

Un geoprocesamiento para la publicación de metadatos geográficos

Sergio Trilles Oliver, José Gil Altaba, Laura DíazSánchez,
Joaquín Huerta Guijarro

Institute of New Imaging Technologies
UniversitatJaume I

{strilles,ldiaz,jose.gil,huerta}@uji.es

Resumen

En la actualidad los metadatos son una de las claves más importantes para el descubrimiento de la información geoespacial publicada en las Infraestructuras de Datos Espaciales (IDE). En este trabajo se presenta un mecanismo de generación y publicación de metadatos mediante un geoproceso con interfaz estándar *Web ProcessingService* (WPS). El proyecto nace en el marco de la directiva europea INSPIRE como solución a la publicación de datos geoespaciales y sus metadatos de forma automática.

Palabras clave: metadatos, asistencia automática, geoprocesamiento, descubrimiento, INSPIRE.

1 Introducción

Hoy en día en las Infraestructuras de Datos Espaciales (IDE) la existencia de los metadatos de los recursos es necesaria para descubrirlos a través de sus características [2]. Por lo que es necesario, satisfaciendo con la directiva INSPIRE [1], generar y publicar metadato por cada dato que se publique en la infraestructura. Estos metadatos generados deben ser validados (cumplir con los estándares) y describir fielmente al dato referido.



Este trabajo tiene como objetivo llenar un vacío en el proceso de tratamiento y publicación de datos geográficos, que es el de la publicación y generación automática de los metadatos asociados.

A fecha de hoy no existe un mecanismo automatizado trivial, que no requiera una alta participación por parte de los técnicos en el proceso. El objetivo perseguido es poner al alcance de todos, un mecanismo que ofrezca de forma fácil la publicación de los datos y sus metadatos. Utilizando para ello servicios de procesamiento geoespacial basados en estándares existentes en las IDE.

2 Estado del arte

Los metadatos son creados por los proveedores de datos y almacenados en catálogos siguiendo los estándares dictados por organismos internacionales como ISO[4] que son adoptados como estándares de facto por la comunidad. La finalidad de dicho proceso es que un usuario pueda encontrar un dato concreto en entornos distribuidos.

El problema de los metadatos recae, principalmente, en su creación y mantenimiento para que el dato sea encontrado de forma efectiva. Hoy en día la creación de los metadatos presenta poca sincronización con el de la creación y publicación de datos, además de una alta complejidad para automatizar la extracción de características del dato y así generar el metadato [5][6][18].

Existen principalmente tres técnicas para la generación de metadatos que son manual, automática y mixta [7][8][9]. Asimismo otros autores proponen cinco métodos[10]: de forma manual, ampliando la información almacenada con los valores obtenidos a través de consultas, utilizando mediciones automáticas y observaciones, mediante datos extraídos y calculados, y finalmente, deducidos a partir de otros elementos.

Para estudiar el estado del arte dividiremos en nuestro caso la generación de metadatos en dos tipos, que son por extracción y por deducción [11].



En primer lugar hablaremos de la generación de metadatos por extracción, ya que el trabajo desarrollado está dentro de este tipo. En el mundo del GIS (Geographic Information System) se pueden encontrar diferentes contenedores de información geográfica como son los ráster y los vectoriales. Estos tipos se caracterizan por tener una estructura claramente definida lo que facilita la extracción de los metadatos. Por otro lado, las pocas herramientas GIS existentes que ofrecen una deducción automática de metadatos para los contenedores ráster y los vectoriales se basan en el análisis de estos formatos específicos, y la aplicación de mecanismos *ad-hoc* que procesan los datos para extraer la información que se utiliza posteriormente en la construcción de los elementos de metadatos [12]. Herramientas como CatMDEdi [13] se utilizan para realizar una extracción automática de metadatos en diferentes formatos. Como se refleja en el trabajo realizado por [14], la cantidad de información que puede extraerse dependen fundamentalmente del modelo de representación utilizado y su propio formato de archivo. De esta manera, hay elementos que sólo pueden ser extraídos de ciertos tipos de datos y archivos, mientras que otros, como el tamaño de los datos, pueden obtenerse en cualquier circunstancia.

Otra herramienta muy conocida y utilizada para la generación automática de metadatos de datos geográficos es ESRI ArcCatalog. Esta herramienta permite la carga automática de un número de campos básicos y la actualización de la sincronización de datos y metadatos. Para mejorar esta herramienta se han creado algunas extensiones, tales como el Editor de metadatos del NEM (Núcleo Español de Metadatos), es una herramienta totalmente integrada con el ArcCatalog, capaz de generar un registro de metadatos que cumpla con el estándar ISO19115:2003 y NEM v1.

Para nuestro trabajo hemos utilizado el estándar WPS (Web Processing Service), estos fueron diseñados para estandarizar la forma de ofrecer cálculos GIS a través de Internet. Este estándar proporciona reglas para describir cualquier cálculo (proceso), así como la forma de realizar peticiones al servicio y la respuesta que ofrece. La especificación de los WPS fue publicada por OGC en 2004 y en 2007 la versión 1.0.0 [19].

Un WPS proporciona acceso a operaciones o cálculos de datos geoespaciales, de distinta complejidad, mediante Servicios Web como interfaces. Puede contener operaciones para tratar tanto datos vectoriales como matriciales, provenientes de la red o del propio servidor.

3Contexto

El trabajo descrito se enmarca y extiende el GeospatialService Factory (GSF) [2] [3], un servicio de publicación que aborda la problemática de facilitar la provision de contenido y el mantenimiento de las IDE por parte de todos los participantes. El La investigación y el desarrollo actual del GSF proporciona un component interoperable que en la forma de servicio basado en la especificación estándar WPS del OGC ofrece la funcionalidad de publicar datos geográficos en servicios estándar, marcados por la directiva europea INSPIRE y desplegarlos en las IDE. GSF propone una solución para la publicación de contenido geoespacial orientada al consumidor de datos. Facilitando la publicación automática de los mismos y proporcionando una herramienta de colaboración para el usuario final.

3.1 Requisitos de usuarios IDE

Considerando el ciclo de vida de todo dato, disponible en un sistema de información, formado por cuatro etapas: publicación, descubrimiento, acceso y procesamiento [2] GSF viene a mejorar la etapa de publicación de los datos informáticos, simplificando y automatizando el proceso de publicación. La publicación forma parte del flujo natural de trabajo para el tratamiento de información. Tras el descubrimiento, acceso y/o procesamiento de los datos, es necesaria la publicación del nuevo contenido generado, para poder ser reutilizado por otros interesados en la IDE.

3.2 Extendiendo una IDE con un servicio de Publicación

El resultado del aplicar GSF sobre los datos devuelve como respuesta los puntos de entrada asociados a los servicios de descarga, visualización y descubrimiento en cumplimiento de la directiva INSPIRE. Esto hace que se enmarque dentro del estándar y ofrezca una alta escalabilidad. La Figura 1 muestra en que parte se encuentra el proyecto desarrollado dentro del contorno GIS.

GSF está implementado como un algoritmo con interfaz estándar de servicios OGC, en concreto como especificación WPS. Y construido estructuralmente como patrón factoría. Permite fácilmente incrementar su funcionalidad e iteroperabilidad, mediante la incorporación de nuevos servicios (elementos de la factoría) a la envoltura de ejecución WPS.



Figura 1. Flujo de trabajo de usuarios IDE

Los elementos factoría utilizados [17] tienen la capacidad de crear objetos del mismo tipo y facilitar la ampliación.

En nuestro caso el tipo de objetos comparten un tema común que son los servicios GIS. De esta forma utilizamos la herencia para definir los estándares de OGC, más específicamente los servicios definidos en INSPIRE y GIS. De acuerdo a los documentos de reglas de implementación correspondientes a estos tipos de servicios se debe utilizar para su implementación la norma EN ISO 19128:2005 Web MapService (WMS) para servicios de visualización y Web FeatureService (WFS) para descarga.

El GSF tiene asociado un perfil de publicación o ServicePublicationProfile (SPP), que se compone de la información necesaria para la publicación y generación de metadatos. El SPP es un documento XML que puede ser enviado como parámetro.

4 Generación y Publicación de metadatos

El actual trabajo tiene como objetivo definir un módulo que sea capaz de generar y publicar metadatos, dentro de GSF. La Figura 2 muestra el diseño del sistema a grosso modo y cómo se relacionan los principales módulos.

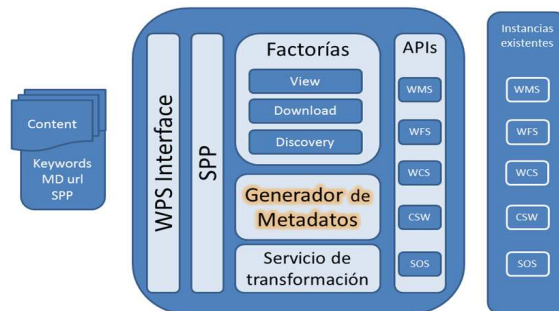


Figura 2. Componentes del GEOSS Service Factory

Para poder generar el metadato es necesario conocer información del dato y del sitio donde está publicado. Parte de esta información la podemos extraer de los servicios en los que se ha publicado el dato. Estos son principalmente servicios de datos disponibles en las IDEs. Estos servicios conformes con las reglas de implementación de INSPIRE, son servicios que implementan interfaces de OGC como WMS (para visualización) y WFS, WCS y SOS (para descarga de datos). Otra información que será más difícil de conocer, es por ejemplo las palabras clave, por eso diseñamos GSF con diferentes parámetros de entrada, como las palabras clave, que pueden ser indicadas por el usuario.

De forma general, el generador de metadatos tendrá que actuar cuando se haya publicado el dato, si no se le indica lo contrario. La factoría encargada de trabajar con el generador de metadatos es la *Discovery Factory*.

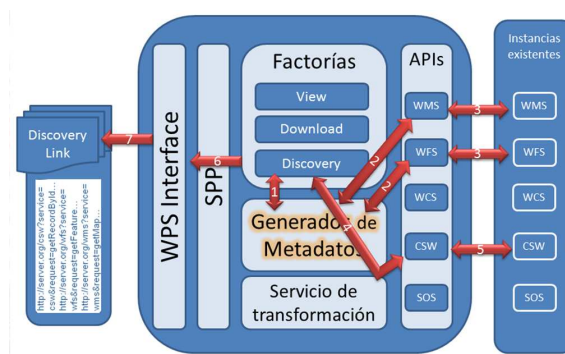


Figura 3. Secuencia de pasos para la publicación y generación de metadatos en el GSF

El primer paso del generador será consultar los servicios WMS y WFS para obtener la información necesaria para crear el metadato. Esto se realizará mediante las peticiones oportunas. Una vez consultada la información necesaria se aplicará el proceso de generación del metadato y procederá a la publicación en el catálogo (CSW). Finalmente el WPS devolverá el enlace resultante de la publicación. Este funcionamiento se puede ver en la siguiente Figura 3.

Cuando se termina de publicar un dato, si el WPS no ha recibido el parámetro MD_URL con un metadato, se llama al generador de metadatos. Éste como parámetros de entrada, necesita por parte del algoritmo de publicación, las urls a los servicios publicados para generar las peticiones, además de la entrada de las palabras clave. Como único parámetro de salida del GSF tenemos una URL de acceso al metadato del servicio de catálogo donde se haya publicado.

Para explicar con más detalle el proceso de generación de metadato, tenemos la imagen siguiente (Figura 4a) donde se aprecia las entradas y salidas del módulo *Metadata Generator*. La Figura 4b muestra los pasos involucrados.

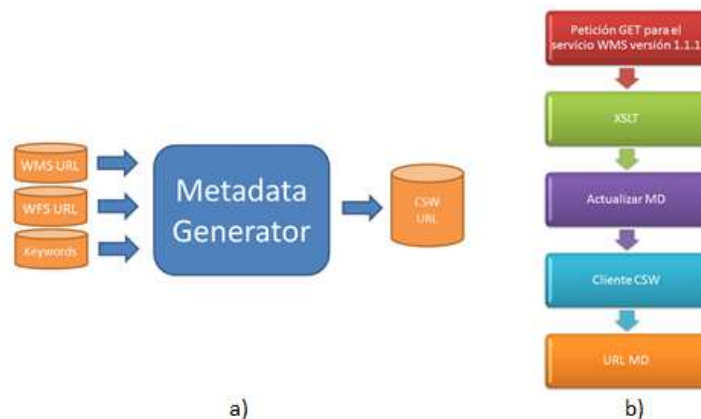


Figura 4. a) Parámetros de entrada y salida del generador de metadatos b) Secuencia de pasos del generador de metadatos

En el primer paso, se realiza una petición GetCapabilities a los servicios de datos donde el dato ha sido previamente publicado. Por ejemplo si el dato es un dato vectorial que ha sido publicado para visualización en un WMS y para su descarga en un WFS, el Generador de Metadatos envía primero un GetCapabilities al servicio WMS. Esta petición devuelve un XML con las características del dato que serán necesarias para la creación del metadato. Segundo, hace lo mismo con el servicio WFS para aportar más información al metadato, este trabajo está en curso en el prototipo actual.

El siguiente paso es aplicar la plantilla XSLT (Extensible Stylesheet Language Transformations) de generación de metadatos que hemos implementado. La transformación tendrá como entrada el XML obtenido de la petición anterior. El resultado que se obtiene tras la transformación es otro documento XML que forma el metadato cumpliendo los estándares ISO 19139 e INSPIRE.

El tercer paso consiste en parsear el XML del metadato para rellenar campos que no han sido posibles llenarlos en la transformación, como keywords o urls a los servicios. Tras este paso el metadato se considerará completado.

A continuación se publicará el metadato en el catálogo que esté definido en el SPP. Este paso se realiza de forma análoga a la publicación de datos pero en este caso al servicio de Catálogo, que de nuevo y siguiendo las reglas de implementación serán normalmente servicios que implementan el interfaz OGC CSW. En nuestro trabajo utilizamos el perfil transaccional (CSW-T) de acuerdo con la especificación de servicios de catálogo del OpenGIS. De esta forma, se tendrán que implementar todas las operaciones que permite el protocolo CSW-T. Éstas son: GetCapabilities, GetRecords, GetRecordById, GetDomain, DescribeRecord y las transaccionales.

El último paso, es obtener la URL que identifica el metadato publicado en el servidor. Tras obtener la URL el generador de metadatos se considerará como finalizado.

Para la generación del metadato se ha optado por crear la plantilla mediante XSLT o Transformaciones XSL. Esta tecnología es un estándar de la organización W3C (World Wide Web Consortium) que nos permite una forma de transforma documentos XML o incluso que no lo sean. Como ya lo hemos dicho un metadato no es nada más que un XML siguiendo un estándar, por lo que podemos aplicar este tipo de transformaciones.

Debemos mencionar que el metadato que se genera no tiene todos los campos de la especificación ya que algunos no se pueden obtener de forma automática. Se han elegido los campos que podemos ver en la siguiente figura (Figura 5). En ella aparecen los principales campos que contendrá el metadato generado.

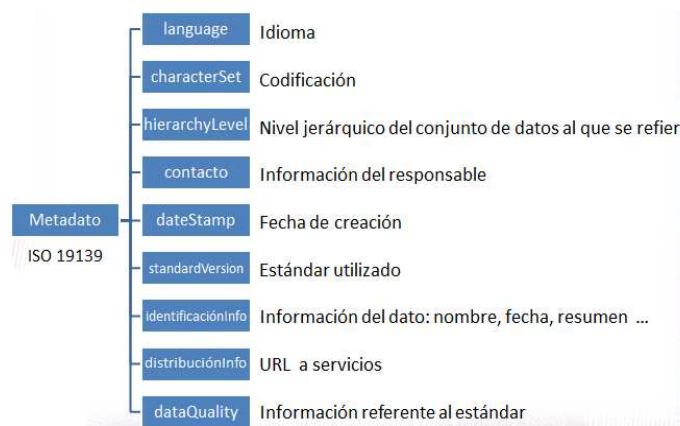


Figura 5. Algunos campos del metadato

5 Experimentación

Para poder realizar la experimentación, únicamente nos hace falta un dato geográfico en formato Shapefile como entrada del servicio WPS. Para ejemplificar el proceso se ha tomado un dato geográfico con los incendios de un año producidos en la provincia de Castellón. El análisis de impacto medioambiental producido por dichos eventos, representa un caso de estudio típico de interacción entre distintos usuarios.

Para que cualquier información geográfica esté disponible para todo el mundo, es necesario que esta esté publicada en un servidor de datos geográficos y que además tenga un metadato que describa el dato y que este sea publicado en un servidor o catálogo de metadatos.

Actualmente el mayor servidor de datos geográficos, es el llamado Geoserver. Siendo este el más usado. Por eso hemos elegido la especificación Geoserver en nuestro servicio de publicación. Por otra parte, el catálogo de metadatos más importante es GeoNetwork. Por lo que también lo hemos elegido para la publicación del metadato.

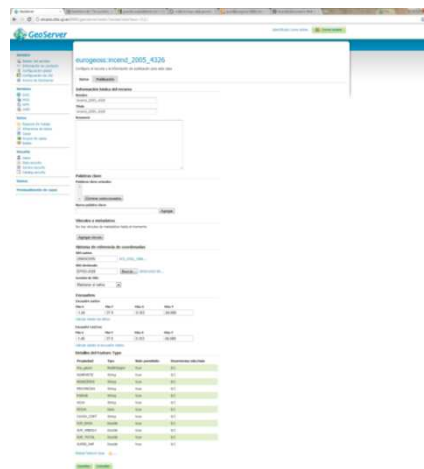


Figura 6. Captura de la información del dato publicado en Geoserver

Se ha dado una posible solución al problema de disponibilidad de datos y metadatos y lo que es más importante su enlace, de forma que cada vez que publicamos un dato y/o metadato estos están enlazados para mejorar no solo la visibilidad y descubrimiento del dato si no también su acceso.

Como trabajo futuro, se han definido diferentes ramas de expansión. La primera de ellas es continuar el trabajo actual para generar un cliente que facilite la ejecución del WPS, tanto en dispositivos fijos como móviles. Otra línea de investigación es incrementar los elementos de los metadatos generados de forma automática.

Agradecimientos. Este trabajo ha estado parcialmente subvencionado por el proyecto España Virtual y el proyecto europeo Eurogeoss, además de la Universitat Jaume I.

Referencias

- [1] INSPIRE EU Directive. 2007. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union, L 108/1, Volume 50, 25 April 2007. <http://eurlex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML>.
- [2] Díaz, L., Granell, C., Gould, M., Huerta, J.: Managing user generated information in geospatial cyberinfrastructures. Future Generation Computer Systems, 2011.
- [3] Díaz, L., Schade, S.: Geoss service factory: Assisted publication of geospatial content. Proceedings 14th AGILE International Conference on Geographic Information Science, 2011.
- [4] International Organization for Standardization, <http://www.iso.org>

- [5] Bulterman, D.: Is it time for a moratorium on metadata? IEEE Multimedia, pages 10–17, 2004.
- [6] Nogueras-Iso, J., Bernabé, M.A., Manso, M.A.: Automatic metadata extraction from geographic information. Automatic metadata extraction from geographic information, pages 379–385, 2004.
- [7] Campbell, T.: Fostering a culture of metadata production. GSDI10: Tenth International Conference for Spatial Data Infrastructure, 2004.
- [8] Barton, J. et al. Currier, S.: Quality assurance for digital learning objects repositories: issues for the metadata creation process. Research in Learning Technology, 2004.
- [9] JORUM. The jisc online repository for [learning and teaching] materials: Jorum scoping and technical appraisal study. Volume V: Metadata, 2004.
- [10] Beard, K.: A structure for organizing metadata collection. third international conference. Workshop on Integrating GIS and Environmental Modelling, 1996.
- [11] Beltran, A., Díaz, L., Granell, C., Huerta, J., Abargues, C.: Description and Publication of Geospatial Information. Universitat Jaume I. Spain.
- [12] Nogueras-Iso, J., Bernabé, M.A., Manso, M.A.: Automatic metadata extraction from geographic information. Automatic metadata extraction from geographic information, pages 379–385, 2004.
- [13] Nogueras-Iso, J., Torres, J., Muro-Medrano, M.P., Zarazaga-Soria, P.R., Lacasta, F.J.: A java tool for creating iso/fgdc geographic metadata. IfGI prints, 2003.

- [14] Taussi, M.: Automatic production of metadata out of geographic datasets. University of Technology, 2007.
- [15] Smyth, P., Hand, D., Mannila, H.: Principles of data mining. MIT P, 2001.
- [16] Lutz, M., Klien, E.: The role of spatial relations in automating the semantic annotation of geodata. In Proceedings of the Conference of Spatial Information Theory, 2005.
- [17] Erich Gamma, J., Helm, R., and Vlissides, J.: Design patterns: Elements of reusable object-oriented software. Addison-Wesley Professional Computing Series. Addison-Wesley Publishing Company, 1995.
- [18] Arturo Beltrán, Christian Martín. Generando descripciones de recursos para gvSIG desde GeoCrawler. 6th gvSIG User Conference (gvSIG 2010). Valencia, Spain, November 2010.
- [19] Schudt, P., (ed) 2007. OpenGIS Web Processing Service Version 1.0.0, Open Geospatial Consortium. <http://www.opengeospatial.org/standards/wps> (accessed 4 Sep. 2008)