

Creación de Metadatos: Metodología y experiencia del Grupo de Catalogadores de la Información Geográfica

Criado, M ⁽¹⁾, Crespo, M ⁽¹⁾, Rodríguez, C ⁽¹⁾, Bravo, M ⁽¹⁾, Ballari, D ⁽¹⁾

(1) Laboratorio de Tecnologías de la Información Geográfica
Universidad Politécnica de Madrid
Autovia de Valencia km 7,5 Campus Sur UPM
913311968, metadatos@topografía.upm.es

Resumen

A pesar de que uno de los principales requisitos para el establecimiento de una Infraestructura de Datos Espaciales es contar con los metadatos de la Información Geográfica, esta no resulta una tarea sencilla. Uno de los motivos de esta dificultad reside en no contar con una metodología de creación de metadatos, que indique el camino a recorrer y asegure los resultados a obtener.

El Grupo de Catalogadores de la Información Geográfica perteneciente al Laboratorio de Tecnologías de la Información Geográfica de la Universidad Politécnica de Madrid, nace gracias a una iniciativa del Instituto Geográfico Nacional, y su misión se centra en superar el obstáculo que supone la falta de personal capacitado y metodologías en la creación de metadatos.

El principal objetivo del Grupo consiste en prestar apoyo a los productores de la Información Geográfica para la elaboración de los metadatos basados en la Norma ISO 19115 y el Núcleo Español de Metadatos, tanto en la Administración General del Estado, como en la Administración Regional y Local.

Así a lo largo de sus dos años de vida, el Grupo de Catalogadores ha dado soporte al Instituto Geográfico Nacional, Ministerio de Medio Ambiente (Banco de Datos de la Biodiversidad), Ministerio de Fomento (Centro de Estudios Hidrográficos - CEDEX, la Dirección General de Planificación y Coordinación Territorial y a la Dirección General de Carreteras), entre otros, siempre gracias al apoyo y financiación del Instituto Geográfico Nacional.

El objetivo del presente artículo es divulgar y acercar a toda la Comunidad IDEE:

- La metodología diseñada y comprobada, para que otros organismos o entidades puedan aplicarla.
- Los nuevos retos en los que se halla inmerso el Grupo de Catalogadores (generación masiva de los metadatos de las unidades de las distintas Series Cartográficas, metadatos para productos raster basados en la Norma ISO 19115-II, formación de expertos catalogadores, etc.)
- La experiencia adquirida tras dos años de trabajo relacionado con la creación de metadatos.

Palabras clave: Información Geográfica, Grupo de Catalogadores, Metadatos, Metodología para la creación de metadatos, Cuestionario de metadatos.

1 Introducción

Uno de los requisitos principales para el establecimiento de una Infraestructura de Datos Espaciales (IDE) es contar con metadatos que describan la Información Geográfica, siendo este un requisito esencial para localizar, describir y evaluar los datos disponibles [1].

A pesar del reconocimiento general de la importancia de los metadatos, su generación dista de ser una tarea sencilla y atractiva. Diferentes obstáculos se interponen a los creadores de metadatos, entre ellos [2]:

- Los estándares de metadatos son demasiado difíciles y extensos de implementar.
- La producción de metadatos requiere tiempo y otros recursos.
- Hay pocos beneficios tangibles e incentivos para producir metadatos.
- Falta de personal capacitado.
- Dificultad de utilizar las herramientas de generación de metadatos.

En base a estos obstáculos se han planteado distintas iniciativas para lograr su superación:[3]

Obstáculo	Iniciativa
Los estándares de metadatos son extensos	Núcleo español de Metadatos[4].
Los estándares de metadatos son complicados.	Guía de Usuario del Núcleo Español de Metadatos[5].
La producción de metadatos requiere tiempo y otros recursos	Diseño de una metodología de creación de metadatos basada en Cuestionarios
Falta de personal capacitado	Creación de un grupo de expertos Catalogadores de la Información Geográfica y Planes de Formación.
Dificultad de utilizar las herramientas de creación de metadatos	Desarrollo y mantenimiento de la aplicación Open Source CatMDEdit.

Tabla 1. Obstáculos e iniciativas relacionadas con metadatos de la geoinformación.

El presente artículo se centra en la iniciativa de diseñar una metodología para la creación de metadatos de la geoinformación. En una primera parte se presenta la metodología diseñada y sus componentes principales. A continuación se describe uno de los retos en el que se halla inmerso el grupo de catalogadores: la creación masiva de los metadatos de las unidades de series cartográficas. Seguidamente se presenta un primer acercamiento a la creación de metadatos de productos raster. Para finalizar se incluyen conclusiones y futuras líneas de actuación.

2 Diseño de una metodología para la creación de metadatos

La carencia de un método o procedimiento que ayude a la creación de metadatos es un factor que actúa como barrera. Por este motivo el Grupo de Catalogadores ha diseñado una metodología para la creación de metadatos, con el objeto de definir un conjunto de fases que guíe a los catalogadores, facilite la comunicación con los productores de datos e impulse la creación y publicación de metadatos dentro de un organismo.

Para llevar a cabo esta tarea se empleó el diseño de “**investigación-acción**”, el cual se caracteriza por efectuar el proceso de investigación con pasos en espiral al mismo tiempo que se interviene. El investigador (el catalogador en el presente estudio) se halla inmerso en la realidad que debe investigar y abordar, por lo que deja de ser un observador externo, para formar parte de la situación estudiada. Los pasos que se realizan en espiral son: Definición del problema, Formulación del plan, Implementación del plan y Retroalimentación[6]. La metodología que a continuación se describe es la obtenida una vez realizados estos pasos.

La metodología diseñada para la creación de metadatos se basa en tres pilares fundamentales:

- la participación de personal capacitado,
- la utilización de un método de recopilación de datos basado en un cuestionario y
- la definición de un proceso de creación de metadatos.

Esta metodología puede adaptarse para casos particulares de creación de metadatos como lo es la creación masiva de unidades de series cartográficas. Este caso se describirá más adelante.

2.1 Participación de personal capacitado.

El personal a intervenir en el proceso es:

- Productores o responsables de la Información Geográfica a catalogar. Poseen un profundo conocimiento sobre los procesos empleados en la creación de los datos, su propósito, contenido, calidad, etc. Serán los encargados de proveer la información necesaria que dará contenido a los metadatos. Estos responsables forman parte del organismo productor de los datos.
- Catalogadores de la Información Geográfica. Poseen un profundo conocimiento en metadatos de la Información Geográfica, normas, perfiles, recomendaciones y herramientas a utilizar para la catalogación. Serán los responsables de solicitar la información necesaria para crear un metadato, organizar la información proporcionada por los responsables de los datos y crear el archivo de metadato que será incorporado al catálogo de metadatos. Estas personas pueden ser parte del organismo o personal externo. En el caso particular de este estudio el grupo de catalogadores es externo al organismo, perteneciendo a la Universidad Politécnica de Madrid – Laboratorio de Tecnologías de la Información Geográfica.

2.2 Utilización de un método de recopilación de datos basado en cuestionarios.

El cuestionario diseñado transforma los rigurosos elementos de metadatos de la norma ISO 19115 y el Núcleo Español de Metadatos (NEM) en preguntas simples y familiares, con ejemplos que guiarán al responsable en el momento de ser completado [3].

De esta forma los responsables de los datos proveen información documental, sin necesidad de

pasar por los rigores de unos metadatos formales completamente estructurados [1], dejando esta tarea a los Catalogadores de la Información Geográfica

Esta forma de recopilar los datos para llevar a cabo la creación de los metadatos de Información Geográfica, ha sido utilizada por el USGS (United States Geological Services) [7] y el Clearinghouse de Uruguay [8].

El cuestionario puede ser personalizado y adaptado de acuerdo al tipo de datos (series, hojas, imágenes satélites, cartografía básica, capas de sistemas de información geográfica, etc.) y de acuerdo con el organismo a catalogar, para simplificar así la labor de su completado.

2.3 Definición del proceso de creación de metadatos.

El proceso de creación de metadatos está compuesto por un conjunto de fases sucesivas, que serán llevadas a cabo por el propio catalogador.

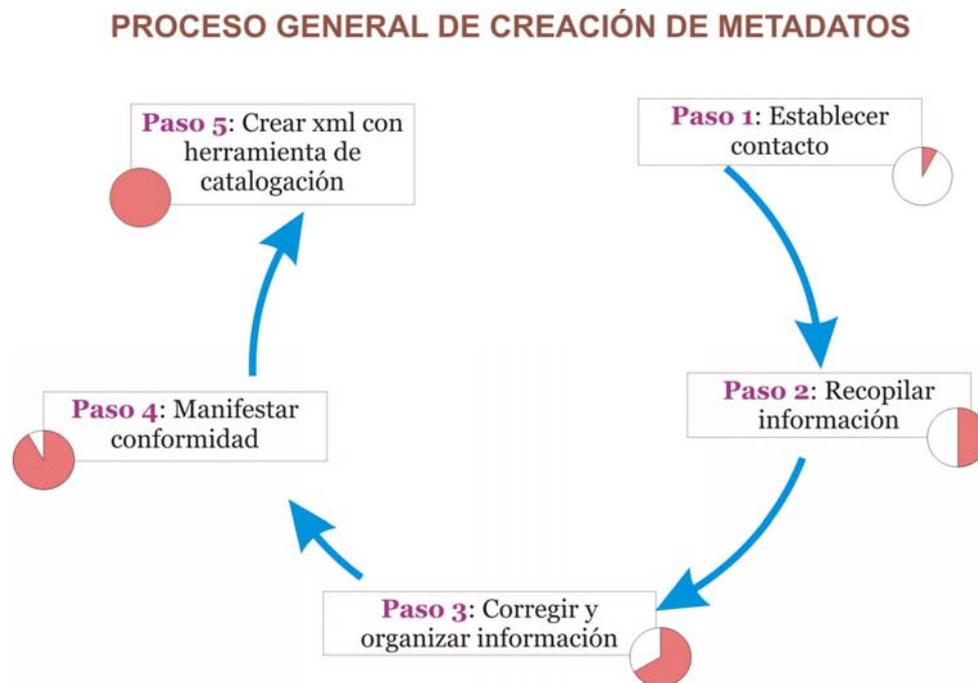


Figura 1. Proceso General de Creación de Metadatos

Paso 1: Establecer contacto con el responsable del conjunto de datos. En esta primera fase el catalogador toma contacto con el responsable de los datos, le explica el objeto de la entrevista, el proceso a seguir y la información que le solicitará, además de concretar una fecha para la realización de la entrevista del próximo paso.

Paso 2: Recopilar información. En este paso se realiza una entrevista en base al **cuestionario** para recopilar la información necesaria que completará el metadato. El cuestionario puede ser completado por el mismo responsable de los datos o por el catalogador, en el supuesto que se le facilite la información necesaria. Al concluir esta actividad la creación del metadato se encuentra en el punto medio del proceso (50%), ya que la mayor parte del volumen de información ha sido recopilada.

Paso 3: Corregir y organizar información. En esta etapa un experto catalogador revisa que el metadato cumpla con la norma ISO 19115, el Núcleo Español de Metadatos (NEM) y demás recomendaciones. En caso de ser necesario se solicita más información al responsable de los datos. Al culminar este tercer paso se estima que el 70% del proceso de creación del metadato ha sido realizado.

Paso 4: Manifiestar conformidad con el contenido. El responsable del producto debe revisar el metadato para dar su conformidad a la publicación de la información contenida en el metadato. En este punto el 90% del camino ha sido recorrido.

Paso 5: Crear XML del metadato con herramienta de catalogación. Por último es necesario generar un documento XML, para ello se utiliza la herramienta de catalogación CatMDEdit u otras.

Esta forma de trabajo permite seguir un proceso sistematizado para la creación de metadatos. Además permite realizar estimaciones sobre el grado de avance del proceso de creación de metadatos del organismo. Un ejemplo de los informes generados sobre el grado de avance de los metadatos de un organismo puede ser el siguiente:

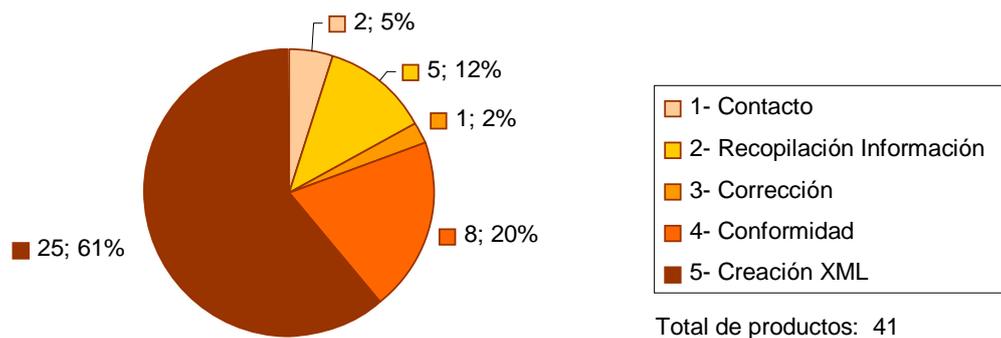


Figura 2. Ejemplo de informe del grado de avance de la creación de metadatos

2.4 Aplicación de la metodología

La metodología diseñada se ha aplicado en los siguientes organismos, limitándose en todos los casos al estudio de series cartográficas y capas de sistemas de información geográfica:

- 1- **Instituto Geográfico Nacional: 41 productos** pertenecientes a las series cartográficas del Mapa Topográfico Nacional (MTN), Modelo Digital del Terreno (MDT), Base Cartográfica Numérica (BCN), Atlas Nacional de España, Base de Datos de Líneas Límites (BDLL), Redes Geodésicas, entre otros. Estos metadatos se encuentran disponibles en el catálogo de metadatos de la IDEE.
- 2- **Ministerio de Medio Ambiente: 28 productos** entre los que se encuentran Propiedad Forestal, Red Natura ZEPA y LIC, Espacios Protegidos, Reservas de la Biosfera, Mapa Forestal de España, Inventario Nacional de la Biodiversidad, etc. Estos metadatos se encuentran disponibles en el Banco de Datos de la Biodiversidad del Ministerio de Medio Ambiente [9].

- 3- **Ministerio de fomento: 24 productos** pertenecientes a la Dirección General de Planificación y Coordinación Territorial, a la Dirección General de Carreteras, al Instituto Geográfico Nacional y al CEDEX. Estos metadatos formarán parte de la Infraestructura de Datos Espaciales del Ministerio de Fomento.

3 Creación de los metadatos de las unidades de las series cartográficas

Otro de los retos a los que se enfrenta el Grupo de Catalogadores es la generación masiva de los metadatos de las unidades de series cartográficas. En este caso la metodología a seguir presenta leves modificaciones respecto a la anteriormente descrita debido al volumen de información a manejar y a la necesidad de generar masivamente los metadatos.

El ámbito de ejecución de esta tarea es el Instituto Geográfico Nacional, el cual cuenta con aproximadamente 27.000 unidades correspondientes a 23 Series Cartográficas, entre ellas se encuentra el Mapa Topográfico Nacional (MTN), el Modelo Digital del Terreno (MDT), las Bases Cartográficas Numéricas (BCN), el Atlas Nacional de España (ANE), Mapas Diversos, etc.

3.1 Adaptación metodológica para la creación de unidades de series cartográficas

Las variaciones metodológicas, respecto a la metodología general, se centran en las etapas de recopilación de la información (Paso 2) y de creación de los archivos XML (Paso 5).

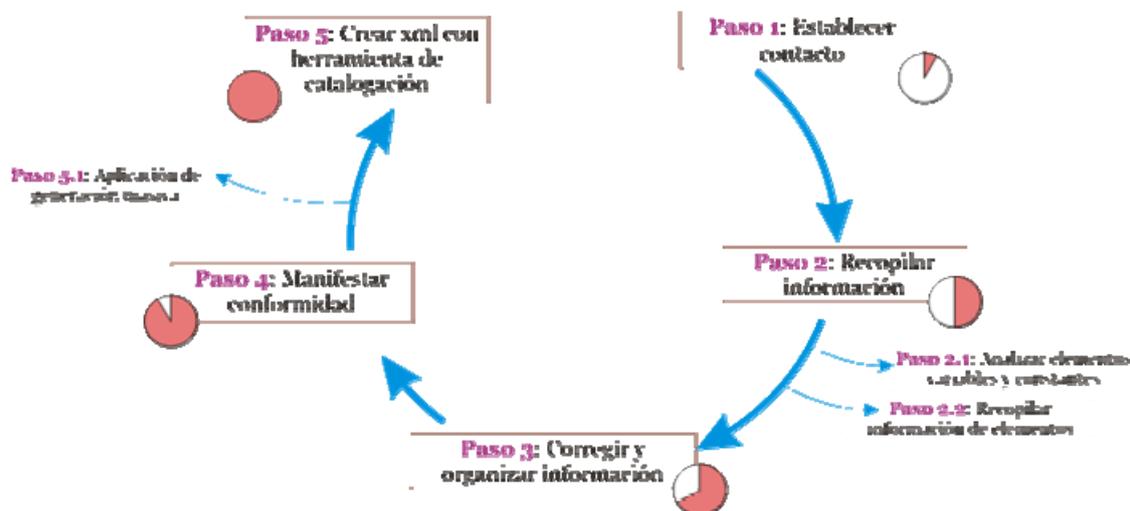


Figura 3. Adaptación metodológica para la creación de metadatos de las unidades de una serie cartográfica.

Paso 2: Recopilar información. Durante este proceso se desarrollarán las siguientes tareas:

2.1 Análisis de elementos constantes y variables.

Primeramente se debe detectar qué elementos de metadatos son variables en las unidades. Para ello el responsable del producto seleccionará, a partir del listado de

elementos de metadatos, cuáles son variables y cuáles permanecen constantes para las unidades de las series analizadas. Ejemplos de campos variables puede ser coordenadas extremas, fechas de creación o revisión, nombre de las unidades, etc.

Los *elementos constantes*, respecto de la serie cartográfica, heredarán su contenido del metadato de la serie. Para el caso de los *elementos variables*, la situación no es tan sencilla ya que existirán distintos tipos de elementos variables:

- a. *Elementos constantes entre unidades*: En este grupo se encuentran los campos para los cuales la información es constante para todas las unidades de la serie, pero variable respecto a la serie. Ejemplo: Nivel jerárquico, unidades de distribución, etc.
- b. *Elementos variables entre unidades*: En este grupo se encuentran los elementos para los cuales la información varía para cada unidad de la serie. En función de las características de cada elemento existirán:
 - i. *Elementos en los que varía todo el contenido*. Ejemplo: Fecha de referencia del conjunto de datos, extensión geográfica, extensión vertical, etc.
 - ii. *Elementos en los que varía parte del contenido*. Ejemplo: Título del conjunto de datos, resumen descriptivo, etc.

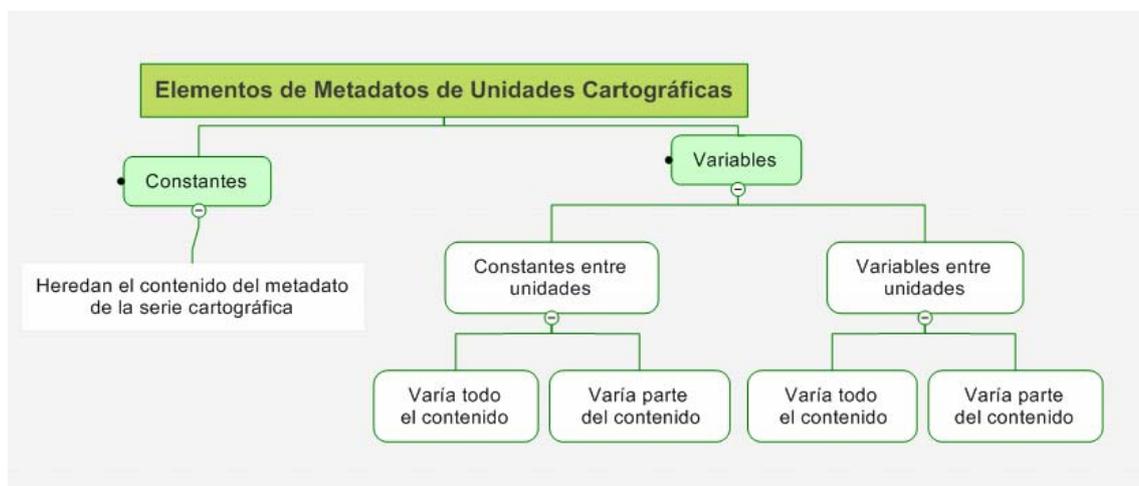


Figura 4. Esquema de tipos de elementos variables y constantes

2.2 Recopilar la información de los elementos.

Para que el Grupo de Catalogadores pueda disponer de la información necesaria para completar cada elemento variable de cada unidad, el responsable de los datos debe proveer dicha información al catalogador. Esta información se almacena en una base de datos Access, para su posterior utilización en el paso 5.

Paso 5: Crear xml de metadatos. Para llevar a cabo este último paso del proceso se ha desarrollado una aplicación informática que permite generar masivamente los metadatos de las unidades basándose en:

- el metadato de la serie en formato XML,
- la identificación de elementos de metadatos constantes y variables ,
- la información variable recopilada y almacenada en la base de datos.

La finalidad de la aplicación es evitar el trabajo manual y repetitivo que supone la creación de los metadatos de las unidades. Para obtener los archivos XML de todas las unidades que componen una serie, es necesario:

1. Seleccionar el formato XML en el cual se van a generar las unidades: ISO 19115 ó ISO 19139.
2. Cargar el archivo XML del metadato de la serie cartográfica, donde se indica mediante una *etiqueta* los elementos de información variable.
3. Cargar el fichero de la base de datos que recoge la información que dará contenido a las unidades.
4. Relacionar los elementos de metadatos variables con los campos de la base de datos.
5. Generar los archivos XML.

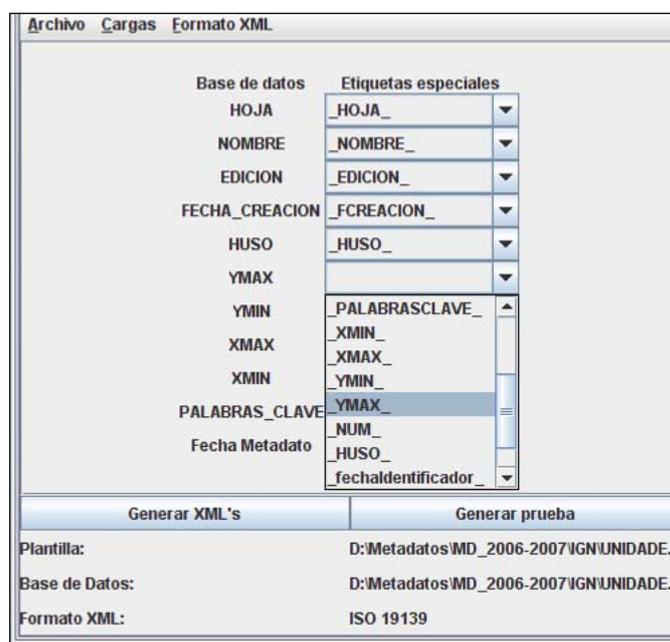


Figura 5. Interfaz de aplicación de creación masiva de metadatos de unidades cartográficas

4 Metadatos de Productos Raster

La Asociación Internacional para la Estandarización (ISO) se encuentra en proceso de revisión de la extensión de la norma ISO 19115 para datos raster e imágenes (imágenes de satélite, ortofotografías, modelos digitales, etc.). Esta extensión *ISO 19115-2 "Geographic Information-Metadata for imagery and gridded data"* [10] intenta suplir la carencia de elementos de metadatos de la norma ISO 19115, en lo referente a:

- las propiedades de los equipos de medición usados para adquirir los datos raster e imágenes,
- la geometría de los procesos de medición empleados por los equipos,
- los procesos de producción usados para obtener los datos raster e imágenes,
- los métodos numéricos y procesos informáticos usados.

En la siguiente imagen puede apreciarse paquetes de metadatos que la norma ISO 19115-2 añade sobre el modelo general de ISO 19115 [11].

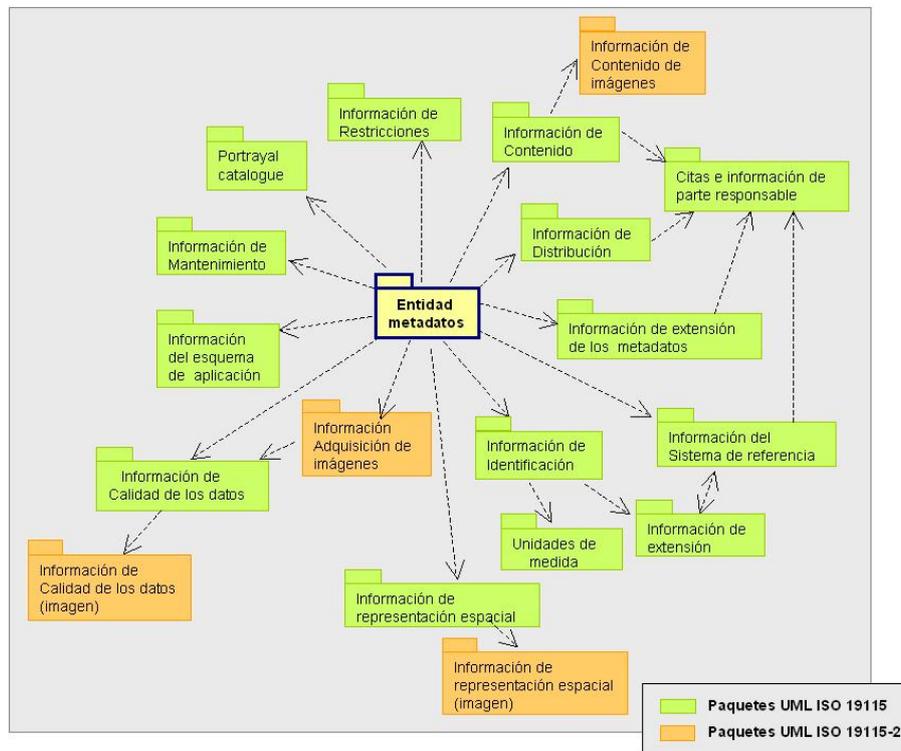


Figura 6: Paquetes de metadatos de la norma ISO 19115-2 y su relación con la norma ISO 19115.

A pesar que esta extensión aún no se encuentra aprobada como norma ISO, ni las principales herramientas de catalogación (CatMDEdit, ArcCatalog, IME, GeoNetwork, etc) implementan el actual borrador, el Grupo de Catalogadores se ha puesto en marcha para analizar la aplicabilidad de la extensión ISO 19115-II en un conjunto de datos de muestra.

Los datos que se están analizando son Cartoimágenes Provinciales, el Plan Nacional de Fotogrametría Aérea (PNOA), el Plan Nacional de Teledetección (PNT), el Corine Land Cover y el Sistema de Información sobre la Ocupación del Suelo en España (SIOSE). Actualmente se están analizando estos productos para determinar los elementos de mayor aplicabilidad de la norma ISO 19115-II. Además se están revisando los manuales y plantillas de metadatos de los proyectos PNOA, PNT y SIOSE que el IGN ha elaborado para ser enviados a las distintas Comunidades Autónomas y así asegurar la uniformidad en la creación de los metadatos.

7 Conclusiones y trabajo futuro

La metodología diseñada y aplicada ha permitido alcanzar el objetivo de creación de metadatos de los productos de Información Geográfica de varios organismos como son el Instituto Geográfico Nacional, Ministerio de Medio Ambiente y Ministerio de Fomento.

Esta metodología puede ser utilizada y adaptada para los casos particulares de otros organismos y supone una base a seguir para aquellos que están iniciándose en esta actividad.

El desarrollo y utilización de una aplicación informática para la creación masiva de los metadatos supone una disminución considerable del tiempo dedicado a la creación de los metadatos de unidades, tarea que sin esta aplicación seria inabarcable.

Como resultado del análisis de la creación de metadatos de productos raster basandos en la norma ISO 19115 e ISO 19115-2, se pretende identificar:

- Los elementos de metadatos de la extensión ISO 19115-II que presentan mayor aplicabilidad para los productos analizados,
- Posibles elementos que no están siendo contemplados por la extensión,
- Elementos claves que puedan se incluidos en la próxima versión del Núcleo Español de Metadatos – NEM.

Es necesario destacar que la no aprobación definitiva de la norma por parte de ISO, supone una barrera debido a las posibles variaciones que su contenido puede sufrir en el futuro.

En trabajos futuros el grupo de catalogadores centrará la atención en la catalogación de las unidades de series raster y en el rol de los metadatos en sensores.

Referencias

- [1] Global Spatial Data Infrastructure. “El Recetario IDE”. Capítulo Tres: Metadatos – Describiendo Datos Geoespaciales “ Versión 2.0. Enero 2004
- [2] Lynda Wayne. “Institutionalize metadata. Before it institutionalizes you.” Federal Geographic Data Committee. Noviembre 2005.
- [3] Ballari D, Sánchez Maganto A, Nogueras-Iso J, Rodríguez A, Bernabé MA, 2006a. Experiences in the use of an ISO19115 profile within the framework of the Spanish SDI. *Proceedings of GSDI-9. Santiago de Chile*, 6-10 nov. 2006, 10 p.
- [4] CSG, 2005. “SGTNEM_2005_01: Núcleo Español de Metadatos”. Infraestructura de Datos Espaciales Española. 2005. Consejo Superior Geográfico (CSG), Ministerio de Fomento, <http://www.idee.es/resources/recomendacionesCSG/NEM.pdf>
- [5] CSG, 2007. Guía de Usuario del Núcleo Español de Metadatos. Consejo Superior Geográfico (CSG), Ministerio de Fomento, <http://www.idee.es/resources/recomendacionesCSG/GuiaUsuarioNEM.pdf>
- [6] Hernández Sampieri R, Fernández-Collado C, Baptista Lucio P, “Metodología de la Investigación”4º Edición. 2006.
- [7] <http://geology.usgs.gov/tools/metadata/tools/doc/ctc/>
- [8]<http://cndg.clearinghouse.gub.uy/documentos/biblioteca/otros/Cursos/index-18.html>
- [9]http://www.mma.es/portal/secciones/biodiversidad/banco_datos/info_disponible/index.htm
- [10] ISO, 2005. ISO/CD 19115-2 Geographic Information –Metadata- Part 2: Metadata for imagery and gridded data. International Organization for Standardization (ISO). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39229
- [11] Sánchez Maganto A, Nogueras-Iso, Ballari D. Normas ISO sobre Metadatos ISO19115, ISO19115-2, ISO19139, ISO 15836 (Dublin Core). Pendiente de publicar en revista Mapping.