

Experiencia en la creación de metadatos en una IDE. Caso práctico de la Confederación Hidrográfica del Duero.

G. Sanz ¹; S. Sánchez ¹; J. Fernández ²; V. Arqued ²; I. Montequi ^{3*}; R. Martínez-Alegría ^{3**}.

¹ Fundación Miguel de Cervantes.
Consultores Independientes.
gsanz@uemc.es
ssanchez@uemc.es

² Confederación Hidrográfica del Duero.
C/ Muro, 5. CP: 47004 - Valladolid
Telf: (+34) 983 215 400 / Fax: (+34) 983 21 54 38
oph@chduero.es

³ Universidad Europea Miguel de Cervantes.
C/Padre Julio Chevalier, 2. 47012 - Valladolid
Telf.: (+34) 983 001 000. Fax: (+34) 983 278 958
*imontequi@uemc.es
**rmartinez@uemc.es

Resumen

En este trabajo se presenta una comparación entre distintos procedimientos de creación de metadatos con las herramientas específicas de mayor difusión disponibles en la actualidad. Su desarrollo se sustenta en la experiencia práctica adquirida por los autores en el proceso de análisis, selección, y creación de metadatos asociados a la información cartográfica de una Infraestructura de Datos Espaciales (IDE).

La información cartográfica de partida procede de la Infraestructura de Datos Espaciales de la Cuenca del Duero (MIRAME IDE-Duero).

El marco de referencia está definido por los principios de interoperabilidad de la Directiva 2007/2/CE INSPIRE, la Infraestructura de Datos Espaciales de España (IDEE), dentro del contexto del Núcleo Español de Metadatos (NEM) y del “Water Information System for Europe” (WISE).

Las herramientas de gestión de los datos cartográficos comparadas han sido: El soporte lógico de libre difusión gvSig. La plataforma privada de mayor difusión dentro de los Sistemas de Información Geográfica ArcInfo. La aplicación específica de gestión de metadatos CatMDEdit. Y las herramientas disponibles en MÍRAME-IDEDuero desarrolladas sobre la plataforma libre GeoNetwork opensource.

Palabras clave: Metadatos. Infraestructura de Datos Espaciales. INSPIRE. NEM. WISE. CatMDEdit. Sistema de Información Geográfica. Duero.

Introducción

Contexto y objetivos

La Confederación Hidrográfica del Duero, por su naturaleza de organismo de cuenca, tiene un amplio rango de competencias relacionadas con el mundo del agua. Para el desarrollo de tales competencias, hoy día, resulta fundamental el uso de sistemas de información geográfica. Además, debido a las necesidades de intercambio de información con otros agentes involucrados, ha sido fundamental la creación de una IDE. Su nombre es MÍRAME-IDEDuero y se acceda a través de página web del organismo (1).

El Sistema MÍRAME-IDEDuero (2) es una herramienta que da soporte continuado al Plan Hidrológico y a la labor técnica que desarrolla el Organismo de cuenca. El sistema es el resultado de la combinación de las correspondientes bases de datos y servicios web. Entre ellos está el servicio de catálogo, que da acceso a los metadatos creados, y que sirve además como puente para acceder a otros servicios de catálogo externos.

El SIG puesto en marcha por la CHD está condicionado por las directrices que establece la Directiva 2007/2/CE del Parlamento Europeo y del Consejo de 14 de marzo de 2007 por la que se establece una infraestructura de información espacial en la Comunidad Europea (INSPIRE) (3), dentro del marco que regula la puesta en común de los datos espaciales por parte de las administraciones públicas relativas a la temática ambiental. Esta directiva presenta, dentro de sus objetivos, la adecuada documentación de la información geográfica, mediante la creación de metadatos.

La elaboración de metadatos está normalizada de acuerdo, entre otras, por la ISO 19115 (4). Aunque este estándar es muy extenso, permite el desarrollo de subconjuntos de menor entidad, tales como el Núcleo Español de Metadatos (NEM) (5) impulsado por el Instituto Geográfico Nacional (IGN), o el núcleo WISE, desarrollado por la Comisión Europea en documentos como “WISE GIS

guidance document” (2008) (6), con la participación directa, como representantes españoles, de los técnicos de la Confederación Hidrográfica del Duero.

El objetivo del presente trabajo es mostrar la experiencia práctica de creación y mantenimiento de metadatos en una infraestructura creada de acuerdo a los principios de INSPIRE. Pretende abordar aspectos que van desde las dificultades del estándar, hasta las particularidades de las principales herramientas existentes en el mercado en el momento del estudio, pasando por la importancia de las singularidades de la cultura de la organización.

La existencia de los metadatos se justifica por la necesidad de los organismos y organizaciones para gestionar la Información Geográfica y confeccionar sus Catálogos. El Catálogo es la componente básica de una IDE, y la más importante, ya que va a facilitar a los usuarios la búsqueda, localización, comparación y utilización de los conjuntos de datos geográficos y, en el futuro, de todo tipo de servicios y recursos. Para que los catálogos procedentes de distintas organizaciones sean interoperables y admitan búsquedas distribuidas, es necesario que satisfagan las especificaciones de interoperabilidad de Open Geospatial Consortium y que estén constituidos a partir de normas de metadatos que sean comunes para todos ellos.

Antecedentes

Los antecedentes históricos de las IDE se remontan a 1994 en Estados Unidos de America cuando se creó el Comité Federal de Datos Geográficos (Federal Geographic Data Committee (FGDC)) (7), responsable de la puesta en marcha y gestión de la Infraestructura Nacional de Datos espaciales (National Spatial Data Infrastructure (NSDI), (8). Ya en ese momento se comienza a evidenciar la necesidad de crear metadatos para la información geográfica.

Posteriormente fue creado “Dublin Core Metadata Initiative” (DCMI) (9) es una organización abierta, sin ánimo de lucro, que desde 1995 participa en el desarrollo de estándares de metadatos interoperables. Este grupo de trabajo ha aportado una visión más práctica, facilitando la labor de la creación de los metadatos. Esta labor de simplificación ha sido continuada por los distintos organismos institucionales de los diferentes países adscritos. En el caso de España dicha labor ha sido desarrollada por el Instituto Geográfico Nacional (IGN) creando el Núcleo Español de Metadatos (NEM).

En 2002, dentro del marco europeo, se definió el primer perfil de metadatos de WISE en la Guidance Document on Implementing the GIS Elements of the WFD (Working Group GIS, Jüergen Vogt 2002) (10), sujeto a un proceso continuo de mejora, basado en su uso y en la norma ISO19115 existente en esa fecha.

Posteriormente, esta guía ha sido actualizada en 2008, a partir de los trabajos realizados por un grupo de expertos y personas implicados en la Directiva Marco del Agua (2000/60/CE), con especial interés en aquellas dedicadas al manejo de Sistemas de Información Geográficos. Este grupo ha contribuido al desarrollo y conocimiento de todas las materias referentes al agua y su normativa dentro de la Unión Europea mediante la guía GIS-WISE 2008 (Geography Information System Water Information System of Europe) (5).

Por otro lado, dentro del contexto internacional, el Comité Técnico 211, de Geomática/Información Geográfica, dentro de la Organización de Estandarización Internacional (ISO), ha desarrollado una familia de normas, las ISO 19100, para la Información Geográfica. Esta familia son un conjunto de normas relacionadas con objetos o fenómenos que están directa o indirectamente asociados con una localización relativa a la Tierra.

Las principales normas de esta familia son: la ISO 19115 – 2003 Geographic information, Metadata (3), la ISO 1915-2 Geographic information -Metadata- Part 2 (11) publicada en la fase 60.60, la ISO 19119 – 2005 Información geográfica. Servicios (12), la ISO 19106 “Geographic Information – Profiles” (13), y la ISO 19139 - 2007 Draft Technical Specification: Geographic information – Metadata – XML schema implementation (14).

Otra norma internacional que se debe tener en cuenta es la ISO 15836:2003 “The Dublin Core Metadata Element Set” (15) y los proyectos más relevantes en curso en materia de metadatos, especialmente la Directiva Marco del Agua y las iniciativas IDE regionales.

A todo este trabajo desarrollado por las administraciones públicas hay que añadir el realizado por los trabajos realizados por otros grupos pertenecientes a distinta instituciones dedicadas a la investigación. Durante el desarrollo de este proyecto se ha tenido un contacto más estrecho con el Laboratorio de Tecnologías de la Información Geográfica (LatinGEO), perteneciente a la Universidad Politécnica de Madrid (16). Quienes desarrollaron los cuestionarios, para recopilar la información necesaria para la creación de los metadatos de información geográfica de acuerdo con las diferentes normas y directivas vigentes relativas a metadatos, diferenciando entre información raster, vectorial y servicios.

1. Metodología seguida para creación de metadatos.

Conviene decir que la calidad de creación de estos metadatos viene determinada por la cantidad y veracidad de la información de la que se disponga. Esta tarea es laboriosa debido al volumen de campos solicitados y a la falta de información.

Inspirados en la guía GIS-WISE (6), en una fase previa a la creación de metadatos, se han realizado las siguientes tareas:

1. Exhaustiva revisión del estado del conocimiento y de la bibliografía, constituyendo esta la primera etapa de la metodología teórica.
2. Análisis de conceptos y metodologías existentes sobre el proceso de creación de metadatos, con el fin de adoptar una metodología a las necesidades cartográficas.

Así, se identifican los tres niveles teóricos de metadatos descritos en “los Protocolos para la generación de metadatos espaciales”, de la Conselleria d’Infraestructures i Transport; y Universidad Jaume I (8), distinguiéndose metadatos de descubrimiento, metadatos de exploración y metadatos de explotación. Posteriormente hablaremos de ello.

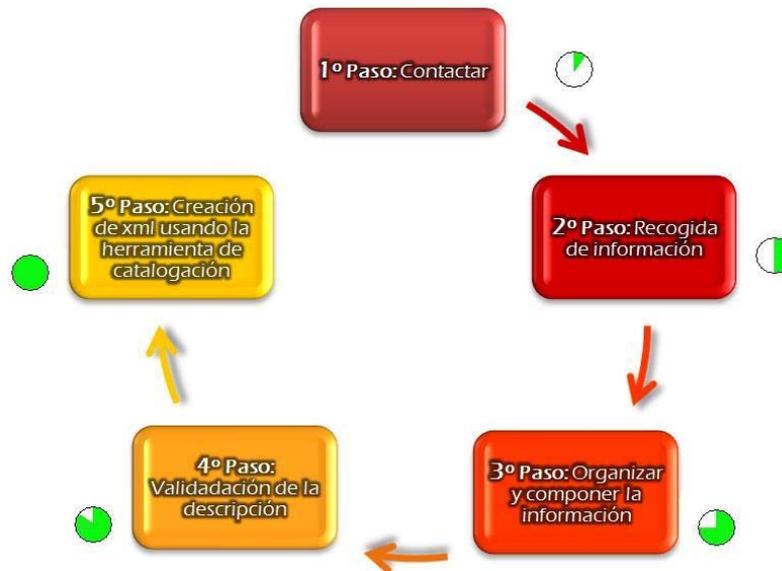


Figura 1. Metodología descrita por LatinGEO y recogida en la guía GIS WISE (6).

2. Herramientas y Justificación de su Uso.

Para la creación de los metadatos para una IDE, conviene buscar una herramienta de edición de metadatos que se ajuste en la medida de lo posible a las necesidades y máximas propuestas por los distintos estándares. Existen multitud en el mercado. En el presente artículo se mencionan las que se valoraron en el momento de crear los metadatos del sistema de información del Duero.

La primera herramienta para la visualización y edición de los metadatos suele ser la propia del programa informático que se esté utilizando para el manejo de la información geográfica. Como consecuencia de la gran difusión del formato “.shp” de representación de la información espacial, la aplicación ArcInfo (manuales) (17) se presenta como la primera opción, aunque los programas informáticos de SIG libre, como gvSIG (manual) (18) y LocalGIS (19), proporcionan un abanico más amplio de herramientas para la edición de los metadatos. Dentro de este grupo de herramientas se encuentra la plataforma GeoNetwork (20), que ha sido la herramienta seleccionada para realizar este trabajo.

Existen también aplicaciones desarrolladas específicamente para la edición de metadatos, entre las que cabe destacar a nivel español el CatMDEdit (manual) (21), desarrollado por el Grupo de Sistema de Información Avanzado de la Universidad de Zaragoza junto y GeoSpatiumLab S.L.

Además, existen otras herramientas desarrolladas específicamente para IDEs concretas, que permiten la creación y edición de metadatos. Éstas, aunque menos versátiles que el CatMDEdit, suelen adaptarse mejor a las restricciones establecidas por los propietarios de la IDE, ya que el portal suele estar desarrollado para lograr los objetivos deseados. Como contrapartida suponen un esfuerzo de desarrollo mucho mayor.

3. Descripción de los procedimientos de creación de metadatos cartográficos según la experiencia práctica con la Confederación Hidrográfica del Duero.

El trabajo comienza con la elección de un núcleo de metadatos acorde con las restricciones acordadas durante el planteamiento del proyecto. En este caso concreto consideraron los subconjuntos NEM y WISE.

En el protocolo de actuación, según la metodología aplicada, se ha establecido tres niveles de metadatos.

Metadatos de descubrimiento. Estado inicial de la información.

Tras establecer la estructura de las capas y su importancia relativa, a efectos de poder priorizar, la fase de obtención de los metadatos de descubrimiento consiste en:

1. Establecer contacto con los responsables de las capas. En nuestro caso, fue necesario que éste fuera siempre directo, debido a las características del proyecto.
2. Determinar el estado de los metadatos, conociendo y publicando el estado en el que se encuentra la información espacial al comienzo del proceso de creación de metadatos.

Metadatos de Exploración. Obtención de la información.

La información restante será obtenida mediante:

- Cuestionarios: recogen toda la información disponible organizada y estructurada. Los cuestionarios han sido proporcionados por el grupo de trabajo de LatinGEO, perteneciente a la Universidad Politécnica de Madrid, se revisaron y adaptaron a las exigencias de la CHD incluyendo los requisitos del NEM y del WISE.
- Reuniones: en ellas se completarán los metadatos mediante el uso de fichas, o directamente en la aplicación informática seleccionada, CatMDEdit, o el propio Servicio de Catalogo del Sistema MÍRAME-IDEDuero, de la CHD.

Una forma de reducir los tiempos en la creación de metadatos es desarrollando un metadato inicial que recoja la mayor cantidad posible de información, con el fin de que los responsables únicamente deban revisarlos y validarlos.

Se ha de intentar obtener toda la información exigida por la norma o estándar sobre el que se trabaja, pero generalmente debido a la gran cantidad de datos, muchos de ellos, de difícil obtención, de dudosa fuente, o escasa veracidad, se ha de optar por pedir a la organización toda la información que consideren oportuna.

En la práctica se ha comprobado que la utilización de cuestionarios, en el caso de la CHD no resulto operativa, ya que todo el conocimiento de la información se centraba en un departamento pequeño, haciendo inviable completar la todos los cuestionarios. Por este motivo se procedió a elaborar directamente los metadatos geográficos directamente empleando la herramienta del Servidor de Catalogo del Sistema MÍRAME-IDEDuero con la información de partida.

Esta medida resulta ser extremadamente eficaz, porque permite avanzar de forma significativa en la ejecución del proyecto, ya que reduce el tiempo que deben invertir los responsables, reduciendo su tarea al de revisión en un entorno en el que están familiarizados.

Metadatos de Explotación. Protocolos de Creación.

Con la información procedente de los niveles anteriores, se elabora el archivo “.xml” con formato estándar en el Servicio de Catalogo del Sistema MÍRAME-IDEDuero. Posteriormente son cargados en dicha aplicación y validados por el responsable de la información pasando a estar a disposición del público de la red.

Durante esta etapa es necesaria la adopción de un tesoro que unifique la terminología, permitiendo hacer agrupaciones temáticas que faciliten la búsqueda de los metadatos. Para ello se han revisado y analizado diferentes tesauros, con el fin de homogenizar y limitar el alcance de las palabras clave, que deben ser capaces de describir y organizar de forma adecuada el contenido de cada capa.

Revisión de Tesauros.

- AGROVOC (22): Tesoro desarrollado por “The Food and Agriculture Organization” (FAO – UN). Posee una estructura jerárquica y temática. Está desarrollado en los principales idiomas, se encuentra disponible en la red, y su actualización es constante.
- GEMET (23): Tesoro de la Agencia de Medio Ambiente de la Unión Europea (EEA). Posee diccionarios de todos los idiomas de los países de la UE, no presenta una estructura jerárquica de términos, sino que se trata de un diccionario, y su principal temática es la Ambiental. Su última actualización se realizó el 21 de Abril de 2009, y también está disponible en la red. Mantiene una relación documentada con la función proporcionada por el geoportal INSPIRE.

En el caso del la CHD se utilizó el AGROVOC de la FAO, pero después de diferentes experiencias y teniendo en cuenta la relación que el GEMET mantiene con el geoportal INSPIRE, se recomienda el uso de este último, de forma preferente respecto a otros. En términos prácticos se uso AGROVOC porque se estaba a la espera de la Actualización y mejora del GEMET, que finalmente se produjo en abril de 2009.

Elaboración de una ficha de actualización.

Con carácter complementario y con el objeto de facilitar el correcto mantenimiento y la continua actualización de los metadatos se elaboró una ficha con este fin. Esta se ha basado en la ficha que maneja el MARM para los metadatos del SIA, adaptándola y simplificándola para hacerla accesible a todas las personas que trabajan con información geográfica, sean especialistas o no.

Esta ficha se adapta al estándar mínimo propuesto por el NEM conteniendo los 27 campos obligatorios, por ello se puede usar esta ficha para la creación de nuevos metadatos, aunque se recomienda para las capas de nueva creación ceñirse, si es

posible, a la totalidad de campos existentes dentro de uno de los estándares. Esto se debe a que esta ficha no responde a ningún estándar, pero permite agilizar y asimilar el proceso de creación de metadatos.

Diagramas de los Procedimientos de Creación de Metadatos Analizados.

La metodología descrita se ha particularizado para cada una de las herramientas de edición de metadatos que se han utilizado para la realización de este estudio. Los diagramas que se presentan a continuación, muestran el proceso de elaboración y los pasos necesarios para su publicación en el Sistema de Catálogo Web de metadatos.

1. Desde ArcGis y CatMDEdit

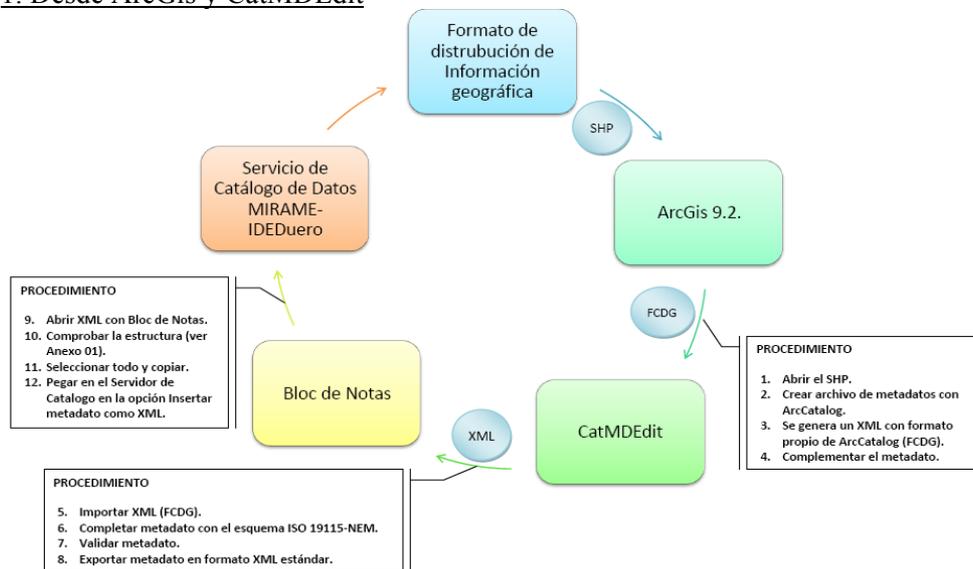


Figura 2: protocolo de creación de metadatos a partir de ArcGis y CatMDEdit.

Directamente desde CatMDEdit

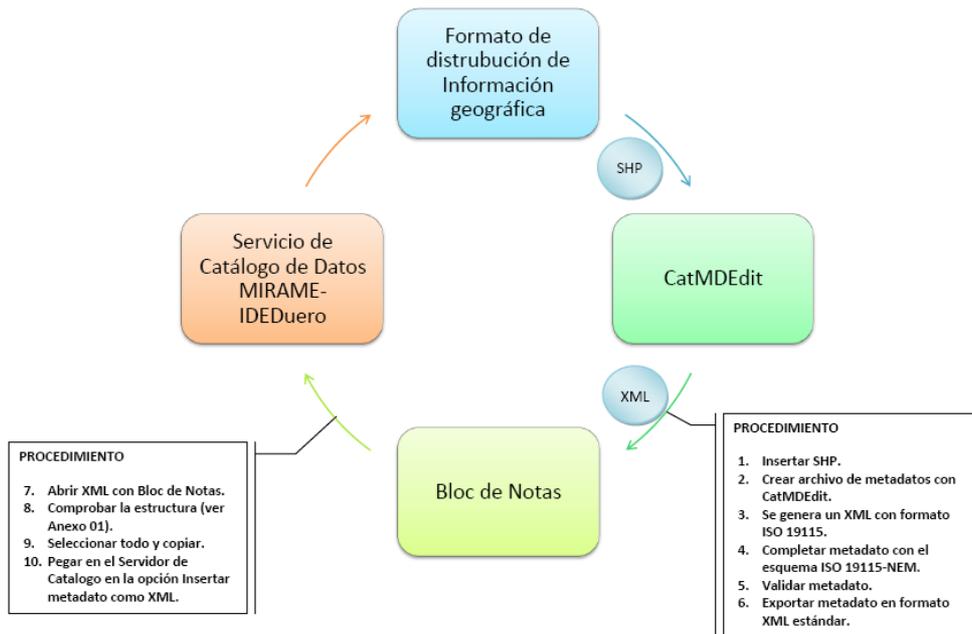


Figura 3: protocolo de creación de metadatos desde el CatMDEdit.

Desde gvSIG

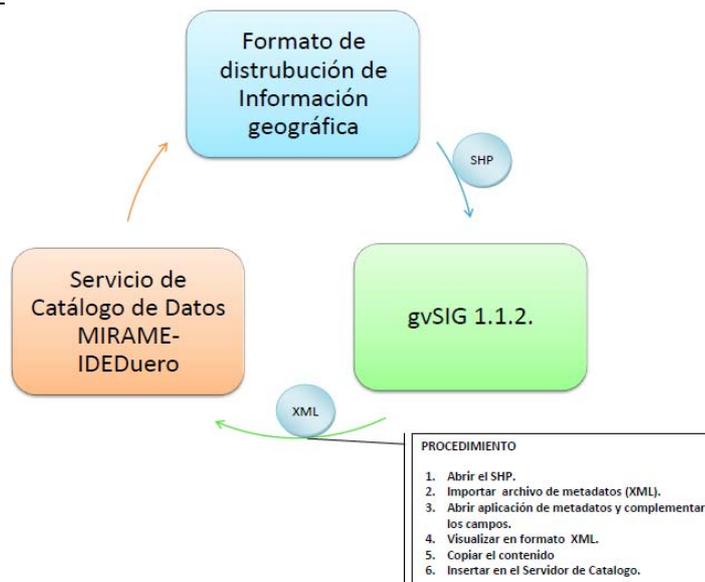


Figura 4: protocolo de creación de metadatos desde el gvSIG

Desde el catalogo Web de la CHD

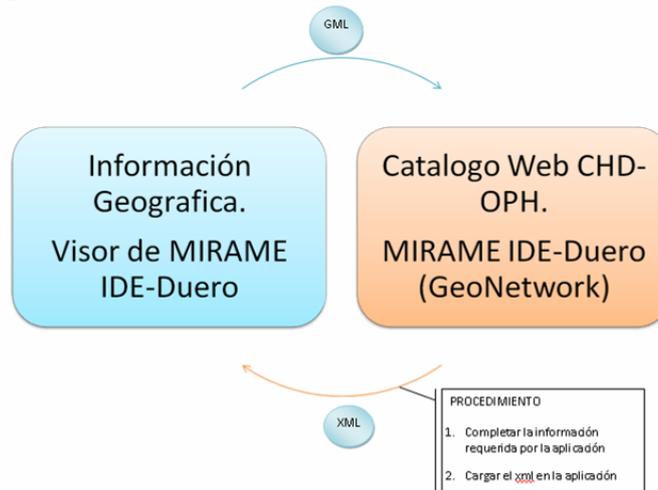


Figura 5: Protocolo de creación de metadatos desde la aplicación del Sistema MIRAME- IDEDuero (GeoNetwork).

CONCLUSIONES

1. La complejidad institucional dificulta el acceso a los datos. Este punto es importante cuando se quiere crear metadatos de capas que derivan de otras originadas por otros organismos. Afortunadamente, la progresiva creación de metadatos por parte de los principales creadores de cartografía está simplificando el problema.
2. Se considera fundamental, antes de comenzar a crear metadatos, acordar el estándar y fijar el núcleo a utilizar. Optimiza esfuerzos.
3. Las herramientas empleadas en este proyecto presentan diferentes niveles de complejidad, desde el más básico e intuitivo del gvSIG hasta el más completo y complejo de CatMDEdit. Es también importante decidir de manera realista es el más apropiado a las necesidades de la organización.
4. Las herramientas de elaboración de metadatos están todavía en la fase de asimilación del estándar ISO 19139, que otorga estructura a los campos del metadato, por lo que pueden existir problemas de incompatibilidad entre ellas.
5. El servidor de catálogo que utiliza la CHD se ha configurado como una herramienta útil y válida para la creación y mantenimiento de metadatos. Finalmente la experiencia del Duero ha sido satisfactoria.

6. La estrategia de cuestionarios es valida para la información de nueva creación, pero pierde eficacia de forma directamente proporcional a la antigüedad de las capas. Para este tipo de información se recomienda simplificar los cuestionarios al máximo, siempre dentro del estándar.
7. El constante proceso de gestión de la información geográfica exige que el proceso de metadataje sea continuo e iterativo.

Referencias

- [1] Página Web de la Confederación Hidrográfica del Duero: <http://www.chduero.es/>
- [2] Directiva 2007/2/CE del Parlamento Europeo y del Consejo de 14 de marzo de 2007 por la que se establece una Infraestructura de Información Espacial en la Comunidad Europea (INSPIRE).
- [3] UNE-ISO 19115. 2005. Versión española de la Norma Europea UNE-ISO 19115 de 2005, que a su vez adopta íntegramente la Norma Europea Geographic information, Metadata ISO 19115:2003. Comité técnico AEN/CTN 148 Información Geográfica Digital. AENOR, Madrid.
- [4] Guía de usuario NEM. 13 de Octubre de 2006. IDEE. Consejo Superior Geográfico.
- [5] Cima, J. (2008). WISE GIS guidance document – second edition. Geography Information System – Water Information System of Europe
- [6] Página Web del MÍRAME- IDEDuero: <http://www.mirame.chduero.es>
- [7] Página Web del The Federal Geographic Data Committee: <http://www.fgdc.gov>
- [8] Gould, M; Madrid, M; López, M. (2005). “Protocolo para la generación de metadatos espaciales”. Infraestructura de datos espaciales de la Generalitat Valenciana. Conselleria d’Infraestructures i Transport; y Universidad Jaume I.
- [9] Página Web The Dublin Core metadata
<http://dublincore.org/>
- [10] Vogt, J. and Working Group GIS. (2002). Guidance Document on Implementing the GIS Elements of the WFD.
- [11] UNE-ISO 19115-2 Geographic information --Metadata -- Part 2: Extensions for imagery and gridded data. Publicada en la fase 60.60
- [12] UNE-EN ISO 19119:2007. Información geográfica. Servicios. Versión española de la Norma Europea UNE-ISO 19119 de 2005.
- [13] UNE-ISO 19106:2004. International Organization for Standardization: Draft Technical Specification: Geographic information -- Profiles
- [14] UNE-ISO/PTDS 19139. International Organization for Standardization: Draft Technical Specification: Geographic information – Metadata – XML schema implementation. Technical. Committee 211 (2004).

- [15] UNE-ISO 15836:2003. International Organization for Standardization: Information and documentation - The Dublin Core metadata element set.
- [16] Página Web de LatinGeo:
http://www.latingeo.net/lab_geo/index.php?
- [17] ArcGIS Desktop Help. 9.2. Data management with ArcCatalog.
(<http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=welcome>)
- [18] Manual de usuario gvSIG 1.1 versión 3 en español. 2007. Conselleria d'Infraestructures i Transport.
- [19] Página Web de LocalGIS:
<http://161.67.130.130/localgis/>
- [20] Pagina Web de GeoNetwork opensource:
<http://geonetwork-opensource.org/>
- [21] Manual de Usuario CatMDEdit 4.0 Metadata Editor. (2008). TeIDE (Universidad de Zaragoza y GeoSpatiumLab S.L.).
- [22] Página Web AGROVOC:
<http://aims.fao.org/website/AGROVOC-Thesaurus/sub>
- [23] Página Web GEMET:
<http://www.eionet.europa.eu/gemet><http://www.ideo.es>