



Universidad
Carlos III de Madrid



Esta es una versión postprint del siguiente documento:

Zarazaga-Soria, F. J., Nogueras-Iso, J., Méndez-Rodríguez, E., Tolosana-Calasanz, R. y Muro-Medran, P. R. (2006). El papel de Dublin Core en el desarrollo de las infraestructuras de datos espaciales. En Carlos Granell y Michael Gould (eds.). *Avances en las infraestructuras de datos espaciales*. Universitat Jaume I: Castelló de la Plana, pp. 51-60.

© Autores, 2006

El papel de *Dublin Core* en el desarrollo de las infraestructuras de datos espaciales

F. J. Zarazaga-Soria*, J. Noguerras-Iso*, E. Méndez-Rodríguez**, R. Tolosana-Calasanz*, P. R. Muro-Medrano*

* Departamento de Informática e Ingeniería de Sistemas Universidad de Zaragoza, María de Luna 1, 50018, Zaragoza - e-mail: iaaa@unizar.es

** Departamento de Biblioteconomía y Documentación Universidad Carlos III de Madrid, 28903, Getafe (Madrid) - e-mail: emendez@bib.uc3m.es

RESUMEN

Las tendencias actuales de caracterización de recursos de información geográfica para su oferta a través de Infraestructuras de Datos Espaciales (IDE) se centran en la información geográfica más tradicional (mapas, coberturas, modelos digitales del terreno, etc.). Para ello han utilizado como soporte de descripción los trabajos del grupo TC 211 de ISO (fundamentalmente ISO 19115). No obstante, todavía queda una ingente cantidad de servicios e información más heterogéneos susceptibles de ser ofrecidos a través de una IDE. El reto que se plantea en estos momentos es la definición de una estrategia que posibilite la incorporación de estas fuentes y servicios de información garantizando la interoperabilidad entre sistemas, sin que ello vaya en detrimento de una descripción o caracterización de recursos, adecuada, completa y suficiente. Es en este contexto donde *Dublin Core* puede jugar un papel fundamental como norma de metadatos (ISO 15836) de propósito general, fomentando la interoperabilidad en distintos dominios informativos, entre ellos también la información geoespacial. El objetivo de este capítulo es presentar un modelo de utilización del conjunto de elementos y principios de *Dublin Core* como base para el proceso de asignación de metadatos asociados a todo tipo de recursos en el contexto de una IDE, así como las decisiones técnicas básicas que deberían tomarse para dar soporte a servicios de creación y búsqueda de información sobre este modelo general propuesto.

PALABRAS CLAVE

metadatos, *Dublin Core*, interoperabilidad.

1. Introducción

Desde que surge el concepto de Infraestructura de Datos Espaciales (IDE) en 1994 [1] se ha llevado a cabo un gran avance en el desarrollo de conceptos, modelos y arquitecturas que han posibilitado la creación de una sólida base. Esta base ha permitido la puesta en funcionamiento de un número cada vez más importante de IDEs a niveles locales, regionales, nacionales y transnacionales. No obstante, basta con echar un rápido vistazo para observar que el eje central sobre el cual se sustenta la mayor parte de estas iniciativas lo constituye la información geográfica más tradicional (mapas, coberturas, modelos digitales del terreno, etc.). Para ello se ha venido utilizando como soporte de caracterización de estas informaciones la norma ISO 19115 y la especificación técnica ISO 19139. En algunos entornos IDE ya se está comenzando a trabajar sobre la idea de caracterizar los servicios geográficos haciendo uso de la norma ISO 19119. Sin embargo se está dejando de lado toda una ingente cantidad de servicios e información más heterogéneos susceptibles de ser ofrecidos a través de una IDE (un ejemplo concreto puede encontrarse en [2]). Ya en 1998 Goodchild explicaba que la comunidad de sistemas de información geográfica había venido trabajando con la información geográfica, mientras que la información geo-referenciada es mucho más amplia al incluir cosas tales como las fotografías, los videos, la música y la literatura que pueden dar una variable de localización [3].

El reto que se plantea en estos momentos es la definición de una estrategia que posibilite la incorporación de estas fuentes y servicios de información garantizando la interoperabilidad entre sistemas, sin que ello vaya en detrimento de una descripción o caracterización de recursos, adecuada, completa y suficiente.

El resto del capítulo se estructura como sigue. A continuación se presenta una rápida revisión de los estándares de caracterización de información y servicios que están siendo utilizados en la mayor parte de las IDEs que se están construyendo. Seguidamente se propone *Dublin Core* como modelo base de caracterización a utilizar en las IDEs, y como este estándar va a posibilitar la integración de nuevas fuentes de información y servicios, así como la compatibilización con los modelos y servicios actualmente en uso. Este capítulo termina con un apartado de conclusiones.

2. Estándares de metadatos en IDEs

Dos son los grandes contribuidores en el ámbito de la estandarización en IDEs: del comité técnico TC 211 de ISO y *Open Geospatial Consortium* (OGC). Como conjunción de los trabajos de ambas entidades se puede identificar el manejo de metadatos en dos grandes áreas: los catálogos de datos y servicios, y los *capabilities* o declaraciones de funcionalidad de los servicios OGC.

2.1. Catálogos de datos y servicios

El OGC define una especificación de servicios de catálogo (*Catalogue Services Specification*) a tres niveles de detalle:

- *Modelo general (General Model)*: Modelo abstracto que especifica un conjunto de interfaces de servicios que soportan la funcionalidad de descubrimiento (*discovery*), acceso (*access*) y mantenimiento y organización (*maintenance and organization*) de catálogos de información geoespacial y sus recursos relacionados.
- *Protocolo de conexión (Protocol Binding)*: Modelo que añade guías para el diseño de la implementación del modelo general. Esto incluye un mapeo entre las interfaces, operaciones y parámetros generales y los constructores disponibles en el protocolo seleccionado.
- *Perfil de aplicación (Application profile)*: Modelo que extiende un protocolo de conexión documentando las decisiones de implementación y seleccionando una representación concreta para los contenidos de los catálogos.

El protocolo de conexión CSW (*Catalogue Services for the Web*) es la primera propuesta de OGC en la cual se incluye referencia explícita a los modelos de metadatos a utilizar en los procedimientos de intercambio de información. Los actuales trabajos de especificación de catálogo presentan tres perfiles de aplicación: CORE, ISO 19115/ ISO 19119 y *ebRIM*:

- *Perfil CORE*: Según la especificación de OGC para este perfil, se establece una serie de términos por los cuales se puede interrogar al catálogo. La propia especificación indica que estos términos están basados en los propuestos por *Dublin Core*. De manera análoga, a la hora de devolver los resultados, se indica que las respuestas seguirán un modelo análogo al propuesto por *Dublin Core*. No obstante, en ninguno de los dos casos se establece claramente cuál debe ser el esquema concreto a utilizar. De hecho, según las propuestas que se presentan en la especificación del perfil, parece ser que se plantea el uso del esquema XML de *Dublin Core*, mientras que el más extendido y completo es el esquema RDF.
- *Perfil ISO 19115/ISO 19119*: Esta perfil trata de cubrir el uso del catálogo OGC tanto como catálogo de metadatos de datos como de metadatos de servicios. Las dos mayores limitaciones con el que se tropieza vienen de la mano de la mezcla de estándares y, fundamentalmente, de la no disponibilidad de modelos estándar de representación de estos estándares en XML (ISO 19139 está a punto de convertirse en estándar, mientras que no se conoce la existencia de un trabajo específico de sustentación de ISO 19119 en XML).
- *Perfil ebRIM*: Según figura en la propia definición del término, *ebRIM* (*ebXML Registry Information Model*) especifica el modelo de información

utilizado por un registro *ebXML* (<http://www.ebxml.org>). Su cometido es describir el modo en el que va a poderse acceder a los servicios a través del registro *ebXML* al cual se encuentra vinculado. En el caso del catálogo OGC, *ebRIM* proporciona la descripción del modelo de metadatos que se está utilizando por la correspondiente implementación a la que se encuentra asociado. Como se puede observar, este perfil no fija ningún estándar de metadatos concreto a utilizar.

2.2. Los *capabilities* de los servicios OGC

Una de las herramientas más utilizadas para la caracterización de los servicios OGC de una IDE viene de la mano de las *capabilities* de sus servicios. En el modelo general de servicios que plantea OGC, todos los servicios heredan de OGC *Web Service*, que es una interfaz en el que se define como operación *GetCapabilities*. Esta operación tiene como cometido el proporcionar una descripción de las capacidades del servidor y suministrando al cliente información de los servicios que ofrece y de los elementos sobre los que ofrece el servicio. En este sentido, las *capabilities* de un servicio OGC constituyen los metadatos del mismo. No obstante, aunque esta descripción recoge muchos de los elementos que aparecen en estándares como *Dublin Core* e ISO 19119, no hay establecida ninguna correspondencia formal con los mismos.

3. Modelo base de metadatos para una IDE

3.1. *Dublin Core*

La iniciativa de metadatos *Dublin Core*, simplemente *Dublin Core* o DC [4], es actualmente el modelo de metadatos más aceptado para describir, recuperar e intercambiar información electrónica, independientemente del dominio científico o disciplinar. En sus orígenes (marzo 1995), el DC surge como un modelo de metadatos dirigido a la descripción embebida en el código (entonces) HTML por parte de los autores de los recursos, para una recuperación más eficaz y cualificada en motores y otras herramientas de búsqueda web, liderando el desarrollo de metadatos estructurales para la recuperación de información en Internet [5]. Con el tiempo, el *Dublin Core* ha ido evolucionando hacia un formato de registro para el intercambio de información y a un estándar básico para la interoperabilidad entre repositorios de información científica, sobre todo gracias a la integración del DC con el protocolo OAI-PMH [6], pero también a la versatilidad del esquema y al nivel de estandarización formal que ha adquirido.

El conjunto de elementos *Dublin Core* (DCMES) [7, 8], implica 15 elementos básicos para la descripción y recuperación de recursos digitales independiente de la disciplina, tipo de información o dominio científico de los recursos. Desde su defi-

nición en 1995 hasta la actualidad, el interés por este modelo ha ido creciendo y su éxito se debe, entre otras razones a:

- El nivel de normalización formal que ha adquirido rápidamente. Uno de los problemas habituales de los estándares para la web es que se desarrollan y utilizan en un nivel de facto o de especificaciones de dominio público, donde son muy pocos los que alcanzan en nivel de reconocimiento como estándar formal (*de jure*) ISO. En el caso del DC, es, desde el año 2003, un estándar internacional ISO 15836, pero alcanzó el rango de norma en el año 2001, entonces reconocida por la autoridad de normalización americana (NISO). Desde su proclamación como estándar ISO, distintos países han mostrado su credibilidad en este esquema de metadatos, reconociéndolo como estándar nacional, por ejemplo en España, se convertirá en una norma UNE a lo largo del año 2006.
- Es un estándar simple y fácil de entender y usar, cuya codificación es independiente de una sintaxis particular y también de una disciplina o dominio científico específico. La mayoría de los estándares de metadatos nacen ligados a una sintaxis de codificación (normalmente XML), el DC se puede codificar tanto en XML (fundamental en su uso con OAI [6]) como en RDF, en estructuras de bases de datos, e incluso en HTML (tanto en el código fuente de la cabecera como a través de microformatos). También, la mayor parte de los esquemas de metadatos surgen en un contexto de aplicación disciplinar o dominio particular (por ejemplo, FGDC-CSDGM –*Federal Geographic Data Committee*–, en el ámbito de la información geoespacial, TEI –*Text Encoding Initiative*–, en el contexto de la información literaria y/o humanística, etc.), sin embargo, el *Dublin Core* es un modelo de metadatos de propósito general, aplicable a distintas disciplinas, distintos objetos de información digital o en distintos repositorios o colecciones digitales.
- La simplicidad del DC no va en detrimento de su especificidad, ya que se puede usar con diferentes matizaciones, vocabularios o esquemas (*Dublin Core Terms* [9]) y establecer perfiles de aplicación específicos para un dominio informativo particular.
- Hoy en día es un formato ampliamente aceptado con grandes usos y aplicaciones en todo el mundo. Por ejemplo, es el formato adoptado por múltiples programas y proyectos de e-Administración (*e-government*) en Australia, Canadá, Dinamarca, Finlandia, Irlanda, Nueva Zelanda y Reino Unido o para proyectos supranacionales como la Red Europea EUKN (*European Urban Knowledge Network*). Al ser un estándar ISO, cada vez más aplicaciones, de carácter libre o comercial, incluyen el *Dublin Core* como modelo de metadatos de salida y/o almacenamiento (por ejemplo, Conexión OCLC, DCPS, *ContentDM*, *Metabrowser*, etc.).

3.2. Nuevas posibilidades funcionales

Imaginemos un escenario con tres sistemas de información distintos que cumplan cada uno de ellos su objetivo concreto sobre la base de caracterizar la información utilizando estándares internacionales: una biblioteca “en línea” que posibilita el acceso a libros digitales, informes y otras publicaciones, todas ellas caracterizadas utilizando como estándar MARC [10]; un servidor de información de eventos culturales de una ciudad y cuyo modelo base es *Dublin Core*; y finalmente una IDE. Suponiendo que cada uno de estos servicios permitiese el acceso a sus funcionalidades de búsqueda por terceros sistemas de información, sería posible desarrollar servicios especializados como un proveedor de información turística (eventos y publicaciones pueden ser unidos con informaciones referentes a las rutas a seguir durante un viaje turístico) o un proveedor de información cultural (publicaciones pueden ser vinculadas a un evento y adicionalmente se puede suministrar información relativa a las vías de acceso y ubicación del lugar donde ocurre el evento) tal y como se muestra en la figura 1.

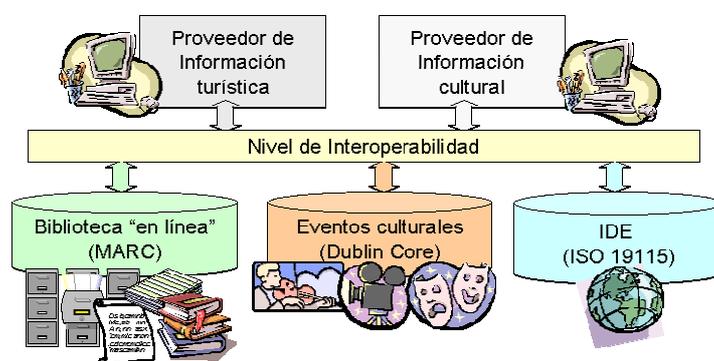


Figura 1. Caso de uso de interoperabilidad

Tal y como se puede observar, la principal limitación con la que se cuenta es la necesidad de entender y hacer que los demás entiendan la información que se intercambia. Para ello es necesario dotarse del nivel de modelo de datos común que puede corresponder con la propuesta hecha anteriormente sobre la base de *Dublin Core*. De este modo, si una IDE es capaz de ofrecer sus informaciones bajo esta perspectiva, se abren las puertas a una mayor explotación de la misma al posibilitar su inclusión como sistema legado en otros sistemas de información.

Por otro lado, si los propios servicios con los que opera la IDE son capaces de moverse sobre los esquemas de *Dublin Core* se nos abren otras dos puertas de suma importancia. Por un lado, se podrán integrar nuevos tipos de contenidos que, separándose un poco de lo que constituyen las clásicas informaciones geográficas, pueden jugar un papel fundamental a la hora de dotar a los usuarios de informaciones de interés. De esta manera, informaciones tabulares de censos, informes electrónicos (en formato PDF, HTML, PS, etc.) sobre determinadas áreas geográficas, o

incluso “hiper-enlaces” con direcciones web relevantes podrían ser incluidas con facilidad en nuestros servicios de catálogo. De otra parte, de manera relativamente sencilla se podrían incorporar las informaciones de otros sistemas legados a los resultados de los servicios de una IDE. Por ejemplo, el servicio de nomenclátor de la IDEE podría enlazar casi de manera automática con los almacenes de fotografías de los puntos de agua con los que cuenta la Confederación Hidrográfica del Ebro al objeto de poder mostrar fotografías de los elementos que cataloga.

3.3. Adaptación a *Dublin Core* de las IDEs actuales

Cuatro son los componentes fundamentales en los que sería necesario trabajar con vistas a sustentar el funcionamiento de una IDE sobre el modelo general de metadatos basado en *Dublin Core*: metadatos de datos y servicios, servicios de búsqueda, y servicios de nombres.

3.3.1. Metadatos de datos y servicios

Dentro de los procesos de creación de metadatos y servicios, sería necesario que los modelos que se manejasen permitiesen la compatibilidad con *Dublin Core* (algunas herramientas ya lo permiten, como es el caso de *CatMDEdit* (<http://sourceforge.net/projects/catmdedit>). DC ha demostrado su capacidad para servir de base al manejo de metainformación de diferentes tipos de fuentes a través de la definición de perfiles de aplicación o *application profiles* adaptando su semántica a un dominio de información particular, o a través del mapeo de elementos con otros esquemas de metadatos (creación de *crosswalks*) que permiten configurar jerarquías de tipos de elementos, utilizando a *Dublin Core* como fundamento de conversión de todos ellos [11]. Para el caso concreto de información geográfica, ya existe una definición de pasarela o *crosswalk* que permite la correspondencia de elementos de ISO 19115 a *Dublin Core* sin pérdidas de información [12]. Esto ha dado pie a la definición de un perfil de aplicación específico para la creación de metadatos de información geográfica en el marco de un proyecto piloto para el lanzamiento de la directiva Europea INSPIRE [13] (*Dublin Core Metadata Application Profile for Geographical Data Mining*). Sería necesario llevar a cabo actuaciones semejantes para el caso de ISO 19119 o de las *capabilities* devueltas por los servicios OGC a través de la operación *GetCapabilities*.

3.3.2. Servicios de búsqueda

La mayor parte de los motores de recuperación de información que dan soporte a los servicios de búsqueda que operan en la actualidad se basan en los modelos XML de los estándares de la serie ISO 19100, o en modelos relacionales propieta-

rios. Por otra parte, a la hora de recuperar por parte del cliente las informaciones, prácticamente en ningún caso se ofrece la posibilidad de obtener la información en *Dublin Core*.

En el caso de los motores de recuperación, sería necesario modificarlos para que fuesen capaces de indexar y buscar sobre el modelo *Dublin Core*. Esto no tiene por qué resultar nada extraño ya que si se estudian los campos que componen DC se puede observar que éstos encajan bastante bien con lo que se denominan metadatos de búsqueda en [14]. Además, ya existen soluciones en la actualidad que trabajan sobre esta filosofía [15].

En la parte de los resultados de las búsquedas, OGC ya ofrece una solución suficientemente satisfactoria mediante el perfil CORE de la especificación CSW del catálogo. Sería necesario forzar a que todos los catálogos que se pongan en funcionamiento cumplieren dicha especificación.

3.3.3. Servicios de nombres (servicio de nomenclátor, *gazetteer*)

Un caso especial de servicios de una IDE lo constituyen los servicios de nombres por el gran volumen de información que maneja (solamente la Xunta de Galicia cuenta con alrededor de un millón de topónimos), la gran cantidad de aplicaciones de los resultados obtenidos y los problemas de duplicados (especialmente a la hora de integrar nombres de varias fuentes). El estándar ISO 19112 [16] propone una especificación de contenidos para este tipo de servicios que se puede aproximar bastante a DC. Sin embargo, este modelo de contenidos deja pendiente toda una serie de flecos que trabajos como el que se está desarrollando en la actualidad para la elaboración del “Modelo Español de Nomenclátor” está poniendo de manifiesto. El uso de DC para la creación de este modelo puede suponer un factor fundamental para paliar estas deficiencias, así como facilitar la incorporación de contenidos de gran relevancia en los servicios de nombres (puntos de agua de la Confederación Hidrográfica del Ebro, elementos relevantes del Patrimonio Industrial [2], etc).

4. Conclusiones

En este trabajo se ha presentado una propuesta de utilización de *Dublin Core* como modelo base de todos los metadatos con los que se trabaje en una IDE. La adopción de este modelo algo más general que los que en la actualidad se están utilizando requeriría unos pequeños ajustes técnicos sobre las actuales soluciones tecnológicas, pero, sin embargo, sentaría unas prometedoras bases en tres líneas fundamentales: un mayor y mejor aprovechamiento de los datos y servicios que en la actualidad se están sirviendo por las propias IDEs, abriría las puertas a la incorporación de nuevos datos y servicios que en la actualidad no se encuentran presentes, y haría más fácil la integración de las IDEs con otros sistemas de información al

objeto de incrementar las posibilidades funcionales que se prestan sobre los servicios de éstos.

Ya se están poniendo en marcha iniciativas destinadas a aprovechar *Dublin Core* como pivote sobre el que hacer girar los contenidos de información y la tecnología y servicios de una IDE construidos sobre esta base. Patrimonio Industrial y Patrimonio Paleontológico son dos áreas de gran interés en las cuales se están dando los primeros pasos.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto TIC2003-09365-C02-01 del Plan Nacional de Investigación Científica y Desarrollo Tecnológico del Ministerio de Educación y Ciencia de España.

Referencias

- [1] U.S. Federal Register, "Executive Order 12906. Coordinating Geographic Data Acquisition and Access: the National Spatial Data Infrastructure (U.S.)", Edition of the Federal Register, 13 abril 1994.
- [2] F. J. Zarazaga-Soria, M. García, M. G. Hernández, P. Biel, F. J. López, J. F. Valero, "La IDEE como Herramienta de Ayuda a la Elaboración del Inventario del Patrimonio Industrial de Aragón", *Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'05)*, Madrid, noviembre 2005.
- [3] M. F. Goodchild, "The Geolibrary". *Innovations in GIS 5*, (eds.) S. Carver, Taylor and Francis, 1998.
- [4] DDCMI (Dublin Core Metadata Initiative): <http://dublincore.org>.
- [5] S. Weibel, T. Koch, "The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions", *D-Lib Magazine*, vol. 6, n° 12, diciembre 2000.
- [6] OAI-PMH (Open Archive Initiative-Protocol for Metadata Harvesting): <http://www.openarchives.org/>.
- [7] DCMES 1.1. (Dublin Core Metadata Element Set): <http://dublincore.org/documents/dces/>.
- [8] E. Méndez, "Dublin Core, Metadatos y Vocabularios". *El Profesional de la Información*, vol. 15, n. 2, marzo-abril, 2006.
- [9] DDCMI Metadata Terms: <http://dublincore.org/documents/dcmi-terms/> (junio 2005).
- [10] U.S. Library of Congress, "MARC standards", *Network Development and MARC Standards Office*, 2004, <http://www.loc.gov/marc/>.

- [11] R. Tolosana-Calasanz, J. Nogueras-Iso, R. Béjar, P. R. Muro-Medrano, F. J. Zarazaga-Soria, "Semantic interoperability based on Dublin Core hierarchical one-to-one mappings", *International Journal of Metadata, Semantics and Ontologies* (en prensa).
- [12] J. Nogueras-Iso, F. J. Zarazaga-Soria, J. Lacasta, R. Béjar, P. R. Muro-Medrano, "Metadata Standard Interoperability: Application in the Geographic Information Domain", *Computers, Environment and Urban Systems* Vol 28, 2004.
- [13] M. A. Latre, F. J. Zarazaga-Soria, J. Nogueras-Iso, R. Béjar, P. R. Muro-Medrano, "SDIGER: A cross-border inter-administration SDI to support WFD information access for Adour-Garonne and Ebro River Basins". *Proc. 11th EC GI & GIS Workshop*, Alghero (Italia), junio 2005.
- [14] D. Nebert, "Developing Spatial Data Infrastructures: The SDI Cookbook v.2.0", *Global Spatial Data Infrastructure*, 2004, <http://www.gsdi.org>.
- [15] R. Tolosana-Calasanz, D. Portolés-Rodríguez, J. Nogueras-Iso, P. R. Muro-Medrano, F. J. Zarazaga-Soria, "CatServer: a server of GATOS". *Proc. 8th AGILE Conference on Geographic Information Science*, Estoril (Portugal), mayo 2005.
- [16] ISO 19112: 2003, "Geographic information – Spatial referencing by geographic identifiers", *International Organization for Standardization (ISO)*.