

## **Spatial Metadata Automation: A Key to Spatially Enabling Platform\***

Hamed Olfat<sup>1</sup>, Mohsen Kalantari<sup>1</sup>, Abbas Rajabifard<sup>1</sup>,  
Hervé Senot<sup>1</sup>, Ian P. Williamson<sup>1</sup>

Centre for SDIs and Land Administration, The University of Melbourne  
[h.olfat@pgrad.unimelb.edu.au](mailto:h.olfat@pgrad.unimelb.edu.au), [saeidks@unimelb.edu.au](mailto:saeidks@unimelb.edu.au),  
[abbas.r@unimelb.edu.au](mailto:abbas.r@unimelb.edu.au), [hsenot@student.unimelb.edu.au](mailto:hsenot@student.unimelb.edu.au),  
[ianpw@unimelb.edu.au](mailto:ianpw@unimelb.edu.au)

### **Abstract**

Metadata is a vital tool for any spatially enabling platform, and helps the users to share, discover, assess and access data and services. However, the spatial industry faces different issues and challenges regarding metadata generation, updating and improvement; which could affect the quality of this crucial component of any sharing platform. The main issue is the lack of an appropriate approach to the automated metadata generating and updating process. Metadata and related spatial data are often managed and maintained separately. This issue involves different aspects, including the lack of proper methodologies to integrate metadata and spatial data in a common environment, the generation and updating of metadata outside the spatial dataset lifecycle and the dependency of metadata creation on the metadata authors' knowledge of the dataset. In addition, the current data discovery services are not user-friendly and sufficiently efficient to serve the end users to easily find the most appropriate datasets and services to meet their needs in a spatially enabling platform. In response to these issues, this paper presents the new approaches to create, update and improve the content of metadata in an automated fashion to facilitate metadata management. The first approach relates to process-based metadata entry which aims at creating the ISO 19115:2003 metadata elements in parallel with the dataset lifecycle. This approach has the potential to overcome the problem of missing or incomplete metadata through identifying the stage to generate and update metadata within the dataset lifecycle. Also, the paper introduces a new synchronization approach to automate the spatial metadata updating process. This approach would aid the data custodians to update

---

\*This work is licensed under the Creative Commons Attribution-Non commercial Works 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/> or send a letter to Creative Commons, 543 Howard Street, 5<sup>th</sup> Floor, San Francisco, California, 94105, USA.

metadata on the fly whenever the dataset is modified. The synchronization is based on the GML technology to couple dataset and metadata and to exchange them over the Web. The paper also presents and discusses the prototype system implemented and based on the conceptual design of the automatic metadata updating. This system has been integrated with the GeoNetwork opensource and is now up and running. Finally, the paper demonstrates the prototype systems which have been designed and developed following the automatic metadata enrichment approach. This approach is based on Web 2.0 and Folksonomy concept and involves improving the content of descriptive keyword (as a metadata element) which is the first gateway for discovering existing datasets in a sharing platform. The prototype systems have been implemented within two different environments: Model Information Knowledge Environment (MIKE) and GeoNetwork opensource.

**Keywords:** Spatial, metadata, automation, enablement, updating, enrichment, SDI, GML

## 1. INTRODUCTION

Metadata is a critical component of any Spatial Data Infrastructure (SDI) initiative in which the aims are to facilitate data discovery, assessment and access. It not only provides users of spatial data with information about the purpose, quality, actuality and accuracy and many more of spatial datasets; but also metadata performs crucial functions that make spatial data interoperable, that is, capable of being shared between systems.

Following the key role of metadata in an SDI, the existence of metadata has been acknowledged as one the fundamental indicators for assessing any spatially enabling platform (Rajabifard, 2007; Ezigbalike and Rajabifard, 2009), so that the lack of tools for metadata management will prevent the emerging of spatially enabled societies (Williamson *et al.*, 2011).

In this regard, to achieve a spatially enabled platform delivering complete, precise, up-to-date and reliable metadata to end users through the spatial data catalog system has been a concern of data providers and custodians. However, the spatial industry has been facing different issues and challenges regarding the generation, maintenance and improvement of metadata which have strongly affected the quality of this critical spatial product.

Metadata is commonly gathered much after spatial dataset is created and is stored separately to the actual dataset it relates to. Separation of storage creates two independent datasets that must be managed and updated: spatial data and metadata, which are often redundant and possibly inconsistent (Rajabifard *et al.*, 2009).

Furthermore, the current manual and semi-automatic methods for metadata creation and updating have been considered as monotonous, time consuming, and a labor-intensive process by the organizations (Olfat *et al.*, 2010a). In addition, the existing methods are unable to maintain metadata current with dataset changes, so that metadata is usually incomplete, out-of-date and in some cases missing. Particularly, following highly increasing amount of spatial data being created, updated and published over the Web the organizations are seeking an automated approach to facilitate metadata management. This automated function should be able to provide the end users with the most up-to-date, reliable and precise metadata concurrent with frequent dataset changes. Exploring the considerations to design and develop such an automated function to create, update and improve the content of metadata is the main aim of the research project titled 'Spatial Metadata Automation' which is currently under investigation in Australia.

As a result, this paper as an output of the research project aims at presenting the state of the art developments to facilitate metadata management for spatially enabling platforms. In this regard, the paper first identifies the barriers to spatial metadata delivery. Next, the paper presents the process-based metadata entry approach designed and developed to improve the current status of metadata creation. It then presents the concept of and the prototype system implemented for the automatic metadata updating approach. Finally, the paper demonstrates the prototype systems developed for the automatic metadata enrichment approach and discusses on the future directions of the research.

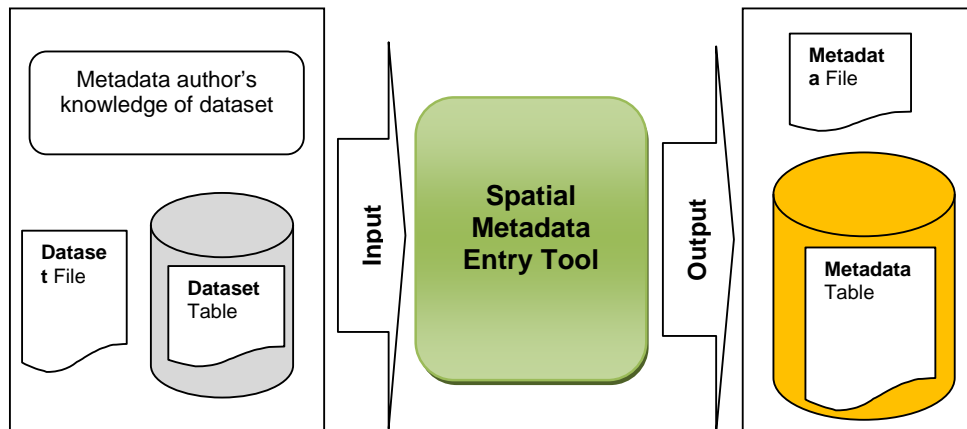
## **2. BARRIERS TO SPATIAL METADATA DELIVERY**

As already mentioned there are different issues and challenges which prevent the reliable, complete and up-to-date metadata to be delivered to end users within an enabling platform. These obstructions have been identified and discussed in more detail in this section.

Due to describing different aspects of the dataset such as identification, quality, citation, extent, constraint, etc. by metadata (ISO, 2003), ideally it should be part of a spatial dataset. Also, metadata values should be generated and updated with any change in the dataset from very first stages of the dataset lifecycle. However, metadata generation is commonly undertaken after the dataset is fully created or is ready to be published over the Web at one point of time which is not an incessant practice parallel to the dataset lifecycle. Collecting metadata later requires considerable effort and not all the information might be available (Timpf *et al.*, 1996) and also the metadata gathered in this way is often missing or incomplete (Rajabifard *et al.* 2009).

Also, as illustrated in Figure 1 metadata and datasets are generally created independently and stored in separate environments, so that metadata cannot be updated at the same time with any change in the actual dataset. Kalantari *et al.* (2009) considered the current data model for metadata storage as a 'detached metadata data model' and then acknowledged a new data model as 'integrated metadata data model' already introduced by Najjar (2006) to overcome the issues remained with the detached data model. In the integrated metadata data model, the spatial data and metadata are stored in a common file or database. This common metadata-spatial data set can be considered to be a 'comprehensive spatial data', so that with any change in the data the metadata will be updated simultaneously.

**Figure 1: Current Flow for Metadata Creation, Storage and Updating**



In addition to these challenges, the current data discovery services are not user-friendly and sufficiently efficient to serve the end users in a spatially enabling platform. These services need more interaction with the users to improve the content of 'search word' metadata element (or 'descriptive keyword' based on ISO 19115:2003) which is the main gateway for discovering and finding datasets over the Web. Kalantari *et al.* (2010) also argued that finding effective keywords to describe the spatial datasets is fundamental within any sharing platform. The right keyword for any spatial dataset means the keyword which is consistent with the content of the dataset and can reveal its essence and applications. In addition, a good keyword should be comprehensive and address the probable queries made by users from diverse categories. Moreover, a keyword should be popular meaning that most of the users agree on that keyword.

Besides, as part of the metadata automation research and in order to explore the current status of metadata management, an online survey was carried out by participation of the major Australian organizations dealing with spatial data. The

results of this survey also confirmed that the organizations in Australia are facing the main issues and challenges mentioned above (Olfat *et al.*, 2010a).

To address the issues and challenges regarding metadata management, an automation framework was designed and developed which has been first introduced by Kalantari *et al.* (2009). According to this framework, the following approaches have been set up to facilitate metadata management which are discussed in more detail:

- Process-based metadata entry approach
- Automatic metadata updating approach
- Automatic metadata enrichment approach

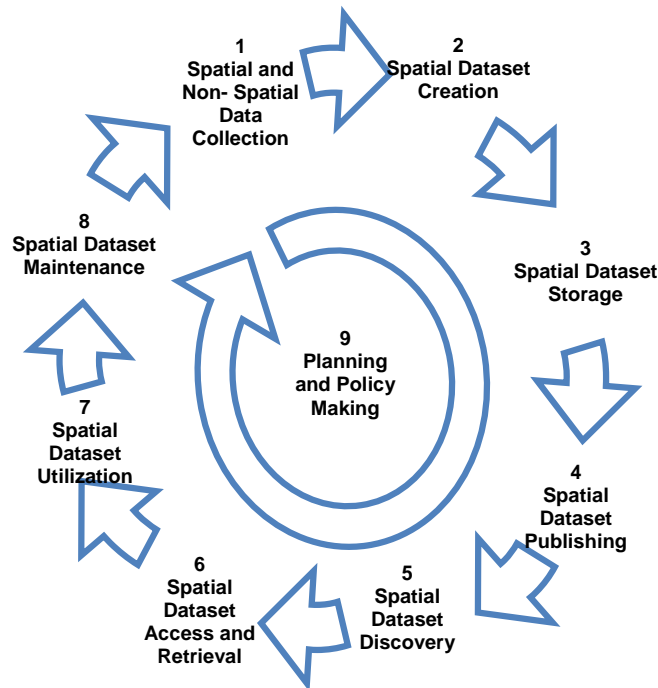
### **3. PROCESS-BASED METADATA ENTRY APPROACH**

Although metadata has always been considered as a component of a spatial dataset, a process separate from the spatial dataset lifecycle has typically been considered to generate and update the metadata file, as already shown in Figure 1. The metadata created in this way is entirely dependent to its author's knowledge of the dataset. Even if the metadata generation and updating are rooted in dataset extraction (product-based metadata entry approach which is based on extracting metadata from dataset); many required metadata values will be missing unless they are authored manually.

Therefore, in this research to create metadata a process-based approach was designed and developed which aims at creating metadata parallel to dataset lifecycle. This approach not only identifies the metadata elements that should be created in any step of the dataset lifecycle, but also aids the organizations to determine the responsible party for generating metadata values in any step of the dataset lifecycle.

The prerequisite to design such a process-based approach was an overall spatial dataset lifecycle. This lifecycle was designed based on the 'Information Lifecycle' recommended by the Australian Government Information Interoperability Framework (AGIMO, 2006) which includes different steps consisting of planning, creation/collection, organization/storage, access, using, maintenance, re-using and sharing the information. However, according to the nature of 'information' in this research, which is 'spatial data', the information lifecycle was modified and an overall lifecycle which also covered the whole 'information lifecycle' was defined. The overall spatial data lifecycle is illustrated in Figure 2.

Figure 2: Overall Spatial Dataset Lifecycle



As illustrated in Figure 2, the first step of the lifecycle would be collecting the spatial and non-spatial data to create the dataset. The related data to create the spatial datasets could be collected through different methods depending on the planning decisions in terms of users' needs, the purpose of collection, the required quality, scale, extent, etc. Following the data collection, different activities such as standardization, aggregation and quality assurance would be undertaken to create the spatial dataset.

Once the dataset is created, it would be stored in the database. Later, the stored dataset would be published to a networked environment to be shared among the end users. The users would be also provided with the facilities to discover the existing and shared datasets as well as to access and retrieve the appropriate ones for their needs. The data catalog system plays a critical role here. As soon as the dataset is retrieved by the user, it would be utilized in relevant spatial activities. According to the dynamic nature of changes to datasets which usually occur in short time frames, an effective maintenance process was considered in the spatial dataset lifecycle as well.

Finally, according to different policies required in each step of the dataset lifecycle, a planning and policy making process was also considered as an ongoing procedure which is parallel to other steps. In this procedure, different planning activities to gather data, create dataset, publish data, update data as well as policies regarding spatial dataset/metadata distribution, responsibilities, rights, restrictions, standards, languages, extensions etc. would be made.

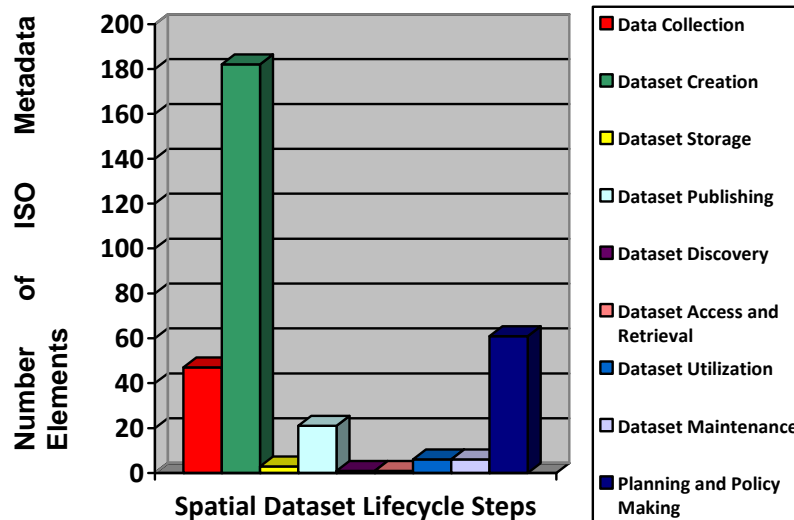
However, according to the variety of organizations dealing with spatial data around the world which have their own specific approaches, this lifecycle would not be wholly identical and could be optimized based on the current activities and responsibilities in different organizations.

Once the lifecycle was designed, the ISO 19115:2003 metadata elements which are expected to be created in each step were identified. Having explored the ISO 19115:2003 Standard, it was realized that the values regarding metadata elements could not be completely achieved unless they are created within different steps of spatial dataset lifecycle. In other words, all the metadata elements cannot be generated at the same time, because the metadata author is usually not familiar with the whole process of dataset lifecycle and needs additional information and details related to different steps of this lifecycle to generate or update metadata. For example, the values regarding dataset identification, quality and content should be created during the 'data collection' and 'dataset creation' processes, or the values regarding dataset restrictions and responsibilities are the values that should be created under a 'planning and policy making' process.

In addition, the distribution of metadata elements related to spatial dataset lifecycle steps was identified and illustrated in Figure 3. According to this Figure, the highest number of metadata elements should be generated and updated in the spatial dataset creation step. Planning and policy making, data collection, publishing, maintenance, utilization, discovery, access and retrieval are respectively the next steps with the highest number of elements.

Following the distribution of metadata elements and dataset lifecycle steps, the process-based metadata entry approach was implemented within the GeoNetwork opensource, the most commonly used metadata catalog among Australian organizations (Olfat *et al.*, 2010c), as a new add-on.

Figure 3: Distribution of Metadata Elements and Spatial Dataset Lifecycle Steps



#### 4. AUTOMATIC METADATA UPDATING APPROACH

Another major contribution of the metadata automation research has been designing and developing a new approach to real-time spatial data and metadata updating. The concept of automatic metadata updating is explored in the next section.

##### 4.1. The Concept of Automatic Metadata Updating

Automatic Spatial Metadata Updating concept has been already introduced by Kalantari *et al.* (2009) and Olfat *et al.* (2010b) referred as 'Synchronization'. This concept is about how to update metadata automatically and at the same time with any change in its related dataset. In fact, the real-time updating concept aims at synchronizing the metadata with frequent dataset changes in an automated fashion and regardless of any human interaction.

This automatic function therefore will benefit the organizations associated with spatial metadata to save time and effort while updating their large number of metadata records, and will also reduce the risk of inconsistency and redundancy in their spatial data and metadata.

According to the aim of automatic metadata updating approach, the prerequisite for taking this approach into account would be designing and building an



integrated data model for storing metadata and dataset. Through this data model, each dataset would be related to its metadata record. Having the relationship between these two sources and accommodating dataset geometries and attributes as well as metadata values into a middleware would result in a comprehensive dataset which can be exchanged over the Web and between different end users as well as spatial systems. Comprehensive dataset means a dataset which contains three fundamental components including geometry (and topology), attribute and metadata.

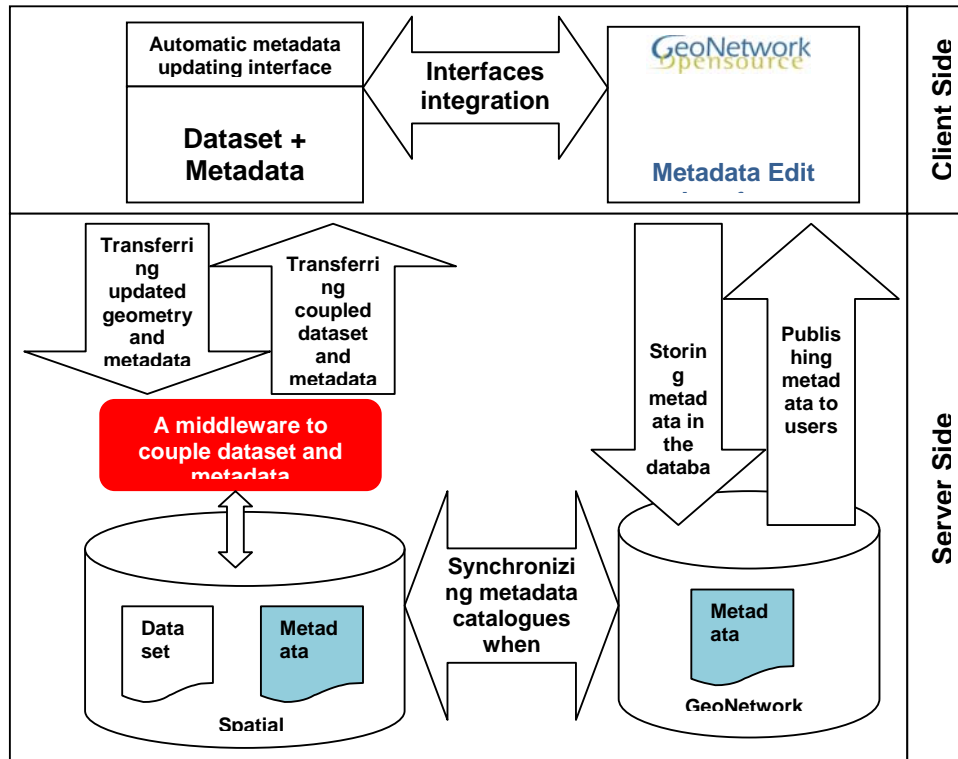
By transferring this comprehensive dataset to users through a user-friendly interface, they would be able to visualize dataset and metadata and also to edit dataset. Another significant expectation here is that user should be able to modify dataset and see its reflection on related metadata values (such as bounding box, date of revision, quality, etc.) simultaneously and automatically. After any modification, the new values for geometry, attributes and metadata would be transferred to backend. Moreover, the history of metadata revisions would be documented in the metadata record.

Integrating the real-time updating approach with one of the most commonly used metadata entry tools would add value to the proposed concept and result in more recognition from users. Therefore, according to the popularity of the GeoNetwork opensource among the users for cataloguing metadata, this tool was chosen to be involved in the whole real-time dataset and metadata updating approach. Following the integration, the metadata record which is updated at the time of dataset modification would be synchronized with its matching record in the GeoNetwork database. In fact, two metadata repositories (spatial database and GeoNetwork database) contain the same set of metadata elements for each metadata record.

As a result, in order to access the main interface for dataset modification two methods have been considered: direct and through the GeoNetwork opensource. Direct access method is for the users who are only responsible party for the dataset maintenance, with any change the user makes on the dataset not only the metadata inside the comprehensive dataset is updated automatically but also the matching metadata record in the GeoNetwork database would be synchronized. The access method through the GeoNetwork is designed for the users who are responsible parties for both dataset and metadata maintenance, so that by accessing the GeoNetwork metadata edit page they would be able to access the dataset modification interface. Then, after any change to the dataset the relevant metadata values would be updated automatically in both comprehensive dataset and GeoNetwork database.

The conceptual design for automatic metadata approach is illustrated in Figure 4.

**Figure 4: Conceptual Design for the Automatic Spatial Metadata Updating Approach**



Once the automatic metadata updating was conceptualized, a prototype system was implemented to prove the concept. The considerations for implementing this prototype system are discussed in the next section.

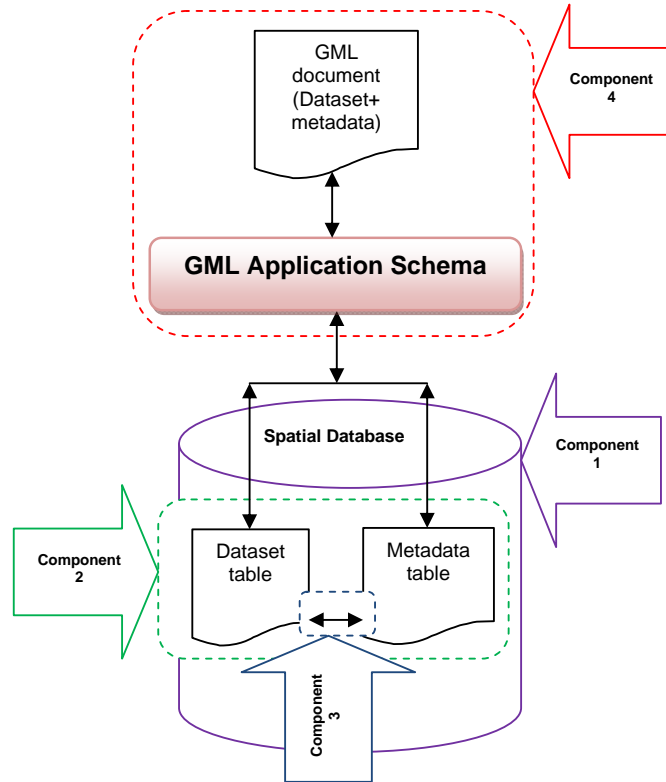
#### 4.2. Implementing the Automatic Metadata Updating Approach

The prerequisite for implementing the conceptual design illustrated in Figure 4 was to design and form an integrated data model to accommodate both spatial data and metadata values and exchange them over the Web. The following components were identified to build such a data model, as illustrated in Figure 5:

1. Spatial database to store both dataset and metadata;
2. Dataset and metadata tables;
3. Defined relationship between dataset and metadata tables; and

4. Middleware to accommodate dataset and metadata values and to include their relationship which is able to exchange them over the Web.

**Figure 5: Components of the Integrated Metadata Data Model**



The first component of the integrated metadata data model would be a spatially enabled database which is capable of storing spatial and non-spatial tables including dataset and metadata. PostgreSQL spatially enabled with PostGIS was selected for this purpose.

The second component of this data model would be the tables to store dataset and metadata. The dataset table would contain the attributes and geometry columns and metadata table(s) would embrace the ISO 19115:2003 elements. The third component would be a relationship between dataset table and metadata table. This relationship indicates that which metadata record in metadata table is related to which dataset in the spatial database.

The fourth component would be a middleware which is able to store both dataset and metadata values and their relationship and exchange them over the Web. This middleware should also be involved in returning the changes (updated values) to the required tables. To identify such an environment, the available data exchange formats (e.g. GML, KML, XML, and MEF (Metadata Exchange Format)) were investigated. Throughout the existing formats, Geography Markup Language (GML) International Standard by the Open Geospatial Consortium (OGC) which met the required conditions was selected for this purpose due to its capabilities to support the encoding, transferring and storage of spatial data characteristics as well as metadata values. GML is an XML grammar for expressing geographical features and serves as a modeling language for geographic systems and also as an open interchange format for geographic transactions on the Internet.

In addition, following Batcheller *et al.* (2009), 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. GML also 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).

Because 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).

In the GML Application Schema, the components of a comprehensive spatial data including geometry, attributes and metadata can be defined as elements conformance with the OGC GML Encoding Standard. Through this schema, the dataset and metadata would be coupled and exchanged in a middleware over the Web.

Also, through the commonly used Web Map Server technologies (such as Geoserver and *deegree*) which support the GML Application Schema, connection to spatial databases, mapping the relationships between dataset and metadata, mapping the spatial and non-spatial metadata elements to the related tables/columns in the database and setting up the WFS-T (Transactional Web Feature Service) capabilities for each element can be managed. Accordingly, a

GML Application Schema based on *deegree* open source technology (deegree, 2010) was developed as part of the integrated metadata data model for the automatic metadata updating.

Following the integrated data model, the output which consists of values from several related tables (dataset and metadata) is known as a complex feature. Complex features contain properties that can include further nested properties to arbitrary depth and are used to represent information as a collection of related objects of different types, not as an XML view of a single table (GeoServer, 2011). As a result, the fourth component of the integrated metadata data model would be a GML clone of the spatial database content (including dataset and metadata tables).

Therefore, another requirement for designing and developing the automatic spatial metadata updating was choosing an OGC compliant Web Server which supports WFS-T for complex features capability. This server should also support the GML Application Schema as the feature type which indicates how dataset and metadata elements have been mapped to different tables and columns in the spatial database. By publishing the complex feature in GML format, the WFS-T Server would enable the client side to read GML as the output of metadata integrated data model and write the changes back to dataset table in spatial database, after any dataset modification. According to deegree open source capabilities to support WFS-T for complex features, this technology was chosen to run the Web Server.

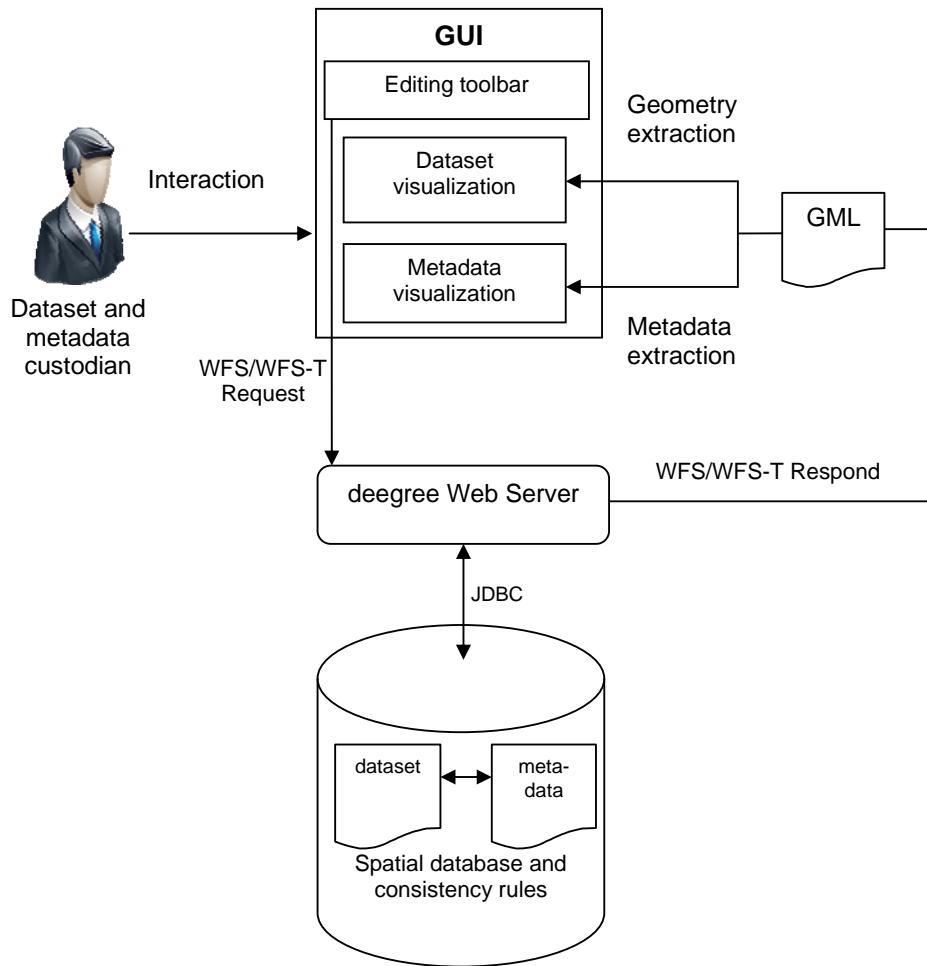
In addition, for setting up the approach to automate the spatial metadata updating a user-friendly Graphical User Interface (GUI) was required to read data in GML format from WFS-T Server, view dataset and metadata and edit the dataset. Not only should this interface be able to read the WFS/WFS-T respond in GML format, but also it should be able to extract the geometry features and metadata separately from the GML and show them in different sections. In addition, the interface should include an 'Edit Toolbar', which enables the users to modify the dataset and save the changes by sending WFS-T request to the Web Server. Figure 6 illustrates the data flow for the automatic metadata updating process.

The GUI was developed using HTML and JavaScript based on OpenLayers and GeoExt libraries/frameworks. Figure 7 shows this interface.

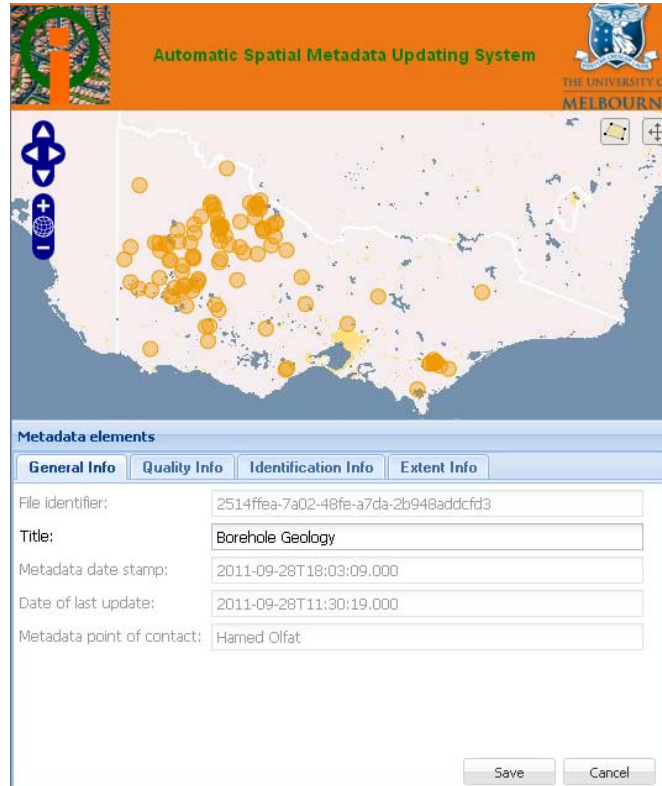
To design the approach to automate metadata updating, another requirement was identifying the metadata elements which could be updated after dataset modification and replacing their old values with the new ones in the metadata table in the spatial database. In this regard, first the metadata elements which were affected by the dataset change were recognized. Some of these elements which were employed in the prototype system are listed below:

- Date of last update (in General Info tab)
- Metadata date stamp (in General Info tab)
- Lineage statement (in Quality Info tab)
- Reference system (in Identification Info tab)
- North bounding latitude (in Extent Info tab)
- South bounding latitude (in Extent Info tab)
- West bounding longitude (in Extent Info tab)
- East bounding longitude (in Extent Info tab)

**Figure 6: Data Flow for the Automatic Metadata Updating Process**



**Figure 7: The Interface of Automatic Metadata Updating Prototype System**



Moreover, there was a need to design algorithms for generating the new values for the above metadata elements based on the changes occurred in the dataset and affecting them in the metadata table. Some of the values such as dates could be generated on the client side; however others had to be created on the server side. Obviously, once the 'Save' button (according to Figure 7) is pressed the new metadata values would be replaced in the interface (in the tabs designated for them) through getting the WFS-T respond from server.

The next consideration was the integration of metadata updating interface with the GeoNetwork opensource interface (Metadata Edit Page). Once this integration was performed, the next requirement was synchronizing the metadata catalogues stored in both GeoNetwork database and spatial database. To do this, a Web Service titled 'SYNC' was developed using PHP programming language to synchronize the metadata catalogues through sending CSW (Catalog Service for the Web) update requests.

As GeoNetwork opensource already includes CSW Server since it installs, another requirement here was to select an OGC compliant Web Server which supports CSW for updating the metadata record stored in the metadata table in the spatial database. The connection between this server and the metadata table would be through the GML Application Schema which indicates the mapping between the ISO 19115:2003 metadata elements and metadata table(s) and columns in the spatial database. *deegree* technology was selected for running the CSW Web Server.

The use of prototype system demonstrated that the automatic metadata updating approach would benefit the spatially enabling platform in terms of:

- Reducing the burden of manual metadata updating after dataset modification;
- Facilitating the interoperability matters by publishing datasets in GML;
- Enabling the custodians to publish and share datasets along with attributes and metadata in a single document;
- Avoiding the metadata to be missing, incomplete, out-of-date and unreliable;
- Giving a peace of mind to data custodians to make sure metadata is always current with dataset changes; and finally
- Giving a better discovery service to users seeking for spatial datasets over the Web by providing them with the most recent version of metadata.

It should be also emphasized that the ESRI through its proprietary ArcCatalog application has developed three synchronizers to synchronize the metadata content when values in the spatial data change: FGDC synchronizer, ISO synchronizer, and Geography Network synchronizer (ESRI, 2002). Although ArcCatalog synchronization is invaluable, it brings forth numerous problems associated with archiving and bibliographic control (Westbrooks, 2004). Also, the ArcCatalog uses the detached data model and is restricted to dataset formats such as file-based items (Shapefiles), ArcSDE geodatabase items, ArcIMS Images, ArcMap Images and Feature Services to synchronize metadata (ESRI, 2010).

In contrast, as already discussed, the automatic metadata updating approach has been designed based on the open sources environment which allows it to be reused and developed by the spatial community. Also, dataset and metadata are stored and published using the integrated data model in this approach. Moreover, this approach has been designed in a way to be dataset format neutral through generating and publishing the GML documents to the client side. In addition,



recording and documenting the history of changes on the dataset has been supported by the automatic spatial metadata updating approach.

## **5. AUTOMATIC METADATA ENRICHMENT APPROACH**

Automatic metadata enrichment involves improving content of the metadata through monitoring tags and keywords that are used by end users for finding resources (e.g. spatial dataset). Creating metadata by monitoring user interaction is based on the Folksonomy concept, a concept that was first introduced by Thomas Vander Wal in 2004 as 'the result of personal free tagging of information and objects for one's own retrieval'. Kalantari *et al.* (2010) introduced two models to implement the automatic spatial metadata enrichment concept: indirect and direct models. Indirect model or system-oriented model relies on the system to monitor and record the most popular keywords and tag them to resources as new metadata values. However, the direct model or user-oriented model is based on users' interaction in the system to tag the resources with the keywords they think best describe those.

Following the automatic spatial metadata enrichment conceptual design and according to the main interests of the research project industry partners, both two models have been implemented as prototype systems within two different environments: Model Information Knowledge Environment (MIKE) and GeoNetwork opensource.

### **5.1. Implementing the Automatic Metadata Enrichment within MIKE**

The MIKE has been developed by the Victorian Department of Primary Industries for management and registry of instances of biophysical and socio-economic modeling work in Victoria. Based on Williams *et al.* (2009), MIKE has been populated with a number of land use change and impact models as reported by Nichol *et al.* (2005). Many of these models have been applied in Victoria to understand: adaptive management of native vegetation, rural land use change, groundwater dependencies and socio-economic conditions. The MIKE is also recently populated with a number of climate change models applied by the Victorian Climate Change Adaptation Program (VCCAP) in South-West Victoria (Olfat *et al.*, 2010a).

Each model in the MIKE is described by a metadata record including different elements; such as classification (or type) of the model, key contact (including full details e.g. address and phone), model limitation, input, abstract, background, history, purpose, programming language, Web page, author, keywords, access, application type, etc. (Williams *et al.*, 2009).

A core aspect of MIKE involves researching solutions for managing the relationships between spatial modeling tools and their associated data inputs and data products. This requires active management in the interoperation of model metadata with spatial metadata. Assisting this goal was the implementation of the automatic metadata enrichment application to enable end users to associate keywords to elements within the MIKE to support interoperation and to make search and discovery of models and modeling activities easier and more intuitive. Figure 8 illustrates the interface of this prototype system.

**Figure 8: The Interface of Automatic Metadata Enrichment Prototype System within MIKE**

The screenshot displays the MIKE web application interface. At the top, there is a navigation menu with links: Home, Map View, Model Query, Model Discovery, Data Discovery, Administration, and Support. Below the menu is a header with the MIKE logo and the text "Model Information Knowledge Environment Department Of Primary Industries - DPI Inc 2008". The main content area is titled "Model Discovery:" and includes a "Tips for this page:" section. The tips section provides instructions for using the Model Discovery Function, offering two options: Option 1 (Using a keyword) and Option 2 (Using Tag Cloud). Option 1 includes a text input field for "Insert Keyword:" and a "Discover Models" button. Option 2 features a "Popular KEYWORDS for Models Discovery" section with a list of keywords, including Analysis, Apsim, Australia, Bas5, California, Cat, Clarke, Climate, Cloud, Computation, Crop, Dpi, Economic, Evaporation, Farming system, Growth, Hamed, Integration, Modflow, Parameterization, Perfect, Rainfall, Rainman, River, Salinity, Simulate, Simulation, Sleuth, SOIL, Statistical, Temperature, Topography, Tree, Urban, Urban growth, Water, Weather, and Wind. The bottom of the page contains a footer with the text "MIKE web application build V2.0.02 - ©Copyright 2007/2008 FFSR, Department of Primary Industries Inc."

## **5.2. Implementing the Automatic Metadata Enrichment within GeoNetwork opensource**

In addition to the implementation of automatic metadata enrichment within MIKE, to more widely prove the concept the following add-ons were developed within the GeoNetwork opensource as illustrated in Figure 9:

- *Suggestion list add-on*: this add-on provides a suggestion list based on previously searched terms. Using this add-on in the GeoNetwork, all subsequent searches benefit from previous searches. This facility provides a user-generated context for metadata creation and automation.
- *Tag Cloud add-on*: this add-on provides a visual representation of user generated search words and tags' relative importance in the GeoNetwork's user interface. The Cloud is enriched by user interactions on metadata records. Through this means a user-generated taxonomy can be established.
- *Tag monitor add-on*: this add-on automatically observes users' interaction with the GeoNetwork and collects users' recognition of and feedback on metadata records. The add-on monitors every interaction of users including downloading, exploring metadata details, etc. It also asks users to submit new keywords, agree or disagree with existing keywords.

The automatic metadata enrichment approach will result in many advantages for spatially enabling platforms, such as:

- Facilitating the spatial dataset discovery process;
- Improving the quality of 'keywords' describing the datasets which are the first gateway to discover the datasets;
- Engaging users in enriching the content of metadata files;
- Aiding the data catalog interface to be more user-friendly through the interaction with users;
- Sharing the knowledge of expert users about datasets among all users; and
- Helping new users when they are not sure what keywords to use for data discovery.

Figure 9: The Automatic Metadata Enrichment Add-ons within the GeoNetwork Opensource



## 6. CONCLUSIONS AND FUTURE DIRECTIONS

A spatially enabled platform cannot emerge without accessing and delivering the most up-to-date, reliable, complete and precise spatial metadata. The metadata empowers the end users to discover the existing datasets, find their required ones, assess the suitability of datasets against their needs and finally access the datasets they are interested in. However, there are some barriers which have limited the high quality delivery of metadata in today's world. Creating metadata in a separate process to dataset lifecycle, storing metadata in a detached environment to its related dataset, updating metadata after dataset modification and not in a real-time process and lack of user-friendly interfaces for dataset discovery are some of the major issues and challenges which the spatial industry faces.

In order to address the current challenges, this paper presented three main approaches including process-based metadata entry, automatic metadata

updating and automatic metadata enrichment. The process-based metadata entry approach has been designed and developed to aid data custodians to create metadata in parallel with dataset lifecycle. This has the potential to overcome the problem of missing or incomplete metadata through identifying the stage to generate and update metadata within the dataset lifecycle. The automatic metadata updating or synchronization approach aims at updating metadata automatically and at the same time with any change in the dataset modification process. This approach would reduce the burden of manual tasks to update metadata each time dataset changes. The automatic metadata enrichment approach improves the quality and content of keyword as a significant metadata element to describe datasets by sharing the knowledge of different users about datasets among others.

Moreover, the paper illustrated the prototype systems implemented to prove the concepts of automatic metadata updating and enrichment. These systems are now up and running and the next phase would be testing and improving them among the research project partners' environments.

#### **ACKNOWLEDGEMENTS**

The authors wish to acknowledge the support of the Australian Research Council Linkage Project (LP0883933) on Automating and Integrating Spatial Data and Metadata Processes, its industry partners and the members of the Centre for Spatial Data Infrastructures and Land Administration in the Department of Infrastructure Engineering, The University of Melbourne, in the preparation of this paper and the associated research. However, the views expressed in the paper are those of the authors and do not necessarily reflect those of these groups.

#### **REFERENCES**

- AGIMO. (2006). *Australian Government Information Interoperability Framework*. Australian Government Information Management Office (AGIMO), Department of Finance and Administration, Australian Government.
- Batcheller, J. K., Gittings, B. M. and Dunfey, R. I. (2009). A Method for Automating Geospatial Dataset Metadata, *Future Internet*, 1: 28-46.
- deegree. (2010). *deegree Web Feature Service v2.5*, at [http://download.deegree.org/deegree2.5/docs/wfs/html/deegree\\_wfs\\_documentation\\_en.html](http://download.deegree.org/deegree2.5/docs/wfs/html/deegree_wfs_documentation_en.html) [accessed 10 June 2011].
- ESRI. (2002). *Creating a custom metadata synchroniser*, at <http://www.esri.com/library/whitepapers/pdfs/metadata.pdf> [accessed 9 September 2009].

- ESRI. (2010). *Metadata synchronization*. at [http://webhelp.esri.com/arcgisDEsktop/9.3/index.cfm?TopicName=Metadata\\_synchronization](http://webhelp.esri.com/arcgisDEsktop/9.3/index.cfm?TopicName=Metadata_synchronization) [accessed 27 October 2011].
- Ezigbalike, C. and Rajabifard, A. (2009). "Indicators for Assessing Spatially Enabled Government Services", *GSDI 11 World Conference*, June 15- 19, Rotterdam, The Netherlands.
- GeoServer. (2011). *Complex Features*, at <http://docs.geoserver.org/stable/en/user/data/app-schema/complex-features.html> [accessed 15 February 2011].
- Huang, C. H., RueyChuan, T., PoDeng, D., and MingLee, H. (2009). Building GML-Native Web-Based Geographic Information Systems, *Computers and Geosciences*, 35(9): 1802-1816.
- ISO. (2003). *Geographic Information - Metadata (ISO 19115:2003)*, International Standardization Organization (ISO).
- Kalantari, M., Olfat, H. and Rajabifard, A. (2010). "Automatic Metadata Enrichment: Reducing Spatial Metadata Creation Burden through Spatial Folksonomies", in Rajabifard, A., et al. (Eds). *Spatially Enabling Society*. Luven University Press, pp. 119-129.
- Kalantari, M., Rajabifard, A. and Olfat, H. (2009). "Spatial Metadata Automation: A New Approach", *Proceedings Spatial Science Conference 2009 (SSC2009)*, September 28- October 2, Adelaide, Australia, pp. 629-635.
- Najar, C. R. (2006). *A Model-Driven Approach to Management of Integrated Metadata- Spatial Data in the Context of Spatial Data Infrastructures*, Doctor of Sciences, University of Rostock.
- Nichol, G. E., MacEwan, R.J., Pettit, C., Dorrough, J., Hossain, H., Suter, H., Cherry, D., Beverly, C., Cheng, X., Sposito, V., McNeill, J., Melland, A., and Shanks, A. (2005). "A Review of Models Applicable to 'Our Rural Landscape'", *International Congress of Modelling and Simulation Advances and Applications for Management and Decision Making (MODSIM05)*, Melbourne, Australia.
- Olfat, H., Kalantari, M., Rajabifard, A., Williamson, I., Pettit, C., Williams, S. (2010a). "Exploring the Key Areas of Spatial Metadata Automation Research in Australia", *GSDI 12 World Conference*, October 19-22, Garden city, Singapore.

- Olfat, H., Rajabifard, A. and Kalantari, M. (2010b). "Automatic Spatial Metadata Update: a New Approach", *XXIV FIG International Congress 2010*, April 11-16, Sydney, Australia.
- Olfat, H., Rajabifard, A. and Kalantari, M. (2010c), A synchronisation approach to automate spatial metadata updating process, *Coordinates Magazine*, VI(3): 27-32.
- Paul, M. and Ghosh, S. K. (2008). Application Schema Mapping based on Ontology: An Approach for Geospatial Data Storage, *Digital Information Management*, 6(1): 51-60.
- Rajabifard, A. (2007). *Towards a Spatially Enabled Society*, Melbourne: The University of Melbourne.
- Rajabifard, A., Kalantari, M. and Binns, A. (2009). "SDI and Metadata Entry and Updating Tools", in Loenen, B. v., Besemer, J. W. J. and Zevenbergen, J. A. (Eds). *SDI Convergence. Research, Emerging Trends, and Critical Assessment*, Delft, pp. 121-136.
- Timpf, S., Raubal, M. and Kuhn, W. (1996). "Experiences with Metadata", *Proceedings 7th International Symposium on Spatial Data Handling*, August 12-16, Delft, The Netherlands, pp. 12B.31–12B.43.
- 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.
- Williams, S., Pettit, C., Hunter, D., Cherry, D. (2009). "A Prototype Metadata Tool for Land Use Change and Impact Models- A Case Study in Regional Victoria, Australia", in Loenen, B. v., Besemer, J. W. J. and Zevenbergen, J. A. (Eds). *SDI Convergence. Research, Emerging Trends, and Critical Assessment*, Delft, pp. 137-150.
- Williamson, I., Rajabifard, A., Wallace, J., Bennett, R. (2011). "Spatially Enabled Society", *FIG Working Week 2011*, May 18-22, Marrakech, Morocco.



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

OLFAT, H; Kalantari Soltanieh, SMKS; Rajabifard, A; Senot, H; Williamson, IP

**Title:**

Spatial Metadata Automation: A Key to Spatially Enabling Platform

**Date:**

2012

**Citation:**

OLFAT, H., Kalantari Soltanieh, S. M. K. S., Rajabifard, A., Senot, H. & Williamson, I. P. (2012). Spatial Metadata Automation: A Key to Spatially Enabling Platform. International Journal of Spatial Data Infrastructures Research, 7, pp.173-195.  
<https://doi.org/10.2902/1725-0463.2012.07.arty>.

**Publication Status:**

Published

**Persistent Link:**

<http://hdl.handle.net/11343/32585>

**File Description:**

Spatial metadata automation: a key to spatially enabling platform