

Perfil de metadatos del Servicio de Teledetección de INTA

Domínguez Barroso A.⁽¹⁾, Amaro Cormenzana A.⁽¹⁾, de Miguel Llanes, E.⁽¹⁾

⁽¹⁾Servicio de Teledetección (INTA), Ctra. de Ajalvir, km.4, 28850 Torrejón de Ardoz, Madrid.
dominguezba@inta.es, amaroca@inta.es, demiguel@inta.es.

Resumen

Como resultado de la próxima implementación de la Directiva INSPIRE en Europa para la armonización en la distribución de la Información Geográfica se creó en España la IDEE, Infraestructura de datos Espaciales de España, que está coordinada por el Consejo Superior Geográfico. Este órgano consultivo del Gobierno está desarrollando entre otras actividades, el Núcleo Español de Metadatos, que pretende servir de referente en España para la generación de información sobre datos geográficos.

El Servicio de Teledetección del INTA como productor de datos de teledetección ha querido adaptarse a esta nueva normativa generando un perfil de metadatos tanto para imágenes aeroportadas como de satélite. Para lo cual en el año 2004 el INTA inició un proyecto autofinanciado en su totalidad.

Este documento recoge los elementos de metadatos de la norma 19115 que el Servicio de Teledetección del INTA ha incorporado a su perfil, y su comparación con las recomendaciones del NEM (Núcleo Español de Metadatos). También analiza las principales dudas que han surgido y las decisiones tomadas en este proceso.

1. Introducción

La Directiva europea INSPIRE que pretende regular y armonizar la distribución de Información Geográfica en Europa, ha dado lugar a la creación de las Infraestructura de datos Espaciales (IDE) en los distintos países de la Unión.

En España la IDE está coordinada por la Comisión de Geomática del Consejo Superior Geográfico (CSG), en el que están representadas diversas instituciones públicas, tanto nacionales como autonómicas, que proporcionan datos geográficos. En 2002 el CSG nombró un Grupo de Trabajo específico a su vez subdividido en varios subgrupos que desarrollan tareas encaminadas a la implementación de la IDEE, acorde con las directrices marcadas por INSPIRE. Uno de esos subgrupos está encargado de desarrollar un Núcleo Español de Metadatos (NEM), que recogerá una

recomendación de elementos de metadatos extraídos de las normas ISO 19115 y *Dublin Core Metadata*.

En el Servicio de Teledetección del INTA, como productor de datos de Teledetección tanto espacial como aeroportada, hemos querido desde el primer momento incorporar la nueva normativa sobre distribución de Información Geográfica, para lo cual hemos desarrollado un perfil de metadatos lo más estandarizado posible para los productos que generamos.

Hemos partido de la norma ISO 19115 y hemos tenido en cuenta las recomendaciones incluidas en la primera versión del NEM. El proceso nos ha planteado una serie de dudas y problemas debidos fundamentalmente a que esta norma ISO 19115 (*Geographic Information – Metadata*) está pensada para otro tipo de información geográfica, y la parte 2 (*Geographic Information – Metadata for imagery and gridded data*) dedicada a imágenes, aún no ha sido aprobada. Sin embargo, no hemos podido posponer esta tarea ya que los proyectos en los que estábamos participando, por una parte nos obligaron a proporcionar metadatos (BACCHUS y SPARC), y por otra nos hicieron ver los problemas que podía suponer no contar con una buena y completa información sobre los datos auxiliares (cartografía, mdt, etc.) para mejorar el proceso de los datos aeroportados (BACCHUS y DONANA).

2. Metodología

Nuestro objetivo era generar un perfil de metadatos que fuera de utilidad para proporcionar información a los usuarios de nuestros datos, tanto de aquellos proporcionados por CREPAD (Centro de Recepción, Proceso, Archivo y Distribución de datos de observación de la Tierra, situado en la Estación Espacial de Maspalomas) como de los datos aeroportados producidos en el laboratorio de Teledetección (LABTEL). Además este perfil tenía que cumplir los requisitos exigidos por la ISO 19115, en cuanto a contenido, y de la ISO 19139 en cuanto a estructura.

Por tanto, la metodología seguida consistió en:

1. estudio de la ISO 19115 y selección de los metadatos de interés,

2. estudio de la ISO 19139 y utilización del esquema XML propuesto hasta la fecha por dicha norma,
3. desarrollo de una aplicación de apoyo para la generación de perfiles de metadatos de acuerdo con la ISO 19139,
4. generación de un perfil de metadatos propio,
5. comprobación de nuestro perfil con el NEM,
6. y elaboración del perfil definitivo.

A lo largo de este proceso encontramos muchas dificultades que hemos intentado solucionar respetando al máximo las reglas establecidas por estas normas internacionales. A continuación se analizan estos problemas y cómo han sido resueltos.

3. Resultados

3.1. Selección de los elementos de la ISO 19115

Esta norma consiste en un núcleo de metadatos, denominado *Core*, que recoge 7 elementos obligatorios, 11 elementos opcionales y cuatro elementos de uso condicionado. En total 22 elementos considerados básicos por la ISO para proporcionar metadatos sobre un conjunto de información geográfica.

En concreto el *Core* proporciona información básica sobre los datos: localización, disciplina que abarca, resolución y acceso a los mismos; y sobre los metadatos: codificación, nombre de fichero, idioma, norma en la que se basa y contacto responsable de los metadatos. Por tanto, lo primero fue incorporar estos elementos a nuestro perfil.

El resto de la norma está constituida por una serie de entidades que proporcionan información más concreta sobre la identificación, el contenido, el área geográfica, la calidad, los procesos, las restricciones y la distribución de los datos, y que constituyen una lista de más de 400 elementos.

Para nuestro perfil seleccionamos las siguientes entidades:

- MD_Metadata->spatialRepresentationInfo para proporcionar información sobre el número de filas y columnas de las imágenes.
- MD_Metadata->referenceSystemInfo para establecer los parámetros de la proyección de las imágenes cuando están georreferenciadas.
- MD_Metadata->identificationInfo donde se describen características tales como propósito de los datos, aplicaciones, restricciones de uso y acceso, resolución o extensión geográfica y temporal.

- MD_Metadata->contentInfo que contiene una clase específica para descripción de imágenes y que permite incluir información sobre el parámetro contenido por los píxeles, sobre las bandas de la imagen (número, rango del espectro, número de bits por píxel, etc.) y también sobre las condiciones de adquisición de la imagen (iluminación, cobertura nubosa) y el nivel de proceso.
- MD_Metadata->distributionInfo que permite informar sobre el formato de distribución entre otras cosas.
- MD_Metadata->dataQualityInfo que incluye información sobre la precisión posicional y temática, así como sobre la fuente original de los datos y la descripción de cada uno de los procesos por los que han pasado hasta llegar al nivel actual. Esto es lo que se denomina linaje (*Lineage*).

Además hemos incluido aquellos elementos de la entidad MD_Metadata que permitan proporcionar información sobre los propios metadatos (nombre del fichero, versión de la norma y el esquema, responsable de los metadatos, etc.).

Las mayores complicaciones han surgido a la hora de completar un ejemplo, intentar compaginar la información útil para datos aeroportados con la de los productos de satélite proporcionados por CREPAD, e incluir información sobre la adquisición de datos aeroportados.

3.2. Aplicación de la ISO 19139 al perfil de metadatos

Esta norma denominada - *Dataset Geographic Metadata Implementation Model* - define la estructura estándar que deben tener los elementos de la ISO 19115. Su objetivo fundamental es proporcionar las bases para la implementación de un esquema XML que codifique los metadatos sobre la información geográfica.

En otras palabras, la ISO 19115 está constituida por una serie de paquetes UML (*Unified Modelling Language*) que contienen una o más clases UML, y la ISO 19139 desarrolla un modelo de implementación de esos paquetes y clases UML en un esquema XML (*Extensible Markup Language*).

Para la implementación de nuestro perfil hemos seguido la estructura de las diferentes versiones del esquema XML que proporciona esta norma en forma de fichero XSD (*Schema Definition Language*). Esta ha sido una tarea ardua ya que el esquema no sigue la estructura exacta de la ISO 19115 en cuanto a distribución de las entidades y de

los elementos dentro de ellas. Para facilitar esta tarea, generamos una herramienta (*IME – ISO Metadata Editor*) que incluye todos los metadatos de la norma ISO 19115 organizados según la estructura del esquema marcado por la ISO 19139 (ver Fig 1). Esta herramienta nos permite seleccionar los metadatos del perfil y generar un fichero XML que posteriormente puede ser validado con la última versión del fichero XSD.

En este caso los mayores problemas que hemos tenido han derivado de dos aspectos:

- El fundamental es que la norma ISO 19139 todavía está en desarrollo y aún no hemos podido contar con un esquema de implementación definitivo.
- Por otra parte hemos podido comprobar que en algunos casos el esquema contradecía la norma, no sólo en la ubicación de determinadas clases, sino en aspectos tan inequívocos como el número de veces que podía aparecer un metadato o un conjunto de metadatos. Esto nos hacía preguntarnos qué tenía prioridad sobre qué, y aún no hemos encontrado ningún foro que nos haya proporcionado una respuesta.

3.3. Comparación del perfil del INTA con el NEM

El NEM (Núcleo Español de Metadatos) pretende constituir una especie de *Core* de referencia para los productores de datos geográficos en España. Compone una serie de clases y atributos seleccionados de la ISO 19115 y del Dublin Core. Actualmente está en revisión la primera versión que ha sido desarrollada por el Subgrupo de Trabajo NEM establecido por la Comisión de Geomática del Consejo Superior Geográfico.

El NEM sigue siendo un conjunto mínimo de metadatos recomendados, pero ha servido para comprobar que el perfil que estábamos desarrollando seguía las pautas correctas, y para ampliar aspectos de los metadatos que no habíamos considerado importantes, tales como la información sobre el *dataset* agregado. También nos ha servido para entender el significado de algunas de las clases y poder elegir la que más se adecua a la naturaleza de nuestros datos. En el NEM hemos encontrado algunas carencias para la descripción de datos raster que hemos comunicado a los responsables para que consideren su incorporación.

3.4. Desarrollo de ejemplos y extensión del perfil

El resultado de todo el proceso aún está inconcluso debido a que algunas de las bases que hemos utilizado, como el esquema de la ISO 19139 y el NEM, no se encuentran en una versión definitiva. A pesar de esta situación, ya tenemos el perfil de metadatos acorde a las normas estándar en vigor, que nos permitirá proporcionar a los usuarios de nuestros datos una gran cantidad de información en ficheros XML. Además, en este momento estamos desarrollando un ejemplo para datos aeroportados y para datos de satélite y una plantilla que mejore la visualización de los metadatos mostrándolos en html.

En cuanto al desarrollo de los ejemplos, hemos encontrado elementos de la ISO 19115 cuyo dominio hace referencia a otras normas que, en algunos casos, no son fácilmente localizables o están sin terminar. Esto ha ocasionado que para un pequeño número de dichos elementos hayamos tenido dificultades para introducir la información. Es el caso de elementos tipo RecordType y AttributeName definidos en la norma ISO 19103 o la clase TM_Primitive definida en la ISO 19108.

También hemos querido incorporar información sobre la adquisición de los datos aeroportados o de satélite, por lo que hemos desarrollado una extensión de la ISO 19115 intentando mantener la misma filosofía que se está utilizando para la realización de la ISO 19115 parte 2, que describe los metadatos para imágenes, y la ISO 19130, que define la forma de describir sensores y modelos de datos. La entidad que hemos creado se denomina MD_Metadata->acquisitionInformation y está constituida por la clase IT_AcquisitionInformation, que a su vez incluye tres clases:

- IT_PlatformIdentification que incluye el nombre y la descripción de la plataforma
- IT_InstrumentIdentification que describe el tipo de sensor y que hemos desarrollado para los instrumentos que tiene el INTA y que están contemplados en la ISO 19130, que son ScanLinearArray (el sensor Multispectral Daedalus 1268 y el sensor Hiperespectral AHS) y DigitalFrameCamera (AMDC). Sobre estos tipos de sensores se incluyen elementos para introducir su nombre, descripción, tipo de colector, fov, ifov, ángulo de muestreo, velocidad de giro o tiempo de integración.
- IT_FlightInformation que permite describir la altura, dirección y velocidad del vuelo.

