

Servicios de Catálogo de Información Geográfica y sus Infraestructuras de Apoyo, un Perfil Java para la Especificación *Coarse-Grain* de OpenGIS

J. Noguera¹, R. López², O.Cantan³, F.J. Zarazaga⁴, J. Gutierrez⁵

Departamento de Informática e Ingeniería de Sistemas
Universidad de Zaragoza

María de Luna 3, 50015 Zaragoza, España

¹jnog, ²lrafa, ³ocantan, ⁵adsogu@ebro.cps.unizar.es

⁴javy@posta.unizar.es

<http://iaaa.cps.unizar.es>

Resumen. El "*OpenGIS Consortium*" usa el término "*Catalog*" para describir el conjunto de interfaces que soportan organización, búsqueda y acceso a información geoespacial. Los servicios de catálogo ayudan a usuarios, ó al software de aplicación, a encontrar información que ya existe en cualquier sitio dentro de un entorno de computación distribuido. Un catálogo de información geográfica se puede considerar como una base de datos especializada en información acerca de recursos geoespaciales que están disponibles para un grupo ó una comunidad de usuarios. Este trabajo presenta el desarrollo de servicios de catálogo de información geográfica, y sus infraestructuras de apoyo, de acuerdo a la especificación "*Coarse-Grain*" de *OpenGIS* utilizando tecnología Java.

1. Introducción

Actualmente existe una gran cantidad de datos geográficos que han sido recogidos desde hace más de 35 años por diferentes instituciones y empresas con diferentes objetivos, y la velocidad de esta recolección de geodatos se incrementa rápidamente con avances en tecnologías tales como los sistemas de imágenes de satélite de alta resolución, los sistemas de posicionamiento global (GPS, Glonass), los sistemas de bases de datos, las nuevas tecnologías de software aplicables al geoprocetamiento y con el creciente número de personas y organizaciones que están recogiendo y utilizando geodatos [1]. A todo esto hay que añadir que alrededor de un 80% de las bases de datos utilizadas en la administración contienen referencias geográficas (direcciones postales, coordenadas cartográficas o distribución por municipios, sectores, barrios, secciones censales, etc.), por lo que puede hacerse un tratamiento de estos datos relacionado con la localización. Sin embargo, todo este ingente volumen de información no esta siendo suficientemente explotado debido, principalmente, al desconocimiento de su existencia por parte de los posibles consumidores. En este sentido, los catálogos son la herramienta necesaria para poder poner en contacto a consumidores con productores de información [2].

El OpenGIS Consortium (en adelante OpenGIS)[3][4] es una organización sin ánimo de lucro, fundada en 1994 y dedicada a la promoción de nuevas aproximaciones técnicas y comerciales para geoprocetamiento abierto e interoperable, fundada por las más importantes entidades industriales, gubernamentales y académicas. Los miembros del OpenGIS comparten una visión positiva de una infraestructura nacional y global de información, en la que los geodatos y los recursos de geoprocetamiento se mueven libremente, completamente integrados con las últimas tecnologías de computación distribuidas, accesibles a cualquiera, "geo-habilitando" una gran variedad de actividades, que se encuentran actualmente fuera del dominio del geoprocetamiento, abriendo de esta forma nuevos mercados. Su objetivo es conseguir acuerdos sobre interfaces de software estándar abiertos que posibiliten la ínteroperación entre sistemas de geoprocetamiento de vendedores diferentes y de diferentes tipos (GIS/SIG, teledetección, cartografía automática, gestión de instalaciones, etc.); comunicando por medio de interfaces derivadas-de-consenso sistemas diferentes accederán a datos y funciones de procesamiento de cada uno a través de redes.

El OpenGIS Consortium usa el término "*Catalog*" para describir el conjunto de servicios de interfaz que soportan la organización, descubrimiento y acceso de información geoespacial. Estos servicios pertenecen a los servicios *Domain Access* que se encuentran dentro de los servicios *Open Exchange*. Todas estas clases de servicios se manejan como *Common Applications* por el *OpenGIS Technical Reference Model* [3].

En este trabajo se presenta el desarrollo que se está realizando destinado a la construcción de un catálogo compatible con "OpenGIS Catalog Services" tal como se especifica en [4] y usando Java como lenguaje de programación. El resto del artículo se estructura de la siguiente forma. La siguiente sección presenta la arquitectura del catálogo y su relación con el proyecto global. La sección 3 muestra la infraestructura de metadatos geospaciales gestionada por el catálogo. En la sección 4, se muestra el diseño y la implementación de un catálogo compatible OpenGIS. El artículo termina con una sección de conclusiones y futuras líneas de trabajo.

2. Contexto y arquitectura del catálogo

El catálogo que aquí se presenta se enmarca dentro de un proyecto que se está realizando destinado a la creación de la tecnología necesaria para poder ofrecer servicios de catalogación y consulta de metadatos a través de Internet. La Figura 1 muestra la arquitectura de alto nivel de este proyecto. En ella se puede observar como el catálogo es el elemento central del mismo y alrededor del cual giran toda un serie de servicios destinados a ofrecer diferentes grados de funcionalidad a los proveedores y consumidores de información geográfica, así como labores de administración y gestión del catálogo tanto en modo local, como remoto.

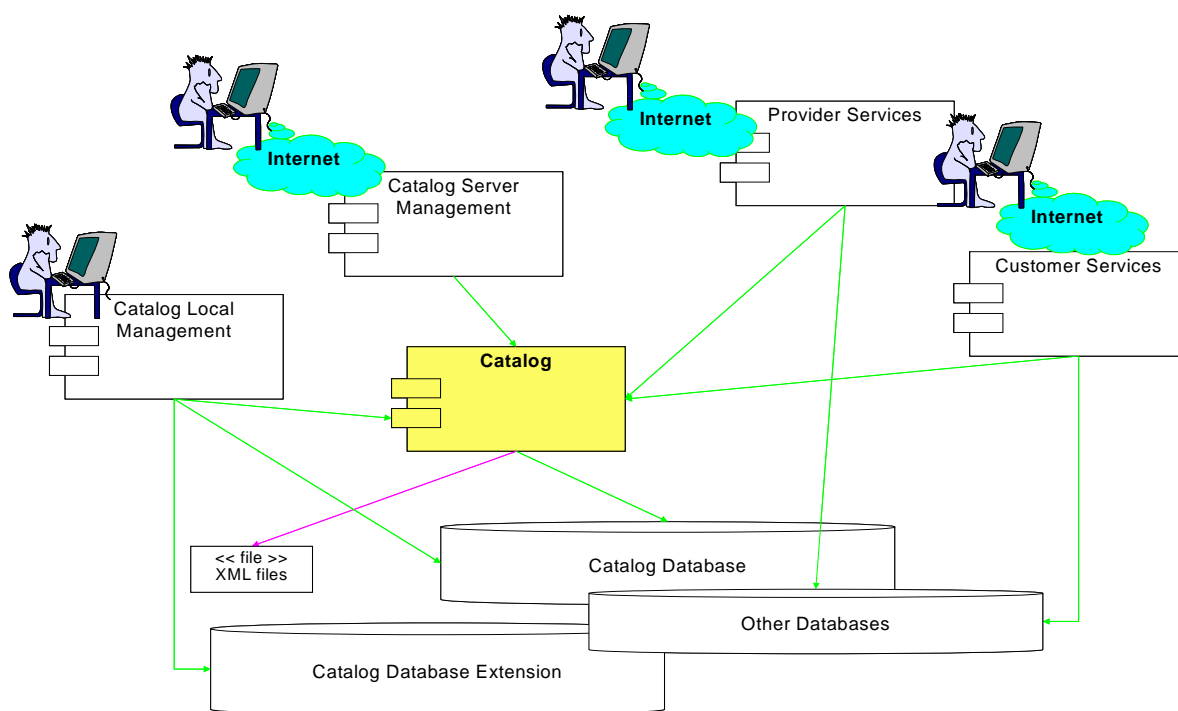


Figura 1. Contexto del catálogo

La especificación de OpenGIS establece la interfaz con la que debe contar un catálogo sin entrar a determinar como se lleva a cabo su implementación. La arquitectura que se está siguiendo para la construcción del mismo (ver Figura 2) propone la implementación de una serie de servidores que son los responsables de ofrecer los aspectos básicos de funcionalidad, y un servidor CORBA que llevará a cabo la implementación de la gestión de sesiones, según la interfaz que propone OpenGIS en el *Coarse-Grain Structural Model* que se presenta en [5], mientras que el resto de servicios que ofrece la interfaz estándar se redirigirán a los correspondientes servidores que integran el catálogo. Estos últimos, además de disponer de los servicios básicos presentados en el estándar, ofrecen una serie de servicios adicionales que incrementan la potencia del catálogo para ser utilizado en aplicaciones específicas.

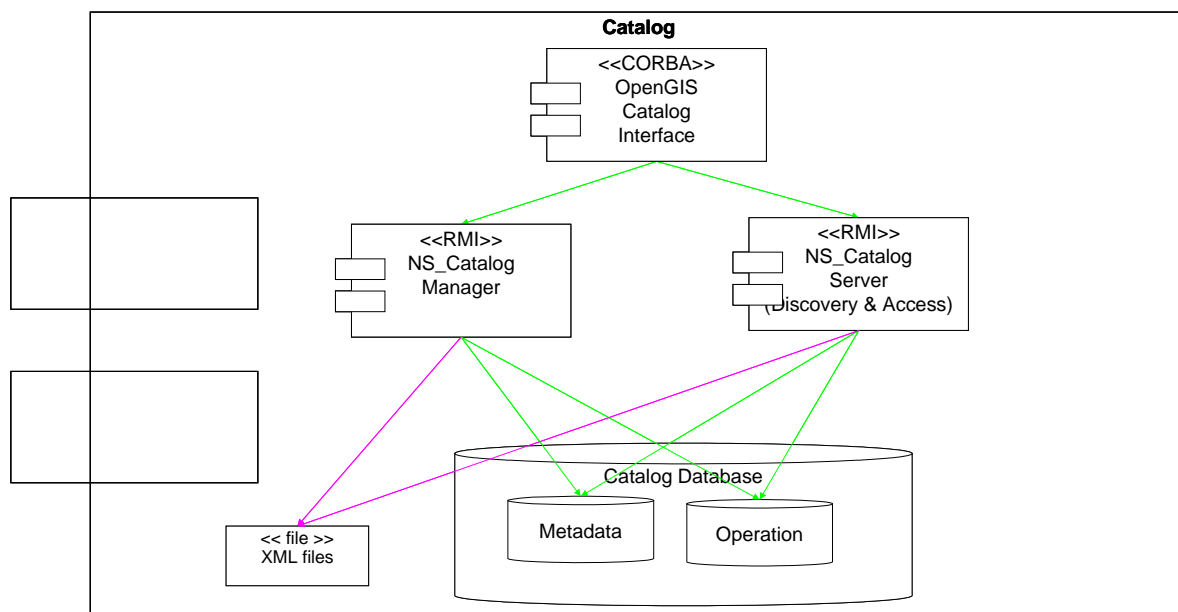


Figura 2. Arquitectura del catálogo

El catálogo trabaja con dos bases de datos. La primera de ellas es la destinada a dar soporte a los metadatos que describen los datos geográficos catalogados y sobre los que se realizan las búsquedas. La segunda permite llevar a cabo el control sobre las operaciones que se ejecutan en el sistema, tanto en los procesos de catalogación como en los de búsqueda, con el fin de tener una serie de informaciones recopiladas que permitan la confección de históricos de evolución de los metadatos y estadísticos de uso del catálogo.

El diseño de la arquitectura del catálogo se completa con un repositorio en el que se almacenan los metadatos en formato XML con el fin de poder facilitarlos eficientemente una vez que se solicitan sin tener que construirlos en su totalidad a partir de la información contenida en la base de datos de metadatos.

3. Metadatos geospaciales

Hay varios estándares ó pre-estándares en el mercado identificando elementos de metadatos que describen consistentemente un recurso de datos geospaciales específico. Algunos ejemplos de estos estándares son el americano *CSDGM* del FGDC [6], el *committee draft ISO CD 15046* propuesto por el comité *ISO/TC 211* (ver [7]), y el europeo *prENV 12657* propuesto por el comité *CEN/TC 287* (ver [8]). Tomando como punto de partida el estándar de la FGDC (*CSDGM*) con algunos elementos adicionales tomados de las recomendaciones del CEO se ha construido un modelo de objetos de referencia (la Figura 3 muestra la parte central del mismo ya que el modelo de objetos cuenta con unas 40 clases distintas) y su persistencia en base de datos relacional. Hay que tener presente que en el momento actual el estándar FGDC Americano es el único que se encuentra consolidado, el ISO es un borrador y el CEN Europeo son un conjunto de recomendaciones. Esto hace que cualquier posible desarrollo deba quedar abierto a las más que previsibles modificaciones que se vayan produciendo en las especificaciones de los metadatos. Nuestro propósito es establecer un modelo de objetos de referencia alrededor del cual se van a construir toda una serie de pequeños traductores o pasarelas que permitan la importación y exportación de metadatos ajustados a diferentes estándares.

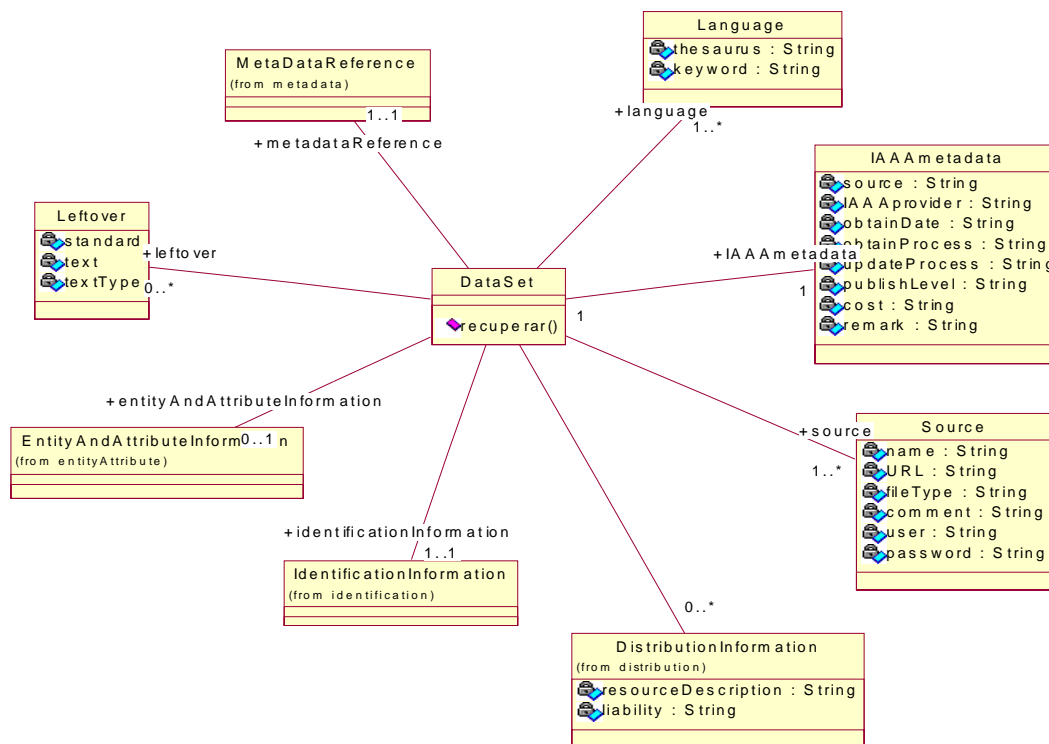


Figura 3. Modelo de Objetos de Metadatos (versión central del mismo)

La persistencia del modelo de metadatos en la base de datos consiste en un esquema entidad-interrelación dando persistencia al conjunto básico de elementos de metadatos. Este conjunto puede ser completado con un campo de texto (vease Leftover en figura 3) conteniendo todas las secciones no incluidas previamente el esquema ER pero disponibles en el proceso de catalogación (P. Ej.: Importando un fichero XML con metadatos acordes al estándar FGDC). El esquema Entidad-Interrelación da persistencia al conjunto de elementos considerados como más probables a la hora de realizar búsquedas, y otras secciones que resultan interesantes para la descripción e identificación de datos geoespaciales. Por otro lado, este campo de texto nos permite dar soporte a todas las secciones de los estándares usados para crear los metadatos y proporcionar en cualquier momento un fichero con las secciones de un estándar construido a partir del conjunto básico de elementos y el resto de secciones almacenadas en Leftover. Debemos remarcar que la forma de recibir/transferir metadatos con los proveedores/clientes del catálogo es a través de ficheros XML que cumplan el DTD (Document Type Declaration) del estándar correspondiente. Con vistas a facilitar las labores de creación y mantenimiento de los metadatos, se han construido un conjunto de herramientas capaces de trabajar directamente con el modelo de objetos presentado. En [9] y [15] puede encontrarse mayor detalle sobre estas herramientas de edición, creación, e importación y exportación de metadatos con XML.

Algunos elementos de metadatos, como la sección de *Keywords* del CSDGM ó secciones similares en otros estándares, consiguen sus valores de listas predefinidas de *controlled keyword*. Estas listas definen un conjunto de términos representando conceptos y categorías dentro de uno ó varios dominios temáticos que permiten clasificar y caracterizar datos provenientes de diferentes fuentes con términos relacionados. El uso de estas palabras clave controladas facilita el mapping entre un vocabulario seleccionado y una gran colección de datos geográficos. Esta capacidad de indexación, característica común de cualquier motor de búsqueda, aumenta el rendimiento de las consultas al catálogo basadas en estas secciones controladas. La implementación de estas controlled keywords está basada en un concepto más amplio denominado thesaurus, que permite la definición de relaciones entre términos como jerarquías, equivalencia (sinónimos) y asociaciones. Estos procesos de clasificación están asistidos por autoridades y organizaciones reconocidas que proporcionan índices y listas. Algunos ejemplos de estas listas son: las listas controladas "*Discipline*" y "*Parameter Type*" proporcionadas por CEO [10]; el "*NASA Master Directory*" [11] propuesto por el FGDC como un tesauro estándar.

dar temático; ó el "ADL Feature Type Thesaurus" [12] para indicar la naturaleza de una localización geográfica. De todas formas, es posible también crear una lista controlada para un dominio específico que no haya sido considerado en listas más generales. Un ejemplo de esto podría ser la clasificación de datos geoespaciales que se ha generado para llevar a cabo la clasificación de los datos que la *Confederación Hidrográfica del Ebro* tiene publicados en su página Web (<http://oph.chebro.es>).

Además de esta estructura jerárquica de palabras clave, hemos extendido las posibilidades de clasificación con sistemas más específicos de clasificación tomando ideas y conceptos del mundo de la Inteligencia Artificial y las Bibliotecas Digitales (en [13] se explican estas posibilidades de clasificación con mayor detalle).

4. Servicios del Catálogo compatible OpenGIS

La interfaz estándar OpenGIS compatible con el *Coarse-Grain Structural Model* propuesto en [5] engloba cuatro núcleos de funcionalidad (ver Figura 4): servicios de control de acceso al catálogo y gestión de sesiones (bajo la denominación de *Catalog Services*), servicios destinados a la gestión de catalogo (denominada *Catalog Manager*), servicios para recuperar la información almacenada a través de la descripción de sus metadatos (denominada *Access Services*), y una tercera para buscar dentro del catálogo (denominada *Discovery Service*).

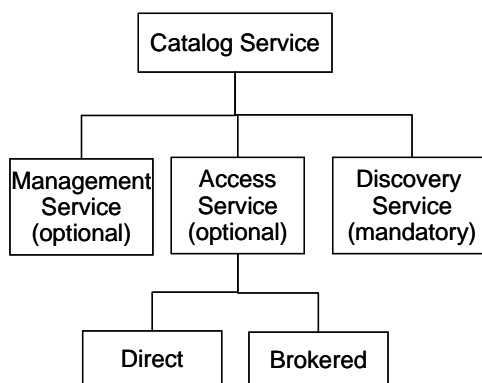


Figura 4. Arquitectura OpenGIS de los servicios del catálogo

A continuación se va a describir el diseño construido para llevar a cabo la implementación de estos servicios dentro de la arquitectura del catálogo reseñada anteriormente.

4.1. Catalog Services

Los servicios generales del catálogo son los responsables de llevar a cabo todas las tareas de gestión de sesiones. Hay que tener en cuenta que cuando se ejecuta una consulta el volumen de los datos que el catálogo devuelve puede llegar a ser equivalente a toda la información contenida en el catálogo (por ejemplo: *dame todos los datos que cumplan TRUE*). Obviamente, resulta completamente inviable el transferir toda esta información de una sola vez. Ante esta situación, OpenGIS realiza una separación de los servicios de acceso a metadatos (*Discovery Services*) en dos pasos, por un lado se ejecuta la búsqueda (*Search Services*), y por otro se solicita que parte de los resultados obtenidos se desea (*Present Services*). No existe, por tanto, una independencia entre servicios, sino que éstos permanecen vinculados basándose en el concepto de sesión.

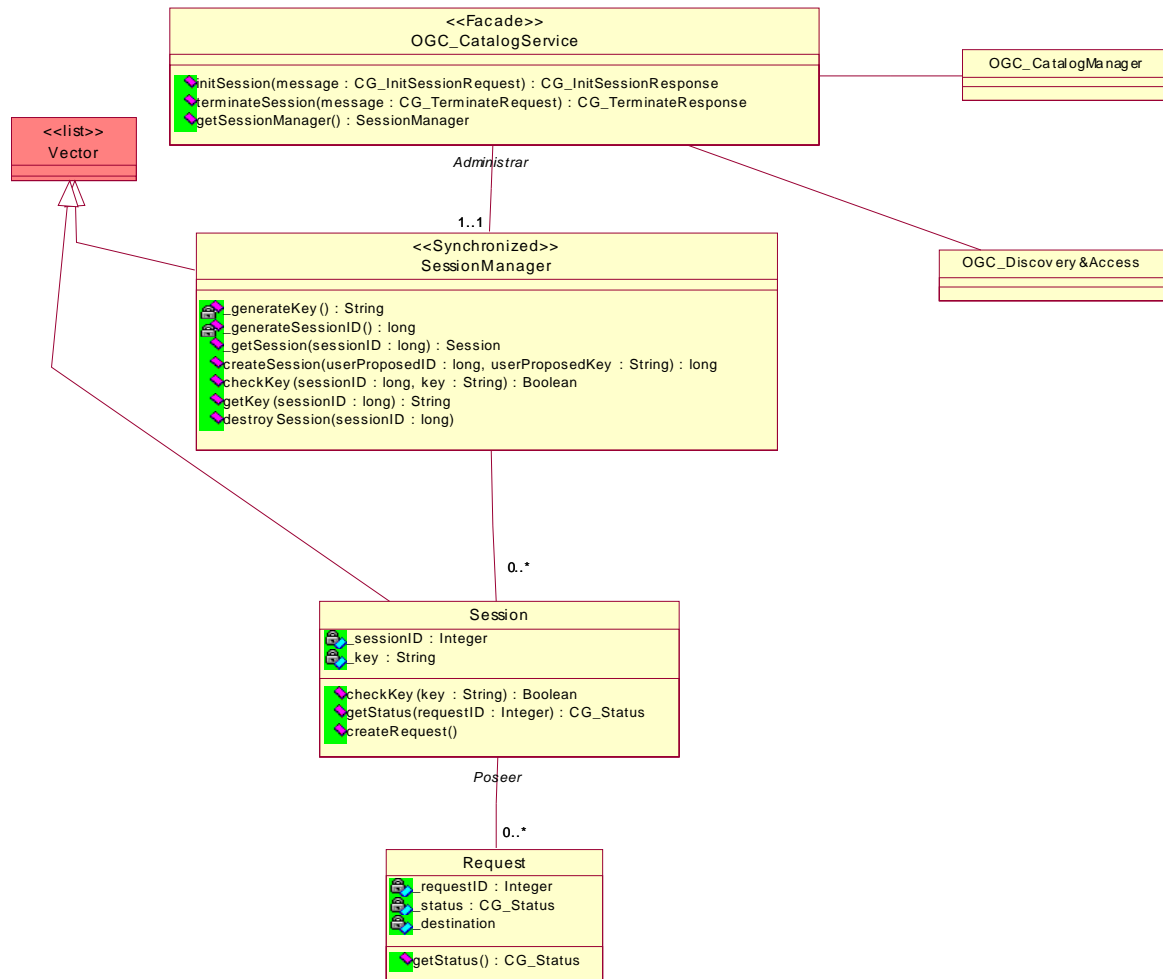


Figura 5. Catalog Services

Para llevar a cabo la construcción de esta gestión de sesiones, se ha recurrido a la utilización de un vector especializado que es el encargado de realizar todo el control de las sesiones. Cada una de las instancias del vector recoge la información asociada a una sesión concreta. Entre otras informaciones, una sesión debe guardar los aspectos referentes a las diferentes consultas asociadas a ella (una sesión, según la especificación de OpenGIS, puede tener cero o más consultas asociadas), por lo que también se ha diseñado como un vector especializado. Por otro lado, cuando un servicio de la clase *OGC_CatalogService* es solicitado, éste se limita a buscar en el vector de sesiones cuál es el servicio capaz de ofrecer la funcionalidad deseada.

4.2. Discovery Services

Los servicios OpenGIS Discovery permiten a los usuarios buscar dentro del catálogo. Las búsquedas de información pueden ser tanto locales, como distribuidas sobre otros catálogos OpenGIS. Para esto último, se dispone de un componente para descomponer las consultas en porciones destinadas a un catálogo compatible con OpenGIS específico, ya sea local ó remoto. Una vez que se obtienen los resultados parciales, este componente es capaz de juntarlos y componer la respuesta a la consulta del cliente. Para realizar una búsqueda distribuida es necesario tener información acerca del catálogo remoto así como su sintaxis de interrogación.

Para llevar a cabo la resolución de las preguntas en el catálogo local, se recurre a la invocación de los servicios necesarios dentro de la amplia interfaz proporcionada por el servidor RMI que actúa como catálogo extendido.

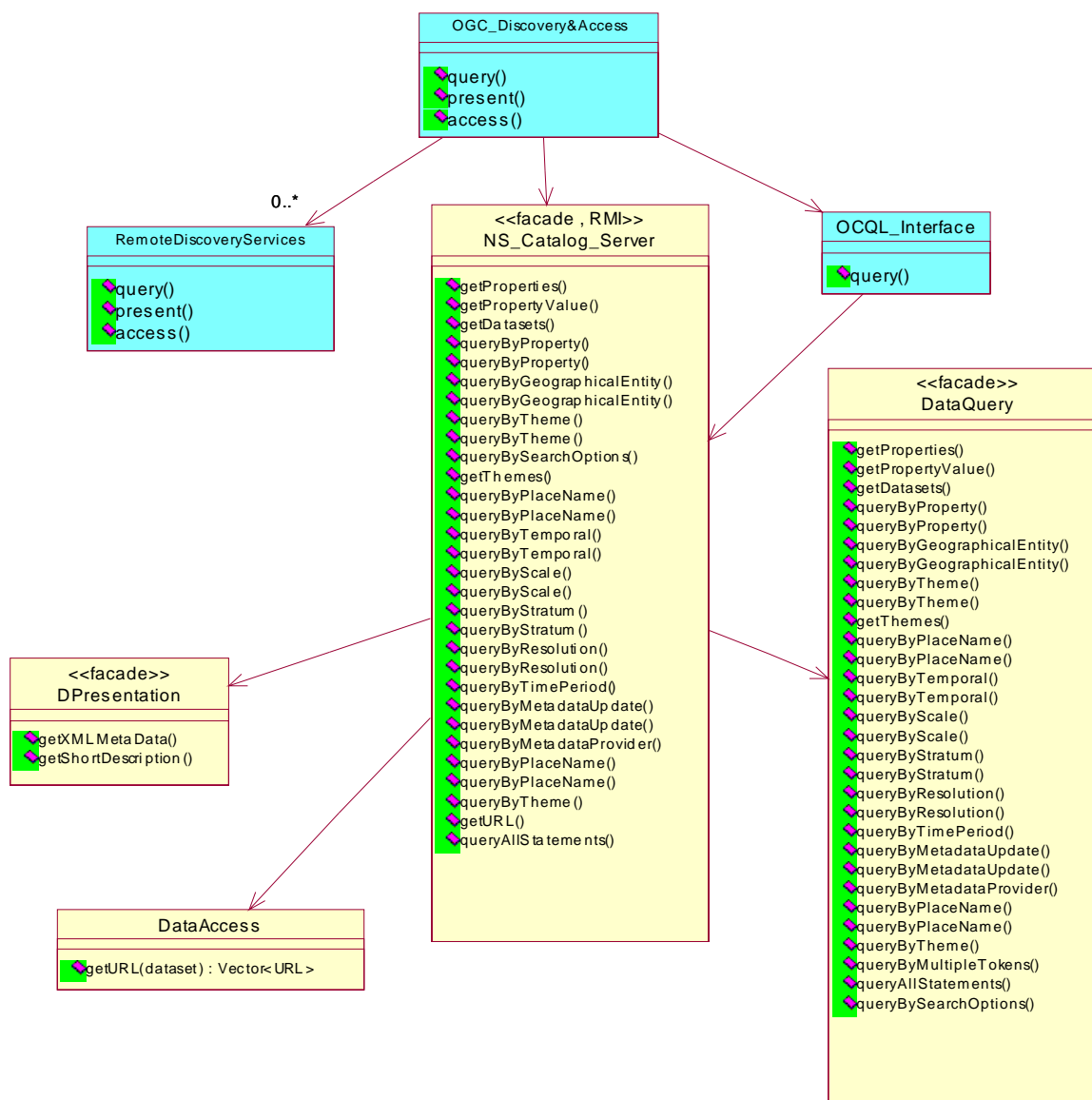


Figura 6. Discovery & Access Services

La Figura 6 muestra el modelo de objetos en UML en el que se recoge el diseño de los servicios de búsqueda y presentación, tanto del catálogo OpenGIS, como del catálogo extendido. La funcionalidad de búsqueda y presentación de resultados está encapsulada en el objeto utilizando el patrón de diseño *facade* (ver [14]) que permite ofrecer una única interfaz para ambas clases de servicios: búsqueda (clase *DataQuery*) y presentación (clase *Dpresentation*).

La consulta del cliente está escrita en lenguaje neutral de interrogación: el *OpenGIS Common Query Language* (OCQL), que es un subconjunto de SQL (ver [5]). La clase *OCQL_Interface* es la que realiza la traducción semántica entre las consultas expresadas en OCQL y las operaciones ofertadas por el catálogo extendido.

Finalmente, la clase *OGC_Discovery&Access* es responsable de descomponer las consultas en porciones destinadas a servicios remotos (vía *RemoteDiscoveryServices*) ó a los servicios locales, y para juntar los resultados parciales con el fin de componer la respuesta a la consulta del cliente.

4.3. Access Services

El servicio de acceso proporciona al usuario los medios para acceder a los datos geográficos localizados a través del servicio de *Discovery*. Este es un servicio cuya implementación no es obligatoria para ser compatible con el *OpenGis Catalog*. OpenGIS divide el servicio de acceso en 2 tipos: *Direct Access* (acceso directo) y *Brokered Access* (a través de un proceso intermedio).

Según la especificación, el acceso directo proporciona al cliente con un *handle* (manejador, referencia) que usado por el cliente permite el acceso a los datos. La definición específica de ese *handle* está fuera del alcance de la interfaz de catálogo de OpenGIS. Nuestra implementación oferta acceso directo a los datos almacenados (datos completos ó una imagen representativa de los mismos) a través de su URL, ó intentando invocar los servicios de acceso de los catálogos OpenGIS remotos. En la Figura 6 se puede observar este diseño. La clase *OGC_Discovery&Access* trata de satisfacer este servicio invocando al correspondiente servicio ofertado por el catálogo extendido. Nuevamente el uso del patrón *facade* permite que este servicio se oferte a través de la misma interfaz que el resto de servicios del servidor. En el caso de que la invocación deba ser remota, es la clase *RemoteDiscoveryServices* la que aporta los mecanismos necesarios para que este servicio se pueda satisfacer.

El otro tipo de servicio de acceso es el *Brokered Access* que proporciona al cliente los métodos para pedir los datos que serán entregados de alguna forma fuera de la interfaz de catálogo. La implementación de esta clase de servicio de acceso depende del proceso específico de cada organización proveedora de datos para distribuir sus productos. Por tanto, este trabajo debería ser realizado si el catálogo se particulariza para un proveedor especial.

4.4. Management Services

El objetivo de este conjunto de servicios es proporcionar las herramientas necesarias para añadir nuevos metadatos al catálogo, así como proceder a la actualización o eventual borrado de los ya catalogados. En la Figura 7 se puede observar el diseño básico de estos servicios.

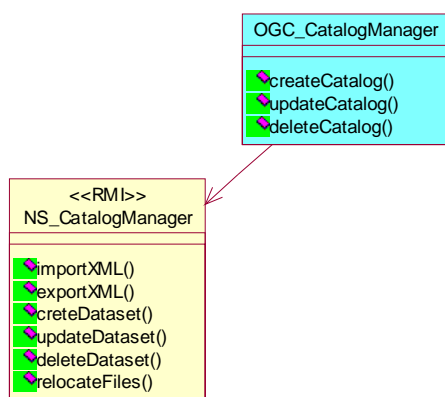


Figura 7. Catalog Management Services

Nuevamente se dispone de un objeto que es el responsable de ofertar la interfaz compatible OpenGIS. Los servicios de esta interfaz resuelven su funcionalidad invocando a los que están siendo ofertados por una versión extendida de la gestión del catálogo. Ésta incrementa la funcionalidad básica propuesta por OpenGIS con servicios destinados a proporcionar una mayor versatilidad y potencia a aplicaciones que se conecten directamente al servidor extendido de gestión del catálogo.

5. Conclusiones

En este trabajo se ha presentado la construcción de un catálogo de metadatos geográficos ajustado al estándar propuesto por OpenGIS. La necesidad de catálogos de metadatos geográficos queda plenamente justificada ante el gran volumen de datos existentes y el pobre aprovechamiento que los mismos se está realizando. En este sentido, la disponibilidad de una especificación de servicios comúnmente aceptada, como es la propuesta por OpenGIS, posibilita la homogeneización de estos catálogos y promueve su interoperabilidad.

Uno de los problemas más relevantes que hemos encontrado ha sido el "des-estandarizado" mundo de los estándares de metadatos geoespaciales. La estandarización propuesta por OpenGIS afecta, de momento, a los servicios de catálogo, pero no a los metadatos que en dicho catálogo se almacenan. Debido a esto, ha sido necesario preparar nuestros componentes para que se adapten fácilmente a nuevos estándares, y nuevas versiones de los actuales.

La utilización de Java en el desarrollo del catálogo esta facilitando el trabajo en aspectos tales como la implementación de procesos servidores (el uso de tecnologías de objetos distribuidos como CORBA y, especialmente, RMI están plenamente integradas dentro del propio lenguaje), multiplataforma del software desarrollado y gran reutilización del código (prácticamente de forma directa, el mismo código puede presentarse como un servidor RMI o integrarse dentro de una única aplicación).

La principal desventaja de Java como lenguaje de programación procede en parte de una de sus mayores ventajas: el hecho de ser un lenguaje multiplataforma. Esto obliga a que, por norma general, el código de Java se interprete, en lugar de compilarse, siendo de esta forma más lento y con tendencia a un mayor consumo de memoria que otros lenguajes cuyo código es compilado. De todas formas, esto es cada vez un problema menor, ya que están surgiendo nuevas herramientas de optimización [16].

Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto TIC98-0587 de la Comisión Interministerial de Ciencia y Tecnología (CICYT).

Bibliografía

1. K. Buehler, L. McKee (eds.). The OpenGIS Guide. Introduction to Interoperable Geoprocessing. Part I of the Open Geodata Interoperability Specification (OGIS). OGIS Project 6 Technical Committee of the Open GIS Consortium Inc. OGIS TC Document 96-001. 1996.
2. J.P. Baron, M.J. Shaw, A.D. Bailey. "Web-based E-catalog Systems in B2B Procurement". *Communications of the ACM*, May 2000. vol 43, n.5. pp 93 – 100.
3. OpenGIS Project Document 99-112, OpenGIS Consortium 1999. The OpenGIS Specification Model. Topic 12: The OpenGIS Service Architecture (version 32).
4. OpenGIS Project Document 99-113, OpenGIS Consortium 1999. The OpenGIS Abstract Specification. Topic13: Catalog Services (version 4).
5. OpenGIS Project Document 99-051s, OpenGIS Consortium 1999. OpenGIS – Catalog Interface Implementation Specification (version 1.0).
6. Metadata Ad Hoc Working Group. Document FGDC-STD-001-1998, Federal Geographic Data Committee (USA) 1998. Content Standard for Digital Geospatial Metadata.
7. ISO/TC 211 committee draft for Geographic Information Metadata. ISO CD 15046 part 15, dated 1998. <http://www.statkart.no/isotc211/>
8. European Committee for Standardization (CEN), The CEN/TC 287 Geographic Information European Prestandards, ENV 12657. Euro-norme Voluntaire for Geographic Information- Data description - Metadata. <http://www.cenorm.be/>
9. F.J. Zarazaga, R. López, J. Noguerras, O. Cantán, P. Álvarez, P.R. Muro-Medrano. "Cataloguing and recovering distributed geospatial data, a Java approach to build the OpenGIS Catalog Services". EC-GIS 2000, 6th European Commission GI and GIS Workshop. Lyon, Francia, 28 – 30 Junio de 2000.
10. A User Guide provided by the Center for Earth Observation Programme (CEO programme) Of the European Commission. February 1999. Recommendations on Metadata. Describing the data, services and information you have available! (version 2.0).

11. NASA Master Directory. Draft geospatial thematic keywords from the NASA Master Directory in short and long format for CSDGM of FGDC. <http://www.fgdc.gov/clearinghouse/reference/refmat.html> .
12. Alexandria Digital Library Project. <http://www.alexandria.ucsb.edu>
13. J.A. Bañares, F.J. Zarazaga, J. Noguerras, J. Gutiérrez and P.R. Muro-Medrano. "Construction and Use of Concept Hierarchies from Word Taxonomies for Searching Geospatial Data". Aceptado para su publicación en las actas de las Geographical Domain and Geographical Information System. EURESCO Conferences. La Londe-les-Maures, France, 22-27 September 2000.
14. E.Gamma, R.Helm, R.Johnson, J.Vlissides, 1996. Design Patterns. Elements of Reusable Object-Oriented Software. Addison-Wesley.
15. F.J.Zarazaga, R.López, J. Noguerras, O. Cantán, P. Álvarez, P.R.Muro-Medrano. "First Steps to Set Up Java Components for the OpenGIS Catalog Services and its Software Infrastructure". 3rd AGILE Conference on Geographic Information Science, Helsinki/Espoo , Finlandia, 25 – 27 Mayo de 2000
16. B.Eckel, *Thinking in Java*. Prentice Hall PTR, 1998.