

Automatic Spatial Metadata Update: a New Approach

Hamed OLFAT, Abbas RAJABIFARD and Mohsen KALANTARI, Australia

Key words: Spatial metadata update, Automation, Synchronisation, GML, XML

SUMMARY

Spatial metadata is a vital tool for spatial data management, retrieval and distribution. It is also a critical component for any spatial data sharing platform which provides users with information about the purpose, quality, actuality and accuracy of spatial datasets. With the amount of spatial data exchanged through the web environment, the demand for automatic spatial metadata creation and updating to describe such resources is increasing. However, automatic spatial metadata updating is still in its infancy and automatic approaches are being explored by researchers.

So far different processes and tools have been developed which generate and update a limited number of spatial metadata elements in different standard schemes automatically, thus a large amount of spatial data elements need to be imported manually. In order to improve this situation, this paper aims at exploring a new synchronisation approach based on XML/GML technologies to automate spatial metadata update process, by which dataset properties are read from the dataset file and written into its metadata file automatically.

The paper first discusses the important role of metadata in Spatial Data Infrastructures (SDIs) as an enabling platform and proposes an architecture to manage spatial metadata. It then compares different methods of spatial metadata generation and presents a spatial metadata automation framework. Based on this framework, the paper finally introduces a synchronisation approach to achieve the spatial metadata automatic update.

Automatic Spatial Metadata Update: a New Approach

Hamed OLFAT, Abbas RAJABIFARD and Mohsen KALANTARI, Australia

1. INTRODUCTION

Metadata is commonly defined as "data about data" and is the key to ensuring that resources will survive and continue to be accessible into the future (NISO, 2004). We now face an increase of spatial datasets being created and exchanged between people or organisations. As more data and information is produced, it becomes more vital to manage and locate such resources (Göbel and Lutze, 1998). The role spatial metadata plays in the management and location of these resources has been widely acknowledged (Tsou, 2002; Limbach et al., 2004).

Furthermore, spatial data can now be easily downloaded from the Internet, e.g., spatial data catalogues (Devilleers et al., 2002). With the increasing distribution of spatial data over the Internet there is a corresponding increasing demand for spatial metadata describing the spatial data in a networked environment.

Metadata also plays a critical role in any SDI initiatives. One of the first steps for the setting up of an SDI is the creation of metadata standards and a corresponding metadata catalogue (Pasca et al., 2009). These not only provide users of spatial data with information about the purpose, quality, actuality and accuracy of spatial datasets, they also perform the 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).

In spite of the numerous benefits of metadata, the remaining issues and obstacles to the creation and updating 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 is commonly viewed as an overhead and extra cost. Also, metadata for spatial datasets is often missing or incomplete and is acquired in heterogeneous ways. Moreover, metadata is usually created and stored separately to the actual dataset it relates to, and is often managed by persons 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 updating processes, this paper aims at exploring a new synchronisation approach as an automated process for updating spatial metadata, by which dataset properties are read from the dataset file and written into its metadata file automatically. This is based on ongoing research by authors on "Spatial Metadata Automation". This paper first proposes an architecture to manage spatial metadata in the SDI context and then compares different methods of spatial metadata creation and updating and focuses on an automation framework. This framework embraces three

streamlines of *create*, *update* and *enrich*. Finally, a new synchronisation approach is introduced to address the automatic updating streamline.

2. SPATIAL METADATA MANAGEMENT ARCHITECTURE

The creation of an enabling platform such as SDI for the delivery of spatial data and tools will allow users from diverse backgrounds to work together with current technologies to meet the dynamic market place (Rajabifard et al., 2005). Within an SDI platform, metadata plays a key role to facilitate accessing up-to-date and high quality spatial data and services (Williamson et al., 2003).

Within the SDI context, developing an architecture which covers the metadata management process from spatial metadata preparation to its publication in a networked environment is fundamental. To achieve this purpose, an overall architecture for spatial metadata management has been developed (figure 1).

The presented architecture includes components such as publishers, registry service, catalogue service, metadata editor, metadata repository, applications and requesters. In this architecture, spatial metadata publishers publish spatial metadata to web environment through web services. Metadata records after publishing are registered in data catalogues through registry services. The catalogue service handles the discovery and publishing of metadata entries and harvests metadata records from other repositories. Spatial metadata repositories store metadata records which are published. Moreover, existing metadata records stored in repositories can be manipulated and updated through metadata editors. Lastly, different applications such as spatial metadata discovery, viewing, and access would be available for requesters (users and value-adders) via web services.

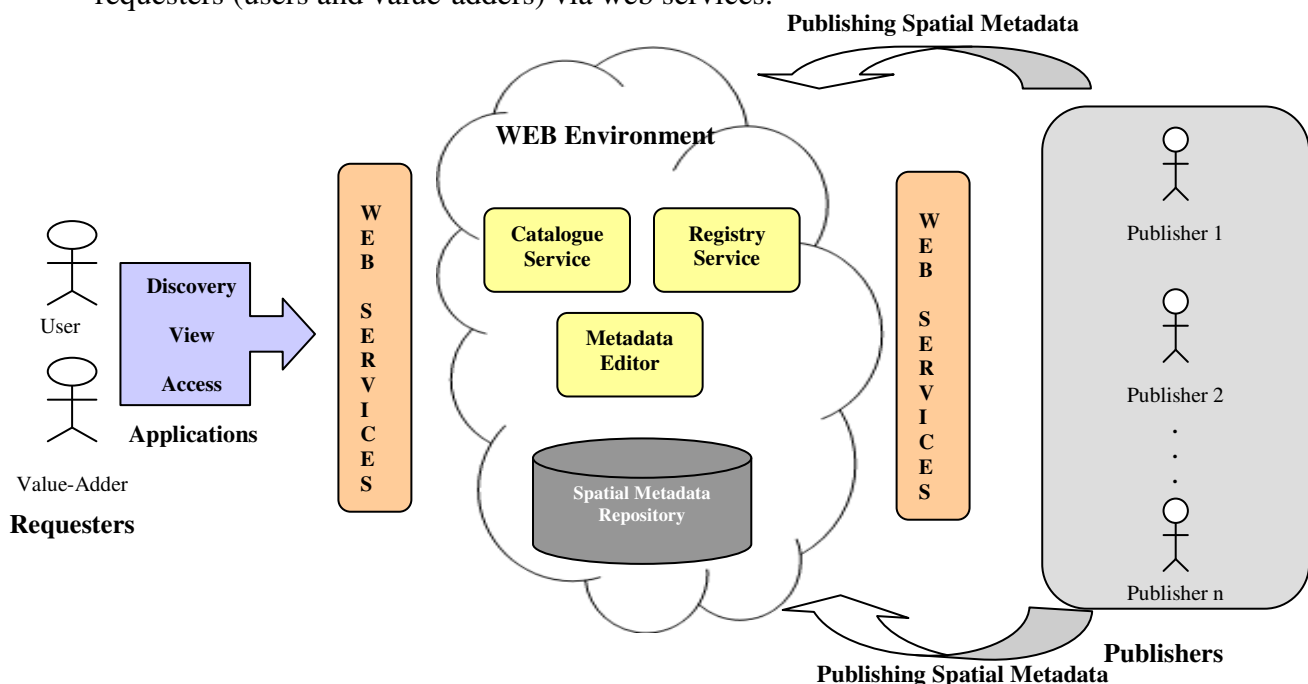


Figure 1. Spatial Metadata Management Architecture

In this architecture, standardised spatial metadata is a powerful tool that enables the requesters to discover and select the relevant spatial datasets quickly and easily. Hatala and Forth (2003) also concluded that metadata can fulfil its purpose only when it complies with some agreed upon standard. A metadata standard provides a set of elements, defines their meanings, and provides guidelines and constraints on how to fill element values. On the other hand, in order for this architecture to be efficient and effective, it is essential that up-to-date spatial metadata be delivered in a specified standard to the requesters. As the demand for standardised metadata increases, spatial industry needs to identify automated metadata production methods that are more efficient and less costly than those practices involving manual production (Greenberg, 2004). Automatic metadata generation can be facilitated when the structure of metadata is based on a selected standard. However, there are different standardisation methods in metadata domains as discussed below.

2.1 Spatial metadata standards

International organisations have been working for several years in order to achieve a common standard regarding metadata for spatial information. The most important ones are ISO 19115 (Kresse and Fadaie, 2004), FGDC (The Federal Geographic Data Committee), DIF (Directory Interchange Format), and Dublin Core (Schindler and Diepenbroek, 2008).

In recent years, the popularity of the ISO 19115 standard which provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data (Moellering et al., 2005), some organisations have decided to adopt profiles of this standard. For instance, the Australia New Zealand Land Information Council (ANZLIC) released an Australian/New Zealand profile of AS/NZS ISO 19115:2005, Geographic information-Metadatas (implemented using ISO/TS 19139:2007, Geographic information-Metadatas-XML schema implementation) in August 2007 (ANZLIC, 2009). Also through efforts seeking for collaboration, U.S. and Canadian scientific volunteers from the International Committee for Information Technology Standards Technical Committee L1 (INCITS/L1) and the Canadian General Standards Board Committee on Geomatics (CGSB-COG) developed the INCITS 453-2009, the North American Profile (NAP) of ISO 19115: 2003, Geographic Information – Metadatas, to meet the requirements of both countries in July 2009 (FGDC, 2009). In addition, the Permanent Committee on GIS Infrastructure for Asia & the Pacific (PCGIAP) has recently released a draft version of Asia-Pacific Geospatial Metadata Profile based on ISO/IS 19115 core metadata elements and other regional profiles as well as the survey results of participating countries.

The standardisation of the spatial metadata structures will bring efficiency to metadata management by providing a common understanding of metadata elements; however the challenge of metadata automation lies in metadata being human readable/understandable rather than machine readable/understandable. The next section discusses metadata generation approaches and the importance of metadata automation with emphasis on metadata being machine interpretable.

3. SPATIAL METADATA GENERATION APPROACHES

The generation of spatial metadata can be separated into automatic, semi-automatic and manual data mining methods (Taussi, 2007) as illustrated in Figure 2. Automatic methods, such as automatic retrieval or searching and sorting of data are based on computerisation. Semiautomatic methods combine automatic and manual methods. Manual methods are based on human reasoning and decision making.

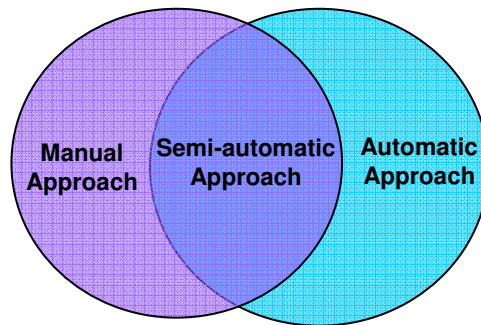


Figure 2. Spatial Metadata Generation Approaches

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 the early 1990s. Figure 3 illustrates the spatial metadata creation approaches and different types of spatial metadata based on technological initiatives.

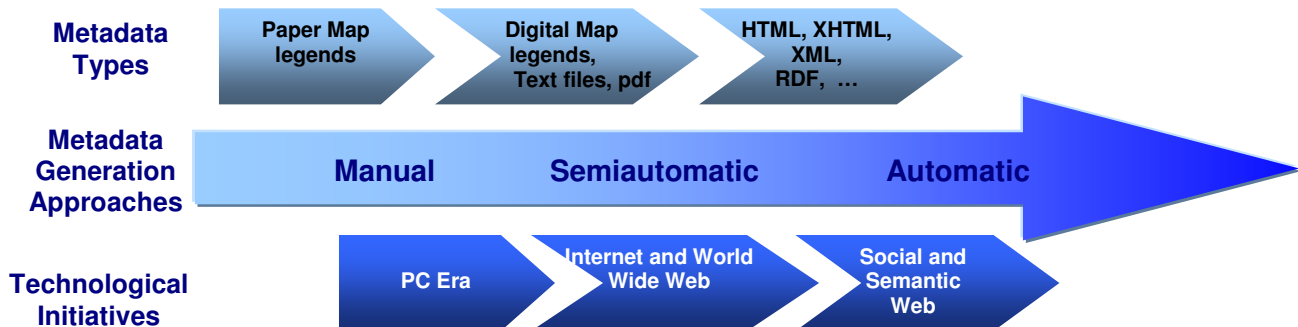


Figure 3. Spatial Metadata Generation Approaches, Types & Technological Initiatives

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 (Guptill, 1999; West and Hess, 2002), resulting in a pervasive outlook which shuns metadata creation (Mathys, 2004). Meanwhile, one of the main obstacles to the widespread adoption of systems which make intensive use of metadata is the time and effort required to apply metadata to multiple resources and the inconsistencies and idiosyncrasies in interpretation that arise when this is a purely human activity (Hatala and Forth, 2003). Moreover, it is commonly viewed by organisations as an overhead and extra cost. Finally, 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 to all practitioners including spatial data producers, vendors, distributors and users. Many organisations are also looking at automated metadata systems to reap automatic metadata generation benefits. This is evidenced by the large number of projects and companies who are creating programs which automate metadata (Baird and Jorum Team, 2006). In the next section a framework for spatial metadata automation is introduced.

3.1 Spatial metadata automation framework

The idea of automatic spatial metadata generation research is rooted in automatic indexing, abstracting, and classification of spatial data content, which began with the need to organise increasing amount of spatial related data and inability of human-authored methods to cope with huge amount of spatial metadata (Rajabifard et al., 2009). 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.

Although automated metadata generation is still in its infancy and there is no conceptual framework to define, several approaches have emerged, including metatag harvesting, content extraction, automatic indexing or classification, text and data mining, social tagging, and the generation of metadata from associated contextual information or related resources (Polfreman and Rajbhandari, 2008). A framework for automating spatial metadata which is based on three main streamlines including automatic creation, enrichment and updating, has been introduced by (Kalantari et al., 2009) as illustrated in figure 4.

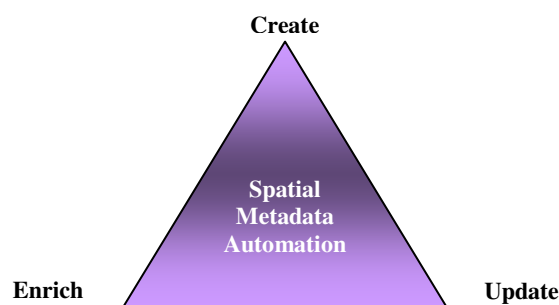


Figure 4. Spatial Metadata Automation Framework

Automatic Creation: When there is no existing metadata associated with spatial data, there is a need for exploring methods to create spatial metadata. Humans create metadata by writing descriptions of resources either in a structured or unstructured form. Computer applications

can extract certain information from a resource or its context. Several automatic metadata extraction methods have been studied, e.g. hand-coded rule-based parsers and machine learning (Han et al., 2003). For highly structured tasks rule-based methods are easy to implement. The resulting rule system is usually domain-specific and cannot be easily translated for use in other domains. Machine learning, on the other hand, is more robust and efficient (Han et al., 2003). Several learning models are available. Among the most popular are the Naïve Bayes model (NB), the Hidden Markov Model (HMM), Support Vector Machines and Expectation Maximization. Supervised machine learning (SML) algorithms include training data and machine self-correction based on errors in machine performance against the training set (Greenberg et al., 2006).

Manso et al. (2009) also discussed how to automatically produce metadata items compliant with ISO 19115 standard to support dynamic interoperability by extracting the information stored in files and databases through computations or by inference. As a result, they stated that of the 151 metadata items providing dynamic interoperability, 54 of them (including raster data, Digital Terrain Model (DTM) and vector data) may be automatically produced (35%). They also concluded that although this value is quite high, it should be cautiously interpreted since it is a “hopeful” value representing the ceiling of the automatic production.

Automatic enrichment: Automatic enrichment involves improving the content of metadata through monitoring tags that are used by users for finding datasets. A tag is a non-hierarchical keyword or term assigned to a piece of information (such as an internet bookmark, digital image, or computer file). Tagging was popularised by websites associated with Web 2.0 and is an important feature of many Web 2.0 services (Mika, 2005). This kind of spatial metadata can help in describing an item and allowing it to be retrieved by browsing or searching. Spatial tags will be chosen informally and personally by the spatial data creator or by its users, depending on their use. On a spatial data directory if many users are allowed to tag many spatial data, this collection of tags can become a spatial folksonomy a method that can collaboratively create and manage metadata to annotate and categorize spatial data (Kalantari et al., 2009).

Automatic updating: Automatic spatial metadata updating 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 metadata to be updated at the same time with its related spatial data update process. Therefore, it will benefit the organisations associated with spatial metadata to save time and effort and will also reduce the risk of inconsistency and redundancy in the spatial data and metadata. Following the predictable advantages of automatic updating, it has increasingly been investigated by researchers. However, the automatic update implementation still faces some obstacles and restrictions which are discussed in the next section.

4. AUTOMATIC SPATIAL METADATA UPDATE – CURRENT METHODS AND CHALLENGES

Automatic updating is one of the main streamlines of the 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 updating practices involving at a minimum two separate applications (Batcheller, 2008). Rajabifard et al. (2009) also state 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 to which it can be used are unclear. They also continued by discussing 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 utilising 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.

However, some elements of metadata obviously cannot be automatically updated. These would not be stored in an integrated fashion with the spatial data. Only those metadata elements that can be automatically updated would be integrated with the spatial data. This will save producers of data both time and money associated with the updating of metadata records, and will also aid data users who require up-to-date metadata to be delivered with data for their use (Rajabifard et al., 2009).

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. ArcCatalog automatically creates metadata for datasets stored in the geo database if none exist. Some of the automatically generated metadata describe the dataset's current properties, i.e coordinate system, entity, and attribute information. Every time the metadata librarian views the metadata, ArcCatalog automatically updates or synchronises dataset properties with its most current values. Of course, the synchronisation ensures that the metadata is perpetually up-to-date according to the changes in the dataset (Westbrooks, 2004).

The process of synchronisation is accomplished using metadata standard specific synchronisers. For example, three synchronisers are provided with ArcCatalog: an FGDC synchroniser, an ISO synchroniser, and a Geography Network synchroniser. Figure 5 illustrates the mechanics of synchronisation process in ArcCatalog.

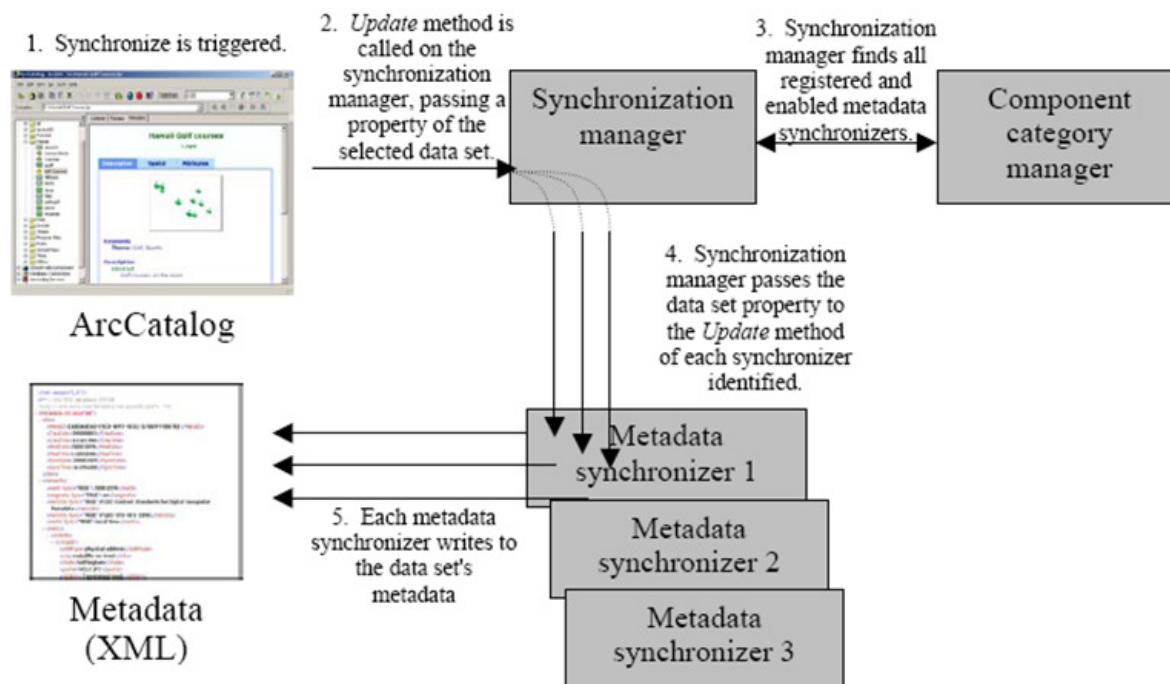


Figure 5. Mechanics of the Synchronisation Process in ArcCatalog (ESRI, 2002)

Although automatic synchronisation is invaluable, it brings forth numerous problems associated with archiving and bibliographic control (Westbrooks, 2004). Making distinctions between metadata versions, editions, and updates is crucial for any type of digital library with archiving responsibilities. The inability of the synchroniser to differentiate a version of a metadata record from an edition or update introduces a new set of challenges.

In addition, the current synchronisation process generates and updates a limited number of spatial metadata elements in different standard schemas automatically and a large amount of spatial data elements should be imported manually. In other words, the current synchronisation process is undertaken semiautomatically.

Moreover, spatial data are usually created and stored by organisations in different formats (e.g. Shp, Dwg, Dxf, Coverage, Dgn, etc.) which make the synchronisation process complex. In fact, complicated algorithms should be provided to support the synchronisation process to update the spatial metadata associated with these diverse spatial datasets.

As a result, in order to implement the synchronisation process especially in terms of automating this process as much as possible and also supporting different spatial dataset formats, a new approach has been proposed in the next section.

5. AUTOMATIC SPATIAL METADATA UPDATING - A NEW APPROACH

Following the requirements for automatic updating or synchronisation implementation, a new approach based on Geography Markup Language (GML) is under development. In fact, using GML as a common standard for which various datasets can be translated to would benefit the

synchronisation process in terms of using less-complicated algorithms and also saving time, resources and efforts.

GML is rapidly emerging as a world standard for the encoding, transport and storage of all forms of geographic information (Lake, 2005). GML is an XML grammar for expressing geographical features and serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet. Indeed the OGC (Open Geospatial Consortium) has proposed GML specifications that take advantage of XML to apply to geographic information sharing. Batcheller et al. (2009) also state that the appearance of GML has helped alleviate many of the concerns relating to data compatibility and interoperability, providing an open dialect for data transfer not bound to specific software offerings.

Whereas GML is a Markup Language, it means that GML document has to follow certain rules in order to be a valid GML document. This set of rules is defined in a schema document. The documents should conform to the requirements in the GML specification. GML version 1.0 uses the Document Type Descriptors (DTDs) for defining the structure, the elements and the associated attributes for a feature. GML version 2.0 and 3.0 use XML schema instead of DTD. GML application schema is also an extension of XML Schema and provides a set of type definitions and element declarations that can be used to check the validity of well-formed GML documents (Paul and Ghosh, 2008).

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 example, 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).

Moreover, once comparing the official GML core schemas by OGC 2007 (e.g. features, geometric primitives, coordinate reference systems, topology, temporal information and dynamic features, units, etc) and ISO 19115, Geographic information- metadata schemas (ISO 2003) such as identification, constraints, data quality, maintenance, spatial representation, content, etc, it has resulted that there are a number of common elements between these two standards which could be mapped from the GML document to the metadata file. The

identification of these common metadata elements has been recently under investigation during the current research and as part of the experience Table 1 shows the potential GML elements which could be mapped to ISO core metadata elements. Table 2 also illustrates an example of mapping the dataset geographic extent element from GML to ISO metadata standard.

Table 1. Mapping GML elements to ISO 19115 metadata elements

Related GML 3.2.1 element(s)	ISO 19115 core elements (M: Mandatory, O: Optional, C: Mandatory under certain conditions)
AbstractGML, AbstractFeature, AbstractFeatureCollection	Dataset title (M) Abstract describing the dataset (M)
AbstractTimePrimitive	Dataset reference date (M)
boundedBy	Geographic location of the dataset (by four coordinates or by geographic identifier) (C)
AbstractGML	Dataset character set (C)
AbstractGeometry	Spatial resolution of the dataset (O)
VerticalDatum, VerticalCS, TemporalCS, TemporalDatum	Additional extent information for the dataset (vertical and temporal) (O)
AbstractCRS	Reference system (O)
TimeTopologyComplex	Lineage (O)
AbstractMetadata	On-line resource (O)
	Metadata file identifier (O)
	Metadata standard name (O)
	Metadata standard version (O)
	Metadata language (C)
	Metadata character set (C)
	Metadata point of contact (M)
	Metadata date stamp (M)
	Distribution format (O)
	Dataset topic category (M)
	Dataset language (M)
Dataset responsible party (O)	

Table 2. Mapping the dataset geographic extension element from GML to ISO

Schema for encoding dataset geographic extension in GML 3.2.1	Schema for encoding dataset geographic extent in ISO 19115
<pre><element name="boundedBy" nillable="true" type="gml:BoundingShapeType"> <annotation> <documentation>This property describes the minimum bounding box or rectangle that encloses the entire feature.</documentation> </annotation> </element></pre>	<pre><xs:element name="EX_GeographicBoundingBox" substitutionGroup="gmd:AbstractEX_GeographicExtent" type="gmd:EX_GeographicBoundingBox_Type" /></pre>

Following the discussed relation of GML standard and metadata elements, the new synchronisation approach is based on XML/GML technologies (figure 6).

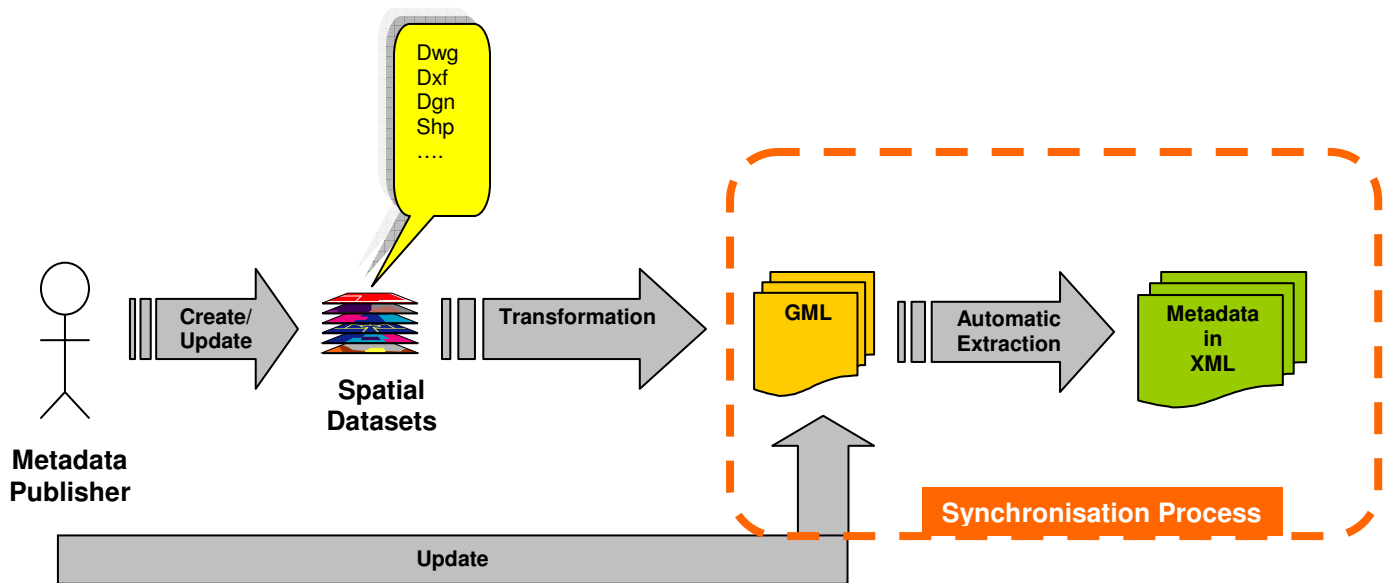


Figure 6. A new approach to automatic spatial metadata update

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 updating through a transformation method. To implement this transformation, proper GML application schema should be designed to encode the maximum range of metadata elements in the schema. Through the transformation, an instance document to contain the actual data and a GML schema to describe the document would be provided.

In order to seamlessly translate, transform, integrate and distribute various spatial data in hundreds of formats (e.g. GML, GIS and CAD formats, raster formats, etc.), some software vendors have developed appropriate solutions. For instance, Safe Software's FME (Feature Manipulation Engine) is a spatial ETL (Extract, Transform and Load) platform that helps the publishers easily solve the complete spectrum of data interoperability challenges, including managing proprietary and evolving data formats, adapting to new schemas and lack of standards and difficulties accessing, restructuring, integrating and distributing data (Safe Software, 2009).

Therefore, after the creation of dataset in the 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. In fact, 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.

The most important consideration in this approach is that most applications so far make use of only a subset of GML core schemas as per the requirement of the domain (Paul and Ghosh, 2008).

In order to better conceptualise the process of mapping the metadata elements from GML to ISO, a dataset (titled “town”) in ESRI shape file format has been transformed to GML by FME Universal Translator. As a result of this, the output GML document has been generated as shown in figure7.



Figure 7. GML document of “town” dataset

Based on this document, and comparing the geographic elements encoded in that and the core metadata elements suggested by ISO 19115, the following elements could be extracted as metadata elements and mapped to the metadata XML file as tagged in figure 7:

- Dataset title: by <fme:town>
- Geographic location of the dataset: by <gml:boundedBy>
- Dataset content: by <gml:featureMember>

The number of identified metadata elements is dependant to the transformer application as well as the GML application schema applied by the transformer. Thus, the new synchronisation approach under development in this research aims to expand the number of metadata elements which could be extracted through GML documents.

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 simplify 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 saving more time, resources and energy through reducing the number of updating processes. Additionally, 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.

6. CONCLUSION AND FUTURE DIRECTIONS

Spatial metadata can be created and updated through manual, semiautomatic and automatic approaches. The two first approaches are considered as monotonous, time consuming, and labour-intensive processes by organisations and they are commonly viewed as an overhead and extra cost. Moreover, metadata for spatial datasets is often missing or incomplete and is acquired in heterogeneous ways. Therefore, automatic metadata creation and update is being explored by researchers due to important efficiency, cost, and consistency advantages over manual and semiautomatic processes.

Along these lines a spatial metadata automation framework has been considered which includes automatic creation, enrichment and update. In this framework, automatic update or synchronisation is a process by which spatial metadata elements are read from dataset and written into its spatial metadata file automatically. However, automatic update currently faces some restrictions. The structure of spatial data and metadata storage in a separate fashion and also variety of spatial dataset formats which should be considered in synchronisation process could be regarded as some of these limits.

Therefore, a new synchronisation approach based on GML as an international standard for geographic data encoding, transfer and storage is proposed to address the automatic update current restrictions. Thus, the spatial data should be transformed to GML via a transformation method before the synchronisation process begins. The output content of GML document also depends on the transformer application design as well as the application schema that is used. Through the synchronisation process which is independent of spatial dataset formats, spatial metadata elements are extracted from GML document and written into XML metadata file. Consequently, spatial metadata could be updated at the same time with spatial data update process. In addition, this proposed approach not only saves time, resources and efforts spent by data publishers to update spatial metadata, but also increases the number of spatial metadata elements which can be updated automatically.

Following the new synchronisation process based on extracting spatial metadata elements from GML document, investigating the relation between GML standard encoding specifications and ISO 19115 metadata elements is proposed as the future research direction for improving automatic metadata update process.

ACKNOWLEDGEMENTS

This paper is part of an ongoing research project titled “Spatial Metadata Automation Project” and is supported by Australian Research Council (ARC) - linkage program. 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, CubeWerx and Logica. 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

- ANZLIC (2009). Spatial metadata, ANZLIC website, http://www.anzlic.org.au/infrastructure_metadata.html (accessed 3 July 2009)
- Baird, K., and Jorum Team (2006). Final report for automated metadata, A review of existing and potential metadata automation within Jorum and an overview of other automation systems.
- Batcheller, J.K. (2008). Automating geospatial metadata generation—An integrated data management and documentation approach, *Computers & Geosciences 2008 - ELSEVIER*, pp.387–398
- Batcheller, J.K., Gittings, B.M., Dunfey, R.I. (2009), A Method for Automating Geospatial Dataset Metadata, *Future Internet 2009*, Volume 1, pp. 28-46
- Devillers, R., Gervais, M., Bédard, Y., Jeansoulin, R. (2002). Spatial Data Quality: From Metadata to Quality Indicators and Contextual End-User Manual, *Proceedings of OEEPEISPRS Joint Workshop on Spatial Data Quality*, 20–21 March, Istanbul, Turkey.
- ESRI (2002). Creating a custom metadata synchroniser, An ESRI White paper, July 2002, <http://www.esri.com/library/whitepapers/pdfs/metadata.pdf>
- FGDC (2009). North American Profile, FGDC website, <http://www.fgdc.gov/nap/metadata> (accessed 10 August 2009)
- Göbel, S., Lutze, K. (1998). Development of meta databases for geospatial data in the WWW, In: *Proceedings of the Sixth ACM International Symposium on Advances in Geographic Information Systems 1998*, ACM, Washington DC, USA, pp. 94–99.
- Greenberg, J. (2004). Metadata Extraction and Harvesting, *Journal of Internet Cataloging*, Volume6, Number4, pp.59 - 82
- Greenberg, J., Spurgin, K., Crystal, A. (2005). Final report for the AMEGA (Automatic Metadata Generation Applications) project. Technical, http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf
- Greenberg, J., Spurgin, K., Crystal, A. (2006). Functionalities for Automatic-Metadata Generation Applications: A Survey of Metadata Experts’ Opinions. *International Journal of Metadata, Semantics and Ontologies*. Vol. 1, No. 1, 2006.
- Guptill, S. C. (1999). Metadata and data catalogues In *Geographical Information Systems*, Vol. Management Issues and Applications (Eds, Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W.) John Wiley & Sons, Inc., pp. 677-692.
- Han, H., Giles, C. L., Manavoglu, E., Zha, H., Zhang, Z., Fox, E. A. (2003) Automatic Document Metadata Extraction using Support Vector Machines. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries*, 37-48

- Hatala, M., Forth, S. (2003). A comprehensive system for computer-aided metadata generation. In: Proceedings of 12th International Conference of The World Wide Web Consortium (WWW2003), Budapest, May 20-24, 2003.
- Huang, C.H., Chuang, T.R., Deng, D.P., Lee, H.M. (2009). Building GML-native web-based geographic information systems, *Computers and Geosciences*, 2009 – Elsevier, pp. 1802–1816.
- ISO (2003). *Geographic Information - Metadata (ISO 19115:2003)*
- Kalantari, M., Rajabifard, A., Olfat, H. (2009). *Spatial Metadata Automation: a New Approach*, SSC2009, Adelaide, Australia.
- Kresse, W., Fadaie, K. (2004). *ISO Standards for Geographic Information*. Springer, Heidelberg.
- Lake, R. (2005). The application of geography markup language (GML) to the geological sciences. *Computers and Geosciences* 2005, 31, pp. 1081-1094.
- Limbach, T., Krawczyk, A., Surowiec, G. (2004). Metadata lifecycle management with GIS context. In: Proceedings of the 10th EC GI & GIS Workshop, ESDI State of the Art, Warsaw, Poland.
- Manso, M., Wachowicz, M., Bernabé-Poveda, M. (2009). Automatic Metadata Creation for Supporting Interoperability Levels of Spatial Data Infrastructures refereed volume of the GSDI-11 Conference, June, 2009, Rotterdam, The Netherlands.
- Mathys, T. (2004). The Go-Geo! Portal metadata initiatives. In: Proceedings of the Geographical Information Science Research UK 12th Annual Conference, University of East Anglia, Norwich, UK, pp. 148–154.
- Mika, P.(2005). Ontologies Are Us: A Unified Model of Social Networks and Semantics. *International Semantic Web Conference 2005*. pp. 522-536
- Moellering, H., Aalders H.J., Crane A. (2005). *World Spatial Metadata Standards*, Elsevier.
- NISO (2004). *Understanding Metadata*. National Information Standards Organisation, USA.
- OGC (2007). *OpenGIS® Geography Markup Language (GML) Encoding Standard*, Open Geospatial Consortium, http://portal.opengeospatial.org/files/?artifact_id=20509
- Pasca, M., Petriglia, L., Mattioni, F., Torchio, M., Mariotti, C. (2009). Experiences in the Creation and Updating of INSPIRE Compliant Metadata Catalogue, GSDI-11 Conference, June, 2009, Rotterdam, The Netherlands
- Paul, M., Ghosh, S.K. (2008), Application Schema Mapping based on Ontology: An Approach for Geospatial Data Storage, *Journal of Digital Information Management*, Vol 6, Number 1, pp. 51-60
- Polfreman, M., Rajbhandari, S. (2008). *Meta Tools-Investigating Metadata Generation tools*, JISC Final report
- Rajabifard, A., Binns, A., Williamson I. (2005). Development of a Virtual Australia Utilizing an SDI Enabled Platform, Proceedings of FIG Working Week/GSDI-8, Cairo, Egypt.
- Rajabifard, A., Kalantari, M., Binns, A. (2009). SDI and Metadata Entry and Updating Tools in SDI Convergence, ed. B.van Leonen, J W J Besemer, J.A. Zevenbergen, Netherlands Geodetic Commission, Delft, pp.121-138.
- Safe Software (2009). FME platform overview, Safe software website, <http://www.safe.com/technology/FMEplatform/overview.php> (accessed 13 September 2009)

- Schindler, U., Diepenbroek, M. (2008). Generic XML-based Framework for Metadata Portals. Computers & Geosciences Submitted.
- Taussi, M. (2007). Automatic production of metadata out of geographic datasets, master thesis, Department of Surveying, Helsinki University of Technology, May 2007.
- Tsou, M.H. (2002). An operational metadata framework for searching, indexing, and retrieving distributed geographic information services on the Internet. In: M. Egenhofer and D. Mark, Editors, Lecture Notes in Computer Science vol. 2478, Springer, Berlin (2002), pp. 313–332.
- Westbrooks, E.L. (2004). Distributing and synchronising heterogeneous metadata in geospatial information repositories for access, in Hillmann, D. and Westbrooks, E.L. (Eds), Metadata in Practice, APA, Chicago, IL.
- West Jr., L.A., Hess, T.J. (2002). Metadata as a knowledge management tool: Supporting intelligent agent and end user access to spatial data (2002) Decision Support Systems, 32 (3), pp. 247-264.
- Williamson, I., Rajabifard A., Feeney M.E. (Eds.) (2003). Developing Spatial Data Infrastructures: from Concept to Reality, London, UK: Taylor & Francis

BIOGRAPHICAL NOTES

Hamed Olfat commenced his PhD on spatial metadata automation as an ARC linkage project in 2009 at the Centre for SDIs and Land Administration at the Department of Geomatics, the University of Melbourne. The aim of this project is to investigate automation of spatial metadata creation, update and enrichment. Hamed also holds a bachelor degree in surveying engineering and master degree in urban and regional planning.

Abbas Rajabifard is an Associate Professor and Director of the Centre for SDIs and Land Administration at the Department of Geomatics, the University of Melbourne. He is President of the GSDI Association, Vice Chair of Working Group 3 of the United Nations supported Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP), a member of ICA-Spatial Data Standard Commission, and a member of Victorian Spatial Council.

Mohsen Kalantari is a Research Fellow at the Centre for SDIs and Land Administration at the Department of Geomatics, the University of Melbourne working on spatial metadata automation. He was previously involved in the CRC project for the assessment of metadata entry tools and their fitness for ANZLIC requirements. Mohsen currently is also the ePlan project coordinator at Land Victoria, Department of suitability and Environment and is a member of the ICSM ePlan working group. He finished his PhD from the University of Melbourne in 2008. Mohsen has a bachelor degree in surveying engineering and master degree in GIS engineering.

CONTACTS

Hamed Olfat
Centre for SDIs and Land Administration
Department of Geomatics, The University of Melbourne, Victoria 3010, Australia
Tel. +61 3 8344 6771
Fax + 61 3 9347 2916
Email: h.olfat@pgrad.unimelb.edu.au
Web site: <http://www.csdila.unimelb.edu.au/people/rteam/hamed.html>

Assoc/Prof Abbas Rajabifard
Centre for SDIs and Land Administration
Department of Geomatics, The University of Melbourne, Victoria 3010, Australia
Tel. +61 3 8344 0234
Fax + 61 3 9347 2916
Email: abbas.r@unimelb.edu.au
Web site: <http://www.csdila.unimelb.edu.au/people/rteam/abbas.html>

Dr. Mohsen Kalantari
Centre for SDIs and Land Administration
Department of Geomatics, The University of Melbourne, Victoria 3010, Australia
Tel. +61 3 8344 3427
Fax + 61 3 9347 2916
Email: saeidks@unimelb.edu.au
Web site: <http://www.csdila.unimelb.edu.au/people/rteam/mohsen.html>