A synchronisation approach to automate spatial metadata updating process

This paper presents a new approach to automate spatial metadata updating process, by which dataset properties are read from the dataset file and written into its metadata file automatically.

Metadata is commonly defined as “data about data” and also plays a critical role in any Spatial Data Infrastructure (SDI) initiatives. Metadata not only provides users of spatial data with information about the purpose, quality, actuality and accuracy of spatial datasets, but also performs vital functions that make spatial data interoperable, that is, capable of being shared between systems. Metadata enables both professional and non-professional spatial users to find the most appropriate, applicable and accessible datasets for use (Rajabifard et al., 2009).

Regardless of numerous benefits of metadata, issues and obstacles to the creation and update of such geospatial surrogates are numerous. Spatial metadata which is created and updated manually or semiautomatically, is considered as monotonous and time consuming, a labour-intensive process by organisations and it is commonly viewed as an overhead and extra cost. Also, metadata for spatial datasets is often missing or incomplete as it relates to, and is often managed by people with a limited knowledge of its value. Separation of storage creates two independent datasets that must be managed and updated - spatial data and metadata. These are often redundant and inconsistent. Thus the reliability of spatial information and the extent it can be used are unclear.

To address some of these issues, particularly relevant to spatial metadata update processes, this paper aims at exploring a new synchronisation approach as an automated fashion for updating spatial metadata which is based on an ongoing research by authors on “Spatial Metadata Automation”. The paper firstly compares different methods of spatial metadata generation and focuses on an automation framework. This framework embraces three streamlines including create, update and enrich. Finally, a new synchronisation approach is introduced to address the automatic update streamline.

Spatial metadata generation approaches

The generation of spatial metadata can be separated into automatic, semiautomatic and manual data mining methods (Taussi, 2007). These approaches have been formed and evolved based on the technological initiatives over time and the characteristics of spatial metadata such as type and format have been influenced by these initiatives. For instance, after the PC Era and Internet initiative the spatial metadata were generated in Markup Languages (e.g. Hyper Text Markup Language (HTML) and eXtensible Markup Language (XML)) since early 1990s. Figure 1 illustrates the spatial metadata creation approaches and different types of spatial metadata based on technological initiatives.

Among these approaches, many people view manual metadata generation as monotonous and time consuming, a labour-intensive process which is a major undertaking in itself (West and Hess, 2002), resulting in a pervasive outlook which shuns metadata creation (Mathys, 2004). Moreover, it is commonly viewed...
by organisations as an overhead and extra cost. Also, metadata for spatial datasets is often missing or incomplete and is acquired in heterogeneous ways (Rajabifard et al., 2009).

The use of automatic processing can, in turn, permit human resources to be directed to more intellectually challenging metadata creation and evaluation tasks. These factors underlie automatic metadata generation research efforts and the desire to build superior and robust automatic metadata generation applications (Greenberg et al., 2005). More importantly, the ability to automatically generate metadata relating to spatial data, and make it available through SDI will have important benefits all practitioners including spatial data producers, vendors, distributor and user. Many organisations are also looking at automated metadata systems to re-automate metadata generation benefits. This is evidenced by the large number of projects and companies who are creating programs which automatically produce metadata (Baird and Jorum Team, 2006). Accordingly, a conceptual framework for spatial metadata automation which has been introduced by (Kalantari et al., 2009) is reviewed as below.

Spatial metadata automation framework

Today, automatic metadata generation should move beyond subject representation to encompass the production of author, title, date, format, spatial extension and many other types of metadata. In addition, thousands of spatial databases are now networked via the Internet, and information resources are frequently rendered in open and interoperable standards (e.g. XML). These developments should enable automatic metadata generation systems to work on far larger spatial data directories.

For that reason, a framework for automating spatial metadata which is based on three main streamlines including automatic creation, enrichment and update is illustrated in figure 2 (Kalantari et al., 2009).

**Automatic Creation:** When there is no existing metadata associated with spatial data, there is a need for exploring methods to create spatial metadata. Several automatic metadata extraction methods have been studied so far, e.g. hand-coded rule-based parsers and machine learning (Han et al., 2003).

**Automatic enrichment:** Automatic enrichment involves improving content of metadata through monitoring tags that are used by users for finding datasets. This kind of spatial metadata can help describing an item and allowing it to be found again by browsing or searching.

**Automatic update:** Automatic spatial metadata update or synchronisation is a process by which properties of a spatial dataset are read from the dataset and written into its spatial metadata. This automatic function will support the spatial data to be updated at the same time with its related spatial data update process. However, the automatic update implementation still faces with some obstacles and restrictions which have been discussed as following.

**Automatic spatial metadata update – Current restrictions**

Automatic update is one of the main streamlines of automation framework which is regarded with some obstructions. The structure of spatial data and metadata data models is an important part of these limitations. Whereas, dataset creation and editing are detached from metadata creation and editing procedures, necessitating diligent update practices involving at minimum two separate applications (Batcheller, 2008). Rajabifard et al. (2009) also stated that separation of storage creates two independent datasets that must be managed and updated - spatial data and metadata. These are often redundant and inconsistent. Thus the reliability of spatial information and the extent it can be used are unclear. They also discussed the significance of an integrated data model for handling spatial metadata by combining spatial data and metadata in a seamless approach. The research in metadata integration should focus on utilise metadata standards and developments in order to combine metadata and spatial data within an integrated package, so that the process of updating or creating spatial data and metadata – where feasible – becomes one process rather than two.

As a result of this, automatic update should provide a synchronised process through which the spatial data and metadata can be updated simultaneously. In other words, this synchronisation process not only should complete as much of the metadata elements as possible automatically but also it should make sure that the metadata is kept up-to-date with changes to the dataset. ESRI Company through ArcCatalog application has developed some algorithms to synchronise the metadata content when values in the spatial data change. For instance, when a change occurs with a spatial data property such as its projection, the metadata will be updated with the new information (Westbrooks, 2004). The process of synchronisation is accomplished using metadata standard specific synchronizers. For example, three synchronizers are provided with ArcCatalog: an FGDC synchronizer, an ISO synchronizer, and a Geography Network synchronizer.

However, the current synchronisation process generates and updates a limited...
A synchronisation approach to automate spatial metadata update

Following the requirements for automatic update or synchronisation implementation, a new approach based on Geography Markup Language (GML) has been developed. GML is rapidly emerging as a world standard for the encoding, transport and storage of all forms of geographic information (Lake, 2005). The OGC proposed GML specifications that take advantage of XML to apply to geographic information sharing. In fact, GML is an XML grammar for expressing geographical features. GML serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet. As with most XML based grammars, there are two parts to the grammar – the schema that describes the document and the instance document that contains the actual data (OGC, 2009).

Using this method, practitioners may decide to store geographic application schemas and information in GML, or they may decide to convert from some other storage format on demand and use GML only for schema and data transport (OGC, 2007). GML provides several objects for describing geography, including features, coordinate reference systems, geometry, topology, time, units of measure, and generalized values. Applications can extend or restrict these GML objects to fit their requirements (Huang et al., 2009).

Although, GML does not provide an information model for metadata, instead a mechanism to include or reference metadata is provided for all object elements. Indeed, GML provides a framework by which arbitrary user-defined metadata can be attached to any GML object and be distinguished from the defining properties of the object. This is supported through the metadata property which can be optionally attached to anything derived from gml:AbstractGMLType. This metadata property points to or contains a metadata package of properties that are the metadata for the object in question. The content of the metadata package is defined by a metadata application schema (a property list), similar in structure to a GML application schema for features (Lake, 2005). For instance, if metadata following the conceptual model of ISO 19115 is to be encoded in a GML document, the corresponding implementation specification specified in ISO/TS 19139 shall be used to encode the metadata information (OGC, 2007).

With this in mind, the new synchronisation approach is developed based on XML/GML technologies (figure 3); as Huang et al. (2009) also claim that no GIS has been built on native XML/GML technologies so far.

In this new approach, metadata publishers continue creating or updating spatial datasets in required formats (e.g. shape files, cad files, etc.). Then each dataset is transformed to GML after creation or update through a transformation method. To implement this transformation, proper GML application schemas should be designed to encode the maximum range of metadata elements in a GML schema. Through the transformation, an instance document to contain the actual data and a GML schema to describe the document would be provided. Therefore, after the creation of dataset in GML format the synchronisation process would start. Through this process, spatial metadata elements which are encoded in GML.
document would be identified based on a specific standard (e.g. ISO 19115) and extracted via an automatic extraction method and finally written into an XML document (based on XML application schema, e.g. ISO 19139) automatically. Actually, the synchronisation process output is metadata related to spatial dataset in XML format. Whenever a spatial dataset in GML format is updated, the synchroniser would be triggered and the spatial metadata would be updated in XML automatically; that is, spatial metadata will be updated automatically with any change in spatial dataset.

Conclusion

To addressing the issues regarding current spatial metadata updating process, a new synchronisation approach based on GML has been proposed. This new approach to updating spatial metadata automatically will benefit the spatial data and metadata publishers in different aspects. Firstly, it encourages the publishers to create spatial datasets in an international open standard which will help solve the interoperability issues relevant to spatial data transfer and storage through the web environment. Secondly, this approach will assist the publishers to update the spatial data and metadata simultaneously, thus more time, resources and energy could be saved through reducing the number of update processes. In addition, the approach based on GML as an open and neutral framework for spatial data will decrease the publishers’ concerns on spatial data creation and update methods and output formats. Moreover, a large number of spatial metadata elements could be updated automatically through the new approach. Furthermore, less-complicated synchronisation algorithms are required in this approach. Finally, this new process will minimize the risk of spatial data and metadata inconsistency and redundancy.

Acknowledgements

This paper is based on an ongoing research project titled “Spatial Metadata Automation” as an Australian Research Council (ARC) linkage project, which aims to develop and demonstrate an approach for extracting, recording, updating and delivering metadata in an automated and integrated fashion. The research is also supported by industry partners; Department of Sustainability and Environment and Department of Primary Industries - Victoria, Department of Lands - New South Wales, AusSoft Solutions Pty Ltd, CubWerx Australia Pty Ltd and Logica CMG. The Authors acknowledge the support of the members of the Centre for Spatial Data Infrastructures and Land Administration, at the Department of Geomatics, University of Melbourne in the preparation of this paper and associated research; however, the views expressed in this paper are those of the authors and not the views of these groups.

References


