

CartoCiudad apuesta por el software libre.

Julián González García ⁽¹⁾, Ana Velasco Tirado ⁽¹⁾, Alicia González Jiménez ⁽¹⁾, José Miguel Rubio Iglesias ⁽¹⁾, Paloma Verdejo Herreras ⁽¹⁾, Ángel García San Román ⁽¹⁾; Sebastián Mas Mayoral ⁽¹⁾

⁽¹⁾ Centro Nacional de Información Geográfica, Calle General Ibáñez de Ibero 3, 28003 Madrid, carto ciudad@ign.es

RESUMEN

CartoCiudad es una base de datos de red viaria, cartografía urbana e información censal y postal generada a partir de datos de organismos oficiales de toda España que permite la localización de direcciones y el cálculo de rutas y áreas de influencia.

El Instituto Geográfico Nacional (IGN), junto con las comunidades autónomas con las que se han firmado convenios de colaboración, es el encargado de integrar información procedente de la Dirección General del Catastro, del Instituto Nacional de Estadística, de la Sociedad Estatal de Correos y Telégrafos y de los organismos cartográficos de las comunidades autónomas con información procedente de sus bases de datos para formar CartoCiudad.

Desde el principio del proyecto ha existido un especial interés en la utilización de software libre de Sistemas de Información Geográfica (SIG). Así lo demuestra la utilización de herramientas libres de código abierto para la publicación de los servicios web estándar de CartoCiudad: WMS utilizando GeoServer, WFS empleando Deegree, WPS utilizando 52° North y WMS-C con TileCache, todos ellos desplegados en servidores Apache y Apache Tomcat.

Después de varios años utilizando las mismas herramientas de control de calidad y debido al cambio del modelo de datos en respuesta a las necesidades de los usuarios, se está trabajando en una nueva herramienta de edición y control de calidad de CartoCiudad sobre gvSIG en colaboración con proDEVELOP. Esto supone una actualización de la metodología de producción en CartoCiudad aprovechando el dinamismo y la flexibilidad de las soluciones de código abierto.

Por último, se están analizando las ventajas e inconvenientes de migrar el repositorio de datos de CartoCiudad desde una base de datos Oracle (Spatial) a una PostgreSQL (PostGIS) en combinación con el módulo pgRouting, con el objetivo de contar con las mismas funcionalidades de las que se dispone en la actualidad.

Palabras clave: *CartoCiudad, callejero, control de calidad, servicios OGC software libre, gvSIG.*

INTRODUCCIÓN

CartoCiudad es la base de datos de red viaria, cartografía urbana y divisiones censal y postal de ámbito nacional, cuyos datos son oficiales y provienen de la Dirección General del Catastro, el Instituto Nacional de Estadística, la Sociedad

Estatal de Correos y Telégrafos, el Instituto Geográfico Nacional (IGN) y otros organismos autonómicos, y que lleva formando durante los cinco últimos años el propio IGN junto con las Comunidades Autónomas de País Vasco, Valencia, Navarra, La Rioja, Murcia, Illes Balears y Andalucía.

Con el objetivo de crear una base de datos lo más completa posible, incorporando la cartografía urbana de todos los municipios de España y asegurando así la continuidad a nivel nacional, esta base de datos se ha ido construyendo y actualizando desde 2006 mediante procesos de producción anual en el que la unidad de trabajo es el municipio. La complejidad del modelo, el tamaño de los datos y el compromiso del proyecto con la calidad hace necesario disponer de un robusto proceso de control de calidad y de edición de los datos.

La explotación de los datos de CartoCiudad se realiza a través de Internet empleando servicios web accesibles desde el geoportal de CartoCiudad (www.cartociudad.es) e implementados siguiendo los estándares de interoperabilidad de OGC [1]. Estos servicios permiten llevar a cabo operaciones de visualización de información, localización de direcciones, unidades administrativas, códigos postales y secciones censales, cálculo de áreas de influencia y de enrutamiento entre direcciones de cualquier lugar de España, entre otras operaciones.

Desde el inicio del proyecto se ha apostado por el software libre para la implementación de los servicios web de CartoCiudad. A estas soluciones libres se les han sumado recientemente las herramientas de edición y control de calidad, hasta ahora proporcionadas por software propietario. De entre las plataformas libres existentes, gvSIG [2] ha sido la solución empleada para su implementación.

Finalmente, y de cara a la futura migración de los datos de CartoCiudad a un Sistema de Gestión de Bases de Datos libre, se ha hecho una prueba piloto de carga de datos sobre PostgreSQL[3] con extensión PostGIS [4], así como de conexión de los servicios de CartoCiudad, cuyos resultados han sido en general satisfactorios.

En el presente artículo se desarrolla en profundidad las experiencias con software libre en el proyecto CartoCiudad, los problemas encontrados así como las soluciones aportadas, planteando el paso en corto o medio plazo a una total migración de CartoCiudad a tecnologías libres de código abierto.

EVOLUCIÓN EN LA IMPLANTACIÓN DE SOFTWARE LIBRE EN EL PROYECTO.

Desde el comienzo del proyecto, a principios del año 2006, existió un especial interés en la utilización de software libre SIG, debido a las ventajas que su uso conlleva, siendo las más valoradas por el equipo del proyecto las siguientes ¹:

- Permite ahorrar en la adquisición, mantenimiento y renovación de tecnologías, lo cual repercute en un menor coste del proyecto en general.
- Las aplicaciones libres tienen mayor calidad, completitud. Las actualizaciones y correcciones suelen darse con mayor rapidez debido a que el código fuente es totalmente accesible, el diseño de nuevas funcionalidades suele orientarse según las necesidades de los usuarios y no de acuerdo con intereses comerciales tal y como sucede con el software propietario.

¹ Fuente CENATIC (www.cenatic.es)

- El software libre es una tecnología de fácil acceso. La mayoría de las aplicaciones están disponibles en la red lo que facilita su consulta, instalación y prueba.
- Las tecnologías libres disfrutan de un soporte técnico más accesible. La mayoría de iniciativas de este tipo disponen de foros de usuarios donde éstos comparten sus experiencias, además de foros o sistemas de tickets para el reporte de fallos o consultas específicas.

En el año 2007 se creó el portal www.cartociudad.es que refleja claramente la ventaja de utilizar de forma combinada software libre y servicios de publicación de datos estándares. Está basado en la tecnología OpenLayers [5] y permite acceder y consumir varios tipos de servicios implementados según los estándares del Open Geospatial Consortium (OGC). En la actualidad dicho portal proporciona acceso a un total de doce servicios estándares OGC y uno según la recomendación Map Service Tile Caching (WMS-C) [6], de los cuales la mayoría se publican desde el proyecto CartoCiudad:

- 8 servicios Web Feature Service (WFS) [7] para la consulta y descarga de fenómenos de CartoCiudad (viales, códigos postales, municipios, etc.)
- 3 servicios Web Map Service (WMS) [8] para la visualización de mapas, de los cuales uno publica la información geográfica generada en el proyecto CartoCiudad y los otros dos son publicados por la Dirección General del Catastro y por la Infraestructura de Datos Espaciales de España (IDEE) respectivamente.
- 1 servicio WMS-C también para la visualización de mapas teselados de CartoCiudad.
- 1 servicio Web Processing Service (WPS) [9] para el cálculo de áreas de influencia y de rutas a partir de los datos del proyecto.

Las soluciones tecnológicas empleadas para la implementación de los servicios de CartoCiudad son soluciones libres y OpenSource: deegree [10] para los servicios WFS, GeoServer [11] para el servicio WMS, TileCache [12] para el servicio WMS-C, 52north [13] para el servicio WPS, utilizando Apache Tomcat [14] como contenedor de servlets y Apache [15] como servidor web.

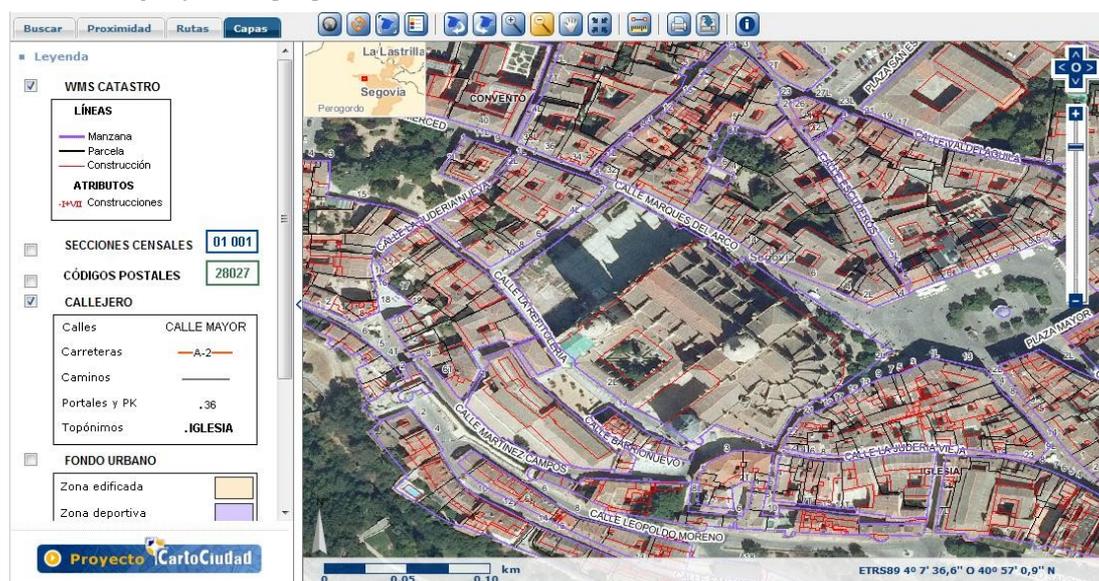


Figura 1: Vista del geoportal www.cartociudad.es

A pesar de la utilización ya consolidada de soluciones libres en la publicación de los datos, el resto del software empleado en el proyecto era propietario. En la Tabla 1 se presenta un resumen del software utilizado en el proyecto en sus primeros años de vida.

Tabla 1: Resumen del software utilizado en los años 2006-2007.

Sistema	Descripción
Sistema operativo	Windows Server [16] para los servidores de bases de datos y de aplicaciones y Windows XP [17] (corporativo) para las estaciones de trabajo
Sistema gestor de base de datos	Oracle Spatial [18]
Sistema de edición	ArcGIS de ESRI [19] para la edición de shapefiles exportados desde la base de datos y una vez editados, importados de nuevo
Sistema de exportación e importación	FME de Safe Software [20]
Sistema de control de calidad	Aplicación de escritorio desarrollada por la E.T.S.I. en Topografía, Geodesia y Cartografía de la Universidad Politécnica de Madrid basada en el lenguaje de programación Java y en el lenguaje de consulta estructurado (SQL)
Sistema de publicación	GeoServer, TileCache, degree y 52north sobre Apache y Apache Tomcat
Geoportal y API	OpenLayers
Ofimática	Microsoft Office [21] (corporativo)

Desde el año 2007 y durante estos cinco años, el proyecto ha ido evolucionando conforme a las necesidades de los usuarios, buscando una mayor integración con los modelos de datos de todas las organizaciones que lo componen y los de las que colaboran con él, así como una mayor agilidad en la actualización, edición y control de calidad de los datos.

La evolución del proyecto se ha visto reflejada en varios cambios del modelo de datos tales como la incorporación de nuevos atributos, la redefinición de entidades o la inclusión de un modelo de estados y ciclo de vida de los fenómenos. A medida que el modelo de datos ha ido cambiando ha sido necesario adaptar los sistemas de edición y control de calidad.

Por ello, en el segundo semestre de 2011, y con el fin de incorporar de forma homogénea en el flujo de trabajo todas las variaciones en el modelo de datos, se comenzó a trabajar en el diseño y la implementación de una serie de herramientas de edición personalizadas así como de control de calidad, en colaboración con proDEVELOP [22], utilizando como plataforma gvSIG.

También en el 2011 se han comenzado a migrar los sistemas operativos de los servidores dedicados al proyecto, de Windows Server a Red Hat Enterprise Linux [23], empezando por los de bases de datos.

El progreso de implantación de software libre en el proyecto para este año será el que muestra la Tabla 2.

Tabla 2: Resumen del software a mediados de 2012.

Sistema	Descripción
Sistema operativo	Red Hat Enterprise Linux para los servidores de bases de datos, Windows Server para los de aplicaciones y Windows XP (corporativo) para las estaciones de trabajo
Sistema gestor de base de datos	Oracle Spatial
Sistema de edición	Plugin sobre gvSIG con edición directa sobre la base de datos.
Sistema de exportación e importación	FME de Safe Software
Sistema de control de calidad	Plugin sobre gvSIG y uso del lenguaje de consulta estructurado (SQL)
Sistema de publicación	GeoServer, TileCache, degree y 52north sobre Apache Tomcat y Apache
Geoportal	OpenLayers
Ofimática	Microsoft Office (corporativo)

Por último cabe destacar que en este año se están realizando una serie de pruebas para estudiar la viabilidad de implantar como sistema gestor de base de datos PostgreSQL junto a la extensión PostGIS para la gestión de datos espaciales.

En los dos siguientes apartados se amplía la información referente a las iniciativas en curso para continuar con la implantación de software libre en CartoCiudad.

HERRAMIENTAS PERSONALIZADAS DE CONTROL DE CALIDAD Y DE EDICIÓN DE CARTOCIUDAD CON GVSIG.

Tal y como se ha adelantado en el apartado anterior, la propia evolución del proyecto hacía necesario desarrollar unas herramientas personalizadas que permitieran la edición de datos de CartoCiudad y una serie de controles automáticos de la calidad de los datos.

Hasta el momento el IGN no disponía de herramientas personalizadas de edición que permitieran, de forma transparente para el operador, garantizar la integridad entre tablas o el versionado de fenómenos,

En cuanto a las herramientas automáticas de control de calidad, en los primeros años del proyecto se colaboró con el grupo de investigación LatinGEO, de la Universidad Politécnica de Madrid, que desarrolló una aplicación Java que actuaba directamente sobre la base de datos Oracle. Al ir evolucionando el modelo de datos [24] estas herramientas se quedaron obsoletas, por lo que posteriormente se decidió utilizar ArcGIS y FME, trabajando directamente con ficheros. Los cambios en el modelo ya indicados, junto a la necesidad de ampliar los controles que se efectuaban y, además, de disponer de una única aplicación que pudiera actuar directamente sobre la base de datos, llevó al planteamiento del desarrollo de todo el entorno de control de calidad utilizando una plataforma SIG libre.

La solución elegida está totalmente basada en software libre con gvSIG como SIG de escritorio y Sextante como biblioteca de análisis espacial. Concretamente, consiste en un plug-in sobre gvSIG que permite conectarse y editar sobre una base de datos Oracle Spatial o PostgreSQL - PostGIS indistintamente.

Esta plataforma fue elegida fundamentalmente por la independencia tecnológica que proporciona al IGN, y por la posibilidad de ser utilizada por otras organizaciones colaboradoras del proyecto (consejerías o institutos cartográficos de las comunidades

autónomas que producen CartoCiudad) sin necesidad de imponerles costes de licencias de otro software.

Herramientas de control de calidad

Las herramientas de control de calidad han sido diseñadas satisfaciendo los requerimientos y los flujos de trabajo específicos del proyecto. Así, los controles de calidad que se realizan a los nuevos datos de CartoCiudad para validarlos se agrupan en tres tipos, que se aplican en tres fases diferentes:

- Fase 1: Controles básicos de aceptación, que consisten en verificar que se cumplen las condiciones mínimas para poder incorporar los datos: que las tablas cumplan el modelo de datos, que existan datos de los municipios objeto de la revisión, que no existan duplicados, etc.
- Fase 2: Contraste de los nuevos datos con los ya existentes en la base de datos: que no se inserten fenómenos que ya existían comparándolos por su identificador, ni se eliminen fenómenos que no existiesen.
- Fase 3: Garantizar que la base de datos consolidada (donde se hayan integrado las actualizaciones con los datos ya existentes) cumple todos los requisitos del modelo de CartoCiudad en cuanto a consistencia conceptual (integridad referencial, listas controladas, etc.) y consistencia topológica (tramos conectados, polígonos postales cerrados, etc.).

Para diseñar el flujo del control de calidad automático, es necesario contar con tres bases de datos o esquemas: el que contiene los datos de partida, o “BD Original”, el que contiene los datos nuevos fruto de la actualización, o “BD Actualización”, y en el que se han integrado las actualizaciones, o “BD CCalidad”.

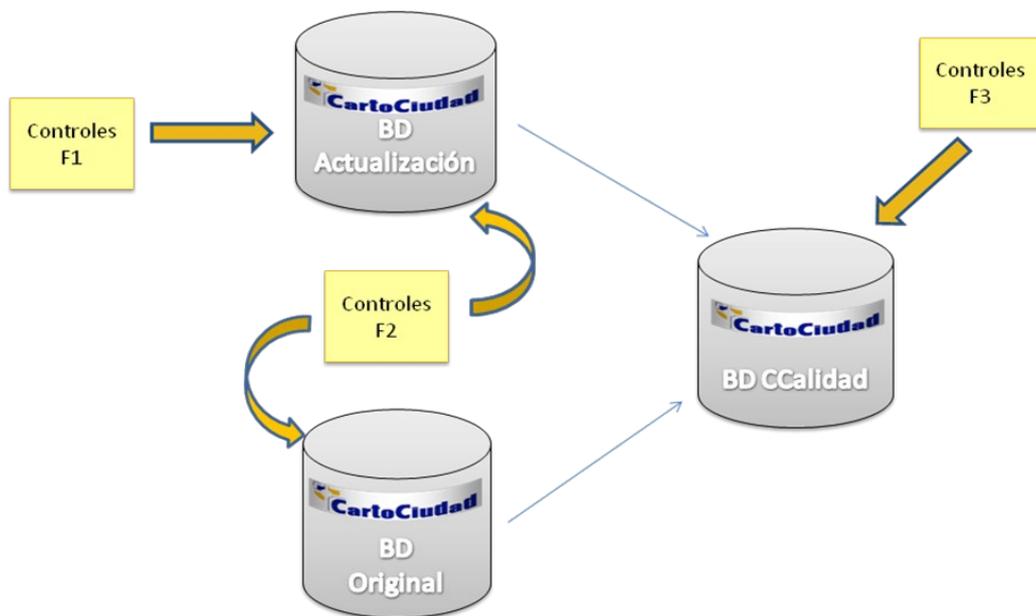


Figura 2: Bases de datos utilizadas en el control de calidad de CartoCiudad

La versatilidad de la herramienta diseñada hace que las validaciones se puedan realizar a ficheros shapefiles o a una base de datos, por lo tanto, la BD Actualización puede ser en realidad una base de datos Oracle Spatial o PostgreSQL o bien un

conjunto de ficheros shapefile. Por otra parte, se puede elegir el área geográfica a la que se desea aplicar las validaciones. Además, cada uno de los controles es independiente, por lo que puede seleccionarse uno de ellos, varios o todos ellos.

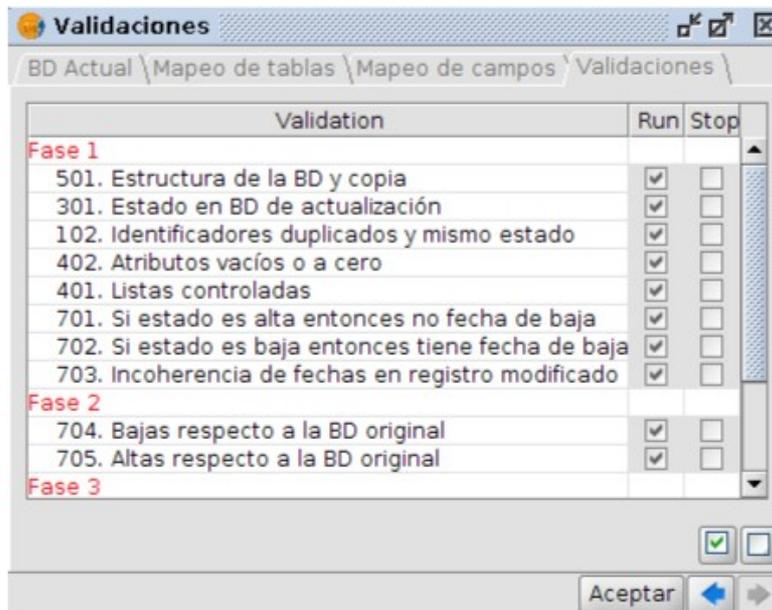


Figura 3: Pantalla de selección de validaciones

Por último, la herramienta ofrece el resultado de las validaciones y unas tablas con los elementos erróneos.

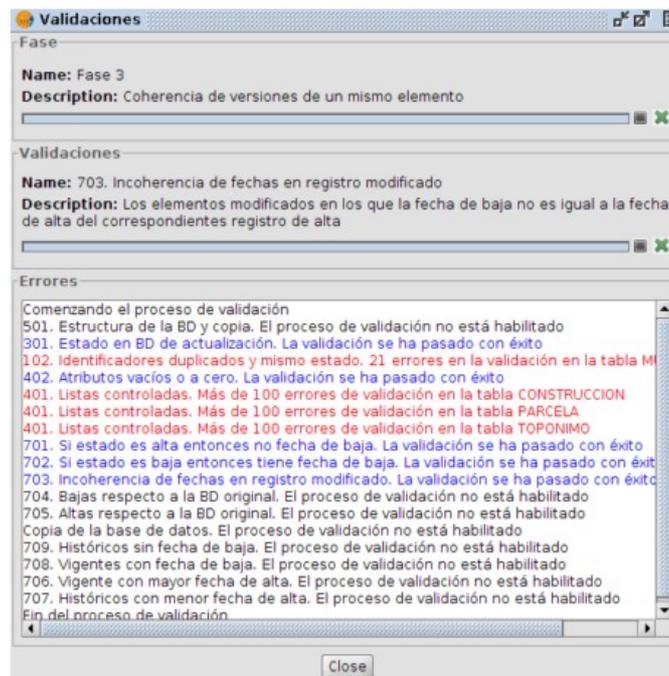


Figura 4: Pantalla de resultado de validaciones

Herramientas de edición

Las utilidades de edición están integradas como una barra de herramienta más en gvSIG. Las funcionalidades destacables de estas herramientas son:

- Permiten mantener las relaciones entre tablas, es decir, los tramos que componen un vial, los portales asignados a cada tramo, etc.
- Respetan el ciclo de vida y el estado de los elementos, almacenando automáticamente la fecha de edición y si se está dando de alta o de baja un elemento
- Permiten trabajar con los dominios y las listas de valores de los atributos de las tablas de CartoCiudad.

Se trabaja en sesiones de edición, de forma que permite aislar cada sesión y efectuar los controles oportunos a los datos introducidos que aún están dados de alta, sin tener que someter de nuevo a toda la base de datos a los citados controles. Si se superan estos controles, los registros pasarán a estado vigente en el caso de las altas o a histórico si se trataba de una baja. Del mismo modo, si el resultado tras la edición no es el deseado, es posible revertir los cambios de una determinada sesión.

Para agilizar el acceso a los datos se incorpora una utilidad de selección de áreas, de forma que no es necesario volcar en edición todos los elementos de la base de datos de una determinada tabla, sino sólo los correspondientes a un área concreta. Ésta se puede definir arrastrando el ratón por el mapa o bien eligiendo un polígono de una de las capas activas.

La herramienta de edición permite realizar operaciones genéricas tales como acceso a la base de datos, inserción, selección y edición de geometrías o eliminación de elementos de la base de datos. Por otro lado, se han diseñado otras funcionalidades *ad hoc* para el proyecto CartoCiudad, algunas de ellas son:

- Proyectar ortogonalmente los portales al borde de las parcelas o puntos kilométricos a los tramos.
- Partir tramos en su intersección con la línea de término municipal o con otros tramos, o bien dividirlos por cambio de valor de alguno de sus atributos.
- Editar polígonos de códigos postales, tratando de forma simultánea polígonos adyacentes, añadiendo polígonos adyacentes a uno dado.
- Establecer relaciones entre elementos: indicar el tramo al que pertenece un portal o el vial al que corresponde un tramo.

PRUEBA PILOTO DE IMPLANTACIÓN DE UNA BASE DE DATOS POSTGRESQL – POSTGIS EN CARTOCIUDAD.

Pruebas realizadas

Con el objetivo de completar la migración a un entorno 100% software libre en el proyecto CartoCiudad, desde finales de 2011 se han estado realizando una serie de pruebas para estudiar la viabilidad de implantar como Sistema Gestor de Base de Datos PostgreSQL junto a la extensión PostGIS para la gestión de datos espaciales.

En el marco del estudio realizado se han llevado a cabo las siguientes pruebas:

- Generación del esquema del proyecto en PostgreSQL: se ha podido reproducir el modelo de datos actual sin ningún tipo de problema.
- Se han revisado todos los tipos de datos de cada campo en cada tabla para almacenarlos de una manera más eficiente, consiguiéndose reducir el tamaño en disco de la base de datos. Carga de los datos desde Oracle utilizando FME: la carga de datos se completó con éxito pese a algunos problemas en el proceso.

Una vez reproducido el modelo de datos se ha pasado a la siguiente fase consistente en poblar de fenómenos la base de datos. La carga, que se ha realizado con FME atacando directamente a Oracle, ha presentado algunos problemas a la hora de ejecutarla en un único proceso, puesto que los recursos del servidor en el que se ejecutaba se agotaban. Ha sido necesario descomponerla en varios procesos consecutivos para evitar el agotamiento de los recursos.

- Conexión del servicio WMS con GeoServer: los resultados han sido satisfactorios.

El cambio de origen de datos se realizó creando un nuevo DataStore mediante la interfaz web del propio GeoServer. Por el cambio de DataStore fue necesario rehacer de nuevo la definición de las entidades, capas y grupos de capas. No obstante, al no variar el modelo de datos no fue necesario ningún cambio adicional.

- Conexión de los servicios WFS con deegree: los resultados han sido satisfactorios pero con algunos problemas en el proceso.

El cambio de origen de datos se realizó al nivel de definición de fenómenos (ficheros xsd de definición de esquema de aplicación), los cuales incluyen llamadas a funciones almacenadas en PL/SQL que hubo que reproducir en PL/PgSQL. Sin embargo, la estructura de las funciones fue fácilmente generada al ser muy parecidos ambos lenguajes. La mayoría de ellas incluían sentencias con llamadas a funciones nativas de Oracle Spatial que hubo que sustituir por sus equivalentes en PostGIS. Al realizar este proceso, se pudo comprobar las diferencias de comportamiento de las funciones nativas de Oracle Spatial y PostGIS, siendo más tolerantes con algunas geometrías inválidas las primeras.

Trabajo futuro

Una vez comprobado que las aplicaciones de los servicios WMS y WFS de CartoCiudad funcionan correctamente sobre la base de datos PostgreSQL – PostGIS, los siguientes pasos a dar en el estudio de viabilidad serán:

- Comprobar la conexión del servicio WPS con 52north, para lo cual será necesario redefinir por completo todas las clases que implementan las operaciones del WPS. Las clases actuales aprovechan la API Java de Oracle Spatial y utilizan como fuente de datos el modelo para redes propio de Oracle. Por lo tanto, además de redefinir estas clases previamente será preciso buscar una alternativa al modelo de red de Oracle Spatial en PostgreSQL – PostGIS. El propio modelo de datos de CartoCiudad implementa un sencillo modelo de red que podría aprovecharse de alguna extensión como pgRouting [25] para el cálculo de rutas.
- Revisar y adaptar las nuevas herramientas de edición y control de calidad, inicialmente diseñadas para la explotación de una base de datos Oracle, para su uso con el nuevo sistema gestor de base de datos.

¿Y DESPUÉS?

Se continuará con el proceso de migración de de Windows Server a CentOS [26] de los sistemas operativos de los servidores de aplicaciones donde se encuentran desplegados los diferentes servicios que publica el proyecto.

Por último se buscarán alternativas a FME entre las herramientas ETL (Extract, Transform, Load) que sean libres, como por ejemplo GeoKettle [27].

No obstante, no se prevén cambios en lo referente a los sistemas operativos de las estaciones de trabajo ni a las aplicaciones ofimáticas por ser parte del sistema corporativo del Ministerio de de Fomento.

Como resumen se muestra en la Tabla 3 un resumen del software que se utilizaría después de llevar a cabo todos los cambios.

Tabla 3: Resumen del software a mediados de 2012

Sistema	Descripción
Sistema operativo	Red Hat Enterprise Linux para los servidores de bases de datos, CentOS para los de aplicaciones y Windows XP (corporativo) para las estaciones de trabajo
Sistema gestor de base de datos	PostgreSQL – PostGIS
Sistema de edición	Plugin sobre gvSIG con edición directa sobre la base de datos.
Sistema de exportación e importación	GeoKettle o similar
Sistema de control de calidad	Plugin sobre gvSIG y uso del lenguaje de consulta estructurado (SQL)
Sistema de publicación	GeoServer, TileCache, deegree, 52north, TileCache sobre Apache Tomcat y Apache.
Geoportal	OpenLayers
Ofimática	Microsoft Office (corporativo)

CONCLUSIONES

La implantación de soluciones libres en el proyecto se ha realizado de forma paulatina, comenzando con el uso de software libre centrado en la publicación de datos (GeoServer, deegree, 52north, TileCache, OpenLayers, Apache y Apache Tomcat), seguido de la migración de los sistemas operativos de los servidores de bases de datos (de Windows Server a Red Hat Enterprise Linux), continuando en el presente con el diseño de nuevas herramientas de edición y control de calidad sobre gvSIG y Sextante. Además en un futuro próximo y como resultado del análisis expuesto en este documento, se prevé migrar el sistema gestor de base de datos a PostgreSQL – PostGIS y continuar con la migración de los sistemas operativos en los servidores asignados al proyecto.

Después de adoptar en el proyecto diferentes tipos de software libre (sistemas operativos, aplicaciones de escritorio, software SIG) y de la realización de las últimas pruebas con PostgreSQL – PostGIS, es posible afirmar, por experiencia propia, que el cambio de software propietario a software libre no implica mayor complejidad, ni trastornos que el que puede suponer el cambio entre software propietarios.

REFERENCIAS

- [1] <http://www.opengeospatial.org/>
- [2] <http://www.gvsig.org/web/>
- [3] <http://www.postgresql.org/>
- [4] <http://postgis.refractor.net/>
- [5] <http://openlayers.org/>
- [6] http://wiki.osgeo.org/wiki/WMS_Tiling_Client_Recommendation
- [7] <http://www.opengeospatial.org/standards/wfs>
- [8] <http://www.opengeospatial.org/standards/wms>
- [9] <http://www.opengeospatial.org/standards/wps>
- [10] <http://deegree.org/>
- [11] <http://geoserver.org/display/GEOS/Welcome>
- [12] <http://tilecache.org/>
- [13] <http://52north.org/>
- [14] <http://tomcat.apache.org/>
- [15] <http://httpd.apache.org/>
- [16] <http://www.microsoft.com/es-xl/servidores/default.aspx>
- [17] <http://windows.microsoft.com/es-ES/windows/products/windows-xp>
- [18] <http://www.oracle.com/es/products/database/options/spatial/index.html>
- [19] <http://www.arcgis.com/home/>
- [20] <http://www.safe.com/>
- [21] <http://office.microsoft.com/es-es/>
- [22] <http://www.prodevelop.es/>
- [23] <http://www.redhat.com/>
- [24] <http://www.cartociudad.es/portal/1024/especificaciones.htm>
- [25] <http://www.pgrouting.org/>
- [26] <http://www.centos.org/>
- [27] <http://www.spatialytics.org/projects/geokettle/>