

Joan Capdevila*, Jose M. Agudo**, F. Javier Zarazaga-Soria**, Jesús Barrera***, Alejandra Sánchez****,
 Carolina Soteres****, Marta Criado****, María Crespo****

Gateway MARC21-ISO19115: definition and reference implementation

Keywords: MARC21; Gateway; ISO19115; metadata; implementation

Summary

In 2008 an interdisciplinary working group to promote the publication of cartographic heritage through the Spatial Data Infrastructure started working. One of the first tasks undertaken was the definition of a gateway computer between MARC21 cataloging format and the metadata standard ISO 19115. Here are the results: the definition of the gateway and a computer implementation using the free software CatMDEdit.

Introduction

The Directorate General of the National Geographic Institute (IGN) is undertaking a major effort to promote the use of Spatial Data Infrastructures (SDI) as a technology for sharing geographic information in Spain. Furthermore, since the year 2002, National Geographic High Council, through the Special Commission of Geographic Information Infrastructure of Spain, established the Working Group Spatial Data Infrastructure of Spain (GT IDEE), where the necessary arrangements for the proper functioning of SDI in Spain are discussed and tasks of dissemination and promotion of the SDI paradigm are carried out (Vandenbroucke and Biliouris 2011).

Historical Cartography is considered of great interest to historians, documentarians, researchers and citizens interested in the use of mapping as a source of information of a historical nature. To facilitate remote access through Internet, the Spatial Data Infrastructure (SDI) is a good alternative as it allows the use of the information in an interoperable way.

In order to promote and facilitate to the libraries the processes necessary to reach the publication of data and documents of a geographical and historical maps on the Internet, using the SDI strategy, meetings were held in 2007 with various people in charge of libraries, map libraries and archives to study how they could use the SDI to publish their documents. As a result a Working Group was established to define and implement projects in order to publish historical maps using SDI strategy: the Interdisciplinary Working Group on Cartographic Heritage in the SDI (GTI PC-IDE), part of GT IDEE.

The first problem considered was to use the MARC21¹ records existing today in Spain so that they could be used as part of a SDI. This required that these records be transformed to ISO²

* Servicio Regional en Cataluña. Instituto Geográfico Nacional, Barcelona, Spain [joan.capdevila@seap.minhap.es].

** Universidad de Zaragoza, Zaragoza, Spain [joselilo@unizar.es] [javy@unizar.es]

*** GeoSpatiumLab, Zaragoza, Spain [jesusb@geoslab.com]

**** Instituto Geográfico Nacional. Madrid, Spain [asmaganto@fomento.es] [csoteres@fomento.es]

***** DMSGGroup, Madrid, Spain [mcriado@dmsgroup.es] [mcrespo@dmsgroup.es]

¹ Profile of format MARC (*Machine Readable Cataloging*). It was decided to work with this since it was enhanced to better describe cartographic documents and objects and has been used by many members of the group.

metadata standard. So, the first job of the Group was to achieve compatibility between MARC21 cataloging formats and the UNE EN-ISO 19115:2006, so that the metadata records and cartographic funds for libraries could be channeled into the scope of development of the SDI.

To solve this problem, the IGN –in collaboration with the DMS Group company, the University of Zaragoza and with the advice of members of GTI PC-IDE– have developed a gateway that allows catalogers to work in both formats interchangeably. There was already some background. Chandler et al. (2000) had developed a Gateway between MARC21 and Dublin Core³. However, this case did not cover the specific georeferencing historical cartography. Leveraging the meeting Digital Technologies in Cartographic Conference Heritage held in Barcelona in June 2008, a round table was organized which supported the creation of the gateway (Montaner 2009).

This paper shows the process followed by the GTI PC-IDE to launch Gateway MARC21-ISO19115, culminating with the possibility to Import/Export files in the CatMDEdit⁴ application, a metadata tool based on free software. The first section describes the steps undertaken to create the Gateway between standards. The following describes how the computer implementation was performed. Finally, in the conclusions, we describe the future work.

Harmonization and semantic mapping

The first step taken to establish the relationship between the standard was to describe them to perform a comparative study of cataloging formats. This was necessary to study and evaluate the elements belonging to MARC21 and ISO19115, Metadata Spanish Core (NEM v.1.0) and the INSPIRE Metadata Regulation. It started from two initial tables, as described by J. Nogueras-Iso et al. (2004), phase of harmonization. The first table of MARC21 had a total of 40 items (including the header, see figure 1) and the initial table of NEM had all the elements of metadata included in this core. As a result of this first phase of harmonization a cross table of items was obtained.

# (1)	Etiqueta	Ind. 1	Ind. 2	Nombre elemento	Restric (3)	Multiplicidad (4)	Definición semántica	Tipo de datos (5)	Valores por defecto (6)
	Cabecera								
	pos. 00-04			Longitud del registro	M	NR	Cadena de 5 caracteres numéricos generada automáticamente que especifica la longitud del registro.	Numérico	
	pos. 05			Estado del registro catalogado	M	NR	Indica la relación del registro con un fichero.	a - Incremento del nivel de codificación / c - corregido o revisado / d - borrado / n - nuevo / p - incremento del nivel de codificación de prepublicación.	
	pos. 06			Tipo de registro	M	NR	Indica las características del registro y define sus componentes.	e - Material cartográfico (mapas, atlas, globos, mapas digitalizados y otros documentos cartográficos) / f - material cartográfico manuscrito	e - Material cartográfico (mapas, atlas, globos, mapas digitalizados y otros documentos cartográficos)
	pos. 07			Nivel bibliográfico del registro	M	NR		a - Parte componente monográfico / b - parte componente seriada / c - Colección / d - subunidad / i - Recurso integrado / m - Monografía/documento / s - publicación en serie	Manual

Figure 1. Table of MARC21.

The next step was considered the most important in the development of the gateway and was the establishment of relations between the two cross-tables. Thus, establishing a semantic corre-

² Specifically, it refers to the ISO 19115 standard, which the Spatial Data Infrastructure of Spain has already defined a profile for use in Spain: Spanish Core Metadata (NEM). Subsequently, we have taken into account the corrections introduced implementing standard metadata that was published as development of the INSPIRE directive.

³ It is a much extended profile of ISO 19115.

⁴ <http://catmdedit.sourceforge.net/>

spondence between elements of the standard source, in this case MARC21, and their correspondence with the elements considered semantically equivalent to the standard target, in this case ISO 19115. The first difficulties appeared in this step because it was necessary to know exactly what was the meaning and use that was given to each field in each table. As noted, not all organizations used the MARC21 standard in the same manner.

To carry out this work, a methodological process was defined: the first sought the minimum correlation, as defined by those metadata elements belonging to the NEM and INSPIRE, which allowed achieving interoperability sought. This study allowed us to detect cases in which there was no correspondence with the elements of mandatory and solutions put forward in the definition phase of conversion rules. In a second step, the minimum mapping was extended with fields of ISO 19115 and ISO19115-2 elements and the items were revised to comply with the NEM Profile v1.1. Thus, an initial breaking of elements was obtained and an evaluation questionnaire was created which was made available to the members of the Group for approval. The responses obtained by different members of the Group were discussed and variations in the initial cross settled, allowing the definition of a final cross in order to increase the lists managed by both standards, to introduce new keywords and raise the use of the thesaurus; in this way both a controlled list of keywords is introduced and the current standards are not modified. In conclusion, we obtained a total of 64 crossover elements from MARC21 versus 60 elements from ISO 19115 (Crespo et al. 2010).

Once the tables were aligned for each format and performed the crossing of elements, it was necessary to define a set of conversion rules that allow to put into practice the Gateway. These rules are specific guidelines necessary to establish the correspondence between the elements of MARC21 and the elements of ISO 19115. Each rule documents and resolves each of the many cases that can occur: the crossing of an item in a format with several of the other format, the existence of unnecessary duplications, the use of controlled word lists or thesaurus; complete automation of the fields "not available" , appropriate treatment dates, treatment and the crossing of abbreviations and so on.

The results obtained in these "theoretical" first steps are available on the portal on Geographic Information Metadata, under *Cartographic Heritage*⁵.

Automated implementation

The fourth step was to provide a computer implementation in order to test the gateway definition and to put it into operation. The integration of any computer proposal into a reference software tool is considered a good practice. In this case, an open source software has been selected with the aim of making the gateway available to everyone interested in its use. CatMDEdit (Zarazaga-Soria et al. 2003) was the selected application for this purpose. CatMDEdit is a metadata editor tool that facilitates the management of resources through their metadata, paying special attention to geographic information resources. It is an initiative of the IGN, which is the result of the scientific and technical collaboration with the Advanced Information Systems Group (IAAA) of the University of Zaragoza with the technical support of GeoSpatiumLab (GSL). The tool, the main aim of which is to promote metadata creation as a mechanism that facilitates the processing of data in a more effective way, is developed in Java and internationalized to six different languages. Some examples of information which can be managed by CatMDEdit are: digital or hardcopy topographic maps, web services, geographic information layers, spatial databases, orthoimages,

⁵ http://metadatos.ign.es/metadatos/Patrimonio_cartografico

satellite products and digital terrain models. This tool is specialized in the creation of geographic metadata according to the ISO 19115:2003 "Geographic Information--Metadata" (ISO 2003) and the ISO 19119:2005 "Geographic information--Services" (ISO 2005) standards, but other metadata standards, such as CSDGM (FGDC 1998; FGDC 2000), Dublin Core (ISO 2003b) or MIGRA (AENOR 1998), are also supported thanks to the ability of CatMDEdit to perform transformations between metadata standards. In the same way, the option to export metadata records in different file formats, such as XML, HTML or Excel, is also offered. These options to import and export metadata records has been extended to integrate the gateway for converting between MARC21 and ISO 19115. Thus, the possibility of recognizing MARC21 records was also added to CatMDEdit. The following formats are now supported: MARC XML⁶, MARC 21 *communications format - ISO 2709*⁷ and MARC 21 tagged display (Fig. 2).

<p>MARC XML</p>	<pre><?xml version="1.0" encoding="UTF-8"?> <collection xmlns="http://www.loc.gov/MARC21/slim"> <record> <leader>01142cam 2200301 a 4500</leader> <controlfield tag="001"> 92005291 </controlfield> <controlfield tag="003">DLC</controlfield> <controlfield tag="005">19930521155141.9</controlfield> <controlfield tag="008">920219s1993 caua j 000 0 eng </controlfield> <datafield tag="010" ind1=" " ind2=" "> <subfield code="a"> 92005291 </subfield> </datafield> <datafield tag="020" ind1=" " ind2=" "> <subfield code="a">0152038655 </subfield> <subfield code="c">\$15.95</subfield> </datafield></pre>
<p>MARC 21 <i>communications format - ISO 2709</i></p>	<pre>01142cam 2200301 a 450000100130000000300040001300500170001700800410003401000170007502000250009204000 180011704200090013505000260014408200160017010000320018624500860021825000120030426 000520031630000490036850000400041752002280045765000330068565000330071865000240075 16500021007756500023007967000021008191 92005291 IDLC19930521155141.91920219s1993 caua j 000 0 eng a 92005291 a0152038655 :lc\$15.95 aDLC cDLC dDLC alcac 00 aPS3537.A618 bA88 1993 00 a811 .52 220 1 aSandburg, Carl, d1878-1967. 10 aArithmetic lcCarl Sandburg ; illustrated as an anamorphic adventure by Ted Rand. a1st ed. aSan Diego : bHarcourt Brace Jovanovich, cc1993. a1 v. [unpaged] : bill. [some col.] : c26 cm. aOne Mylar sheet included in pocket. aA poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone. 0 aArithmetic xJuvenile poetry. 0 aChildren's poetry, American. 1 aArithmetic xPoetry. 1 aAmerican poetry. 1 aVisual perception. 1 aRand, Ted, eill. </pre>

⁶ <http://www.loc.gov/standards/marcxml/>
⁷ <http://www.loc.gov/marc/umb/um11to12.html>

MARC 21 tagged display	<p>01142cam 2200301 a 4500</p> <p>000</p> <p>001 92005291</p> <p>003 DLC</p> <p>005 19930521155141.9</p> <p>008 920219s1993 caua j 000 0 eng</p> <p>010 \$a 92005291</p> <p>020 \$a0152038655 . \$c\$15.95</p> <p>040 \$aDLC\$cDLC\$dDLC</p> <p>042 \$alcac</p> <p>050 00\$aPS3537.A618\$bA88 1993</p> <p>082 00\$a811/.52\$220</p> <p>100 1 \$aSandburg, Carl,\$d1878-1967.</p> <p>245 10\$aArithmetic /\$cCarl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.</p> <p>250 \$a1st ed.</p> <p>260 \$aSan Diego :\$bHarcourt Brace Jovanovich,\$c1993.</p> <p>300 \$a1 v (unpaged) :\$bill. (some col.) ;\$c26 cm.</p> <p>500 \$aOne Mylar sheet included in pocket.</p>
------------------------	---

Figure 2. Examples of MARC 21 formats extracted from <http://www.loc.gov/standards/marcxml/>

In order to add a MARC21 record, the import utility of CatMDEdit must be launched. At this point, the user can select a local file which can be encoded according to one of the three MARC 21 formats mentioned before. Once the selection is done, CatMDEdit applies the gateway rules and a new ISO 19115 record is added to the internal browser. The user will then be able to manage this record in the same conditions as any other ISO 19115 record (Fig. 3).

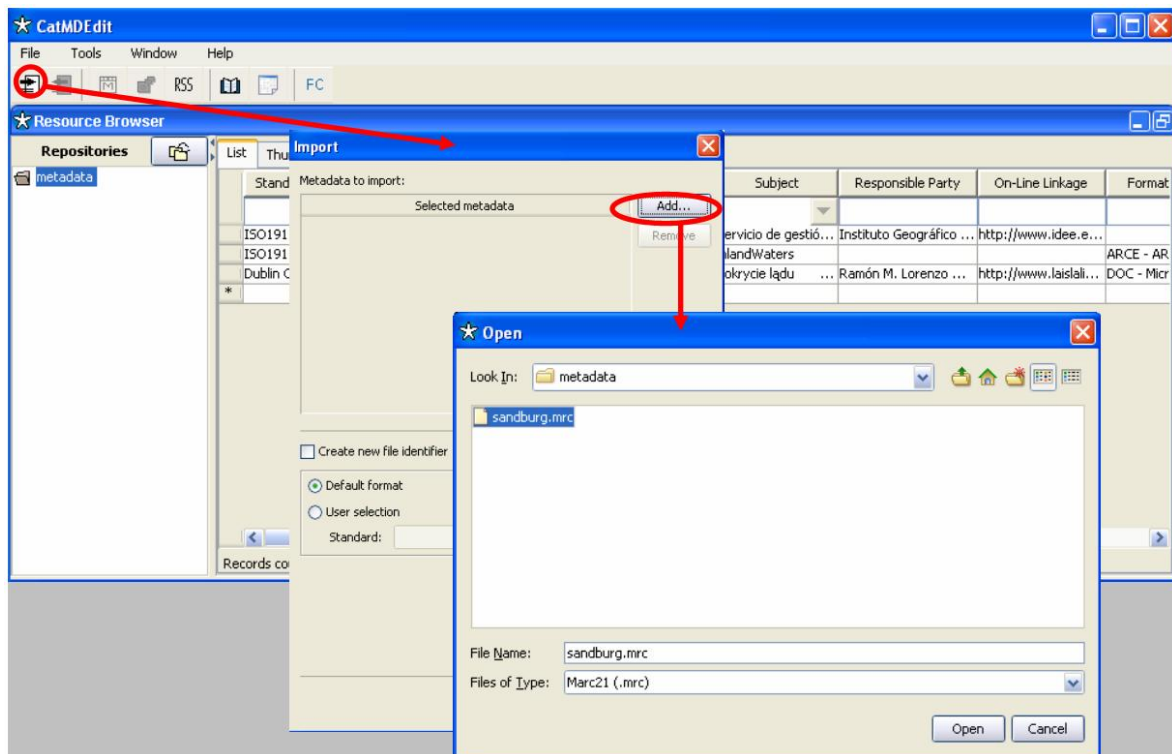


Figure 3. Importing a MARC 21 record in CatMDEdit.

Figure 4 shows the implementation process of transforming from a MARC 21 record to an ISO 19115 record. Once the user has selected the file to be converted, the first step is to transform its contents into a common model. This will make it easier to process the information. For this purpose the model provided by the MARC XML format has been selected. The main reasons for this choice are its ease of use and the existence of an xml schema that normalizes its structure. Transformations to this format are a relative easy task thanks to the existence of an open source library called *marc4j*⁸.

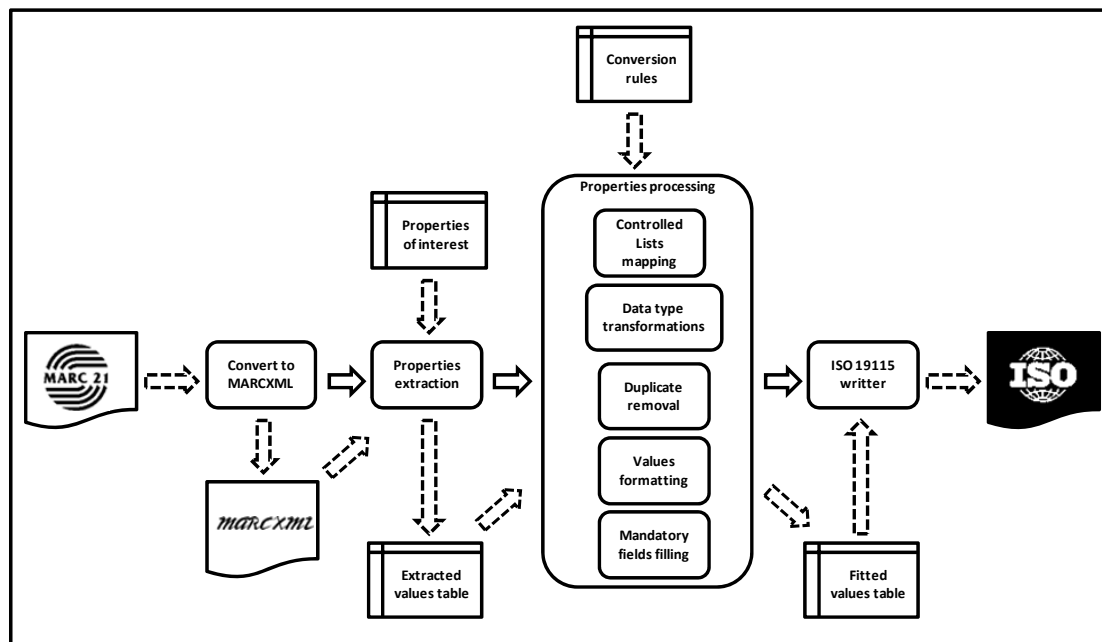


Figure 4. MARC21 to ISO 19115 transformation process.

Having the MARC 21 record in XML format, the next step to take is the extraction of the fields which have been considered of interest in the gateway definition. In this way, a table with the mappings between the ISO 19115 destination properties and the MARC 21 extracted values is obtained. However, these extracted values cannot be directly written in an ISO 19115 record because their representation can be different between a standard and the other (for example, different ways to express a same datatype or special codes for a controlled value). In addition to that, there can be several MARC 21 properties which are mapped to the same ISO 19115 element, so an additional process is needed in order to delete duplicated values or to structure the information in a readable way. Finally, the generated ISO 19115 record must be valid in conformance with the XML Schemas defined by the 19139:2007 (ISO 2007) standard and with INSPIRE Metadata Implementing Rules (European Commission, 2008). All these processes are executed and a table containing the fitted values is generated as a result. The last step is writing all the fitted values in a XML record following the ISO 19115 model and storing it in the local metadata repository managed by CatMDEdit.

In this CatMDEdit distribution, not only the MARC 21 to ISO 19115 transformation has been developed, but also the conversion in the opposite direction is available (Fig. 5). In this case, the export utility must be launched in order to transform from an ISO 19115 record to a MARC 21 file. The implemented process is not detailed in this paper because it is very similar to the one fol-

⁸ <http://marc4j.tigris.org/>

lowed to transforms from MARC 21 to ISO 19115 with the only difference that mapping and transformations are applied in reverse order.

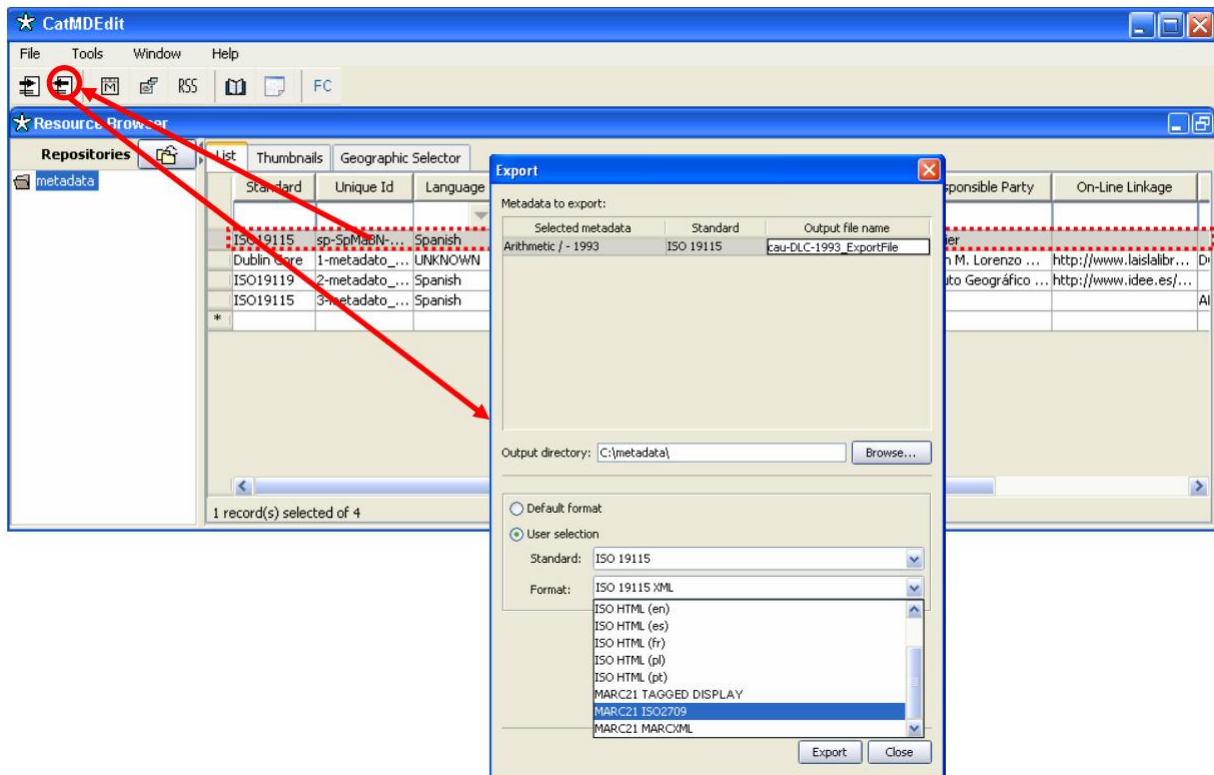


Figure 5: Exporting an ISO 19115 record to a MARC 21 file in CatMDEdit.

Conclusions

The computer implementation of the gateway MARC21-19115 in the tool CatMEdit concluded much of the project designed. A thorough testing of the robustness of the gateway and its implementation still has to be conducted. The test will consist of the implementation of the gateway within an institution that uses a catalog in MARC21. We expect to see in practice what the potential loss of information is due to the process of transformation and what are the problems that end users, documentary and catalogers find. The experience will also serve as an example for other institutions to understand both the advantages and difficulties in using the gateway. We hope to use these lessons to improve both the gateway and its computer implementation.

Finally, it is expected that the gateway will end up being adopted by the National Geographic High Council as a recommendation and serve as a starting point to incorporate cartographic heritage into the development of SDI. As seen, this is an essential element for the publication of historical maps on Internet with the network using interoperable services. But it is not enough, so for the GTI PC-IDE is still a long way from bringing to the world the valuable map treasures kept in our archives.

References

- AENOR (1998). *Mecanismo de Intercambio de Información Geográfica Relacional formado por Agregación (MIGRA): UNE 148001 EXP 1998, versión 1*. Asociación Española de Normalización y Certificación (AENOR). Comité Técnico de Normalización 148 de AENOR(AEN/CTN 148).
- Chandler, A.; Foley, D. and A.M. Hafez (2000). Mapping and Converting Essential Federal Geographic Data Committee (FGDC) Metadata into MARC21 and Dublin Core. *D-Lib Magazine* 6 (1)
- Crespo, M.; Criado, M.; Capdevila, J.; Sánchez, A.; Soteres, C.; Juanatey, M.; Bravo, M. J. and C. Rodríguez (2010). El patrimonio cartográfico en las infraestructuras de datos espaciales. *Revista Catalana de Geografia* 41 [<http://www.rcg.cat/articles.php?id=199>]
- FGDC (1998). Content Standard for Digital Geospatial Metadata, version 2.0: Document FGDC-STD-001-1998. Federal Geographic Data Committee, Metadata Ad Hoc Working Group.
- FGDC (2000). Content Standard for Digital Geospatial Metadata Workbook, version 2.0. Federal Geographic Data Committee (FGDC), Metadata Ad Hoc Working Group.
- ISO (2003). *Geographic information - Metadata. ISO 19115:2003*. International Organization for Standardization (ISO).
- ISO (2003b). *Information and documentation - The Dublin Core metadata element set. ISO 15836:2003*. International Organization for Standardization (ISO).
- ISO (2005). *Geographic information - Services. ISO 19119:2005*. International Organization for Standardization (ISO).
- Montaner, C. (2009). Disseminating Digital Cartographic Heritage: Standards and infrastructures. *e-Perimtron* 4 (1): 53-54.
- Nogueras-Iso, J.; Zarazaga-Soria, F.J.; Lacasta, J.; Béjar, R. and P.R. Muro-Medrano (2004). Metadata standard interoperability: application in the geographic information domain. *Computers, Environment and Urban Systems* 28 (6): 611-634.
- Vandenbroucke, D. and D. Biliouris (2011). Spatial Data Infrastructures in Spain: State of play 2011. *Spatial Data Infrastructures in Europe: State of play Spring 2011*. Leuven: K.U.Leuven. [<http://inspire.jrc.ec.europa.eu/reports/stateofplay2011/rcr11ESv120.pdf>].
- Zarazaga-Soria, F.J. Lacasta, J., Nogueras-Iso, J., Torres, M.P., Muro-Medrano, P.R. (2003). A Java Tool for Creating ISO/FGDC Geographic Metadata. In *Geodaten- und Geodienste-Infrastrukturen - von der Forschung zur praktischen Anwendung. Beiträge zu den Münsteraner GI-Tagen*. IfGI prints. 2003, vol. 18, pp. 17-30.