

Sistema de Apoyo a la Formalidad de Procesos Geográficos

R. Carolina Medina Ramírez

UAM Iztapalapa, Av. San Rafael Atlixco No 186, Vicentina, Iztapalapa, D.F., 09340, 01 (55) 5804-4636
cmed@xanum.uam.mx

Inés Marlén Jaramillo Rosas

UAM Iztapalapa, Av. San Rafael Atlixco No 186, Vicentina, Iztapalapa, D.F., 09340, 01 (55) 5804-4636
cbi205215864@titlani.uam.mx

Yolanda M. Fernández Ordoñez

Colegio de Posgraduados, Carr. México-Texcoco km 36.5, Montecillo, Edo. de México, 01 (55) 5804-5900
yfernand@colpos.mx

Resumen

En este artículo se describe una aplicación web para facilitar la formalidad de procesos geográficos. La propuesta se apoya en tecnologías semánticas (RDF, datos abiertos enlazados), así como en el estándar ISO 19138 para la calidad de los datos y linaje de un objeto geográfico (mapa) considerando metadatos mínimos. Finalmente, se termina este artículo proporcionando algunas conclusiones trabajo futuro.

Palabras Claves: RDF, XML, Datos Abiertos Enlazados, ISO 191138.

1. Introducción

Las bases de datos geoespaciales (GeoBases) se originan a través del análisis de imágenes y la manipulación de datos geográficos. Los conjuntos de datos correspondientes se describen a través de diversas estructuras de metadatos.

GeoBase L9 es un proyecto de investigación para construir una base de datos geoespacial para apoyar a un área de investigación en geomática aplicada enfocada en la gestión de los recursos agrícolas y naturales [1]. La geomática es una ciencia que se encarga de la captura, tratamiento, análisis, interpretación, difusión y almacenamiento de información geográfica a través de la informática.

El área de investigación se sitúa en el Laboratorio de Geomática del Colegio de Postgraduados en Montecillo-Texcoco, Edo. De México. Esta institución se dedica a la investigación y enseñanza a nivel posgrado en ciencias agrícolas. Elementos geoespaciales son proporcionados por los investigadores de esta línea para ser archivados en formato digital para su posterior uso por otros colegas o instituciones. Existe una primera versión de una GeoBase (geobase L9) que fue construida para proporcionar servicios básicos de navegación para localizar conjuntos de datos. Esta geobase está siendo evaluada y enriquecida por los usuarios [2]. Una versión evolucionada de la misma se vislumbra para atender las necesidades de los investigadores como es el localizar y compartir las imágenes de satélite de acuerdo con los criterios exigidos por sus actividades de investigación, en particular en lo que se refiere a la teledetección (Remote Sensing). El uso de los datos geoespaciales, el análisis y representación de los procesos aplicados a las imágenes de satélite así como los elementos de información que figuran en la sección del linaje y la calidad de los datos son solo algunos aspectos a considerar. El construir un centro de consulta semántica de datos geoespaciales obedece no solo a mejorar el intercambio y la reutilización de los datos por los usuarios humanos, sino también para apoyar la apertura de las geobases a los navegadores de la web semántica mediante la incorporación del significado del recurso geográfico en los metadatos.

La comprensión del usuario común de las características de los objetos geoespaciales facilita la reutilización de la información, aspecto importante para la reducción de los costos, tiempo y esfuerzo en la investigación. A nivel de sistemas, un entendimiento común del significado de los metadatos que describen un recurso geográfico (imagen, mapa) son un requisito para lograr la interoperabilidad. Para que los usuarios de este dominio principalmente puedan compartir información necesitan saber acerca de las tareas básicas de pre-procesamiento que se han aplicado a una imagen (por ejemplo, los niveles de corrección) y sobre otros procesos que han producido un mapa de imágenes (algoritmos de inversión específicos, linaje y calidad de los datos) [3]. Información básica se encuentra disponible en las cabeceras de los archivos de imagen, que se suministra de manera sucinta en forma de texto. Sin embargo ese texto es completamente abierto y no se apega a alguna estructura que permita obtener información sobre el proceso y software utilizados, la descripción del proceso, descripción de la fuente de información, entre otros. Si bien, técnicas de minería de texto podrían aplicarse en estos campos textuales, sería recomendable primero apearse a estándares para caracterizar los mapas generados y aplicar enseguida solamente a los metadatos necesarios de tipo textual dichas técnicas. Por otro lado, al carecer de una estructura para documentar el proceso seguido para generar un mapa, gran parte de la información complementaria a la cabecera de la imagen se encuentra disponible de forma externa en documentos de investigación. Por ejemplo, se necesitan características del sensor que afectan a la formación de imágenes, tales como la orientación geométrica o calibración radiométrica para integrar datos de diferentes sensores, una tarea común en la investigación en teledetección [4]. La localización de información detallada para cotejar y fusionarla con coherencia es una tarea de investigación, especialmente para los asistentes y estudiantes investigadores novatos. Cabe mencionar que diferentes fuentes de información pueden generar datos e información dispares. Varios métodos han sido investigados en ciencias de la computación para resolver formatos heterogéneos y hacer frente a diferencias semánticas.

2. Metodología

Se propuso una metodología de cinco pasos para sacar un conjunto básico de metadatos para caracterizar un mapa considerando el linaje y calidad de los datos.

1. Adquisición del conocimiento existente en los mapas
2. Análisis de estándares (ISO 19138) y trabajo existente ad hoc a la Línea GeoBase L9
3. Proponer el conjunto de metadatos de prueba para el linaje y calidad de los datos.
4. Implementar una aplicación web para la formalidad de procesos geográficos
 - 4.1 Caracterizar los mapas en función de la calidad de datos y linaje
 - 4.2 Representar los metadatos capturados en un formato estándar (RDF)
5. Valoración de la aplicación por expertos del dominio (GeoBase L9)
6. Es una propuesta modulable que puede ajustarse a diferentes casos de uso. Lo anterior conlleva a considerar los requerimientos de los usuarios para agregar nuevos atributos apeguándose al estándar ISO y hacer las modificaciones pertinentes en la aplicación.

3. Metadatos geoespaciales

La creación de recursos cartográficos en especial los mapas se realizan con base a un conjunto de actividades que en su mayoría están guiadas por la experiencia del creador del mismo y dependiendo del tipo de mapa a generar. Sin embargo existen algunos procedimientos mínimos necesarios que deben cumplirse de acuerdo a estándares (ISO 19115). Las necesidades de información de los usuarios potenciales deben de cubrirse en función de los casos de uso de la información que se tengan. Son estos últimos quienes guiarán la caracterización de un mapa (metadatos) para soportar futuras consultas a través del vínculo entre los metadatos (búsqueda exploratoria) o a través de una búsqueda guiada/basada en ontologías incluyendo razonamiento.

En una búsqueda exploratoria se ofrecen rutas guiadas que arrojan datos en los mapas.

	ISO [5]	METADATOS COLPOS [2]	SMAP NASA [6]	Sistema de Apoyo a la Formalidad de Procesos Geográficos
Calidad de los datos	<ol style="list-style-type: none"> 1. Id archivo 2. Lenguaje (SP,EN) 3. Conjunto de caracteres (encoding) 4. Nivel de jerarquía 5. Nombre de la jerarquía 6. Contacto (experto de contacto) 7. Fecha 8. Estandar (ISO-) 9. Versión del estándar 	<ol style="list-style-type: none"> 1. Precisión posicional 2. Precisión temporal 3. Precisión temática 4. Completitud 5. Consistencia lógica 	<ol style="list-style-type: none"> 1. Attribute 2. attributeType 3. collectionHardware 4. collectionSession 5. Dataset 6. Series 7. nonGeographicData 8. dimensionGroup 9. Feature 10. featureType 11. propertyType 12. fieldSession 13. Software 14. Service 15. Tile 16. Model 	<ol style="list-style-type: none"> 1. Atributo. 2. Tipo de Atributo 3. Serie 4. Base de datos no geográfico. 5. Dimensión 6. Característica 7. Tipo de característica 8. Tipo de propiedad 9. Sesión de campo 10. Modelo
Linaje	<ol style="list-style-type: none"> 1. Enunciado 2. Pasos del proceso 3. Fuente de información 	<p>Información de los orígenes (campo textual)</p>	<ol style="list-style-type: none"> 1. Statement 2. Identifier 3. softwareReference 4. procedureDescription 5. Documentation 6. runTimeParameters 7. Description 8. scalaDenominator 9. sourceReferenceSystem 10. sourceExtent 11. sourceCitation 12. processedLevel 13. Resolution 	<ol style="list-style-type: none"> 1. Identificador del proceso. 2. Referencia del software. 3. Descripción del procedimiento. 4. Documentación. 5. Descripción de la fuente. 6. Sistema de referencia de la fuente. 7. Grado de la fuente. 8. Nivel procesado.

Tabla 1. Tabla de comparación de metadatos.

El enfoque de Datos Abiertos Enlazados (DAE) puede adoptarse para facilitar una búsqueda semántica exploratoria [8]. DAE es una iniciativa del World Wide Web Consortium que tiene como objetivo hacer que los datos semánticos sean accesibles y reutilizados libremente. Proporciona puntos de partida (lineamientos) para expresar los

datos de forma abierta y establecer relaciones (vínculos) entre los recursos de información. La versión de la aplicación descrita en este documento considera por el momento la captura de los metadatos necesarios para describir el linaje (8 metadatos) y la calidad de los datos (10 metadatos) ambos almacenados en un formato RDF. Los vínculos entre metadatos así como otras características del lineamiento que establece el movimiento de Datos Abiertos Enlazados están en desarrollo. Se hizo el análisis de estándares ISO 19115, 19138 así como la GeoBase L9 disponible para obtener un conjunto de metadatos de prueba para implementar la aplicación. Las Tabla 1 muestra estos metadatos, se somborean aquellos metadatos tomados para la implementación.

4. Sistema de apoyo a la formalidad de procesos geográficos

En esta sección se describe la aplicación desarrollada para capturar los datos correspondientes al linaje (ascendencia/descendencia de los datos) y calidad de los datos existentes en mapas. La necesidad de comprender con precisión el origen y la historia posterior al procesamiento de los datos se ha destacado como un requisito en la investigación científica [7]. Se requiere una herramienta de software para la anotación de estos mapas con el fin de *a) codificar dichos metadatos en un formato estándar y abierto y b) validar los metadatos de acuerdo a ontologías geográficas externas con el objetivo a largo plazo de construir grandes corpus de mapas con anotaciones que podrían permitir búsquedas tanto exploratorias, como semánticas.*

Existen iniciativas internacionales sobre la construcción de conjuntos de datos geográficos abiertos [8,9]. Sin embargo para el contexto mexicano no existe alguna propuesta de impacto global. El mecanismo propuesto para hacer uso de los datos geográficos almacenados en un formato estándar RDF (Resource Description Framework) es por un lado, construir un GeoLinkedData (.mx) de acuerdo con los lineamientos de los Datos Abiertos Enlazados e interrelacionarlo con otras bases de conocimiento que pertenecen a dicha iniciativa. Por otro lado, contribuir a satisfacer las necesidades de búsqueda de información geográfica en México.

La aplicación descrita en este documento utiliza un formato web con el fin de capturar algunos metadatos de mapas considerando la calidad de la información y el linaje. Estas anotaciones son almacenadas en formato RDF con el fin de seguir la iniciativa de los Datos abiertos Enlazados (LOD) [8,9]. El idioma por defecto es el español, pero otros idiomas podrían ser incluidos. La Fig. 1 muestra el formulario sugerido para controlar la captura de datos que describen a un objeto geográfico (mapa) para a) calidad de los datos y b) linaje. Los metadatos considerados para la calidad de la información fueron tomados de la norma ISO 19138 (Medidas de calidad de datos). Para esta versión de la herramienta se consideraron 18 metadatos e implementaron los diez primeros de la siguiente lista: Atributo, Serie, Dimensión, Tipo de característica, Sesión de campo, Tipo de Atributo, Base de datos no geográfica, Característica, Tipo de propiedad, Modelo, Nivel de alcance, Descripción del nivel, Identificador de la medida, Nombre de la medida, Descripción de la medida, Evaluación del tipo de método, Descripción de la evaluación del método y Fecha. La Fig. 2 muestra la parte del formulario correspondiente a la calidad de la información.

localhost/ProcesoCartogr x
localhost/ProcesoCartografico/index.html

30 ANIVERSARIO **ie**
departamento de ingeniería eléctrica

SISTEMA DE APOYO A LA FORMALIDAD DE PROCESOS CARTOGRAFICOS

CALIDAD DE LA INFORMACION

Atributo. <input type="text"/>	Tipo de atributo. <input type="text"/>
Serie. <input type="text"/>	Bases de datos no geografica. <input type="text"/>
Dimension. <input type="text"/>	Caracteristica. <input type="text"/>
Tipo de caracteristica. <input type="text"/>	Tipo de propiedad. <input type="text"/>
Sesion de campo. <input type="text"/>	Modelo <input type="text"/>

LINAJE

Descripcion de la fuente.
Escribe aqui la descripcion

Sistema de referencia de la fuente. <input type="text"/>	Identificador del proceso. <input type="text"/>
Grado de la fuente. <input type="text" value="COMPLETO"/>	Nivel procesado. <input type="text" value="100%"/>
Serie. <input type="text"/>	Referencia del Software. <input type="text"/>

Descripcion del procedimiento.
Describe el procedimiento

02:31 a.m.
24/05/2014

Fig. 1. Formulario para la captura de la Calidad de los datos y linaje de un objeto geográfico (mapa).

CALIDAD DE LA INFORMACION

Atributo. <input type="text"/>	Tipo de atributo. <input type="text"/>
Serie. <input type="text"/>	Bases de datos no geografica. <input type="text"/>
Dimension. <input type="text"/>	Caracteristica. <input type="text"/>
Tipo de caracteristica. <input type="text"/>	Tipo de propiedad. <input type="text"/>
Sesion de campo. <input type="text"/>	Modelo <input type="text"/>

LINAJE


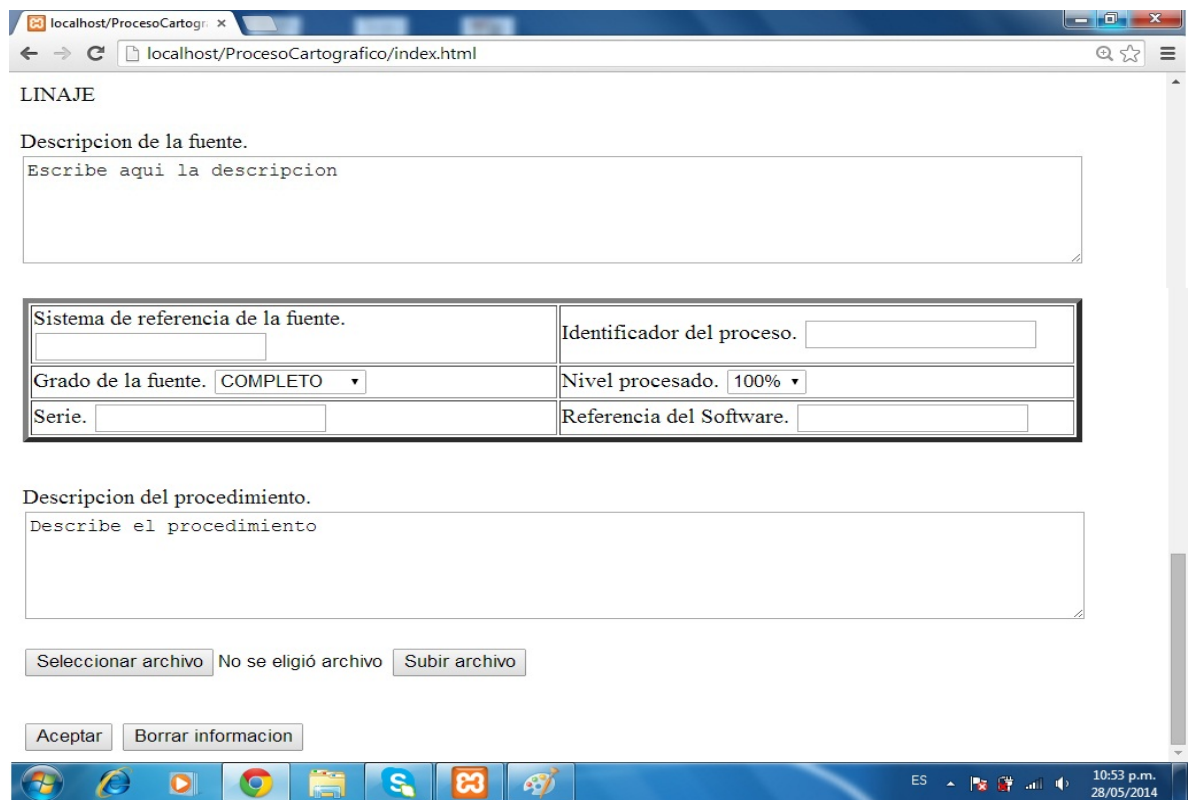


Fig. 2. Calidad de la información: metadatos.



localhost/ProcesoCartogr... x

localhost/ProcesoCartografico/index.html

LINAJE

Descripcion de la fuente.

Escribe aqui la descripcion

Sistema de referencia de la fuente. <input type="text"/>	Identificador del proceso. <input type="text"/>
Grado de la fuente. COMPLETO ▾	Nivel procesado. 100% ▾
Serie. <input type="text"/>	Referencia del Software. <input type="text"/>

Descripcion del procedimiento.

Describe el procedimiento

Seleccionar archivo No se eligió archivo Subir archivo

Aceptar Borrar informacion




Fig. 3. Linaje: metadatos.

Para la sección correspondiente al linaje se describirá la fuente de la información, el procedimiento que realizó con respecto a la fuente y una opción para indicar archivos utilizados. Se consideraron dos aspectos a) Proceso de la información y b) Fuente de información. Los metadatos propuestos para el proceso de información fueron: Identificador del proceso, Referencia del software, Descripción del procedimiento, Documentación. Para la Fuente de información tenemos: Descripción de la fuente, Sistema de referencia de la fuente, Grado de la fuente, Nivel procesado. La Fig. 3, muestra la parte del formulario correspondiente al linaje.

5. Conclusiones

La caracterización es una fase importante en la gestión del conocimiento. Los datos existentes en un mapa son un factor importante para vincularlos, procesarlos e incluso hacer razonamiento sobre ellos. Los términos de descripción más comunes son utilizados en esta propuesta y se apegan a la norma ISO 191138, en particular se hace énfasis en aquellos metadatos que proporcionan información sobre la historia de los procesos de linaje y calidad de los datos. En este trabajo, hemos descrito una aplicación web para la caracterización de mapas apegándose a estándares y que ha sido desarrollada en XML/RDF y tecnologías asociadas. Esta aplicación esta en valoración por los investigadores de la línea GeoBase L9 en el Colegio de Posgraduados, Montecillo- Texcoco estado de México. Se considera ampliar la aplicación incorporando un módulo de minería de texto para el procesamiento de campos de textuales existentes, así como continuar con los lineamientos del movimiento de datos abiertos enlazados.

6. Referencias

- [1] Y.M. Fernández-Ordoñez, J. Soria-Ruiz, R.C. Medina-Ramírez. Building a Conceptual Model and Ontology for Heterogeneous Geodatabases, International Conference Hellenic Association for Information and Communication Technologies in Agriculture Food and Environment (HAICTA), Athens, Greece, pp. 392-396,2008.
- [2] H. I. Camacho Vázquez. Bases de Datos Geo-espaciales para Investigación en Agricultura y Recursos Naturales: Especificaciones Técnicas de Diseño e Implementación, Colegio de Posgraduados, Montecillo, Texcoco, Edo. De México. 2013.
- [3] S. Twomey. Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements, Dover Publications, 2002.
- [4] K. Jacobson. Satellite Image Orientation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part I. Beijing. Pp. 703-709, 2008.
- [5] Guía de normas (Edición en español). Comité ISO/TC 211 Información Geográfica / Geomática. Instituto Panamericano de Geografía e Historia. México. 2010. Accesible en: http://www.isotc211.org/Outreach/ISO_TC_211_Standards_Guide_Spanish.pdf
- [6] B. Weiss, H. Hua y V. Haemmerle, Practical Application Using ISO Metadata, Propulsion Laboratory California Institute of Technology Pasadena, CA. Pag. 34-36
- [7] R. Bose, J. Frew. Lineage retrieval for scientific data processing:a survey, Computing Surveys, ACM, N.Y.(37), pp1-28, 2005
- [8] C. Stadler, J Lehmann, K. Höffner, S. Auer. Linkedgeodata: A core for a web of spatial open data, Semantic Web, 3(4), 333-354, 2012.
- [9] A. De León, V. Saquicela., L. M. Vilches, B. Villanzón-Terrazas, F. Priyatna, O. Corcho. Geographical linked data: a Spanish use case, Proceedings of the 6Th International Conference on Semantic Systems, p. 36, ACM, NY, USA, 2010.

7. Autores

C.Inés Marlén Jaramillo Rosas. Estudiante de la Licenciatura en Computación de la Universidad Autónoma Metropolitana Unidad Iztapalapa.

Dra. Reyna Carolina Medina Ramírez. Obtuvo su grado de Doctor en Ciencias (especialidad Computación) de la Universidad de Nice-Sophia Antipolis, Francia en 2003. Desde septiembre de 2005 es profesora adscrita al Departamento de Ingeniería Eléctrica en la UAM unidad Iztapalapa. Actualmente es jefa del Área de investigación de Redes y Telecomunicaciones. Sus intereses de investigación incluyen web semántica y datos abiertos enlazados.

Dra. Yolanda M. Fernández-Ordoñez. Obtuvo su grado de Doctor en Nouveau Régime en Systemes Informatiques del Instituto Nacional Politécnico de Grenoble, Francia en 1989. Desde 1975 es Profesora Investigadora Titular en el Instituto de Socioeconomía estadística e Informática en el Colegio de Posgraduados. Sus intereses de investigación incluyen Geomática aplicada y ontologías.