TRANSITION BETWEEN METADATA MODELS: APPROACH BASED ON THE USE OF OPEN SOURCE SOFTWARE.

A. Mogollón Díaz[1] J. Nogueras Iso[2], F.J. Zaragoza-Soria[2], J. López Pellicer[2],
J. Lacasta[2] , P.R.Muro-Medrano[2]
[1]Departamento de Informática Programa SIG
Universidad Pontificia de Salamanca
Paseo Juan XXIII, 3, 28040
Campus Madrid (España),
alexdiaz@unizar.es , alxmog@upsam.net
Http:// www.upsam.com
[2]Departamento de Informática e Ingeniería de Sistemas
Universidad de Zaragoza
C/María de Luna 1, 50018 - Zaragoza
Tlf: 976762358 Fax: 976761914.
{jnog , javy, fjLopez Jlacasta, Prmuro}@ unizar.es
Http://iaaa.cps.unizar.es

ABSTRACT

This article shows the work developed for adapting metadata conform to the official Colombian metadata standard NTC 4611 to the international standard ISO 19115. CatMDedit, an open source metadata editor, is used in this task. CatMDedit is able of import variants of CSDGM such as NTC 4611 and export to the stable version of ISO 19139 (the XML implementation model of ISO 19115).

Key words: *Spatial Data Infrastructure (SDI), Metadata Model, FGCD, CSDGM, ISO, NTC 4611, CatMDedit, Open Source.*

INTRODUCCION

Metadata standards: Need for the globalization of the IG

Most commonly defined as "structured data about data" or "data which describes attributes of a resource" or, more simply, "information about data", the concept of metadata is not new: map legends, library catalog cards and business cards are everyday examples. Basically, metadata offers description of content, quality, condition, authorship, and any other characteristics of the resources. It also provides a standardized representation of information. That is, similar to a bibliographical record or map legend, it provides a common set of terminology to define the resource or data. Metadata constitute the mechanism to characterize data and services in order to enable other users or applications to make use of such data and services. Metadata records, each one describing a specific resource, are grouped into catalogs thus providing the users with the possibility of finding the resources of their interest. Therefore, these catalogs are the tool that puts in touch consumers with information producers.

In order to extend the use and understanding of metadata through different communities of users, e.g. to enable distributed searches across a network of catalog servers, it is necessary to use well-defined contents and thus adjust them to a metadata standard. In this way, there are several standard proposals to describe consistently a geographic resource, which have arisen at national or global level and with different scopes. Some of the most extended ones are:

- The Content Standard for Digital Geospatial Metadata (CSDGM, [1]). It was carried out in 1994 by the Federal Geographic Data Committee (FGDC) of the United States to give support for the construction of the National Spatial Data Infrastructure. And although it is a national standard, it is the oldest one and has been incorporated into many GIS tools and networks (e.g. the Clearinghouse project), thus becoming the most widely used in GIS world (e.g. adopted in countries like South Africa or Canada).
- The international standard ISO19115, [2]. The organization responsible for this standard is the International Organization for Standardization which created in 1992 the committee 211 (ISO/TC 211) with responsibilities in "geomatics" (http://www.isotc211.org) This committee is now preparing a family of standards that, in the near future, will obtain the rank as official international standard. One of these standards is the 19115, released as standard in May 2003, which defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. This standard is applicable to: the cataloguing of datasets, clearinghouse activities, and the full description of datasets; geographic datasets, dataset series, and individual geographic features and feature properties. Furthermore, though ISO19115 is applicable to digital data, its principles can be extended too many other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data.

Apart from the geographic-context metadata standards, a good example of a simple general purpose metadata standard is the one proposed by Dublin Core Metadata Initiative (DCMI, http://www.dublincore.org). This initiative, created in 1995, promotes the widespread adoption of interoperable metadata standards and the development of specialized metadata vocabularies that enable more intelligent information discovery systems. The Dublin Core metadata element set is a standard for the description of cross-domain information resources, i.e. any kind of resource, regardless of the media format, area of specialization or cultural origin. This set consists of 15 basic descriptors which are the result of an international and interdisciplinary consensus. Nowadays, the Dublin Core metadata element set has become an important part of the emerging infrastructure of the Internet. Many communities are eager to adopt a common core of semantics for resource description, and the Dublin Core has attracted broad ranging international and interdisciplinary support for this purpose. The Dublin Core now exists in over 20 translations, has been adopted by CEN/ISSS (European Committee for Standardization / Information Society Standardization System), and is documented in two Internet RFCs (Requests for Comments). It has also official standing within the WWW Consortium and the Z39.50 standard. Dublin Core metadata has been approved as a U.S. National Standard (ANSI/NISO Z39.85) [3]), formally endorsed by over seven governments for promoting discovery of government information in electronic form, and adopted by a number of supranational agencies such as the World Health Organization (WHO). Numerous community-specific metadata initiatives in library, archival, educational, and governmental applications are using the Dublin Core as their basis. Moreover, since April 2003, the Dublin Core Metadata Element Set standard has been adopted as ISO standard (ISO 15836) [4]. This approval is the culmination of an incremental process to bring the Dublin Core metadata element set into a worldwide audience. As an international standard, it will be easier for many organizations to adopt and promote the use of Dublin Core to enhance resource discovery on the Internet.

Metadata Application Profiles

Although the conditionality of elements enable certain flexibility of geographic metadata standards, they are very detailed. CSDGM and ISO19115 standards comprise more than 300 elements distributed in sections and subsections. The problem is that in order to complete metadata records in accordance with such detailed standards, metadata creators must be highly qualified and spend quite a lot of time. That is the reason why the own document defining ISO19115 standard also includes a profile called "Core metadata for geographic datasets" that only includes 22 elements, the minimum number of metadata elements required to identify a dataset for catalogue purposes. Other organizations involved in the cataloguing of geographic resources go even further and propose the use of more generic standards like Dublin Core. The idea is to provide at least discovery level metadata. In addition, once the elements to be filled have been selected, it is necessary to establish a set of rules or guidelines for determining their values (ranges, formats, vocabulary controlled lists, etc). That is to say, it is necessary to specify an application profile.

According with Dublin Core, an Application Profile is a declaration specifying which metadata terms an organization, information provider, or user community uses in its metadata and how those terms have been customized or adapted to a particular application. By definition, a Dublin Core Application Profile is based in part on Dublin Core and follows DCMI Grammatical Principles.

In general, an application profile should have in their objectives to facilitate the metadata creation process by: Providing an specification of the subset of the metadata standard terms, choosing the ones that are more relevant for a specific domain; offering guidelines for filling the fields of the metadata according with the specific domain; providing specific keyword controlled lists, thesaurus and ontologies for the context where the application profile is used. In addition, application profiles should offer the base for creating, or adapting, metadata creation tools in order to do the metadata creation process easier.

NTC 4611

In Latin America, the development of information systems, which would replace the traditional system of acquisition, organization, distribution and diffusion of geographic data, has received considerable attention by Governmental agencies, private sector, academic organizations and the user community in general. This interest is reflected in the Cartographic Conference of United Nations for the Americas 1996 [5].

In Colombia the creation of the national SDI, known as ICDE, started from the experience obtained in other countries, specially the work made by the Federal Geographic Data Committee (FGDC) in the definition of the concept of the National Spatial Data Infrastructure (NSDI) of U.S. Perhaps the greatest contribution of this organization was its standard CSDGM. The CSDGM is not an international norm but it was spread quickly among organizations of several countries as a de facto international standard. For example in 1998, Australia, South Africa, Colombia, Uruguay among other countries adopted CSDGM profiles as national standard.

With the U.S. experience started the work of standardization in Colombia. The ICONTEC (National Organism of normalization and certification Colombian) [6] with the help of the National Technical Secretariat of the Austin Codazzi Geographic Institute (IGAC) [7] defined the National Norm of Geographic Metadata denominated NTC 4611. The NTC 4611 is the result of the analysis of norms such as CSDGM, the CEN and ISO TC/211 group. The result is one mixture of the three norms where the influence of the FGDC philosophy prevails clearly.

In the 2004, the American National Standards Institute (ANSI) adopts the ISO 19115 as an American standard National [8]. As a consequence, the FGDC will substitute the CSDGM by a profile of the ISO 19115. For the users community of CSDGM was designed a FGDC to ISO Metadata crosswalk [9], that establishes an equivalence between the elements of both standard. Nevertheless, this is not a solution in Colombia because when using this crosswalk in the Colombian metadata, it has been observed that the Colombian metadata model NTC 4611 is not transformed into a model consistent with ISO 19115. This shows the necessity of building an ISO 19115 profile of Colombian Metadata in order to obtain the interoperability with other Latin American systems.

Metadata Creation support

The creation of metadata is an arduous labor that must be facilitated by the adequate tools. The one selected has been the one provided by the Open Source project *CatMDEdit* (see [10], available at http://catmdedit.sourceforge.net). It is a metadata editor tool that facilitates the documentation of resources, with special focus on the description of geographic information resources.

The tool has been implemented in Java and has the following features:

- Cross-platform (Windows, UNIX). As it has been developed in Java and the storage of metadata records is managed directly through the file system, the application can be deployed in any platform with the minimum requirement of having installed a Java virtual machine.
- Multilingual. The application has been developed following the Java internationalization methodology. Nowadays, there is a Spanish, English, French, Polish, Portuguese and Czech version. With little effort, other languages can be supported.
- Selection and filtering of metadata records stored in the local metadata repository.
- Metadata edition in conformance with the "ISO19115. Geographic Information - Metadata" standard. Four interfaces are provided for the edition of metadata records:
  - A detailed interface following the ISO19115 comprehensive profile.
  - A reduced interface following the "Núcleo Español de Metadatos" (NEM). NEM, a subset of ISO19115, is a recommendation under development which has been defined by the Spanish National Geographical High Board ("Consejo Superior Geográfico"). This subset includes all the elements defined for the ISO19115 Core metadata profile (“Core metadata for geographic datasets”).
  - An interface following the SDIGER - INSPIRE metadata profile (see [12]).
  - An interface following the SDIGER - WFD metadata profile (see [13]).
- Metadata edition in conformance with the SDIGER - Dublin Core Metadata Application Profile for geographical data mining (see [14]).
- Exchange of metadata records according to different standards and formats. Interoperability with other metadata standards apart from ISO19115. The application allows input and output XML files in conformance with the standards CSDGM (Content Standard for Digital Geospatial Metadata, defined by U.S. FGDC), Qualified Dublin Core, SDIGER - Dublin Core Metadata Application

Profile for geographical data mining, or MIGRA (Spanish standard for geographic information exchange).

- Different styles for presentation of metadata records
  - For CSDGM: FGDC HTML (es, en), FAQ HTML (en), Geography Network HTML (en), ESRI HTML (es, en).
  - For ISO19115: HTML (es, en, fr, pl, pt), Excel (format used for both input and output files).
  - For MIGRA: HTML (es).
  - For Dublin Core: HTML (es, en, fr, pl, pt).
- Additional tools to facilitate the edition of metadata:
  - Contacts repository tool. It permits the reuse of contact information (e.g. name, address, telephone …) of organizations and individuals.
  - Thesaurus tool. It enables metadata creators to use thesauri in order to fill in some metadata elements. The use of controlled keywords facilitates the mapping between a selected vocabulary and a large collection of records. This tool is based on ThManager, an Open Source application for thesaurus management (see [15], available at http://thmanager.sourceforge.net.
  - Additional tools for the selection of bounding box coordinates: coordinate conversion between different coordinate reference systems, and graphic selection of coordinates over maps.
  - Automatic Metadata Generation. CatMDEdit enables the automatic generation of metadata for some data formats like Shapefile, DGN, ECW, FICC, GeoTiff, GIF/GFW, JPG/JGW, PNG/PGW.
  - On-line help by means of PDF visualization.
- A Contact Management Tool, allowing reusing contact information (e.g. name, address, telephone…), which is needed in several metadata fields. Thanks to this tool, the contact information about a person is only inserted once and used whenever it is required.

TRANSFORMING METADATA FROM NTC 4611 TO ISO 19115

Transformation Context

In Colombia Bogotá city the D.C council proposed a SDI, as a challenge of modernization and change considering that a great percentage of the institutions and administrative departments of the city handles geographic information that is useful for their decision process.

The decisions that these organisms take depend to great extent on the quality, exactitude and the level of update of these data. The main objective of the SDI is to improve the competitiveness of the administration by fomenting the production and the maintenance of the geographic data and allowing the possibility of a real interchange between the different organizations. This challenge involves the data interchange between different nodes using the current computer, software and communication infrastructure.

In the year 2002 Bogotá council established the political and technical committees to be in charge of the local SDI and it made them responsible of the decisions. The political committee gather departments which mission is closely related to geographic data: "Departamento Administrativo de Catastro Distrital" (DACD, cadaster office), "Departamento Administrativo de Planeación Distrital" (urban planning), "Departamento Administrativo del Medio Ambiente" (environmental data), "Instituto de Desarrollo Urbano" (urban equipment), "CODENSA" (electricity network data), "Empresa de Telecomunicaciones de Bogotá" (communication network data), "Empresa de Acueducto y Alcantarillado de Bogotá" (hydrographic network data), "Gas Natural de Bogotá" (oil network data), "EPM-Bogotá" (communication network data), "Jardín Botánico de Bogotá José Celestino Mutis" (leisure thematic area), "Dirección para la Prevención y Atención de Desastres" (risk management office), "Departamento Administrativo de Bienestar Social" (poverty data), "Departamento Administrativo de Defensoría del Espacio Público" (DADEP, public owned real estate data), "Caja de Vivienda Popular" (social housing data) and "Secretaría de Hacienda Distrital" (tax data). The main task of this committee is the agreement on the data framework, which entity is responsible of each thematic dataset and what are the resources that can be applied to.

The main task in the development of the SDI has been the metadata creation using the CSDGM standard. Consequently, in 2006 there are a high percentage of documented cartographic products of the Bogotá city in this standard. As CSDGM is now deprecated, this is a cause of interoperability problems whose solution is the creation of a new crosswalk that will be able to transform the old NTC 4611 metadata to ISO 19115 metadata.

Theoretical Transformation Model

Although the NTC 4611 is the result of the analysis of norms such as CSDGM, the CEN and ISO TC/211 group, a shallow analysis reveals that this standard is mainly a CSDGM subset with Colombian code lists. This allows the mapping of the controlled lists of both standards. By this way we will obtain metadata in standard CSDGM from metadata NTC 4611. Then we can apply to the result an existent metadata crosswalk from CSDGM to ISO. The resultant metadata will be a Colombian subset of ISO 19115 (which has the same relation to ISO 19115 as Spanish NEM has respect to ISO 19115). This chained transformation is described in the figure 1.

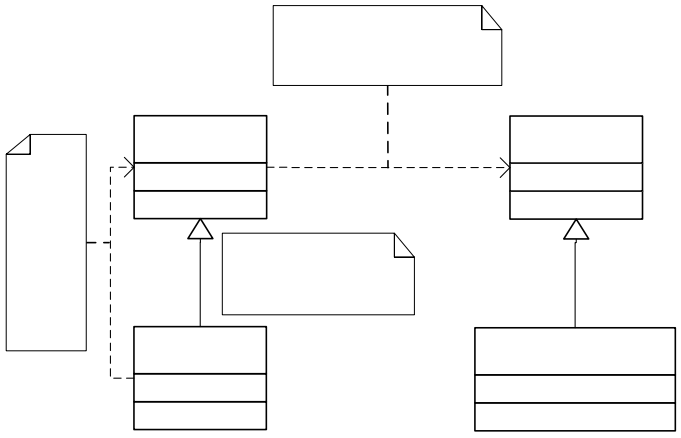

Figure 1: *Sub 19115 Colombian = NTC 4611 x mapping A x Mapping B.*

Making the transformation

We have used the CatMDEdit tool to implement this transformation process. The first and time expensive step is the development of a matching between the NTC 4611 and the CSGDM standard. This is an ad-hoc mapping between NTC 4611 and CSGDM (see figure 2a). This mapping was not implemented in CatMDEdit tool but as open source tool allows modifying the code to support this mapping. The second step was already implemented in the CatMDedit tool: it is a generic transformation between the standards CSGDM and ISO 19115 (see figurate 2b). By this way ISO 19115 compliant metadata was built from NTC 4611 metadata.

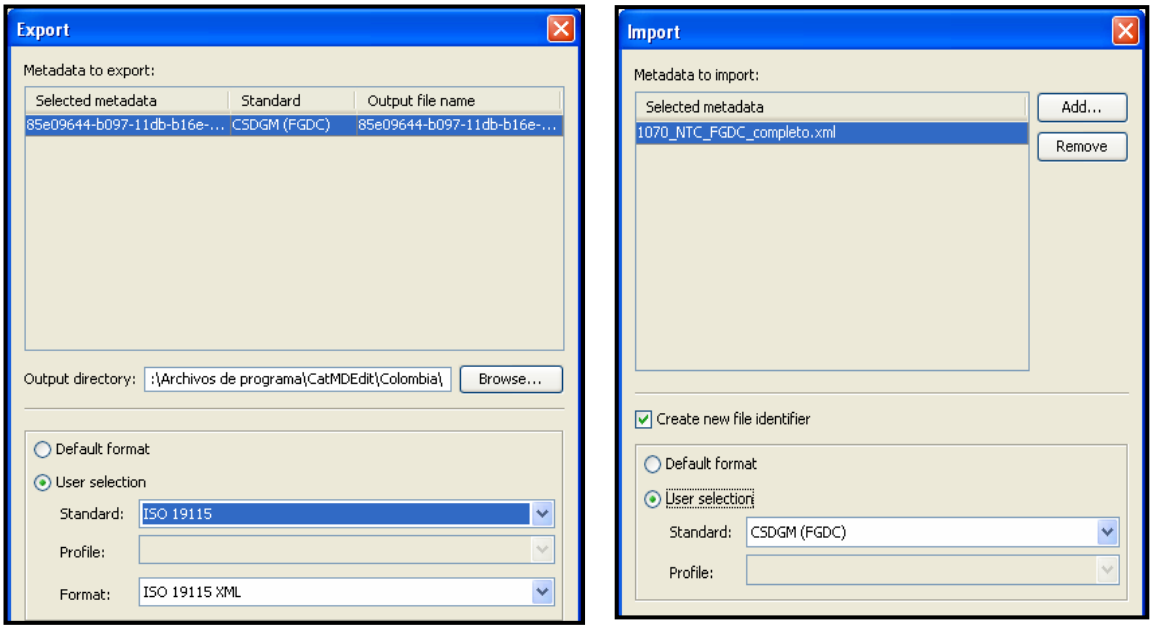

Figure 2: a) To the left, interface of development F1, b) To the right, interface of development

After the application of the referred operations, the transformation did not work. During this process, some problems were found. Following, the characterization of these problems is presented using some metadata examples provided by the Bogotá IDEC@ prototype project.

Controlled lists mapping.

There are some CSDGM items that should be completed using controlled lists; The theoretical model for the transformation suggested the creation of mapping between the NTC 4611 and the CSDGM values. The reality says that in many cases this is not possible.

```
<metadata>
  <idinfo>
    <status>
      <progress>Continuo</progress>
      <update>Continuo</update>
    </status>
  </idinfo>
</metadata>
```

```
<metadata>
  <idinfo>
    <status>
      <progress>IN WORK</progress>
      <update>CONTINUALLY</update>
    </status>
  </idinfo>
</metadata>
```

Figure 3: *NTC 4611 controlled list (left), CSDGM controlled list (right).*

For instance, figure 3 shows an XML extract of NTC 4611 and CSDGM files. The path is the same in bath cases, but the values in the NTC 4611 do not correspond with the CSDGM standard. The FGDC standard proposes for the item *<progress>* on of the following values: *Complete, In work, Planned*. And for the item *<update>*, the values proposed are: *Continually, Daily, Weekly, Monthly, Annually, Unknown, As needed, Irregular, None planned, Free text*. In the second cases, the matching can be done using one of the controlled values of the list or using the *free text* option. But in the case of *<progress>* the value included in the NTC 4611 example cannot be matched with any of the CSDGM values.

Non CSDGM items

In some cases, NTC 4611 included new items that are not included in the FGDC standard and do not follow its extension rules. In other cases, the items have been put in a different place inside the standard.

```
<metadata>
  <idinfo>
    <citation>
      <imageinf>
        <project />
        <pathnum />
        <rownum />
      </imageinf>
```

```
<metadata>
  <spref>
    <horizsys>
      <planar>
        <mapproj/>
          <mapprojp />
            <pathnum />
```

Figure 4: *Non standard items: NTC 4611 (left), CSGDM (right).*

For instance, figure 4 presents the elements *<pathnum>*, *<project>* and *<rownum>*. The first one is included in the CSGDM, but it is not in section *<idinfo>*. The other elements are not part of the FGDC norm.

```
<metadata>
  <idinfo>
    <spdom>
      <geogext>337839.9 m2</geogext>
      <bounding>
        <westbc />
        <eastbc />
        <northbc />
        <southbc />
      </bounding>
      <coorplan>
        <westpc>99085.9</westpc>
        <eastpc>99820.2</eastpc>
        <northpc>106274.5</northpc>
        <southpc>105023.1</southpc>
        <origdatum>LOCALES</origdatum>
      </coorplan>
      <scres>1000</scres>
    </spdom>
  </idinfo>
</metadata>
```

```
<metadata>
  <idinfo>
    <spdom>
      <bounding>
        <westbc>-6.78</westbc>
        <eastbc>-1.83</eastbc>
        <northbc>43.09</northbc>
         <southbc>40.11</southbc>
      </bounding>
      <scale>1000</scale>
    </spdom>
  </idinfo>
</metadata>
```

Figure 5: *Non standard items: NTC 4611 (left), CSGDM (right).*

Figure 5 presents another example of the related problem. All the items inside the section <coorplan> are not part of the FGDC standard. The same problem occurs with the item <*geogext*> and with <*scres*> (related with the scale). The CSDGM offers several possibilities for including information about the scale, but it is never and stand-alone path (in lineage: <*srcscale*>; in spatial reference systems: <*sfequat*>, <*sfctrmer*>; …).

Path modified

Finally, in some cases the NTC 4611 does not follow the CSDGM scheme for the XML paths.

```
<metadata>
  <idinfo>
      <timeperd>
         <begdate>2004/01/01</begdate>
         <enddate>2006/12/31</enddate>
      </timeperd>
  </idinfo>
</metadata>
```

```
<metadata>
   <idinfo>
      <timeperd>
         <timeinfo>
            <rngdates>
               <begdate>2004/01/01</begdate>
               <enddate>2006/12/31</enddate>
            </rngdates>
         </timeinfo>
      </timeperd>
   </idinfo>
   </metadata>
```

Figure 6: *Path modified: NTC 4611 (left), CSGDM (right).*

For instance, figure 6 presents a situation where the elements <*timeinfo*> and <*rngdates*> have been eliminated from the XML path.

CONCLUSIONS

With the recent establishment of an international standard in geographic information domain (ISO 19115 and ISO 11939 for Metadata XML Schema Implementation) its is now time for converging metadata created following the schemes of deprecated metadata standards, pre-standards, and regional norms to this new international norm. The diversity of metadata profiles with their national variations, and without correspondence with well-established standards, causes that the transformations should not be easy to do tasks as the application of existed crosswalks. In most cases, the matching process is very hard and never should assume that there always exists a direct connection between terms.

The influence of the CSGDM in the creation of Latin American metadata models based and the evolution of these models to adapt the local needs (new or different terms and parameters) have generated an exponential complexity in the transformation, that has been only possible mitigated after the transformation of metadata from these profiles to the original CSGDM standard.

This work has shown that transformation from CSGDM Latin America profiles, such as the Colombia NTC 4611, to ISO 19115 is not an easy task. In most cases, these profiles have information that does not correspond to the standard CSGDM. Therefore, current cross-walk between CSGDM and ISO 19115 are not directly applicable.

Open Source tools for the administration edition and update of the metadata such as CatMDEdit with access to its source code allows developers for its modification in order to provide specialized mappings such as the mapping from Colombian NTC 4611 to ISO 1 9115.

ACKNOWLEDGEMENTS

This work is partially supported by Ministry of Education and Science of Spain grant TIN2006-00779. The work of Alexander Mogollón Díaz has been partially financed by the postgraduate program II-E “Scholarships for PHD Latin America Students MAEC-AECI 2005-06”. The work of F.J. López-Pellicer (ref. B136/2006) has been partially financed by a scholarship of the Government of Aragón (Spain).

REFERENCES

[1] Federal Geographic Data Committee (FGDC) (1998),'Content Standard for Digital Geospatial Metadata, version 2.0'(FGDC-STD-001-1998), Technical report, Metadata Ad Hoc Working Group.
[2] Geographic Information - Metadata'(19115:2003), Technical report, International Organization for Standardization (ISO).
[3] ANSI/ASQ E4-1994. Specifications and Guidelines for Quality Systems for Environmental Data Collection and Environmental Technology Programs, Technical report, American National Standards Institute.
[4] INSPIRE (2004),'Homepage of the INfrastructure for SPatial InfoRmation in Europe (INSPIRE): http://www.ec-gis.org/inspire/'.
[5] Banco Mundial, El conocimiento al servicio del desarrollo, Informe sobre el desarrollo Mundial 1998-1999.
[6] Organismo Nacional de Normalización y Certificación en Colombia ICONTEC: Geographic Information. Metadata Geografic.NTC 4611 (2002)..
[7] Instituto Geográfico Agustín Codazzi IGAC: Experiencia y definiciones de la ICDE, Bogota Febrero 2000
[8] Federal Geographic Data Committee: US National Profile of ISO 19115, Geographic Information – Metadata Washington (2005).
[9]. Federal Geographic Data Committee: ISO-FGDC-METADATA-CROSSWALK-V4 (2003)
[10] 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 'Proc. of the GI-days 2003 conference', pp. 17-30.
[11]. Nogueras-Iso, J.; Muro-Medrano, P. & Zarazaga-Soria, F.J. (2005),'SDIGER - INSPIRE Metadata Profile', Technical report, SDIGER: A cross-border inter-administration Spatial Data Infrastructure to support WFD information access for Adour-Garonne and Ebro River Basins. Contract Number - 2004 742 00004, Available at http://www.sdiger.net.
[12] Nogueras-Iso, J.; Muro-Medrano, P. & Zarazaga-Soria, F.J. (2005),'SDIGER - WFD Metadata Profile', Technical report, SDIGER: A cross-border inter-administration Spatial Data Infrastructure to support WFD information access for Adour-Garonne and Ebro River Basins. Contract Number - 2004 742 00004, Available at http://www.sdiger.net.
[13] Nogueras-Iso, J.; Muro-Medrano, P.; Zarazaga-Soria, F.J. & R.Rioja (2005),'SDIGER - Dublin Core Metadata Application Profile for geographical data mining', Technical report, SDIGER: A cross-border inter-administration Spatial Data Infrastructure to support WFD information access for Adour-Garonne and Ebro River Basins. Contract Number - 2004 742 00004, Available at http://www.sdiger.net.
[14] Lacasta, J.; Nogueras-Iso, J.; Lopez-Pellicer, J.; Medrano, P.M. & Zarazaga-Soria, F.J. (2007), 'ThManager: An Open Source Tool for creating and visualizing SKOS', Information Technology and Libraries.