs.geoserver.org/stable/en/user/styling/sld-extensions/rendering-transform.html

[16] A Technique for Maximizing Details in Numerical Weather Map Analysis, Stanley L. Barnes: http://journals.ametsoc.org/doi/abs/10.1175/1520-0450%281964%29003%3C0396%3AATFMDI%3E2.0.CO%3B2

[17] Librería de Web Scrapping JSOUP : http://jsoup.org/

AUTORES

Nacho BRODÍN
ibrodin@prodevelop.es
Prodevelop S.L.
Area GIS

Carlos SÁNCHEZ
csanchez@prodevelop.es
Prodevelop S.L.
Area GIS

Jorge Gaspar SANZ
jsanz@prodevelop.es
Prodevelop S.L.
Area GIS

Pedro VIVAS
pedro.vivas@cnig.es
CNIG
Area soporte IDE/SIG