

Extracción semiautomática de metadatos: hacia los metadatos implícitos

L. Díaz, C. Granell, A. Beltrán, A. Llaves, M. Gould

Centro de Visualización Interactiva, Departamento de Lenguajes y Sistemas Informáticos.
Universitat Jaume I, Avda. Vicent Sos Baynat, s/n. Castellón, laura.diaz@uji.es,
gould@uji.es

RESUMEN

Esta ponencia explica el trabajo realizado hasta la fecha en la extracción semi-automática de metadatos geoespaciales dentro de la plataforma cliente gvSIG. Mostramos el ciclo completo de extracción parcial de la fuente de datos, la adición de elementos subjetivos a través de un editor integrado en gvSIG, y la publicación semiautomática de los metadatos, en formatos estándares, a un servicio de catálogo basado en GeoNetwork. Posteriormente hablamos de la nueva versión del gestor de metadatos, que se ajusta a una nueva arquitectura interna de gvSIG. Cerramos con ideas para llegar a capturar más metadatos de forma implícita, objeto de futuras fases del proyecto.

Palabras clave: *software libre, metadatos implícitos, extracción metadatos, infraestructura de datos espaciales, geoindexación, gvSIG.*

INTRODUCCIÓN

Partimos de la base que los metadatos ya son aceptados como una pieza clave de una IDE, según su arquitectura actual de tres capas, y que los metadatos alimentan al servicio de catálogo ubicado en la importante capa intermedia [2]. La creación y publicación de metadatos según reglas (INSPIRE) y estándares (ISO) internacionales vigentes es sin duda la manera más segura de llegar a una IDE operativa, por lo menos en lo que llamamos una *IDE versión 1.0*. Creemos, sin embargo, que las metodologías y arquitecturas *status quo* –que definen la IDE de hoy– no sean suficientes para la masiva escalabilidad que necesitaría la IDE en su versión 2.0 (siguiendo la terminología web 2.0). La IDE 2.0 necesitaría guardar y dar acceso no sólo a los geo-recursos oficiales de proveedores convencionales (agencias cartográficas), sino también a las contribuciones de unos 6000 millones de sensores, de millones de seres humanos moviéndose por el mundo potencialmente portadores de sensores como por ejemplo receptores GPS, cámaras, telefonía y demás electrónica (Goodchild [3]). Los números realmente son mucho mayores, porque Goodchild no tiene en cuenta los otros innumerables micro-sensores y nano-sensores que, repartidos como polvo por todo el planeta, tendrán la capacidad de provisión de

datos con componente espacial. En la prevista transición de la versión 1.0 a la 2.0 el objetivo será preservar los comportamientos básicos deseados por la comunidad IDE –descubrimiento, descripción, decisiones sobre validez– y trabajar en mejorar su implementación, aumentando a la vez la eficacia y simplicidad. Un reto fundamental para la IDE 2.0 será facilitar (radicalmente) el proceso de la colección y explotación de metadatos, para que sea notablemente más transparente y escalable, y consecuentemente menos tedioso.

Existen tres papeles fundamentales para los metadatos: facilitar el descubrimiento del recurso, facilitar su evaluación, y facilitar la consumición o ejecución del recurso. Es importante notar que el agente que debe realizar el descubrimiento, evaluación y consumición puede ser humano o máquina (otro software). Debemos hacer hincapié en la importancia de que estos metadatos sean computables, es decir, tratables por máquinas. Este hecho evitará mucho trabajo rutinario y mecánico, tanto de creación como de clasificación y en consecuencia en el posterior descubrimiento a toda la comunidad de usuarios de la IDE.

Los metadatos, tradicionalmente, han sido definidos simplemente como ‘datos sobre datos’, sin embargo Nogueras et al [4] llegan indirectamente a una mejor definición, recordándonos que la información consiste en “datos más el contexto para su uso (documentación)”. Si efectivamente el objetivo de la IDE es de mejorar la explotación de la información geográfica (quizás es desafortunado el nombre IDE, primando los datos en lugar de la información) y si con los metadatos estamos creando contexto para arropar y fortificar los geodatos, deberíamos estudiar si este contexto debe ser necesariamente explícito o si métodos implícitos podrían resultar más eficaces, escalables, e igual de útiles (o suficientemente útiles) respecto a los explícitos.

En consecuencia, recomendamos la puesta en marcha de una acción paralela de investigación y desarrollo, por un lado preservando y maximizando la efectividad de la arquitectura “IDE 1.0” y de la metodología actual de recopilación de metadatos, y por el otro lado probando nuevas técnicas para intentar pavimentar un camino más escalable para el futuro. A continuación proporcionamos más argumentos a favor del uso de los metadatos implícitos, con ejemplos de los primeros desarrollos en esta dirección desde el proyecto de software libre gvSIG.

METADATOS HASTA AHORA

La Wikipedia (versión inglesa) presenta múltiples definiciones de múltiples tipos de metadatos, demostrando la gran diversidad de perspectivas y campos de aplicación. Bodoff y otros [5] lamentan la sobrecarga de usos previstos para algunos metadatos: según algunas normas los mismos metadatos deben facilitar a la vez la documentación, la configuración y el punto de acceso al recurso. Según el prefacio del estándar ISO-19115 estos metadatos deben servir para describir geodatos y también servicios, sin embargo en la práctica no nos permiten describir de forma correcta ambas cosas.

Los metadatos deben suponer la pieza clave en la arquitectura IDE, sin embargo problemas prácticos con su creación y mantenimiento están limitando su efectividad para tareas como descubrimiento y evaluación de la utilidad de un recurso geográfico dado. Los metadatos son descripciones o anotaciones textuales de recursos creados por proveedores de datos, generadas manualmente y almacenadas (separadas del recurso) en catálogos según la tradición de las bibliotecas digitales para ser encontrados posteriormente con propósitos informativos. Estos metadatos se definen

de acuerdo a estándares de jure (ISO 19115, INSPIRE IR, Dublin Core, etc.) y pueden ser extraídos automáticamente por software GIS –como la extensión geográfica o la fecha de creación– o son introducidos manualmente en el caso de metadatos subjetivos como el *abstract* o el *lineage*.

Métodos de recopilación de metadatos

Beard [6], en un taller sobre datos geo-ambientales, repasa los posibles métodos para la compilación de metadatos, destacando cinco:

1. introducción manual por teclado
2. búsqueda (*look-up*) desde una tabla de referencia
3. medición del valor
4. computación del metadato
5. inferencia del metadato

El primer método lo conocemos bien, ya que es el método por defecto en la mayoría de los casos hoy en día: el usuario edita una ficha (XML) empleando un editor de metadatos del tipo MetaD o CatMDedit. La tesis de este trabajo se centra en el hecho de que, hasta ahora, el campo de las IDEs se queda enganchado en este primer método.

El segundo método supone que un elemento de metadatos se crea a través de una correspondencia con otro, como en el caso de la derivación de la caja envolvente (4 coordenadas del BBOX) de un topónimo, a través de un servicio de nomenclator (*gazetteer*).

El tercer método supone que, en el proceso de colección de geodatos, un sensor podría medir algo como elevación o temperatura, y colocar ese valor en la ficha de metadatos de forma automática.

El cuarto método se centra en al cálculo de un elemento de metadatos empleando los geodatos en sí. Un ejemplo sería la determinación automática de la provincia de un pueblo por cálculos topológicos.

El quinto método es la inferencia de metadatos a partir de otros metadatos o de los geodatos. Según Beard supone el mejor método –de hecho en algunas situaciones el único– para la creación de metadatos *post hoc*, es decir, documentando geodatos ya existentes. Un ejemplo sería inferir la época de los geodatos por el metadato temperatura, a lo mejor recogido por el método 3, de manera que una regla establecería que para temperatura inferior a 15 grados en Tenerife supongamos invierno. Beard también señala (en 1996) algo que debe ser obvio hoy en día, de hecho está captado en [3]: que la creación de estos metadatos inferidos solapa ampliamente a los campos de investigación de la minería de datos y de la recuperación de datos (por ejemplo ver [3]). Estos metadatos inferidos son los que aquí llamamos implícitos.

METADATOS IMPLÍCITOS

Nuevas tendencias en el campo de la recuperación de información (IR) permiten descubrir y tomar cierto nivel de decisiones sobre los recursos empleando metadatos implícitos según convenciones, pero con mínima intervención humana. Distinguimos tres tipos de metadatos implícitos: 1) los extraídos de la fuente binaria del recurso, 2)

los creados según el contexto del recurso, y 3) los utilizados por robots o crawlers para encontrar los recursos según convenciones básicas del concepto de encontrabilidad (*findability*)[1].

Los metadatos implícitos son aquellos que son creados automáticamente usando el razonamiento inductivo o probabilístico, técnicas informáticas del campo de la IR, además de ideas del campo de los sistemas contextuales. Los metadatos implícitos pueden de esta forma ser descripciones formuladas usando técnicas parecidas a Google. Concretamente, Google usa técnicas IR para determinar (aproximadamente) de forma razonable el contenido y la importancia de recursos web indexándolos proporcionalmente, y ofrece una interfaz de fácil utilización para su descubrimiento. En este modelo ni los proveedores de datos ni el usuario son obligados a crear metadatos: las técnicas y rutinas de Google hacen lo posible para que este proceso funcione usando solo métodos implícitos. Esto parece un modelo interesante a emular para nuestros futuros buscadores de servicios IDE, que pretenderían ser más potentes y amigables.

El World Wide Web Consortium (W3C), en un taller sobre el uso de metadatos para ayudar con la provisión de servicios a medida (en web), recomienda (ver abajo) la derivación automática de los metadatos desde la fuente de datos (contenido *markup*, en su caso) para no sobrecargar al usuario, quién, dicen, sólo debe intervenir en el proceso cuando sea absolutamente necesario, y donde haya claro beneficio.

“Inferred metadata

Where possible, metadata should be derived from the existing markup, in order to reduce the burden on the author. It should be possible for the author to add metadata but should only be necessary where there is a clear need and a tangible return on investment (of effort). “

<http://www.w3.org/2004/06/DI-MCA-WS/excreport.html>

De no ser así, la creación de la famosa Web 2.0 no se producirá jamás por la falta de escalabilidad en el proceso de creación y anotación de contenido.

Tipología de metadatos implícitos

Extraídos

Gran parte del software geográfico de licencia libre se basa en librerías de lectura/escritura de múltiples formatos de datos, siendo GDAL/OGR el más popular por ser capaz de trabajar con la mayoría de formatos GIS conocidos, por lo que resultará sencillo extraer información de forma automática a partir de estos geodatos.

Por otra parte, el propio sistema operativo puede proporcionarnos otra información relevante, como por ejemplo fechas o información acerca del propio sistema de ficheros (nombre del archivo, tamaño...).

Inferidos

Podemos inferir metadatos a partir de otros metadatos o de los geodatos, aplicando diversas técnicas de minería de datos, recuperación de datos, uso del contexto que envuelve al metadato, técnicas de razonamiento, etc. Podríamos saber los límites

administrativos de un dato geo-espacial a partir del conocimiento de su caja envolvente y un nomenclator, podríamos inferir la época de los geodatos por el metadato temperatura si existe en la información del contexto de este metadato, podríamos inferir un resumen más o menos adecuado a partir de información del nombre, de la leyenda de una capa, de su posición geográfica, etc.

Buscados por crawler

Los buscadores de WWW funcionan a través de búsquedas optimizadas sobre datos indexados. Estos índices se forman por agentes 'robot' (*crawlers*) que rastrean continuamente todos los recursos interconectados disponibles en la red.

La clave del futuro éxito de estos métodos es su óptima combinación; no se trata de uno sobre otro, sino el empleo de varios o todos en armonía. Pero en primer lugar hay que entender el problema y las posibles soluciones. Para citar sólo un ejemplo, Forsyth y Wilinsky [7], expertos en bibliotecas digitales pero evidentemente no en las nuevas tecnologías, lamentan que de una foto se pueda inferir quienes son las personas retratadas, pero no cuando la foto fue tomada.

“This class of strategies is likely to be successful only for metadata that can be inferred from the object itself. For example, it may be possible to determine the names of those present from a picture, but it is likely to be impossible to determine the time and data at which the picture was taken”

Pensando en el siglo XXI, las fotos ya no son elementos en papel sino ficheros digitales, y mientras el reconocimiento automático de formas y contenido (quienes son las personas en esta foto) se nos resiste de momento (aunque ya hay soluciones parciales, ver Chelappa y otros [8]), casi todas las fotos digitales (formatos JPEG, TIFF) actuales llevan metadatos internos creados por la propia cámara, que indican entre otras cosas la fecha, hora, y hasta localización geográfica (provista hace tiempo por el formato EXIF) de la foto. Para un buen ejemplo ver:
<http://es.wikipedia.org/wiki/Imagen:Geotagging.png>

IMPLEMENTACIÓN INICIAL

La disponibilidad de un marco de software libre para la prueba de estos conceptos, en concreto un cliente de IDEs extensible (www.gvsig.gva.es [9]), es ideal para que la implementación inicial consista en desarrollar una extensión a dicho cliente dotándolo de la funcionalidad necesaria para la gestión de metadatos. La gestión de metadatos involucra la extracción semiautomática de metadatos implícitos, la gestión de los mismos y la posibilidad de publicación en un servidor de catálogo en una IDE [10].

La extracción semiautomática de metadatos implica que el gestor sea capaz, directamente desde la fuente de datos y sin intervención del usuario, de extraer metadatos, que ofrecerán al usuario información de los datos sobre los que trabaja. Los metadatos no implícitos, los que no puedan ser extraídos de forma automática, podrán ser, en una implementación inicial, editados por el usuario y gestionados por el gestor de metadatos de forma que no sea necesaria la intervención del usuario de forma reiterativa, si no que sean almacenados en plantillas junto al propio dato incluyendo información adicional para posteriores usos.

Arquitectura inicial

Los autores de este trabajo llevan los últimos meses desarrollando un primer gestor de metadatos que estará disponible como una extensión, o *plug-in*, para gvSIG. El Gestor

de Metadatos tendrá una arquitectura que, como se aprecia en la figura 1, interactuará de forma directa con los principales módulos (*core*) de gvSIG. Esta extensión amplía la funcionalidad del modelo de objetos, dotando a los recursos de gvSIG de descripciones sintácticas de forma que posteriormente puedan ser almacenadas en el sistema para posibles intercambios entre usuarios y/o incluso publicadas en Servicios de Registro (catálogos) de Infraestructuras de Datos Espaciales.

En la primera fase (piloto) de implementación los recursos a documentar serán limitados a capas, contemplando en fases posteriores la posibilidad de la documentación de otros recursos como pueden ser vistas (composición de capas), *features*, procesos, etc. La arquitectura y jerarquía del modelo de objetos de la próxima versión del Gestor de Metadatos tendrá en cuenta la extensibilidad de los recursos a documentar.

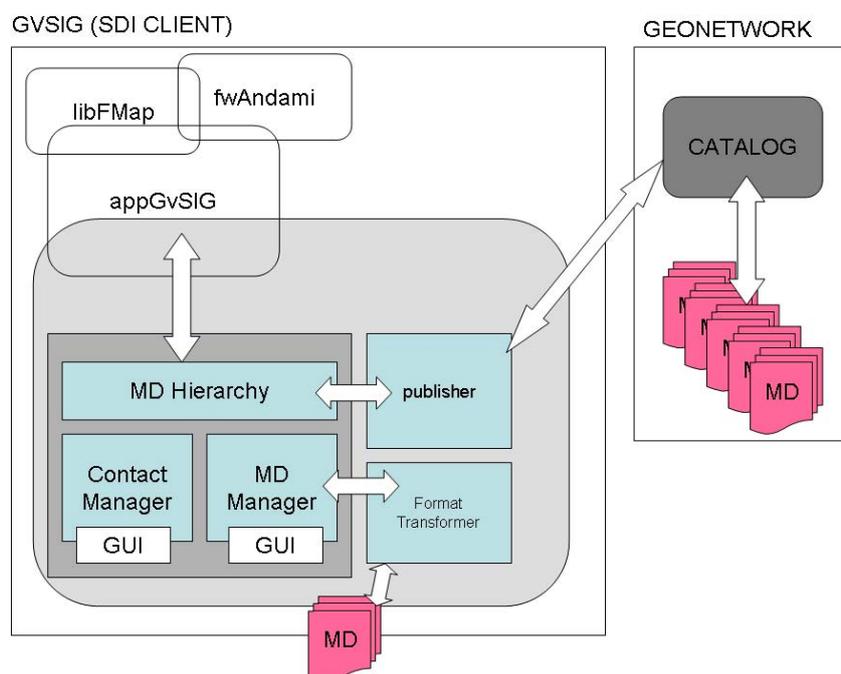


Figura 1. Arquitectura del piloto del Gestor de Metadatos en gvSIG

Funcionalidad

Niveles de documentación.

Hablando de metadatos implícitos a los datos, a nivel de implementación, en la fase de extracción pueden clasificarse en distintos niveles: podemos hablar de metadatos asociados a la fuente de datos; metadatos proporcionados por el sistema; metadatos asociados a un conjunto de datos, como puede ser una vista compuesta por un diseñador de cartografía para un uso específico; o de metadatos de una misma fuente de datos en diferentes momentos.

En la fase inicial de la implementación extraemos y mantenemos metadatos asociados a una capa de la vista de gvSIG, que esta directamente relacionada con una fuente de datos, así como la información que el sistema nos proporciona sobre ella.

Extracción de metadatos.

Hablamos de extracción semiautomática porque los metadatos implícitos son extraídos automáticamente, y los no implícitos en una primera fase serán gestionados por el usuario. En una segunda fase los metadatos no implícitos a la fuente de datos serán inferidos a partir del conocimiento de otros metadatos implícitos relacionados y otras técnicas de recuperación de información. La extracción de los metadatos se realiza como sigue:

- Extracción automática de metadatos directamente desde la fuente de datos y sin intervención del usuario, el gestor será capaz de extraer metadatos, que ofrecerán al usuario información de los datos sobre los que trabaja utilizando la funcionalidad de los drivers de gvSIG de lectura de ficheros con formato conocido, de igual forma se está estudiando la posibilidad de integrar parte de la funcionalidad del wrapper desarrollado por M. Manso de la UPM, sobre la librería GDAL.
- Edición manual de metadatos que no puedan ser extraídos de forma automática o por requisito de usuario podrán ser editados manualmente y gestionados por el gestor de metadatos de forma que no sea necesario la intervención del usuario de forma reiterativa, si no que sean almacenados en plantillas junto al propio dato incluyendo información adicional para posteriores usos.

Generación de un formato de metadato interno

La extensión de metadatos almacenará y mantendrá los metadatos junto a los datos, como hemos comentado anteriormente, adoptando la filosofía de que los metadatos forman parte de los datos, o por lo menos deben ir siempre juntos, pasando así de una separación a una integración de los datos con sus metadatos.

Para la persistencia de los metadatos en el piloto se utiliza un esquema XML para la creación de un formato de fichero XML. La extensión de este fichero XML de metadatos es .MDML (MetaData Mark-up Language), el esquema de los ficheros mdml es un superconjunto de los formatos de metadatos más utilizados (ISO, Dublín-Core, NEM).

Actualmente en gvSIG existe el formato RMF (raster meta file) que contiene metadatos de las capas raster, este fichero está orientado al almacenamiento de información para garantizar un rendimiento óptimo cuando se trabaja con datos raster. También se pretende realizar un estudio para la integración de estos metadatos con el formato MDML, por ejemplo añadiendo en el MDEXchanger la capacidad de transformación de un formato a otro mediante, por ejemplo, el uso de hojas de estilo.

En futuras versiones del Gestor, desaparecerá el formato MDML, a favor del uso de un objeto Java para guardar directamente estos metadatos. Sólo a la hora de exportar los metadatos creados, pasarán por una hoja de estilos y saldrán en formato XML para su posterior envío a un servidor de registro.

Visualización y edición de metadatos.

La extensión del Gestor de Metadatos incluye un módulo de visualización y edición manual de metadatos, el usuario podrá visualizar en cualquier momento los metadatos asociados a una capa y hacer las modificaciones que crea convenientes. Este editor informará de forma gráfica el estado de cada metadato así como su carácter dentro de la ISO y NEM, es decir, podremos saber de forma rápida y visual qué metadatos de carácter obligatorio por ejemplo faltan por rellenar en nuestro conjunto de datos. El

visor-editor de metadatos permite la visualización de metadatos de acuerdo a los diferentes perfiles de metadatos soportados por gvSIG, de forma que un usuario puede alternar las vistas y ver los metadatos según el perfil MDML, ISO, etc. Esta funcionalidad permite al usuario saber a simple vista si dispone, por ejemplo, de todos los metadatos obligatorios según el perfil al que quiera exportar o publicar sus metadatos. Si el usuario desea visualizar o modificar sus datos desde otros editores de datos, como por ejemplo CatMDEdit, el editor de metadatos incluido dentro de la extensión permite exportar los metadatos a formato ISO-NEM u otros estándares para que puedan ser utilizados desde otras aplicaciones que trabajen con estos estándares. Y volver a cargar los metadatos posteriormente a gvSIG importando el archivo generado por el editor externo de acuerdo a alguno de los estándares soportados por ambos.

Importar y Exportar metadatos.

En el caso de que los usuarios quieran compartir metadatos asociados a los datos sobre los que trabajan, como hemos comentado en el apartado anterior, gvSIG podrá tanto exportar sus metadatos a formatos conocidos de metadatos como importar metadatos externos en formatos conocidos y asociarlos a sus conjuntos de datos. Esta funcionalidad está implementada en el módulo MDEXchanger, que tras validar los metadatos realizará la transformación a los diferentes formatos estándar mediante hojas de transformación.

Publicación de metadatos a un servidor de catálogo.

En última instancia el usuario puede publicar sus datos en un servidor de catálogo (por ejemplo GeoNetwork). Si el usuario quiere publicar sus datos en un servicio de catálogo, el Gestor de Metadatos validará los metadatos existentes para dichos datos. Dicha validación consiste en confirmar que el conjunto de metadatos contiene, al menos, el subconjunto de datos obligatorios en el estándar ISO. En el caso en que la validación falle, el usuario podrá utilizar el Editor interno de metadatos para completar los metadatos.

Una vez el conjunto obligatorio de metadatos esté completo, el usuario puede mediante el módulo de publicación publicar su conjunto de datos en un servicio de catálogo disponible.

El MDPublisher, módulo encargado de la publicación de los metadatos, podrá ser invocado desde el visor-editor de metadatos o directamente desde una opción en el menú, de forma que el usuario puede elegir la capa actual para ser publicada o bien, elegir un fichero de metadatos existente en su sistema de ficheros.

Este módulo realizará la conexión al servicio de catálogos elegido por el usuario, realizando una validación de usuario que compruebe que se tienen suficientes permisos sobre el servicio de catálogos. Tras la elección del formato de metadatos se publicarán los metadatos y se les asignarán (de momento por defecto) los permisos y la categoría pertinentes para que estos queden disponibles de inmediato, obteniendo el estado de las diferentes peticiones e informando al usuario del éxito o fracaso de toda la transacción.

De este modo, el usuario puede, utilizando el cliente de Catálogo de gvSIG, realizar la búsqueda de los datos recién publicados cerrando así el ciclo completo de vida de los metadatos.

En una primera fase de desarrollo se utilizará GeoNetwork como Servicio de Catálogo de pruebas, en su versión 2.1. Se estudia la posibilidad de la publicación de archivos MEF de GeoNetwork, formato que empaqueta datos y metadatos. Esto supondría ya otro paso más en el intento de unir cada vez más los recursos y su documentación: dato con metadato.

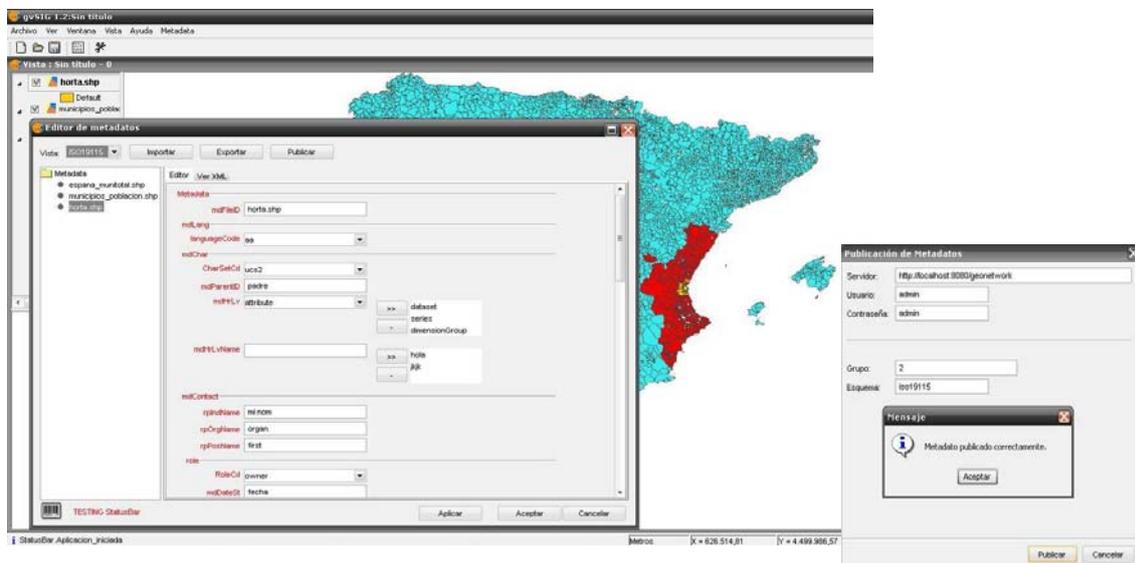


Figura 2: Piloto gestor de Metadatos dentro de gvSIG.

CONCLUSIONES

Esta implementación del concepto de extracción y gestión automática de metadatos facilita a los usuarios la creación y edición de imágenes y datos cartográficos que se desean publicar en una infraestructura de datos espaciales. La naturaleza integrada de esta solución facilita la creación de metadatos muy importantes para el buen funcionamiento de las IDEs, por lo que queda demostrada la validez del concepto en este entorno real y operativo.

En futuras fases se contemplará la documentación de otros niveles, tratando diferentes recursos del sistema como por ejemplo *features*, leyendas, mapas, proyectos, geo-procesos, etc., así como la obtención de la mayor información posible para otros formatos de datos. Del mismo modo se contemplará también la integración con el almacenamiento de metadatos de capas raster, las cuales necesitan almacenar metadatos especiales para el rendimiento óptimo del sistema. Y evidentemente se trabajará para incluir la incorporación de habilidades para importar metadatos externos desde cualquier fuente.

Pero quizás la vía más interesante de futuros desarrollos consiste en otras formas más inteligentes de extracción de metadatos como puede ser metadatos inferidos y/o razonados a partir del conocimiento de metadatos y datos asociados, extracción intuitiva de metadatos intrínsecos(basados en contexto) a la fuente de datos, siguiendo la "metodología Google", incluyendo la extracción automática, usando métodos deductivos, obteniendo así información suficiente para generar textos libres "creíbles" y útiles para describir conjuntos de datos. Otras líneas de investigación interesantes son trabajar con catálogos y metadatos no estándares pero más

sencillos, donde la indexación de información y el uso de metadatos podría ser más eficiente.

Estos retos, a la vez que cambios bastante radicales en la arquitectura en la próxima versión de gvSIG, prometen grandes mejoras en la posibilidad de extraer, y explotar, los metadatos de una forma más integrada, cómoda, y eficaz.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado dentro del proyecto gvSIG, por la Generalitat Valenciana y los fondos FEDER.

REFERENCIAS

- [1] P. MORVILLE, "Ambient findability" O'Reilly Media, 2005.
- [2] C. GRANELL; M. GOULD; M.A. MANSO Y M.A. BERNABÉ, "Spatial Data Infrastructures". En: H. Karimi (ed): Handbook of Research on Geoinformatics. Information Science Reference. 2008.
- [3] M. GOODCHILD, "Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0", International Journal of Spatial Data Infrastructures Research, 2007, Vol. 2, 24-32.
- [4] J. NOGUERAS; F.J. ZARAZAGA Y P. MURO, "Geographic information metadata for spatial data infrastructures", Springer-Verlag, 2005.
- [5] D. BODOFF; HUNG, P.C.K Y M. BEN-MENCHEM, "Web metadata standards: observations and prescriptions", IEEE Software, Enero-Febrero 2005, pp 78-85.
- [6] K. BEARD, "A Structure for Organizing Metadata Collection", The Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico. NCGIA, University of California at Santa Barbara. 1996. http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/beard_kate/metadatapaper.html
- [7] DA FORSYTH Y R WILENSKY, "Research issues for digital libraries", NSF Post-DL Futures Workshop, Chatham, MA, Junio, 2003. http://www.sis.pitt.edu/~dlwshop/paper_wilensky.pdf
- [8] R. CHELLAPPA; WILSON, C.L. Y SIROHEY, S. , "Human and machine recognition of faces: a survey", Proceedings of IEEE, vol, 83, num 5, 1995, 705-741.
- [9] gvSIG. Proyecto gvSIG. <http://www.gvsig.qva.es/> (accedido por última vez Enero 2008)
- [10] L. DÍAZ; C. MARTÍN; M. GOULD; C. GRANELL Y M.A. MANSO, "Semi-automatic Metadata Extraction from Imagery and Cartographic data". International Geoscience and Remote Sensing Symposium (IGARSS 2007). Barcelona, Julio 2007. IEEE CS Press, pp. 3051-3052.