

Estudios de Relación y Conexión entre los Elementos de Metadatos de las Normas ISO 19115 y DUBLIN CORE, e ISO 19115 y MARC21.

Jordi Escriu Paradell

Institut Cartogràfic de Catalunya, Parc de Montjuïc s/n
08038 Barcelona, España - Telf: 34-93 567 15 00

<http://www.icc.es>

jescriu@icc.es; jordiescriu@ya.com

Abstract. El *Institut Cartogràfic de Catalunya* está realizando un proyecto interno para el desarrollo de un sistema de información documental integrado y multiusuario que permita la gestión de la información referente a los registros del fondo documental de la Cartoteca de Catalunya. El proyecto no sólo contempla el desarrollo de este sistema, sino la catalogación definitiva de los más de 250.000 documentos cartográficos que constituyen su fondo documental, así como su digitalización mediante procesos de escáner.

El objetivo es ofrecer al usuario un servicio acorde a las nuevas tecnologías mediante un catálogo en línea que permita la búsqueda de cualquier documento en función de sus metadatos y también la visualización del mismo. Para lograr la máxima difusión e interoperabilidad el sistema ha de compatibilizarse con las normas de metadatos *ISO 19115*, *DUBLIN CORE* y *MARC21*, haciéndose necesaria la realización de ciertos estudios de relación y conexión entre sus elementos de información.

1 Introducción

El mundo actual está presenciando una verdadera revolución de la información, en la que gobiernos, organizaciones, empresas y ciudadanos demandan nuevas modalidades de acceso a la misma acordes a las posibilidades que vienen ofreciendo los avances tecnológicos y la informática.

En este contexto, se hace necesario adaptar las metodologías de trabajo para ofrecer nuevos servicios a los potenciales usuarios, de forma que el intercambio de la información se produzca de forma transparente, rápida y eficiente. Pero, además de ello, dicha información ha de someterse a un proceso controlado de creación, modificación e intercambio, de manera que su formato y contenido sea fácilmente entendido y pueda utilizarse de forma correcta por cualquier usuario distinto del creador de la misma. Toda información debe ir acompañada de una serie de parámetros o metadatos que identifiquen de forma unívoca sus características y que resulten comunes a todos los usuarios. Esto implica la creación de estándares de creación y manipulación de metadatos, establecidos por consenso de un amplio

espectro de profesionales, de forma que se recojan en un esfuerzo común las características definitorias de la información para su posterior tratamiento y utilización.

Las organizaciones productoras y usuarias de la información han de introducir en sus flujos y metodologías de trabajo una serie de costosos pero fructíferos cambios para adaptarse a la nueva situación de manera que el intercambio de la información se realice de forma fluida y eficaz con el fin de satisfacer las demandas de la sociedad.

El *Institut Cartogràfic de Catalunya (ICC)*, consciente de esta situación, realiza importantes esfuerzos para abordar este tipo de transformaciones en todas sus cadenas de producción. Como ente asociado al mismo, la Cartoteca de Catalunya está realizando actualmente un gran esfuerzo para ofrecer a los usuarios un servicio en línea que pueda ser consultado de forma remota, que permita la búsqueda y la consulta de información relativa a cualquiera de los registros de su fondo documental, y su visualización de forma georreferenciada. Estos servicios pretenden integrarse en una aplicación de tipo web programada en lenguaje *HTML4*, *Javascript* y *PHP4*, conectada a una base de datos *ORACLE 9i*, en la cual se almacenará toda la información. La característica fundamental de esta aplicación es su carácter multiusuario, de forma que a través de ella se podrán realizar tanto operaciones de inserción y modificación de datos (por parte de los catalogadores de la Cartoteca) como consultas (por parte de los usuarios finales). Además de facilitar la consulta de la información a través de Internet se desea hacerla accesible desde el portal de *Infraestructura de Dades Espacials de Catalunya (IDEC)*, con objeto de asegurar la máxima difusión de la misma.

Los documentos que constituyen el fondo de la Cartoteca tienen una vinculación directa con el entorno geográfico, puesto que definen o modelizan con distinto grado de exactitud y precisión la realidad espacial existente en una región limitada del planeta mediante su contenido.

Este es sólo un ejemplo de cómo un sector profesional concreto, como el de la biblioteconomía y la documentación, ha de adaptarse también a las nuevas demandas para actualizar sus servicios de forma que no queden obsoletos. Todos los sectores profesionales deben abordar profundas transformaciones para adaptarse a los nuevos tiempos.

En el caso de la aplicación que se desarrolla para la Cartoteca de Catalunya y especificados ya los objetivos requeridos, confluyen tres normas de estandarización diferentes, cada una de ellas orientada a un determinado sector o ámbito profesional: *ISO 19115*, *DUBLIN CORE* y *MARC21*.

1.1 La Norma de Metadatos Geográficos ISO 19115

El objetivo principal de esta norma es proporcionar un esquema formal y una estructura organizada para la descripción de datos geográficos en formato digital mediante la definición de un conjunto de elementos de metadatos que describan sus características básicas. Ha sido desarrollada por el comité técnico *ISO/TC 211* de *ISO*.

Es utilizada por profesionales vinculados al ámbito de la información geográfica, así como por aquellos analistas, desarrolladores y programadores que pretendan implementar sistemas informáticos compatibles con el estándar.

Su aplicación está vinculada a cualquier tipo de información relacionada con el ámbito geográfico.

La implementación de esta norma de estandarización permite:

- a) Describir correctamente los datos geográficos por parte de sus productores.
- b) Organizar y mantener los metadatos asociados a los datos geográficos.
- c) Usar eficientemente los datos geográficos por parte de los usuarios, ya que éstos conocen sus características básicas.
- d) La búsqueda, recuperación y reutilización de los datos geográficos.
- e) Averiguar si los usuarios pueden utilizar los datos para los fines que persiguen.

Todo ello se consigue mediante la definición de:

- a) secciones, clases y elementos de metadatos de uso condicional u obligatorio.
- b) un conjunto mínimo y obligatorio de metadatos (*CORE*) para describir de forma básica los datos geográficos, de tal manera que permita su búsqueda, localización, la consulta de sus propiedades y el uso que puede hacerse de los mismos.
- c) elementos de metadatos de carácter opcional, para describir de manera más precisa los datos geográficos en vistas a aumentar la interoperabilidad.
- d) un método para la extensión o creación de nuevos metadatos, adecuados a la norma, para satisfacer las necesidades concretas de una organización.

1.2 El Conjunto de Metadatos DUBLIN CORE

Es un simple pero efectivo conjunto de elementos de metadatos pensado para la descripción y catalogación de una amplia gama de recursos accesibles desde Internet, llamados comúnmente recursos electrónicos (ficheros, mapas, libros, catálogos, artículos, publicaciones, páginas Web, programas, aplicaciones, servicios, fuentes de datos, etc.), permitiendo su búsqueda en la red así como la consulta de sus propiedades más básicas.

El conjunto está formado por 15 elementos definidos por la *Dublin Core Metadata Initiative (DCMI)* en Dublín, mediante el consenso de un grupo internacional de profesionales de disciplinas diversas, como la biblioteconomía o la informática, abiertos a su vez a las colaboraciones de terceros. El documento *CEN Workshop Agreement CWA 13874*, constituye una muestra fehaciente de este hecho. Actualmente el conjunto de metadatos *DUBLIN CORE* queda establecido por la norma *NISO Standard Z39.85*, que corresponde a la versión 1.1 de este conjunto de metadatos.

Todos los elementos de *DUBLIN CORE* son opcionales y pueden repetirse (poseen cardinalidad *N*). Además, tienen un conjunto limitado de calificadores, es decir, de atributos, que pueden ser usados opcionalmente para afinar el significado concreto de cada elemento.

DUBLIN CORE permite:

- a) Crear y mantener de forma sencilla un reducido conjunto de metadatos para posibilitar una búsqueda efectiva de recursos en Internet.
- b) Poner en común el significado de sus elementos entre todos sus usuarios a nivel internacional.
- c) Extender el conjunto de metadatos que lo constituye, para la correcta definición de recursos de nueva aparición y solventar necesidades relativas a su descripción.

1.3 El Formato de Catalogación MARC21

MARC21 es un formato estructurado, desarrollado por la *Library of Congress* de Estados Unidos, que sirve para almacenar e intercambiar de forma automatizada datos bibliográficos relativos a un amplio abanico de recursos: documentos escritos, música, documentos cartográficos, documentos proyectables, microfilmes, registros sonoros, documentos gráficos bidimensionales no proyectables, ficheros de datos en lenguaje máquina, recursos multimedia, objetos tridimensionales, etc.

Un archivo en formato *MARC21* se compone de una sucesión de códigos *ASCII*, aparentemente sin orden, delimitados por una serie de posiciones y etiquetas estandarizadas (indicadores, campos y subcampos) que sirven para detectar elementos de metadatos específicos.

Los registros de *MARC21* se componen de tres elementos:

- a) Estructura del registro: Es una implementación de la norma internacional *ISO 2709 – 1973*, que rige la estructura básica de los formatos *MARC*.
- b) Designación del contenido: Son los códigos, etiquetas y convenciones definidas de forma explícita para identificar los elementos de información de un registro, los cuales permiten su manipulación.
- c) Contenido del registro: Son los contenidos de información de cada uno de los metadatos usados en el formato. La mayoría de estos contenidos están controlados por normas externas al formato, como las normas *ISBD*, las *AACR2*, etc. u otras normas o códigos usados por el centro o por el sistema que crea el registro. Estos contenidos incluyen títulos, nombres, materias, notas, información de publicación, descripción física, etc.

El formato *MARC21* pretende ser un vehículo estándar de intercambio de información bibliográfica a nivel mundial, y sus contenidos son un compendio de los elementos de información comunes existentes en distintas variantes de formatos *MARC*, que suelen corresponderse con distintos países o zonas geográficas (*USMARC*, *UKMARC*, etc.).

2 Necesidad de los Estudios de Relación y Conexión

Para cada proyecto específico, según las disciplinas técnicas implicadas en su desarrollo, será necesario compatibilizar el proceso con una serie de normas de estandarización concretas, que afecten a ciertos aspectos del mismo.

En el caso del proyecto para el desarrollo de una aplicación o sistema de información documental en la Cartoteca de Catalunya, resulta necesario implementar los estándares *ISO 19115*, *DUBLIN CORE* y *MARC21*, por las razones especificadas a continuación.

La norma *ISO 19115* es esencial a la hora de describir de forma adecuada la información geográfica relativa a documentos cartográficos (que constituyen la mayoría de los fondos de la Cartoteca), y por otro lado, si se quieren dar accesibilidad a los registros de la Cartoteca desde el portal de *IDEC* resulta obligatorio compatibilizar el sistema con esta norma de metadatos. La implementación de *DUBLIN CORE* es esencial a la hora de posibilitar una eficiente y correcta búsqueda de la información en Internet, siendo esta norma ampliamente utilizada para conseguir estos propósitos. El formato *MARC* constituye el formato nativo de la información usualmente utilizada en el ámbito de la biblioteconomía y la documentación. Su uso es esencial en cualquier biblioteca (o cartoteca) para intercambiar información con otras entidades similares. Por ejemplo, este formato es la base en la que se fundamentan los portales de búsqueda de recursos bibliográficos, que acceden a los registros de distintos organismos o centros bibliotecarios.

A la hora de abordar el estudio de los estándares de metadatos para aplicar en el proyecto de desarrollo del sistema de información es necesario tener en cuenta dos aspectos fundamentales:

- a) La documentación e inclusión del mayor número de metadatos posible de un estándar incrementa la interoperabilidad del sistema.
- b) La información almacenada en la base de datos del sistema final ha de evitar redundancias de información con objeto de ocupar el mínimo espacio posible y mantener la coherencia interna del mismo.

Identificados los estándares a implementar y teniendo en cuenta las dos premisas anteriores, el objetivo es estudiar el conjunto de elementos de información que es necesario introducir en el sistema. Este conjunto es la suma de los metadatos que resultan obligatorios para lograr la compatibilización con los estándares y la de aquellos que resultan necesarios para los objetivos finales de la aplicación. En relación a los primeros, considerando la segunda premisa anteriormente citada (evitar las redundancias), es vital establecer qué elementos de metadatos son comunes a las distintas normas de estandarización, de tal manera que no se almacene más de una vez un mismo elemento de información en la base de datos. Por este motivo hay que analizar la relación y conexión entre los metadatos de los distintos estándares a implementar, para establecer los elementos que resultan equivalentes y contienen la misma información. Teniendo en cuenta la primera premisa, es conveniente seguir la política de documentar el máximo número de metadatos de cada norma que sea posible, con objeto de favorecer la interoperabilidad.

Estos razonamientos justifican la necesidad de realizar estudios de relación y conexión entre los elementos de las normas, siguiendo en lo posible criterios de economía de recursos y de tiempo. Para ello conviene analizar en cada proyecto la situación existente, que para el caso de la Cartoteca, puede resumirse en el siguiente esquema.

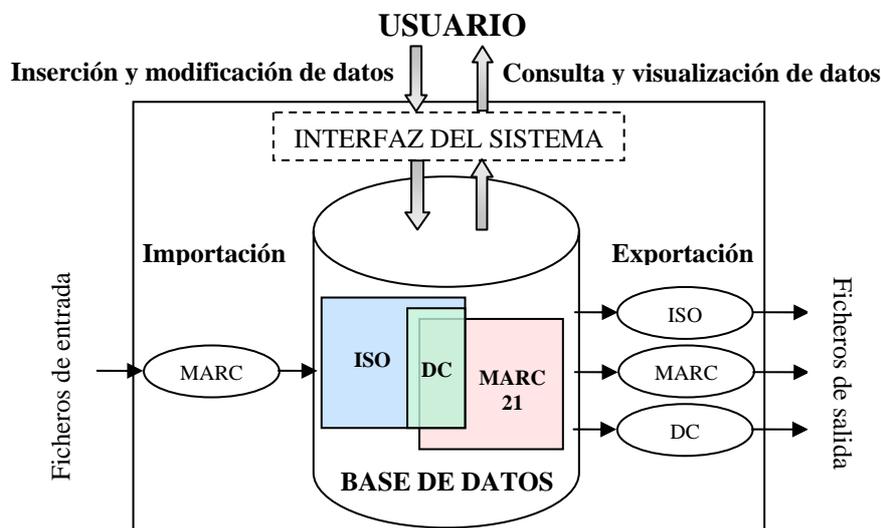


Fig. 1. Esquema del sistema de información documental de la Cartoteca de Catalunya

Como puede apreciarse en la figura 1, los diferentes tipos de usuarios pueden acceder al sistema para realizar operaciones de inserción, modificación, consulta y visualización de datos, según los privilegios que tengan asignados en su perfil. Por otro lado, el sistema prevé la existencia de herramientas de importación y exportación de los datos.

Los elementos de la base de datos pueden estar relacionados, como se observa en la figura, con el estándar *ISO 19115*, el formato *MARC21* o *DUBLIN CORE*.

Analizando con más detenimiento los metadatos de cada estándar, se da la situación de que los elementos comunes entre *ISO 19115* y *MARC21* coinciden casi con la totalidad de los elementos de *DUBLIN CORE*. Esto es debido a que este último conjunto de metadatos, aunque sencillo y escueto, define de forma adecuada una gran variedad de recursos electrónicos, pudiéndose relacionar sus elementos de forma sencilla con los otros dos estándares. Este concepto queda representado en la figura 1, donde vemos que dentro de la base de datos, la zona de intersección entre *ISO* y *MARC21*, es ocupada en su totalidad por la mayoría del conjunto *DUBLIN CORE*. Tan sólo dos elementos de este sencillo conjunto (*DC_Relation* y *DC_Rights*) no pueden relacionarse con ningún elemento existente de *MARC21*, pero sí con *ISO*; por este motivo, vemos que la totalidad de *DUBLIN CORE* se encuentra incluida en esta norma.

La situación descrita anteriormente simplifica los estudios de relación y conexión a realizar, ya que permite encontrar los elementos de *ISO 19115* y *MARC21* que resultan equivalentes entre sí, a través de las relaciones de estas normas con *DUBLIN*

CORE; es decir, encontradas las relaciones entre *ISO 19115* y *DUBLIN CORE*, y entre *MARC21* y *DUBLIN CORE*, obtenemos de forma implícita la relación entre *ISO 19115* y *MARC21*. Este es el motivo por el cual este artículo presenta los resultados obtenidos en los dos estudios citados.

3 Resultados de los Estudios de Relación y Conexión

A continuación se presentan dos tablas (Tabla 1 y Tabla 2) en las cuales se muestran los elementos equivalentes o relacionados entre las normas *ISO 19115* y *DUBLIN CORE*, y entre *MARC21* y *DUBLIN CORE*, respectivamente.

La relación entre *ISO 19115* y *MARC21*, tal y como se ha argumentado en el apartado 2 de este artículo, queda definida de forma implícita con los resultados de los dos estudios anteriores, por mediación de los elementos de *DUBLIN CORE*.

En las tablas siguientes se ha obviado la descripción de los elementos de metadatos implicados y otra serie de propiedades que pueden resultar de interés para el lector. Se ruega a éste se ponga en contacto con el autor para recibir más información.

Tabla 1. Relación y conexión entre *ISO 19115* y *DUBLIN CORE*

ISO 19115	DUBLIN CORE
MD_Metadata > MD_Identification.citation > CI_Citation.title	DC.TITLE
MD_Metadata > MD_Identification.credit	DC.CREATOR
MD_Metadata > MD_Identification.pointOfContact > CI_ResponsibleParty.individualName	
MD_Metadata > MD_Identification.pointOfContact > CI_ResponsibleParty.organisationName	
MD_Metadata > MD_DataIdentification.topicCategory > MD_TopicCategoryCode	DC.SUBJECT
MD_Metadata > MD_Identification > MD_Keywords.keyword	
MD_Metadata > MD_Identification.abstract	DC.DESCRPTION
MD_Metadata > MD_Identification.pointOfContact > CI_ResponsibleParty.individualName	DC.PUBLISHER
MD_Metadata > MD_Identification.pointOfContact > CI_ResponsibleParty.organisationName	
MD_Metadata > MD_Identification.credit	DC.CONTRIBUTOR
MD_Metadata > MD_Identification.pointOfContact > CI_ResponsibleParty.individualName	
MD_Metadata > MD_Identification.pointOfContact > CI_ResponsibleParty.organisationName	
MD_Metadata > MD_Identification.citation > CI_Citation.date > CI_Date > Date	DC.DATE
MD_Metadata.hierarchyLevel > MD_ScopeCode	DC.TYPE
MD_Metadata > MD_Distribution > MD_Format.name	DC.FORMAT
MD_Metadata > MD_Distribution > MD_Format.version	
MD_Metadata > MD_Identification.citation > CI_Citation > MD_Identifier.code	DC.IDENTIFIER
MD_Metadata > MD_Identification.citation > CI_Citation.isbn	
MD_Metadata > MD_Identification.citation > CI_Citation.issn	
MD_Metadata > DQ_DataQuality > LI_Lineage.statement	DC.SOURCE
MD_Metadata > DQ_DataQuality > LI_Lineage > LI_Source.description	
MD_Metadata > MD_DataIdentification.language	DC.LANGUAGE

MD_Metadata > MD_DataIdentification > MD_AggregateInformation.associationType > DS_AssociationTypeCode	DC.RELATION	
MD_Metadata > MD_ServiceIdentification > MD_AggregateInformation.associationType > DS_AssociationTypeCode		
MD_Metadata > MD_DataIdentification.extent > Ex_Extent.description	DC.COVERAGE	
MD_Metadata > MD_DataIdentification.extent > Ex_Extent > Ex_GeographicBoundingBox.westBoundLongitude > Angle		
MD_Metadata > MD_DataIdentification.extent > Ex_Extent > Ex_GeographicBoundingBox.eastBoundLongitude > Angle		
MD_Metadata > MD_DataIdentification.extent > Ex_Extent > Ex_GeographicBoundingBox.southBoundLatitude > Angle		
MD_Metadata > MD_DataIdentification.extent > Ex_Extent > Ex_GeographicBoundingBox.northBoundLatitude > Angle		
MD_Metadata > MD_DataIdentification.extent > Ex_Extent > Ex_GeographicDescription.geographicIdentifier > MD_Identifier.code		
MD_Metadata > MD_DataIdentification.extent > Ex_Extent > Ex_VerticalExtent.minimumValue		
MD_Metadata > MD_DataIdentification.extent > Ex_Extent > Ex_VerticalExtent.maximumValue		
MD_Metadata > MD_Identification > MD_LegalConstraints.accessConstraints > MD_RestrictionCode		DC.RIGHTS
MD_Metadata > MD_Identification > MD_LegalConstraints.useConstraints > MD_RestrictionCode		
MD_Metadata > MD_Identification > MD_LegalConstraints.otherConstraints		

Tabla 2. Relación y conexión entre *MARC21* y *DUBLIN CORE*

MARC21	DUBLIN CORE v.1.1
001	DC.IDENTIFIER
222 \$a	DC.TITLE
245 \$a	
130 \$a	
240 \$a	
740 \$a	
830 \$a	
100 \$a	
110 \$a	DC.DESCRPTION
111 \$a	
255 \$a	
507 \$a	
342 \$a	
255 \$b	DC.PUBLISHER
255 \$c	
034 \$d	
034 \$e	
034 \$g	
034 \$f	
260 \$a, 260 \$b	DC.PUBLISHER
008 (pos. 7 - 14)	DC.DATE
033 \$a	DC.TITLE
362 \$a	
260 \$c	
440 \$a	
490 \$a	
800 \$a	

810 \$a	DC.TITLE
811 \$a	
500 \$a	DC.DESCRPTION
501 \$a	
545 \$a	
504 \$a	
505 \$a	
508 \$a	
520 \$a	
522 \$a	
525 \$a	DC.DESCRPTION
516 \$a	
546 \$a	DC.LANGUAGE
518 \$a	DC.DESCRPTION
020 \$a	DC.IDENTIFIER
022 \$a	
017	
086 \$a	
856 \$u	
080 \$a	
008 (pos. 35 - 37)	DC.LANGUAGE
041 \$a	
546 \$a	
008 (pos. 24 - 27)	DC.FORMAT
300 \$f	
856 \$q	
700 \$a	DC.CONTRIBUTOR
710 \$a	
711 \$a	
600 \$a	DC.SUBJECT
610 \$a	
611 \$a	
630 \$a	
043 \$a	DC.COVERAGE
650 \$z	
651 \$a	
752 \$a, \$b, \$c, 6 \$d (según entidad geográfica)	
072 \$a	DC.SUBJECT
650 \$a	
655 \$a	DC.TYPE

4 Conclusiones

Las principales conclusiones de este documento son:

- a) Todos los ámbitos y disciplinas científicas han de adaptar su organización y sus esquemas de trabajo para actuar en consonancia con los estándares de información, aplicándolos de forma eficiente. Para ello es indispensable realizar fuertes esfuerzos humanos, técnicos y económicos en la modernización, la reingeniería de procesos y la creación de aplicaciones automáticas que gestionen la información.

b) Es necesario establecer vías de conexión entre normas de estandarización relacionadas si se quiere llegar a producir sistemas inteligentes que permitan intercambiar información entre sus distintos formatos convenientemente. Son muchos los estándares que pueden incidir en una determinada disciplina científica o técnica, siendo necesario compatibilizarlos e intercambiar información entre ellos.

c) La creación de estándares de información debe corresponder a grupos de trabajo integrados por profesionales de muy diversa índole, de forma que se recojan todas las interpretaciones posibles para establecer un lenguaje común válido y efectivo. Igualmente, los estudios para establecer la relación y conexión entre elementos de metadatos de distintos estándares han de realizarse por grupos multidisciplinares que recojan por consenso un punto de vista común considerado como válido para un caso general.

d) Los resultados de los estudios de relación y conexión presentados en este documento constituyen una primera aproximación a un problema arduo y complejo. Es necesario ahondar en él con la colaboración de todos aquellos que estén interesados en resolverlo y poner en práctica las soluciones halladas para desarrollar herramientas efectivas de importación y exportación con cambio de formato. Aspectos importantes a estudiar son las posibles discrepancias de cardinalidad u obligatoriedad de inclusión entre elementos de información relacionados pertenecientes a los distintos estándares implicados.

Referencias

1. Infraestructura de Dades Espacials de Catalunya (IDEC): *Informació Geogràfica – Metadades. Esquema IDEC*. – <http://www.idec.es>
2. Infraestructura de Dades Espacials de Catalunya (IDEC): *Interoperabilitat*. – <http://www.idec.es>
3. Manso Callejo, Miguel Ángel: *Metadatos en los Sistemas de Información Geográfica*. – <http://gauss.euitto.upm.es/m.manso>
4. Zaragoza-Soria, F.J. y otros: *Integrating Geographical and non Geographical Data Search Services Using Metadata Crosswalks – Serving the User. 9th EC-GI&GIS Workshop*
5. Technical Committee ISO/TC 211: *Geographic Information – Metadata – ISO 19115:2003 Final Draft*. – <http://www.iso.org>
6. NISO: *The Dublin Core Metadata Element Set v.1.1 - ANSI/NISO Z.39.85-2001* – <http://www.niso.org>; <http://www.dublincore.org>
7. The Library of Congress – Bibliographic Standards: *MARC21 Concise Format* – <http://www.loc.gov/marc/bibliographic/ecbdhome.htm>
8. Estrader, Marcel. Generalitat de Catalunya (Departament de Cultura): *Manual de CATMARC – 2^a Edició. Sistema per a Registres Bibliogràfics*. – <http://www.gencat.net/bc>