

Interoperabilidad entre estándares de meta-datos geográficos¹

D.Anaya, O.Cantán, J.Lacasta, J.Nogueras, F.J.Zarazaga

Depto. Informática e Ingeniería de Sistemas
Universidad de Zaragoza
María de Luna, 3. 50018- Zaragoza. España.
danaya, ocantan, jlacasta, jnog, javy@posta.unizar.es
<http://iaaa.cps.unizar.es>

Abstract. Para una eficiente gestión de la información geográfica es necesario poder disponer de una completa y consistente documentación de ésta. Los meta-datos (“datos acerca de los datos”) constituyen el mecanismo para caracterizar esta información geográfica y posibilitar que usuarios y aplicaciones puedan hacer un uso adecuado de esta información. No obstante, para que los meta-datos pueda ser aprovechados de manera general es necesario que sigan unas reglas y organización predeterminadas, es decir, es necesario que se ajusten a un estándar. El problema con el que se topan los usuarios de información geográfica en estos momentos es que, aunque existe un consenso prácticamente universal sobre el estándar a utilizar en un futuro próximo, hoy por hoy el trabajo de documentación de información geográfica se está realizando sobre distintos estándares, con los correspondientes problemas de interpretación de esta información. Una posible solución es la creación de mecanismos que permitan traducir esta información para adecuarla a los diferentes estándares, a estos mecanismos los denominaremos de manera genérica “pasarelas”. El objetivo de este trabajo es presentar el proceso de creación de pasarelas entre estándares que se ha utilizado por un equipo de la Universidad de Zaragoza para llevar a cabo la creación de los mecanismos que permitan traducir información entre algunos de los estándares de meta-datos geográficos más extendidos.

1 Introducción

La información geográfica, también denominada datos geo-espaciales, es la información que describe fenómenos asociados directa ó indirectamente con una localización respecto a la superficie de la Tierra. Esta información resulta vital para la toma de decisiones y la gestión de recursos en temáticas muy diversas (recursos naturales,

¹ **Agradecimientos:** La tecnología de base de este proyecto ha estado parcialmente financiada por el proyecto TIC2000-1568-C03-01 del Plan Nacional de Investigación Científica y Desarrollo Tecnológico del Ministerio de Ciencia y Tecnología de España, y por el proyecto P089/2001 del Consejo de I+D del Gobierno de Aragón. El trabajo de O.Cantán (ref. B119/2001) ha estado parcialmente financiado por una beca del Gobierno de Aragón y el Fondo Social Europeo.

instalaciones, catastros, economía...) y a todo tipo de niveles (local, regional, nacional ó incluso global) [1]. Actualmente, gran cantidad de datos geográficos son recogidos por diferentes instituciones y empresas. De hecho, se puede afirmar que alrededor de un 80% de las bases de datos utilizadas en la administración contienen referencias geográficas (códigos postales, coordenadas cartográficas,...). Sin embargo, en muchos casos las empresas o personas “consumidoras” de información geográfica no encuentran la información que necesitan y terminan optando por la adquisición de datos contruidos “a medida”. Por otro lado, resulta bastante habitual encontrar que dentro de las propias organizaciones se tiene desconocimiento de los datos con los que cuentan el resto de divisiones de la misma. Esta carencia de organización y sincronización provoca que se realicen consecutivas reconstrucciones de datos con las mismas características.

Esta problemática del desconocimiento sobre “lo que ya existe” no es nueva en otros campos como el de las bibliotecas (especialmente en el caso de las bibliotecas digitales) y para solventarla se requiere el uso y creación de meta-datos. Los meta-datos (“datos acerca de los datos”) constituyen el mecanismo para caracterizar los datos y aplicaciones, lo que posibilita que otras aplicaciones puedan hacer uso de dichos datos o invocar sus servicios. Los registros de meta-datos, cada uno de ellos describiendo un recurso específico, se agrupan en catálogos, que facilitan al usuario la posibilidad de identificar los recursos que le puedan ser de interés. Así pues, los catálogos de datos geoespaciales (contienen meta-datos con perfiles geoespaciales) constituyen la herramienta necesaria para poder poner en contacto a consumidores con los productores de información.

No obstante, todavía no existe un consenso mundial sobre cuáles son los meta-datos más apropiados para describir los datos geográficos. En la actualidad se está procediendo a la creación de estos meta-datos siguiendo un variado número de estándares, pre-estándares y normativas de diverso alcance, con el consiguiente problema para compartir y utilizar esta información. El objetivo de este trabajo es presentar el proceso seguido para llevar a cabo la construcción de una serie de pasarelas que permiten inter-operar entre algunos de los estándares que más se están utilizando en el mundo de la información geográfica.

El resto del artículo se estructura como sigue. A continuación se presentan distintas propuestas de estándares de meta-datos y se profundiza en la problemática de la interoperabilidad entre ellos. En el punto 3 se procede a especificar el proceso de creación de las pasarelas entre estándares de meta-datos. Seguidamente se explican las pasarelas desarrolladas y se muestra la aplicación del proceso para la construcción de una de ellas a modo de ejemplo. Por último, este trabajo finaliza con una sección de conclusiones y trabajo futuro.

2 Estándares de meta-datos geográficos y el problema de su interoperabilidad

Los meta-datos geográficos describen el contenido, calidad, condición y otras características de los datos que permite que una persona localice unos datos y los pueda

entender. Tal y como se menciona en [2], la creación de meta-datos tiene 3 principales objetivos:

- El primero es organizar y mantener la inversión en geodatos realizada por una organización. Conforme el tiempo pasa ó cambia el personal de una organización, los nuevos trabajadores probablemente desconocen el contenido y el uso que se debe hacer de los datos que se crearon con anterioridad y normalmente desconfiarán de los resultados que se puedan generar de esos datos. Por esta razón, las descripciones sobre el contenido y corrección de los datos geoespaciales proporcionadas por los meta-datos propiciarán una reutilización apropiada de los datos. Además, tales descripciones protegerán los intereses de la organización productora si surge algún conflicto debido a un uso incorrecto de los datos.
- El segundo objetivo es proporcionar información referencial para catálogos de datos geoespaciales. Los sistemas de información geográfica requieren a menudo la utilización de datos de diferentes temáticas. Sin embargo, pocas organizaciones se pueden permitir el lujo de crear todos los datos por sí mismos. Frecuentemente, los datos creados por una organización pueden ser también útiles a otros y si estos datos se hacen accesibles a través de catálogos, las organizaciones conseguirán encontrar: datos para usar, organizaciones colaboradoras para compartir y mantener colecciones de datos, así como posibles clientes.
- Finalmente, el tercer objetivo de los meta-datos es proporcionar información que ayude a la propia transferencia de geodatos. De hecho, los meta-datos deberían acompañar a la transferencia de los geodatos. De esta forma, los meta-datos facilitan que la organización receptora pueda procesar e interpretar los datos; incorporarlos a sus almacenes y actualizar los catálogos internos que describen sus datos.

Para extender el uso y entendimiento de los meta-datos por diferentes comunidades de usuarios, por ejemplo para permitir las búsquedas distribuidas a través de una red de catálogos, es preciso utilizar contenidos bien definidos, y por lo tanto es preciso un estándar de meta-datos. En este sentido, ha habido muchas propuestas de estándares para describir consistentemente un recurso geográfico que han ido surgiendo a escala nacional ó internacional y desde distintos ámbitos. Algunos de las más extendidas son:

- El "Content Standard for Digital Geospatial Metadata" (CSDGM) del Comité Federal de Datos Geográficos (FGDC) [2, 3]. Esta iniciativa americana es la única que actualmente tiene el rango de estándar y que fue llevada a cabo en los Estados Unidos por el FGDC y aprobada en 1994. Es un estándar nacional para meta-datos espaciales desarrollado para dar soporte a la construcción de la Infraestructura Nacional de Datos Espaciales de los Estados Unidos [4]. Este estándar ha sido adoptado en otros países como por ejemplo Sudáfrica o Canadá.
- El documento "Recomendaciones sobre meta-datos" del Comité de Observación de la Tierra de la Comisión Europea [5].
- La norma voluntaria europea prENV 12657 [6].
- Dublin Core [7]. Se trata de un estándar de meta-datos de alcance general muy utilizado en el mundo de las bibliotecas digitales que está siendo incorporado al mundo de la información geográfica con el objeto de permitir su compatibilidad con otros sistemas de catalogación de información.

- El borrador de estándar internacional DIS 19115 [8] . En 1992, la Organización Internacional de Estándares (ISO) creó el comité 211 (ISO/TC 211) con responsabilidades en “geomática”. Este comité se encuentra preparando una familia de estándares que en un futuro próximo cobrarán rango de oficialidad y entre ellos figura el 19115 encargado de estandarizar los meta-datos geoespaciales.
- Y en el ámbito español, la sección de meta-datos contemplada dentro de la norma "Mecanismo de Intercambio de Información Geográfica Relacional formado por agregación" (MIGRA). Esta norma fue propuesta por el Comité Técnico de Normalización 148 de AENOR (AEN/CTN 148) y su implantación quedaba limitada por “la implantación de una norma europea -CEN- o internacional -ISO” según se recoge en los objetivos de la propia norma. Actualmente, se está ya trabajando en la búsqueda de la compatibilidad entre MIGRA y los estándares ISO.

Aparte del estándar elegido, los sistemas de catalogación de meta-datos deben soportar (reconocer) tres formatos de meta-datos [1]: el formato de implementación (dentro de una base de datos ó sistema de almacenamiento), el formato de exportación ó codificación (diseñado para la transferencia de meta-datos entre distintos sistemas y computadores), y el formato de presentación (un formato apropiado para ser leído por las personas). Para los dos últimos formatos, hay un consenso general respecto al uso de XML (eXtensible Markup Language [9]) dado que es un lenguaje de marcado con reglas estructurales forzadas a través de un fichero de control (Document Type Definition ó DTD) que permite validar la estructura del documento, es decir, comprobar la conformidad respecto al DTD establecido para un estándar de meta-datos. Además, a través de una especificación complementaria (eXtensible Stylesheet Language ó XSL [10]), un documento XML puede ser usado junto a una hoja de estilo (expresada en XSL) para crear presentaciones ó informes de contenido según los requerimientos del usuario.

La intención de las distintas organizaciones que han propuesto estos estándares, muchos de ellos todavía en estado borrador, es de homogeneizar todas las iniciativas hacia ISO tan pronto sea aprobado como estándar internacional. Sin embargo, en estos momentos el estándar del FGDC sigue siendo el más extendido y soportado por herramientas dentro del mundo GIS. Así pues, hasta que llegue el momento de poder disponer de un único estándar reconocido a escala mundial, y para llegar a la comunidad más amplia de usuarios, los sistemas de catalogación de meta-datos deberían ofrecer esos meta-datos de acuerdo a varios de los estándares de meta-datos más populares ó generalizados. A medida que el número, tamaño y complejidad de los estándares de meta-datos crece, la tarea de facilitar los meta-datos en distintos estándares resulta más larga, difícil y tediosa. Con el objetivo de minimizar la cantidad de tiempo empleado para crear y mantener los meta-datos y para maximizar su utilidad hacia el conjunto de usuarios más amplio posible, se necesita que los meta-datos se creen y se mantengan en un único estándar y que estos sean accesibles a través de otros estándares de meta-datos relacionados. Según esta filosofía, la tendencia de los sistemas de catalogación actuales es intercambiar meta-datos en XML conformes a distintos estándares según la necesidad del usuario, es decir, proporcionando vistas de los meta-datos en ISO, FGDC ó estándares más genéricos como Dublin Core. Para mantener la accesibilidad a través de estándares de meta-datos relacionados se deben crear sistemas informáticos que hagan comunicar varios sistemas hablando varios

dialectos de meta-datos, mediante pasarelas y puentes, apoyados en tecnologías emergentes como el XML y el XSL, tal y como se verá más adelante.

Sin embargo, la construcción de pasarelas (*crosswalks*) entre estándares es mucho más que el uso de una serie de tecnologías del mundo de la programación. Una pasarela especifica la correspondencia entre un estándar de meta-datos y otro, proporcionando la capacidad de hacer accesible para comunidades que utilizan un estándar el contenido de los elementos definidos en otro. Por desgracia, la construcción de pasarelas constituye una tarea difícil y propensa a errores que requiere un profundo conocimiento y una buena experiencia en el trabajo con los estándares de meta-datos que se pretenden asociar. La obtención del conocimiento que se requiere para construir una pasarela es particularmente problemática puesto que los estándares de meta-datos son con frecuencia desarrollados de forma independiente, y en ellos se utiliza diferente terminología, métodos y procesos especializados. Por lo demás, mantener una pasarela entre estándares de meta-datos sujetos al cambio es todavía más problemático debido a la necesidad de mantener una perspectiva histórica. Por ello, la homogeneización en la especificación consistente de estándares de meta-datos relacionados resulta vital para el desarrollo de pasarelas. Ésta tiene como resultado la capacidad para crear y mantener un solo conjunto de meta-datos, y hacer corresponder éstos con cualquier número de estándares de meta-datos relacionados. Como consecuencia, y a través del uso de terminología, métodos y procesos comunes, se simplifica mucho el desarrollo, la implementación y la utilización de estándares de meta-datos relacionados. La siguiente sección propone una forma de generalizar y formalizar los meta-datos y sus pasarelas.

3 Construcción de pasarelas entre estándares de meta-datos

Esta sección presenta los pasos del proceso que se ha seguido para la construcción de una serie de pasarelas entre estándares (estas pasarelas se presentarán más adelante) y que permiten simplificar su implementación mediante el uso de especificaciones formales y procesos automáticos. El proceso sigue los siguientes pasos:

- Homogeneización. En esta fase se busca la especificación formal y homogénea tanto del estándar de meta-datos origen como del destino.
- Emparejamientos semánticos. Para determinar la correspondencia semántica de elementos entre los estándares de meta-datos se requiere un profundo conocimiento de los estándares de meta-datos origen y destino. Como resultado de esta fase se crea una tabla de emparejamientos.
- Cuestiones a resolver en la conversión de meta-datos. Aparte de la tabla de emparejamientos, una pasarela completa debe proporcionar una especificación de la conversión de meta-datos que incluya reglas de correspondencia entre elementos, resolución de las diferencias entre las jerarquías y conversiones del contenido de los meta-datos.
- Implementación automatizada de la pasarela. El último objetivo del proceso es conseguir una pasarela completamente automatizada mediante la aplicación de algún tipo de herramienta. De esta forma, manteniendo un único conjunto de me-

ta-datos, se pueden proporcionar búsquedas y vistas según las distintas familias de meta-datos.

En los siguientes sub-apartados se entra a presentar en mayor detalle cada uno de estos pasos.

3.1 Homogeneización

Muchos estándares de meta-datos utilizan características comunes en la definición de sus elementos de meta-datos. Algunos ejemplos de características comunes podrían ser: un único identificador por cada elemento de meta-datos (Ej.: *tag*, *label*, *identifier*); una definición semántica de cada elemento; el carácter obligatorio u opcional de un elemento basándose en ciertas condiciones; la multiplicidad; la organización jerárquica respecto al resto de elementos; restricciones sobre el valor del elemento (Ej.: texto libre, rango numérico, una fecha o un conjunto de valores predefinido). Una vez establecidas, las características comunes se pueden expresar y usar de manera similar en cada estándar de meta-datos. De esta forma, se pueden aplicar procesos similares a estándares de meta-datos relacionados, con lo que se simplifica tanto la implementación de los estándares de meta-datos como el desarrollo de nuevas pasarelas entre ellos.

La generalización y formalización de la especificación de las propiedades de los meta-datos de un estándar son posibles usando una representación canónica o lenguaje de especificación de meta-datos. Este procedimiento es análogo a la especificación de la sintaxis de un lenguaje de programación usando la conocida notación Backus-Naur-Form (BNF [11]). De hecho, dado el uso de XML en el formato de exportación y presentación de los estándares de meta-datos, se dispone para la mayoría de ellos de un DTD para especificar formalmente la sintaxis de dicho estándar.

Sin embargo, solamente una descripción sintáctica directa de un estándar de meta-datos dado no es apropiada para almacenar toda la información necesaria para automatizar el desarrollo de pasarelas. Es necesario definir un conjunto mínimo de tipos de datos que será usado para derivar a partir de él todos los tipos de datos que se requieren para representar todos los elementos en el estándar destino. Y además, al igual que ocurre con BNF, un lenguaje de especificación de meta-datos no contiene información sobre la semántica de los elementos. Por ello, en este paso se propone la creación de una tabla (que podría implementarse mediante el uso de una hoja Excel) describiendo los elementos de cada estándar aparte del DTD disponible por cada estándar. En esta tabla, cada elemento de meta-datos vendrá definido mediante cinco campos:

- Nr: Numeración que tiene asignada en su propio estándar de meta-datos según su nivel en la jerarquía.
- Longname: nombre largo del elemento estructurado jerárquicamente por niveles.
- Shortname: nombre "corto" del elemento, correspondiente al valor del identificador que se le asigne en su estándar
- Car: multiplicidad junto a restricción de obligatoriedad que impone el estándar sobre ese elemento.
- Descripción: definición semántica del elemento.

3.2 Emparejamientos semánticos

La tarea más importante del desarrollo de pasarelas es la que se ocupa de la determinación de la correspondencia semántica de elementos entre los estándares de meta-datos [12]. Esta tarea implica especificar un mapeo entre cada elemento del estándar de meta-datos origen y aquel que equivalga semánticamente a éste en el estándar de meta-datos destino. Para lograr una tabla de correspondencias coherente es necesario una definición clara y precisa de los elementos de cada estándar.

Muchos estándares de meta-datos proporcionan ya un mapeo semántico con estándares de meta-datos relacionados, con frecuencia este mapeo aparece en forma de tabla en un anexo del estándar. En el proceso que aquí se presenta, al final de esta fase, se construye una tabla de emparejamientos (que se podría almacenar en una hoja Excel).

3.3 Cuestiones a resolver en la conversión de meta-datos

Una pasarela es un conjunto de transformaciones que se aplican al conjunto de elementos del estándar de meta-datos origen, almacenándose como resultado el contenido convenientemente modificado en los elementos análogos del estándar de meta-datos destino. Una pasarela completa (o completamente especificada) está constituida por una tabla de emparejamientos semánticos y una especificación de la conversión de los meta-datos. Esta especificación contiene las transformaciones adicionales requeridas para convertir el documento de meta-datos cuyos contenidos cumplen con el estándar de meta-datos origen en un documento cuyos contenidos cumplen el estándar de meta-datos destino y que no hayan quedado convenientemente especificadas en la tabla de emparejamientos. En los siguientes subapartados se detallan esas cuestiones adicionales a resolver, que en el proceso que se presenta aparecen de modo descriptivo en una columna adicional de la tabla de emparejamiento (comentando las decisiones tomadas para resolver casos especiales) .

3.3.1 Conversión de contenido

Normalmente, los estándares de meta-datos restringen los contenidos de cada elemento a un tipo de datos, rango de valores o vocabulario controlado en particular. Y en algún caso, puede ocurrir que dos elementos que deben emparejarse contengan restricciones distintas en cuanto a su contenido en cada estándar. Por ejemplo, puede darse el caso de tener que transformar texto en valores numéricos ó texto en un valor de tipo fecha. Por tanto se requieren reglas específicas para establecer la correspondencia entre un elemento origen cuyo valor está especificado como texto libre y un elemento destino cuyo valor está restringido a un vocabulario controlado, así como para relacionar elementos origen y destino cuyo contenido está sujeto a diferentes vocabularios controlados.

3.3.2 Correspondencia elemento a elemento

Todos los estándares de meta-datos especifican un número de características asociadas con la definición de algunos elementos. Algunos estándares califican cada

elemento como repetible o no repetible, también pueden indicar cuando un elemento es opcional u obligatorio, aunque en algunos casos como FGDC, se incorporan estos atributos en una sola propiedad, indicando el número de ocurrencias mínimo y máximo de un elemento. Cuando el número de ocurrencias mínimo es cero, se está indicando que dicho elemento es opcional, mientras que si es uno, se indica que es obligatorio. Durante el desarrollo de una pasarela, estas propiedades deben tenerse en cuenta para el mapeo de cada elemento. El caso más sencillo o trivial se da cuando los elementos tienen propiedades idénticas, por ejemplo cuando los dos elementos análogos en cada uno de los estándares son obligatorios y no repetibles, y ambos tienen el mismo tipo de contenido. A continuación se detallan algunos casos de mayor interés y que requieren una resolución más compleja:

- Uno a muchos. En la mayoría de los casos la correspondencia uno a muchos es trivial ya que la ocurrencia del elemento origen corresponde a una sola ocurrencia del elemento destino. Sin embargo existen situaciones particulares en las que el mapeo requiere una solución más específica. Por ejemplo, el estándar origen puede contener un elemento “*keywords*” no repetible cuyo contenido está constituido por una o más palabras clave separadas por comas. Y sin embargo, este elemento se corresponde en el estándar destino con un elemento repetible representando a cada palabra clave. En este caso, el mapeo requiere un conocimiento especializado de la composición del elemento origen, y cómo se expande en múltiples elementos destino. Otro caso de interés es la correspondencia de un elemento origen a dos elementos destino. Por ejemplo, una pasarela de GILS a DIF debería emparejar el elemento de GILS “*Contact Name*” con los elementos destino de DIF “*First Name*” y “*Last Name*”. En este caso, las reglas generales deben especializarse para extraer correctamente el nombre y apellidos del contenido de “*Contact Name*” y emparejarlos con los elementos DIF correspondientes.
- Muchos a uno. La correspondencia muchos a uno debe especificar qué hacer con los elementos extras. Si la solución es relacionar todos los valores del elemento origen con el valor único del elemento destino, se requieren reglas explícitas para especificar cómo deben concatenarse estos valores en el elemento destino. Sin embargo, si la resolución es vincular solamente un valor del elemento origen, con la consecuente pérdida de información posible, se debe indicar el criterio de selección del elemento que efectivamente se va a emparejar (por ejemplo, el valor de la primera ocurrencia del elemento, o el valor del último elemento añadido) y la consecuente pérdida de información.
- Elementos extras en el origen. Otro importante caso que requiere una solución específica es la gestión de un elemento del estándar origen que no corresponde con ningún elemento apropiado en el estándar destino. Muchos estándares de meta-datos proporcionan la capacidad de almacenar información adicional, de forma que la solución adoptada debe especificar de manera precisa cómo debe ser añadido el valor del elemento.
- Elementos obligatorios del estándar destino no resueltos. En algunos casos, hay elementos obligatorios en el estándar destino que no son el correspondiente de ningún elemento del estándar origen. Como el estándar destino requiere un valor para estos elementos, la pasarela debe proporcionar un valor para ellos a partir de los recursos disponibles.

3.3.3 Jerarquía

La mayoría de los estándares de meta-datos organizan sus datos jerárquicamente (mediante secciones y subsecciones). La pasarela debe tratar las posibles diferencias entre las jerarquías de los estándares de meta-datos origen y destino. En el proceso que se presenta, la propia tabla de emparejamiento muestra los elementos organizados jerárquicamente en cada estándar, aunque por supuesto no se profundiza en el emparejamiento de aquellas secciones que se definen recursivamente (por ejemplo sección *Citation* en FGDC) y que hacen que el nivel de jerarquía sea ilimitado.

3.4 Implementación automatizada de la pasarela: uso de hojas de estilo

Teniendo en cuenta que los estándares de meta-datos que se han comentado en la introducción utilizan XML como formato de exportación y presentación, se ha considerado que la tecnología más adecuada para llevar a cabo la implementación de pasarelas es el uso de XSL (eXtensible Stylesheet Language [10]), cuyo objetivo es precisamente la manipulación y transformación de XML. XSL es un lenguaje para expresar hojas de estilo que integra a su vez dos lenguajes relacionados: un lenguaje de transformación (XSL Transformations ó XSLT); y un lenguaje de formateado (XSL Formatting Objects) de documentos XML equivalente al CSS (Cascading Style Sheets) para las páginas HTML. El lenguaje de transformación (XSLT) proporciona elementos que definen reglas para transformar un documento XML en otro documento XML, que puede usar el mismo marcado que el documento original y estar asociado al mismo DTD o puede utilizar un conjunto de elementos completamente diferente.

Así pues, el método para realizar transformaciones consistirá en construir la hoja de estilo que aplicada al documento XML de entrada (conforme el estándar de meta-datos correspondiente) genere como salida un documento XML cuyos elementos cumplan con el estándar destino, y que contenga la misma información representada en el documento de entrada. A continuación se detalla la metodología general que se ha seguido durante la implementación de las hojas de estilo que actúan de pasarela entre los distintos estándares de meta-datos. La metodología seguida se basa en la transformación sucesiva de cada una de las secciones aplicando las tablas de emparejamiento que se han establecido con anterioridad. En particular, se siguen los siguientes pasos para ir completando la hoja de estilo:

- Establecer la declaración del tipo de documento que aparecerá en el documento de salida, y que incluirá la ruta del DTD correspondiente al estándar de meta-datos destino.
- A continuación, por cada sección a emparejar en el estándar de metadatos destino:
 - Se creará una plantilla (basándose en la tabla de emparejamientos) cuyo patrón es el elemento padre de la sección a relacionar del estándar origen, que genere los elementos correspondientes en el estándar destino. En esta plantilla se aplicarán las reglas de transformación necesarias en el estándar origen para cumplir la especificación en cuanto a propiedades y contenido en el estándar destino.

- Una vez que se tiene una primera versión de la hoja de estilo, se aplica sobre un documento XML que cumpla el estándar origen y contenga todos los elementos posibles de la sección ya emparejada. El procesador de hojas de estilo (Ej.: parser XML de Oracle para Java en <http://technet.oracle.com/>) genera como salida un nuevo documento. Aunque probablemente este documento no validará el DTD correspondiente al estándar destino (ya que únicamente contendrá los elementos de las secciones relacionadas hasta ahora), se debe comprobar que las transformaciones se han realizado correctamente. Mediante el uso de una herramienta de edición de XML se debe chequear visualmente sobre el árbol de nodos generados: la ausencia de algún elemento obligatorio; el orden de los elementos generados posición; y la corrección del contenido. En caso de detectar algún error se volvería al paso 1.
- Adicionalmente, se debe comprobar que no se pierde información aplicando la hoja de estilo inversa sobre el documento del estándar destino y obteniendo, como resultado, un nuevo documento que cumple el estándar origen y que únicamente contiene completas las secciones emparejadas hasta ahora. Si existiese alguna diferencia entre el documento inicial y este nuevo documento generado, habría que comprobar en la tabla de emparejamientos si algún elemento del estándar origen no tiene mapeo posible en el estándar destino, y si es así, comprobar si la pérdida de información se debe a esto. En caso contrario, se puede concluir que la pérdida de información se debe a un error en la hoja XSLT, por lo que hay que revisarla y volver al paso 1.
- Una vez se ha comprobado que la transformación de la sección anterior se realiza correctamente, se vuelve al paso ii para comenzar el mismo proceso con la sección siguiente hasta llegar a transformar completamente el documento origen en un documento destino que cumpla con el estándar correspondiente.

4 Pasarelas realizadas

Siguiendo el proceso de construcción de pasarelas que se ha especificado en el apartado anterior, se han obtenido aquellas que realizan la transformación entre el estándar FGDC y los estándares ISO, DC y MIGRA. La **Fig. 1** muestra las pasarelas realizadas donde las flechas en color gris representan las pasarelas realizadas entre el estándar de meta-datos origen y el estándar de meta-datos destino hacia el que apunta la flecha. Adicionalmente, las flechas en blanco representan las pasarelas derivadas que se obtienen a partir de aplicación dos pasarelas intermedias.

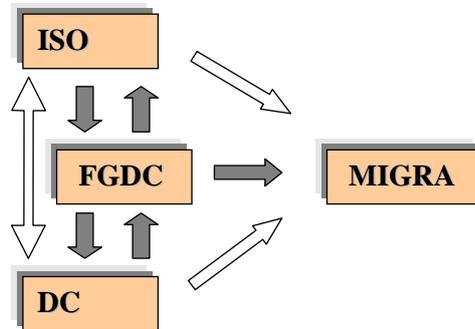


Fig. 1. Pasarelas realizadas y pasarelas derivadas

Como ejemplo de la construcción de pasarelas, este apartado detalla la conversión de FGDC a ISO. Aunque el objetivo de los dos estándares es documentar los recursos de información geográfica, presentan algunas diferencias reseñables.

<i>Secciones FGDC</i>	<i>Paquetes ISO</i>
<i>Secciones principales</i>	
<i>Identification information</i>	<i>Identification information</i> (incluyendo también referencias a secciones <i>Constraint information</i> , <i>Maintenance information</i>)
<i>Data quality information</i>	<i>Data quality information</i>
<i>Spatial data organization Information</i>	<i>Spatial representation information</i>
<i>Spatial reference information</i>	<i>Reference System Information</i>
<i>Entity and Attribute Information</i>	<i>Content information</i>
<i>Distribution Information</i>	<i>Distribution Information</i>
<i>Metadata Reference Information</i>	<i>Metadata entity set information</i> (incluyendo también referencias a secciones <i>Constraint information</i> , <i>Maintenance information</i> , y <i>Metadata extension information</i>)
	<i>Portrayal catalogue information</i>
	<i>Application schema information</i>
<i>Secciones reusables</i>	
<i>Citation information, Contact Information</i>	<i>Citation and responsible party information</i>
<i>Time period information</i>	<i>Extent information</i>

Tabla 1. Emparejamiento de secciones entre FGDC e ISO

En cuanto a la documentación y organización del estándar, el FGDC se presenta estructurado en 10 secciones (7 principales y 3 reutilizables) y contiene 469 elementos diferentes, 119 de los cuales existen únicamente para contener otros elementos. El documento del estándar utiliza reglas de producción en BNF para expresar la sintaxis y por cada elemento se proporciona una descripción del contenido que se debe asignar a cada elemento. Sin embargo, el estándar ISO utiliza ya metodología orientada objeto y se especifica mediante el uso de diagramas UML para establecer las relaciones y la organización de la información. Aquí las secciones principales del FGDC serían equivalentes a los paquetes en los que se estructuran las distintas clases representando distintas entidades de meta-información. En la Tabla 1 se puede observar como ambos estándares poseen algunas secciones comunes a un primer nivel, aunque

esto no es así en niveles más profundos donde elementos análogos se encuentran en puntos diferentes de la jerarquía. Además, el estándar ISO proporciona un diccionario donde se recogen los nombres (509 elementos) de todas las clases y atributos describiendo su contenido y dominio de valores.

Con referencia a la información semántica, el estándar ISO, dado su carácter más novedoso y conciliador, trata alguna de las deficiencias que se pueden encontrar en el estándar FGDC. Por ejemplo, en el estándar ISO se dispone de tipos de datos *Raster* e *Imagery*), mientras que en FGDC sólo se dispone del primero. Asimismo, estos estándares contienen diferencias en cuanto a la terminología de los elementos. Por ejemplo, el elemento *bounding box* del estándar FGDC contiene cuatro elementos, cuyos nombres cortos son, respectivamente: *westbc*, *eastbc*, *northbc* y *southbc*. El elemento correspondiente en el estándar ISO también contiene cuatro elementos: *westBL*, *eastBL*, *northBL* y *southBL*. La única diferencia entre estos elementos y los del estándar FGDC estriba en su nombre, puesto que semánticamente son iguales.

A pesar de las diferencias, uno de los puntos comunes en ambos estándares es que el formato de codificación y presentación más aceptado es XML. Los documentos que pretendan cumplir los respectivos estándares deben mantener conformidad respecto al DTD establecido para el estándar de meta-datos correspondiente. Esto permite que se puede implementar la construcción de la pasarela mediante una hoja de estilo. Como resultado, la hoja de estilo FGDC -> ISO creada permite transformar 4 de las 7 secciones principales del FGDC (incluyendo toda la parte que resulta obligatoria en ambos estándares). Las 3 secciones de las que no se realiza la correspondencia son *Spatial_Reference_Information*, *Entity_and_Attribute_Information* y *Data_Quality_Information*. Establecer la correspondencia de las dos primeras resultó impracticable ya que la organización y filosofía de las secciones en ambos estándares son totalmente dispares. A modo de ejemplo, mientras que los nombres de las subsecciones de *Spatial_Reference_Information* del FGDC se corresponden con los distintos sistemas de coordenadas (*Transverse Mercator*, *Mercator*, *Equidistant Conic*, ...), en ISO lo que se utilizan son códigos mantenidos por una organización (Ej. European Petroleum Survey Group) que se encarga de catalogar los distintos sistemas de coordenadas, elipsoides, datums, En cuanto a la tercera de ellas (*Data_Quality_Information*), esta sección no es obligatoria en ambos estándares y no se estimó oportuno realizar la transformación al no disponer tampoco de registros de meta-datos con dichas secciones completadas que facilitasen su interpretación semántica, además de posibilitar un estudio de la validez de los datos transformados.

5 Conclusiones y trabajo futuro

Este trabajo ha presentado el proceso seguido para llevar a cabo la construcción de una serie de pasarelas que permiten inter-operar entre algunos de los estándares que más se están utilizando en el mundo de la información geográfica, ilustrándolo con el caso concreto de una de las pasarelas realizadas.

Actualmente todas las organizaciones que catalogan meta-datos geográficos (conformes a estándares como FGDC, CEN o MIGRA) pretenden converger hacia el estándar internacional ISO. Pero además de esto, cada día cobra más auge poder pro-

porcionar una visión más genérica de estos meta-datos conforme a estándares de facto como Dublin Core, que progresivamente se va imponiendo en las instituciones públicas de distintos países. Ante estas exigencias, resulta desaconsejable la utilización de distintos editores para conseguir catalogar los meta-datos según todos los estándares que necesitemos. Por el contrario, es preferible para una institución mantener los meta-datos catalogados conforme a un único estándar utilizando una herramienta de catalogación estable y construir las pasarelas que permitan proporcionar vistas de esos mismos meta-datos acordes al estándar que interese en cada momento. No obstante, estas pasarelas deben construirse de acuerdo a métodos formalizados que permitan verificar los resultados de las transformaciones de información, controlando las posibles pérdidas de contenido que puedan derivarse de dichas transformaciones.

Como se ha indicado, el proceso presentado ha sido utilizado para la elaboración de una serie de pasarelas que permiten llevar a cabo la transformación entre algunos de los estándares más populares. Aunque resulte extraño, en estos momentos no se conoce la existencia de pasarelas entre estos estándares que estén accesibles para su uso (libremente o mediante pago) . De hecho, se han establecido contactos con FGDC e ISO para que el trabajo que se ha desarrollado sea utilizado por estas organizaciones para la creación de la pasarela oficial entre ambos estándares. De igual modo, también se va a participar en un proyecto a nivel Europeo que tiene entre sus objetivos el establecer la semántica de los campos de Dublín Core para su uso como estándar de meta-datos geográficos. En este caso, las pasarelas realizadas servirían como primera propuesta de trabajo, o versión cero del resultado esperado.

Una vez construidas estas pasarelas se pretende probar su utilidad en la construcción de aplicaciones de búsqueda, sobre catálogos de información geográfica, donde se permitan establecer restricciones y presentar resultados acordes a un estándar específico. El cliente de búsqueda accederá a un portal ó entrada de una red de catálogos distribuidos en la que cada uno de ellos podrá mantener meta-datos conformes a distintos estándares. Sin embargo, tanto en las peticiones de búsqueda como en la presentación de resultados se aplicarán las pasarelas necesarias para obtener los meta-datos según el formato (estándar) deseado por el usuario.

Bibliografía

1. Nebert D. (eds), 2001. *“Developing Spatial Data Infrastructures: The SDI Cookbook v.1.1”*. Global Spatial Data Infrastructure (<http://www.gsdi.org>), dated May 2001.
2. *“Content Standard for Digital Geospatial Metadata Workbook”*, version 2.0. Federal Geographic Data Committee (USA), 2000.
3. Document FGDC-STD-001-1998 Content Standard for Digital Geospatial Metadata. Metadata Ad Hoc Working Group, Federal Geographic Data Committee (USA), 1998
4. *“Homepage of the National Geospatial Data Clearinghouse : <http://www.fgdc.gov/clearinghouse/clearinghouse.html>”*, National Spatial Data Infrastructure, dated May 2002.
5. *“A User Guide provided by the Center for Earth Observation Programme (CEO programme) Of the European Commission. Recommendations on Metadata. Describing the data, services and information you have available! (version 2.0)”* Center for Earth Observervation.

6. *ENV 12657. "Euro-norme Voluntaire for Geographic Information-Data description- Metadata"*. European Committee for Standardization (CEN), The CEN/TC 287 Geographic Information European Prestandards, <http://www.cenorm.be>.
7. "Homepage of the Dublin Core Metadata Initiative : <http://www.dublincore.org> ", dated May 2002.
8. "Draft International Standard ISO/DIS 19115, Geographic information — Metadata". ISO/TC 211 (<http://www.isotc211.org>), dated September 2001.
9. *Extensible Markup Language (XML) 1.0 (Second Edition)*. W3C Recommendation, <http://www.w3.org/TR/2000/REC-xml-20001006>, dated October 2000.
10. *The Extensible Stylesheet Language (XSL)*. <http://www.w3.org/Style/XSL/>
11. Naur P (ed.), 1963, Revised report on the algorithmic language Algol 60, Comm. ACM 6:1 pp1-17.
12. St. Pierre M., LaPlant W.P., 1998. "Issues in crosswalking Content Metadata Standards". National Information Information Standards Organisation, <http://www.niso.org/press/whitepapers/crswalk.html>, dated 1998.