

SOBRE LA APLICACIÓN DEL ESTÁNDAR OGC GMLJP2 PARA LA DOCUMENTACIÓN DE METADATOS INCLUIDOS EN FICHEROS JPEG2000

A. Zabala (*) y X. Pons (*) (**).

(*) *Dep. Geografía. Ed. B Universidad Autónoma de Barcelona. 08193 Cerdanyola del Vallès (Barcelona).
alaitz.zabala@uab.es*

(**) *CREAF. Centre de Recerca Ecològica i Aplicacions Forestals. Ed. C. Universitat Autònoma de Barcelona.
08193 Cerdanyola del Vallès (Barcelona).*

RESUMEN

La iniciativa GMLJP2 aprobada por el *Open Geospatial Consortium* (OGC) trata sobre la codificación y descodificación de imágenes en formato JPEG2000 que contienen información en GML sobre su georeferencia y cómo esta información en GML debe ser codificada y empaquetada en el fichero JPEG2000. El primer papel que se da a GML dentro del estándar GMLJP2 en relación con el formato JPEG2000 es que GML describa las características radiométricas y geométricas de la imagen comprimida. Además, se puede incorporar un conjunto completo de metadatos que describan estas imágenes, por ejemplo siguiendo ISO 19139 que implementa en XML el estándar abstracto de metadatos ISO 19115.

La comunicación expone las características necesarias que este fichero GML debe contener así como su relación con los estándares de metadatos existentes (ISO 19115, ISO 19115-2). Para desarrollar esta tarea es necesario generar un esquema de aplicación GML que describa la imagen y sus metadatos y empaquetar esta información en el fichero JPEG2000 a través de 'xml boxes'.

ABSTRACT

GMLJP2 initiative approved by the Open Geospatial Consortium (OGC) deals on the coding and decoding of images in JPEG2000 format that contain information in GML about their georeferencing and how this information must be coded in GML and packed on the JPEG2000 file. The first role given to GML in GMLJP2 standard format is that it describes the radiometric and geometrical characteristics of the compressed image. Besides, a complete set of metadata describing the images can be incorporated on the GML file, for example following ISO 19139 (XML implementation specification of the abstract metadata standard ISO 19115).

The paper exposes the necessary characteristics that this GML file must contain as well as its relationship with the existing metadata standards (ISO 19115, ISO 19115-2). It is necessary to generate a GML application schema describing the image and its metadata and to pack this information in the JPEG2000 file through 'boxes xml'.

Palabras clave: metadatos, GMLJP2, compresión imágenes, JPEG2000, estandarización.

INTRODUCCIÓN

En los últimos años, la compresión de imágenes está siendo cada vez más utilizada en Teledetección debido al creciente volumen de datos disponibles. Diversos estudios abordan los efectos de la compresión con pérdida en los usos posteriores de las imágenes y concluyen que hasta razones de compresión razonablemente elevadas los efectos en los usos posteriores de las imágenes son negligibles y en algunos casos beneficiosos (Zabala *et al.* 2006; Zabala *et al.* 2007; Zabala *et al.* 2008).

El formato JPEG2000 se convirtió en un estándar ISO en 2000 y fue revisado en 2004 (ISO/IEC 2004) y presenta diversas características interesantes para el ámbito de Teledetección. Acepta compresión con o sin pérdida, permite la compresión multi-componente (diversas bandas

comprimidas en un mismo fichero JPEG2000), permite alcanzar elevadas razones de compresión con mayor calidad, permite recuperar la imagen a diversas resoluciones y permite definir Regiones de Interés (ROI) comprimidas con mayor calidad.

GML (*Geography Markup Language*) es una gramática XML para expresar entidades geográficas. Sirve de formato de codificación y transferencia de información geográfica a través de Internet. Hay diversas versiones del estándar de codificación. La versión 3.2.1 se adoptó como estándar ISO en 2007 en (ISO 2007). La versión de GML utilizada para ser empaquetada en JPEG2000 es la versión 3.1.1 (OGC 2004). Como en la mayoría de gramáticas XML existen dos partes: el esquema de aplicación que define la estructura de la información (XSD), y el documento "instancia" que contiene los datos en sí (XML).

GMLJP2

La iniciativa GMLJP2 aprobada por el *Open Geospatial Consortium* (OGC 2006) trata precisamente sobre la codificación y descodificación de imágenes en formato JPEG2000 que contienen información en GML sobre su georeferencia y cómo esta información en GML debe ser codificada y empaquetada en JPEG2000. El uso de un esquema de aplicación propio para cada imagen permite la descripción de la geometría y la radiometría de la imagen. Además se puede usar el documento GML para indicar metadatos adicionales de la imagen así como para incluir otras entidades complementarias: entidades geográficas y anotaciones, visualización de estas, definición de Sistema de Referencia y unidades de medida. El documento GMLJP2 fija que es necesario crear los esquemas de aplicación en base a un perfil de GML específico denominado *gmlJP2Profile*, que permite utilizar todos los elementos mencionados.

La inclusión del esquema de aplicación y el documento GML de descripción de la cobertura en el propio fichero JPEG2000 aporta la deseada unidad de los datos y los metadatos. Esta unidad de datos y metadatos es más importante, si cabe, en una imagen de Teledetección, en la que los valores de los píxeles *per se* no son autoexplicativos (¿imagen corregida geoméricamente, radiométricamente, con qué procesos, etc.?).

Estructura del documento GML

Un fichero JPEG2000 puede contener diversos *codestreams*, y por este motivo el estándar define que cada uno de ellos debe estar representado, en el documento GML, por una *featureCollection* (denominada *codestream specific feature collection, csfc*) que es a su vez un miembro de la *featureCollection* raíz del documento.

Cada *csfc* debe contener un *rectifiedGrid* para contener la descripción la geometría de la imagen. También debe contener un elemento *rangeSet* para describir los valores de la imagen, es decir, la radiometría de ésta. Adicionalmente la *csfc* puede contener metadatos (*metadataProperty*) otras entidades (p.ej. anotaciones, etc). Esta información debe ser definida al crear el esquema de aplicación de la capa (XSD). La figura 1 muestra la estructura de un fichero GML instancia (XML) basado en el esquema de aplicación (XSD) de ejemplo del Anexo C del documento GMLJP2.

La versión actual del documento sólo se refiere a la creación de esquemas de aplicación para imágenes georeferenciadas, y se define que se debe usar el elemento *rectifiedGrid* para definir su

geometría. Este elemento requiere de la definición el origen, los vectores directores, los límites y el nombre de los ejes del *rectifiedGrid*. (Figura 2).

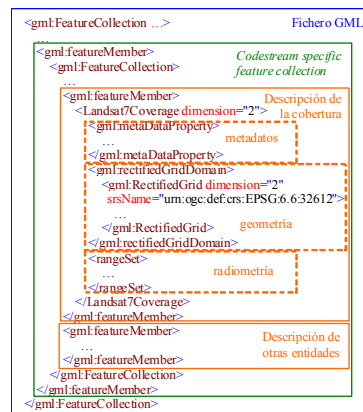


Figura 1.- Estructura de un fichero GML.

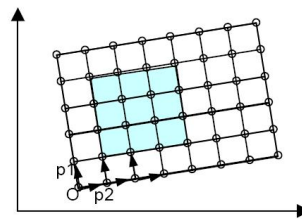


Figura 2.- Definición de los límites (en azul), el origen (punto O) y vectores p1 y p2 del georeferenced Grid de la figura.

La descripción de la radiometría se realiza con el elemento *rangeSet*. En GML este elemento puede estar formado por una de estas cuatro opciones: *gml:ValueArray*, *gml:ScalarValueList*, *gml:DataBlock* o *gml:File*. Esta última posibilidad es la que permite referenciar los valores codificados en una fuente externa, por ejemplo un fichero binario. En el caso de GMLJP2 se utiliza este elemento para referenciar el *codestream* que se está describiendo. Es necesario referenciar el *codestream* usando referencias URI dentro del fichero, por ejemplo "gmljp2://codestream/0" (en *fileName*).

El estándar también describe la estructura y las etiquetas que deben tener las diversas *xml boxes* necesarias que incorporan en el fichero JPEG2000 la información GML (Figura 3). Es obligatorio que exista como mínimo una instancia de fichero GML que describe la cobertura en una *box* etiquetada "gml.root-instance" (el contenido de la cual correspondería al GML de la Figura 1).

Adicionalmente otras XML *boxes* pueden incorporar los esquemas necesarios (*schema data*) así como diccionarios de sistemas de referencia y/o de unidades (*dictionary data*).

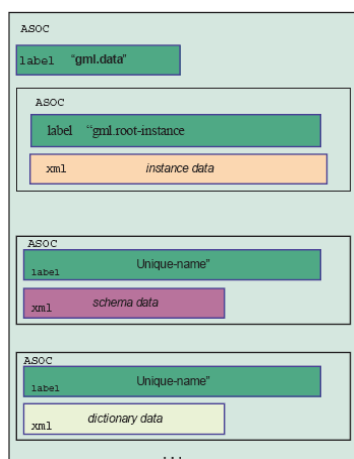


Figura 3.- Estructura de las “xml boxes” para incorporar el GML en el fichero JPEG2000.

METADATOS

Estándares ISO de metadatos

La necesidad de documentar correctamente los metadatos de la información geográfica es, desde hace ya algunos años, indiscutible. Existen diversos estándares de metadatos, pero el más ampliamente utilizado es el estándar ISO 19115, desarrollado por la *International Organization for Standardization* (ISO 2003). Este estándar es abstracto, es decir, que define los metadatos a documentar en cada caso, pero no indica cómo se debe realizar la implantación (en una base de datos, en un fichero de texto, etc.). El estándar ISO 19139 define una implementación XML del estándar ISO 19115 (ISO 2007).

Desgraciadamente los estándares de metadatos generales no recogen adecuadamente todos los metadatos requeridos para la documentación de datos ráster. El recientemente aprobado estándar ISO 19115-2 (ISO 2009) define una ampliación del anterior estándar para poder documentar metadatos necesarios en datos ráster como por ejemplo la información sobre la plataforma o el sensor que ha captado la imagen, la resolución nominal espacial, etc. No existe aún un estándar que defina la implementación XML de este estándar. Pero basándose en el modelo UML del estándar y en las reglas definidas en ISO 19139, es

posible generar los esquemas XML necesarios para trabajar con metadatos ISO19115-2 (disponibles en: <http://eden.ign.fr/xsd/isotc211>, IGN francés). Bravo *et al.* (2008) realizan un análisis de dicho estándar abstracto y su aplicabilidad a productos del Plan Nacional de Observación del Territorio, aunque en este artículo se usan los esquemas definidos en ISO 19139 incorporando la información nueva en el elemento “supplementalInfo”.

GMLJP2 y Metadatos generales de toda la cobertura

GML contiene una propiedad denominada *gml:metaDataProperty* que se puede usar para describir los metadatos generales de un objeto GML, en particular, de la cobertura descrita por el esquema de aplicación (ver Figura 1). Es posible definir que este elemento XML contenga los metadatos en un estándar basado en XML, por ejemplo, usando el tipo *gmi:MI_Metadata_Type* que permite la definición de metadatos siguiendo ISO 19115-2.

GMLJP2 y Metadatos que describen la cobertura

Un segundo nivel de metadatos cuando nos enfrentamos a un esquema de aplicación GML, son los metadatos que describen la geometría y radiometría de la imagen en el documento GML (elementos *rectifiedGrid* y *rangeSet* de la Figura 1).

GMLJP2 y Metadatos de cabeceras JPEG2000

El documento GMLJP2 reconoce todavía un tercer nivel de metadatos, aquellos incorporados en las cabeceras del fichero JPEG2000. Estos metadatos definen algunos metadatos mínimos (p.ej. número de filas y columnas de la imagen).

Recomendaciones GMLJP2

Como el propio documento GMLJP2 define, los metadatos en este contexto deben entender desde un punto de vista jerárquico. La norma especifica que si la información radiométrica y geométrica está en los dos primeros niveles, la descrita en el GML prevalece sobre la descrita en los metadatos generales. En cualquier caso y para evitar errores e incoherencias, la aplicación compresora que genera el fichero JPEG2000 con un esquema de aplicación GMLJP2 incrustado, debe procurar por la coherencia de los metadatos.

Metadatos JPX en JPEG2000

En el documento de extensiones del formato JPEG2000 (ISO/IEC 2004b) se describe un conjunto de metadatos disponibles así como el esquema XML que lo valida (anexo N del documento). La escritura y lectura de metadatos siguiendo este esquema es opcional. Su enfoque está más vinculado con la fotografía que con la información geográfica. Al no ser el estándar

habitual en metadatos de información geográfica se desaconseja su uso.

CONCLUSIONES

Es posible usar GML para describir las características geométricas y radiométricas de una imagen georeferenciada. Además, se puede añadir al esquema de aplicación de la imagen un conjunto completo de metadatos siguiendo el estándar ISO-19115-2 (usando los esquemas de los grupos de trabajo de metadatos).

La incorporación de los metadatos en el propio fichero JPEG2000 requiere del uso de extensiones del propio estándar, pero permite la deseada unidad datos-metadatos, por lo que es altamente recomendable.

Es necesario que los programas de descompresión de imágenes en formato JPEG2000 sean capaces de extraer los metadatos de las mismas para que el usuario pueda aprovecharlos.

El trabajo futuro requiere la verificación de los esquemas del estándar ISO 19115-2, así como la generación de metadatos para imágenes sin corrección geométrica. Este caso no se contempla en el estándar GMLJP2 pero podría ser implementado usando un esquema de aplicación adecuado (que utilice *gml:Grid* en vez de *gml:RectifiedGrid*.)

BIBLIOGRAFÍA

Bravo, M.J., Rodríguez, C., Domenech, E. 2008. Análisis de la norma ISO 19115-2 y su aplicación en proyectos de información raster españoles y europeos. In: *Actas de las V Jornadas de la Infraestructura de Datos Espaciales de España. IDE, aplicaciones al planeamiento y la gestión del territorio*. Ed. Cartográfica de Canarias, S.A. (GRAFCAN), ISBN: 978-84-691-7698-6.

ISO. 2003. *19115:2003: Geographic information – Metadata*. Geneva, Switzerland.

ISO. 2007. *19136:2007: Geographic information – Geography Markup Language (GML)*. Geneva, Switzerland.

ISO. 2007. *19139:2007: Geographic information – Metadata – XML schema implementation*. Geneva, Switzerland.

ISO. 2009. *19115-2:2003: Geographic information -- Metadata -- Part 2: Extensions for imagery and gridded data*. Geneva, Switzerland.

ISO/IEC. 2004. *15444-1:2004: Information technology – JPEG 2000 image coding system: Core coding system*. Geneva, Switzerland.

ISO/IEC. 2004b. *15444-2:2004: Information technology – JPEG 2000 image coding system: Extensions*. Geneva, Switzerland.

Open Geospatial Consortium (OGC). 2004. *OpenGIS® Geography Markup Language (GML) Implementation Specification*. Simon Cox, Paul Daisey, Ron Lake, Clemens Portele, Arliss Whiteside (ed). OGC 03-105r1. February 2004.

Open Geospatial Consortium (OGC). 2006. *GML in JPEG 2000 for Geographic Imagery (GMLJP2). Encoding Specification*. R. Lake, D. Burggraf, M. Kyle, S. Forde (ed). 05-047r3. January 2006.

Zabala, A., Pons, X., Moré, G., Serra, P. 2006. Efectos de la compresión de imágenes con pérdida (JPEG Y JPEG2000) en la clasificación digital de cultivos y zonas forestales. In: Camacho et al (ed.), *El acceso a la información espacial y las nuevas tecnologías geográficas*, 1333-1345, ISBN: 84-338-3944-6.

Zabala, A., Pons, X., Aulí, F., Serra, J. 2007. Implications of JPEG2000 lossy compression on multiple regression modeling. In: Ehlers M, Michel U (ed.), *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VII*. ISBN: 9780819469076

Zabala, A., Pons, X., Aulí, F., Serra, J. 2008. Image compression effects in visual analysis, In: *SPIE International Conference on Satellite Data Compression, Communication and Processing. Society of Photo-optical Instrumentation Engineers (SPIE)*, vol. 7084, August, ISSN: 9780819473042.

AGRADECIMIENTOS

La realización de este trabajo ha sido posible en parte gracias a la financiación aportada por los proyectos de investigación TIC2003-08604-C04-03 (Ministerio de Ciencia y Tecnología, fondos FEDER) y TSI2006-14005-C02-02 (Ministerio de Ciencia e Innovación, fondos FEDER). También deseamos expresar nuestro agradecimiento a la Agència Catalana de l'Aigua y al Departament de Medi Ambient i Habitatge de la Generalitat de Catalunya por su política de inversión y disponibilidad de datos de Teledetección, la cual ha permitido abordar este estudio con plenas garantías.