

METADATOS: EXTRACCIÓN Y DERIVACIÓN AUTOMÁTICA DE ATRIBUTOS.

**MANSO CALLEJO, M.A.
BERNABÉ POVEDA, M.A.**

Universidad Politécnica de Madrid, ETSI en Topografía, Geodesia y Cartografía: Grupo Mercator

Autovía de Valencia Km 7, 28031 Madrid (España)

m.manso@upm.es
ma.bernabe@upm.es

RESUMEN

Para que una Infraestructuras de Datos Espaciales sea útil a una amplia comunidad de usuarios es necesario que existan catálogos o repositorios sobre los que poder realizar búsquedas o consultas relacionadas con información geográfica (I.G.) o con servicios Web que se apoyan en la I.G.. Para que los usuarios puedan realizar consultas simples o complejas, los catálogos deben de contener un inventario bien detallado de los productos y servicios ofrecidos por quien produce los datos, distribuye los productos o presta dichos servicios. Esta descripción detallada de productos o servicios son los Metadatos.

Uno de los componentes fundamentales de las IDES son los estándares, que actúan como mecanismos reguladores, cuyo objetivo es garantizar la interoperabilidad entre sistemas y aplicaciones. En el contexto de los Metadatos de productos el estándar Internacional ISO19115-2003 ha sido asumido por la Agencia Española de Normalización (AENOR) como estándar de metadatos para la I.G. Éste documento recoge de una manera formal los nombres de los atributos que pueden formar parte de un registro de Metadatos y sus significados.

ISO19115-2003 define aproximadamente 400 atributos organizados en clases y tipos de datos que cubren todas las facetas previstas para los metadatos: identificación, información espacial, sistemas de referencia espacial, calidad, distribución, aplicación y la referencia al propio estándar de metadatos usado en la descripción. ISO19115-2003 define cuales de los atributos son obligatorios, cuales son optativos y cuales están condicionados en función del tipo de I.G. que se esté describiendo. De esa enorme colección de metadatos, los países, regiones, comunidades y organizaciones pueden definir libremente subconjuntos (llamados perfiles o plantillas) de metadatos, más simples o más elaboradas, en función de sus necesidades. Este es el caso del Núcleo Español de Metadatos (NEM).

Independientemente de que las organizaciones utilicen el estándar completo o perfiles de metadatos, lo cierto es que el número de atributos que se deben cumplimentar para completar un registro de metadatos es elevado. Este hecho incrementa los costes económicos derivados del tiempo destinado a crear los metadatos. La administración, agencia o compañía que desee poner en marcha una IDE debe prever por tanto, recursos económicos iniciales para la creación de metadatos y debería prever también recursos humanos con conocimientos avanzados en el dominio del conocimiento (I.G.) que permitan crear metadatos de una forma manual y supervisada y mantener actualizados dichos metadatos.

Desde un punto de vista práctico, la creación de metadatos es un proceso lento, meticuloso, sistemático y no libre de errores. En la mayoría de los casos la información que se debe describir es digital y es necesario disponer de aplicaciones informáticas que puedan visualizar los datos, acceder a los atributos que deben de ser almacenados como atributos en el registro de metadatos, realizar transformaciones de coordenadas, interpretar los Sistemas de Referencia por Coordenadas usados, etc... A estas actividades se ha de añadir la búsqueda de información relativa a la razón por la que se generó el conjunto de datos, el pliego de prescripciones técnicas al que debería ajustarse el producto, los resultados de los controles de calidad aplicados a los productos, los procesos de revisión y mantenimiento previstos, etc... El personal especializado en la creación de metadatos, debe destilar toda esta información para trasladarla de forma legible y estructurada sobre un documento de texto utilizando un lenguaje de marcas (XML) o haciendo uso de una aplicación informática de edición.

Estos trabajos de edición de metadatos no están libres de errores. En unos casos los errores son meramente mecanográficos (olvidar un separador decimal, cambiar de lugar la coma), en otros casos los errores son por conmutación de valores entre atributos duales (ancho-alto, este-oeste, norte-sur) y finalmente también se pueden interpretar erróneamente los Sistemas de Coordenadas y como consecuencia las transformaciones de coordenadas serán erróneas. Los errores mencionados ocurren tanto en los documentos en papel como en los documentos digitales ya que en ambos casos han de buscarse los datos y transcribirlos sobre la herramienta de edición de metadatos.

Desde el punto de vista productivo, otro aspecto importante es el tiempo que utiliza un operador y por lo tanto, el coste económico que supone utilizar una herramienta informática para abrir el documento digital que se está describiendo, leer los datos necesarios, transcribirlos sobre el editor de metadatos y procesar algunos cálculos (transformación de coordenadas).

El resto del documento se estructura de la siguiente forma:

Se presenta un estudio realizado sobre los errores tipográficos cometidos en la edición manual de metadatos en lo referente a las extensiones geográficas y los Sistemas de Referencia por Coordenadas.

Se presenta un estudio realizado para determinar el tiempo, y por tanto coste de operador, ocupado en la recuperación de los atributos de la I.G. almacenados en los propios soportes digitales de almacenamiento.

En tercer y último lugar se presenta un estudio en el que se comparan y cuantifican los atributos de la I.G. almacenados en los archivos o bases de datos desde un punto de vista teórico de formato de archivo, con los datos que pueden extraerse usando herramientas informáticas de distintos tipos (ArcCatalog, Gdal/OGR, otras de desarrollo propio).

Para finalizar el documento se presentan las conclusiones, los agradecimientos y las referencias bibliográficas.