EXPLORING THE KEY AREAS OF SPATIAL METADATA AUTOMATION RESEARCH IN AUSTRALIA

Hamed Olfat[1], Mohsen Kalantari[1], Abbas Rajabifard[1], Ian P. Williamson[1], Christopher Pettit[2], Stephen Williams[2]

[1] Centre for SDIs and Land Administration
The University of Melbourne, Victoria 3010
h.olfat@pgrad.unimelb.edu.au
saeidks@unimelb.edu.au
abbas.r@unimelb.edu.au
ianpw@unimelb.edu.au

[2] Department of Primary Industries (DPI) – Victoria
christopher.pettit@dpi.vic.gov.au
steve.williams@dpi.vic.gov.au

Abstract

Improving the spatial metadata management process to facilitate data discovery, access and retrieval through an SDI platform has been the goal of a number of organizations at different jurisdictional levels in Australia. A current linkage research project titled "Spatial Metadata Automation" is being conducted at the University of Melbourne in conjunction with some industry partners. This research project aims to explore different approaches for automating spatial metadata so that the process of creating and updating spatial metadata – where feasible – becomes automatic.

As part of the project an online questionnaire was designed and distributed within the major organizations dealing with spatial data in Australia to assess the users' needs regarding metadata automation and also the current status of the activities in metadata creation and updating.

This paper presents the results of the assessment process and explores the key areas of spatial metadata automation research in Australia. It also reviews some of the more important initiatives regarding spatial metadata in this country and explains the characteristics and framework of the current research. The paper then discusses the structure of the questionnaire and the results of the responses analysis. Finally, the findings, future needs and research questions are presented.

Keywords: spatial, metadata, automation, questionnaire, 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 increasing number of spatial data sets being created and exchanged between people or organizations. 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).

Metadata plays a critical role in any Spatial Data Infrastructure (SDI) initiatives of which the aims are to facilitate data discovery and access. 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). The metadata not only provides users of

spatial data with information about the purpose, quality, actuality and accuracy of spatial data sets, it also performs 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 data sets for use (Rajabifard *et al.*, 2009).

On the other hand authoring and compiling metadata for spatial data sets are often labor intensive and time consuming. Methods and approaches to overcome these issues are welcomed by the spatial industry. The idea of automatic spatial metadata generation research is rooted in automatic indexing, abstracting, and classification of spatial data content, and which began with the need to organize increasing amounts of spatial related data and the inability of human authored methods to cope with a huge amount of spatial metadata (Rajabifard *et al.*, 2009). However automatic spatial metadata generation research efforts are new in the spatial arena and automated metadata generation is still in its infancy (Kalantari *et al.*, 2009).

Currently a spatial metadata automation research project is under investigation in Australia which is focusing on three main streamlines including automatic creation, updating and enrichment. As part of research, the project has also assessed the users' needs regarding metadata automation and the current status of the activities regarding metadata creation and updating within the Australian organizations dealing with spatial data through an online questionnaire. The results have shown and identified some important innovations in metadata automation and areas for improvement.

This paper as an output of the assessment process aims at exploring the key areas of spatial metadata automation research and its related future needs and questions in Australia. Following the introduction, the paper reviews some of the most important activities and initiatives regarding spatial metadata in different jurisdictions. Then, the characteristics and streamlines of the current research are described. The paper then discusses the structure of the questionnaire and the results of the responses analysis. Finally, the findings, future needs and research questions are presented.

2. SPATIAL METADATA ACTIVITIES AND INITIATIVES IN AUSTRALIA

In recent years Australia as one the first innovators of SDI has been paying more attention to improving the efficiency and effectiveness of this spatial data sharing platform. Different governments and organizations have been involved in this activity and have contributed much. These bodies include the Victorian Government, the New South Wales Government, the Western Australia Government, ANZLIC, Geoscience Australia and PSMA Australia Limited among others.

Following the critical role of metadata in any SDI, a major part of the efforts and initiatives mentioned above has been related to the creation, updating and publishing this data about data. Thus to aid the assessment of the current status of activities regarding spatial metadata in Australia, some of these major initiatives are reviewed in the following section.

2.1 ANZLIC Metadata Initiatives

ANZLIC is the peak intergovernmental organization for the collection, management and use of spatial information in Australia and New Zealand. Also the ANZLIC is the main body in Australia contributed to the key efforts in terms of

standardization, generation, publishing and accessing spatial metadata. Some of these initiatives have been discussed as following:

Publishing Metadata Standards

ANZLIC works closely with Standards Australia, Standards New Zealand, the Intergovernmental Committee on Surveying & Mapping (ICSM), and PSMA Australia Limited to raise the awareness and uptake of geospatial standards. There was extensive input by Australia during the development of ISO 19115:2003, particularly from interests associated with ANZLIC, and as a result this standard incorporates several core metadata elements from the previously published ANZLIC Metadata Guidelines V.2. In August 2007 the ANZLIC Metadata Guidelines V.2 were superseded with the release of the ANZLIC Metadata Profile Version 1.1. This metadata standard complies with ISO 19115:2003 (ANZLIC, 2010b).

The Australian Spatial Data Directory (ASDD) Initiative

The ASDD is a national initiative supported by all governments under the auspices of ANZLIC which aims to improve access to Australian spatial data for industry, government, education and the general community through effective documentation, advertisement and distribution. The ASDD has provided search interfaces to discover spatial metadata throughout Australia. The ASDD pilot GeoNetwork gateway sponsored by the Office of Spatial Data Management (OSDM) with the URL http://asdd.ga.gov.au/geonetwork/ has been set up for testing (ASDD, 2010).

BlueNet Metadata Entry and Search Tool (MEST)

The development of the BlueNet Metadata Entry and Search Tool (MEST) has provided a platform for the development of the ANZLIC Metadata Entry Tool (MET). The MEST is based on GeoNetwork; a standards based, free and open source catalogue application to manage spatially reference resources through the web (Rajabifard *et al.*, 2007). The MEST allows users to search for data and assess its quality through a metadata description. Data collected under BlueNet, where appropriate, will be immediately available via the MEST. Data available on request from government agencies is also discoverable using the MEST's 'Remote Search' option (BlueNet, 2010b).

ANZMet Lite

This product has been produced by the OSDM on behalf of ANZLIC's Spatial Resources Discovery and Access Program Steering Committee. This tool is the metadata collection tool to facilitate the creation of an ANZLIC-compliant metadata record (ANZLIC, 2010a).

Reviewing ISO 19115 Standard

Recently, ANZLIC has participated in the ISO 19115 Geographic Information – Metadata review through Standards Australia's IT-004 Geographic Information Committee.

2.2 BlueNet Initiative - The Australian Marine Science Data Network

The BlueNet is a project funded by Department of Education, Science and Training's (DEST's) Backing Australia's Ability program (BlueNet, 2010a). This project is designed to establish a nationally distributed marine science data network linking universities to the Australian Ocean Data Centre (AODC), to support the long term data access needs of Australia's marine science researchers (Rajabifard *et al.*, 2007). BlueNet builds an infrastructure to enable the discovery, access and online

integration of multi-disciplinary marine science data on a very large scale, to support current and future marine science and climate change research, ecosystem management and government decision-making (BlueNet, 2010a).

2.3 Model Information Knowledge Environment (MIKE) - Department of Primary Industries (DPI) – Victoria

The MIKE provides a spatially enabled registry of system model tools. It 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. The spatial component of the system enables it be applied nationally and internationally. The MIKE is currently populated with a number of climate change models applied by the Victorian Climate Change Adaptation Program (VCCAP) in South-West Victoria. The MIKE is an important component to the SDI associated with ecoinformatics[1].

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 are future development efforts in the application of automated metadata tagging to enable end users to associate keywords to elements within the MIKE to assist interoperation and to make search and discovery of models and modeling activities easier and more intuitive. Earlier limitations of MIKE in its ability to directly supporting model services (Williams *et al.*, 2009) are now being addressed. The trial inclusion of the Open Geospatial Consortium (OGC) endorsed products such as Geoserver into the technology stack for the MIKE is one such step towards improving service orientation.

2.4 The Common Spatial Information Initiative (CS2i) – New South Wales

The New South Wales Spatial Council has recently endorsed the distribution of the Draft NSW Spatial Information Metadata Policy for consultation. The CS2i Program Office has produced this policy based on the analysis of metadata policies operating in various jurisdictions across Australia and the standards that govern metadata in Australia and international jurisdictions. The CS2i is the result of collaboration across NSW state government agencies and private industry to provide common e-business functions that will enable customers to access spatial and related information online (CS2i, 2010).

2.5 Interragator+ - Western Australia

Interragator+ is the first step in finding and accessing spatial information held by private and public organizations in Western Australia. Interragator+ is a directory of metadata for Western Australian spatial data. This directory is similar to a library catalogue. The directory helps the users search for data, view data details using metadata, view live data and map resources, create, upload and manage metadata, create and manage user profile and manage maps and queries (Interragator+, 2010).

[1] Ecoinformatics integrates environmental and information sciences applying computer based tools to assist in the access and analysis of environmental data.

3. SPATIAL METADATA AUTOMATION RESEARCH

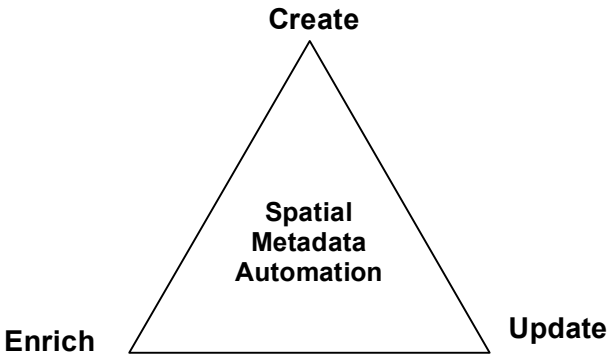
Beside the activities of different organizations in Australia, an Australian Research Council (ARC) Linkage Project is currently coordinated by the Centre for SDIs and Land Administration (CSDILA) at the University of Melbourne. This research project titled "Spatial Metadata Automation" involves researchers from the Centre and industry partners including Victorian Departments of Primary Industries (DPI) and Sustainability and Environment (DSE), the NSW Department of Lands, AusSoft Solutions Pty Ltd, CubeWerx Australia Pty Ltd and Logica CMG.

The project aims at exploring different approaches for automating spatial metadata and the impact of these approaches on critical components of metadata such as standards and data model so that the process of updating or extracting spatial metadata – where feasible – becomes automatic. To address the aim of the project the following objectives have been considered (CSDILA, 2010):

- Study on metadata ontology
- Study on current practices in Spatial Metadata Tools
- Case study within project partners
- Development of the concept of Metadata Automation
- Investigation on technological trends and their impact on metadata: Web 2.0, Semantic Web, etc.
- Design and development of a prototype system for metadata automation.

To better conceptualize this project, a framework for spatial metadata automation which includes automatic creation, updating and enrichment of spatial metadata is defined by Kalantari *et al.* (2009) as illustrated in Figure 1. This framework defines three fundamental but complementary approaches that can be employed for metadata automation as described below:

Figure 1: Spatial Metadata Automation Framework



3.1 Automatic Update

Automatic spatial metadata updating or synchronization is a process by which properties of a spatial data set are read from the data set file and written into its spatial metadata file automatically. Some software vendors (such as ESRI) conceptualize metadata automation as synchronizing the metadata content when values in the spatial data change. In this type of model, for instance when a change occurs of a spatial data property (such as its projection), the metadata will be updated with the new information. Olfat *et al.* (2010) proposed a Geography Markup Language (GML) based approach to facilitate automatic metadata updating. In this

approach GML is used as a medium for transferring metadata details from one file to another and monitor changes for automatic updating of the metadata content.

3.2 Automatic Creation

While automatic updating and synchronization is suitable for updating an existing metadata record, there is a need for other methods when there is no existing metadata associated with spatial data. Operators create metadata by writing descriptions of resources either in a structured or unstructured form. Computer applications can then extract certain information from a resource or its context. This may involve simply capturing information that is already available, such as the format of the file, or running an algorithm to determine the subject of a textual resource by counting keywords or by checking and analyzing pointers to the resource.

3.3 Automatic Enrichment

The third approach is automatic enrichment which involves improving content of metadata through monitoring tags and keywords that are utilized by users for finding data sets. Creating metadata by monitoring user interaction is based on the Folksonomy concept – that is, the process by which many users add metadata in the form of keywords to shared content.

To implement the automation framework described above, there exists a need to assess the current status of spatial metadata management as well as the users' needs regarding the automatic metadata creation and updating within the major participants of spatial data management in Australia. The next section considers the selected approach for the assessment process.

4. METADATA NEEDS ASSESSMENT

According to the wide distribution of organizations dealing with spatial data in Australia and as part of the research methodology, the research team designed and put the survey into practice through an online questionnaire. The questionnaire was designed with aim of identifying current processes, issues and needs when creating and updating metadata records. The questionnaire structure and contents are here described.

4.1 Questionnaire Structure

The first part of the questionnaire was an introduction to the survey and background of the research project. Thereafter, the questions were classified into different categories such as participants' information, organizational spatial data and metadata characteristics, application of metadata, creating and updating metadata, metadata tools, and sharing metadata. Table 1 illustrates the details of the questionnaire.

Table 1: Elements of the metadata research questionnaire

Topics	Sub-topics
Introduction	• Introduction to questionnaire
Background Information	• Research Aims and Objectives
Participants' Information	• Name, Email address and Organization name
Organizational Spatial Data and Metadata Characteristics	• The number of spatial data sets • Activities regarding metadata for spatial data sets • Activities regarding spatial metadata • Metadata input methods
Application of Metadata	• The main applications complementing data discovery, exploration and exploitation
Creating and Updating Metadata	• The amount of manual metadata creation • The amount and format of metadata created by third parties • Standards used for metadata creation • The formats of metadata storage • Existence of any automatic metadata creation application • The updating time interval for spatial data sets • The mechanism for updating spatial data • The updating time interval for spatial metadata • The responsible body for updating spatial metadata
Metadata Tools	• The tools used in metadata entry, creation and updating
Sharing Metadata	• The possibility of sharing metadata with other parties • The amount of metadata accessed by other parties • The mechanisms used to publish spatial metadata • The mechanisms used to access consumers to the digital resources

The survey duration was two months to collect the participants' responses. Once the online survey was closed, the responses were analyzed using qualitative and quantitative analysis methods and the major results were discussed to identify the main areas of interest for metadata automation research.

5. ANALYSIS AND RESULTS

The analysis of the responses has been categorized into five main areas. The results have been described by statistical data and diagrams in those main areas. In most cases the participants were permitted to tick all options that applied for any question.

5.1. Organizational Spatial Data and Metadata Characteristics

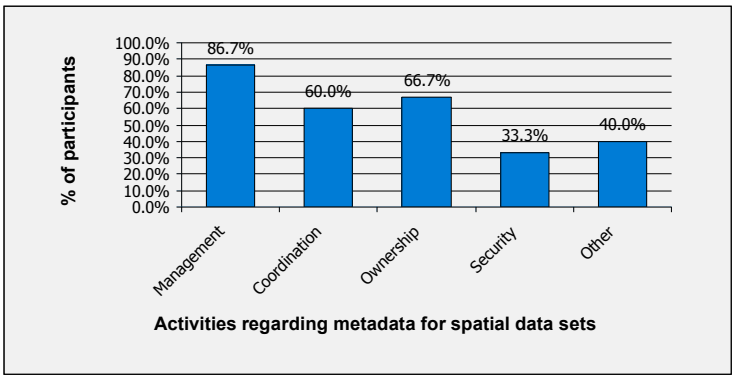
This part was aiming at finding the organizational activities which can be facilitated or should be served by the spatial metadata automation process.

The survey gathered data on general characteristics of spatial data and metadata existing in the participating organizations. The study found that the number of spatial data sets maintained in the organizations varied from a few to indefinite (e.g. data sets for meteorology activities). The issue thus arose that in some organizations the data sets are too various to quantify; hence managing these data and related metadata sets posed a real challenge.

Information was also gathered on activities regarding metadata for spatial data sets in which the participants were involved (Figure 2). Most participants (86.7%) were dealing with management activities. Over half of the participants were involved in ownership (66.7%) and coordination (60%) matters. About 33% of them were dealing with security activities. 40% of the participants also noted in "other" activities category that they were dealing with some other activities such as the creation and

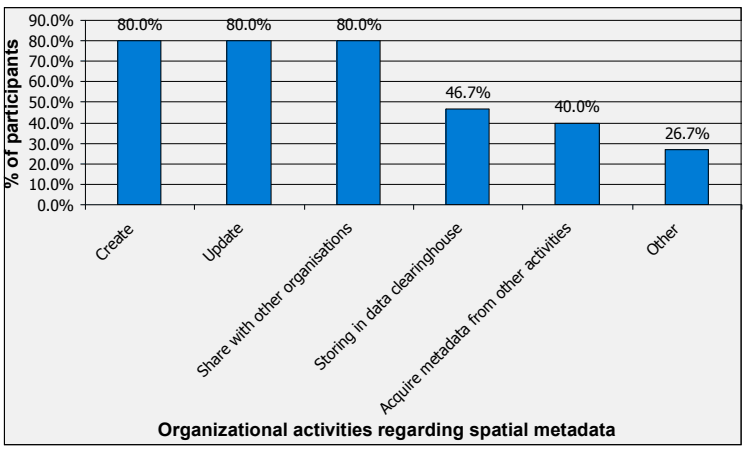
maintenance of metadata publication, managing the ANZLIC Spatial Resource Discovery and Access Program, supporting metadata requirements of all areas of the spatial community within the Australian Government, administrating a node of the ASDD, discovery and accessing internally and externally, and setting policies and management frameworks for metadata.

Figure 2: Distribution of activities regarding metadata for spatial data sets in organizations



In addition, the data was collected on organizational activities regarding spatial metadata (Figure 3). The results show that 80% of the organizations are involved in similar activities related to spatial metadata in terms of creation, updating and sharing with other organizations. Less than half of them (46.7%) are storing metadata in clearinghouses and 40.0% of them acquire metadata from other activities. More than a quarter of the participants (26.7%) also noted that the other activities included developing the frameworks and applications to create and manage metadata, recording metadata manually, search, share, store in a clearinghouse, share with other clearinghouses and embed metadata links within services. These results show that although most of the activities regarding metadata lifecycle are implemented by the organizations internally, there are still multiple parties involved in metadata creation and updating within some organizations.

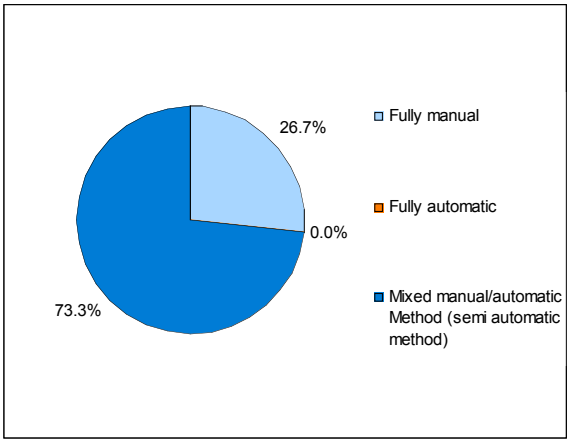
Figure 3: Distribution of organizational activities regarding spatial metadata



The input methods of spatial metadata records were also sought in the survey. As illustrated in Figure 4, no fully automatic method was used. Instead, most of the organizations (73.3%) were using the mixed manual/automatic (semiautomatic) method and the remainder used the fully manual method. Based on these results the current metadata creation methods with manual process involved are labor intensive;

thus they can be viewed as time consuming and expensive processes by the organizations.

Figure 4: Distribution of metadata input methods in organizations



5.2. Application of Metadata

The participants were provided with three popular spatial metadata applications including data discovery, exploration and exploitation and they were requested to explain any other type of application used in their organizations. This part of the survey was aimed at collecting the participants' expectations of spatial metadata and finding how these expectations can be satisfied through the automation process.

The results show that 62.5% of participants agreed with the three applications and the remainder added some other applications consisting of data management by the data custodian or delegated data manager to know when the data set was last updated and when it should be next updated, quality assurance to acknowledge and use the same capture methods, historical recording to know what was captured in this area over time, and limiting liability of the data provider by making the users aware of the latest status of spatial data sets.

5.3. Creating and Updating Metadata

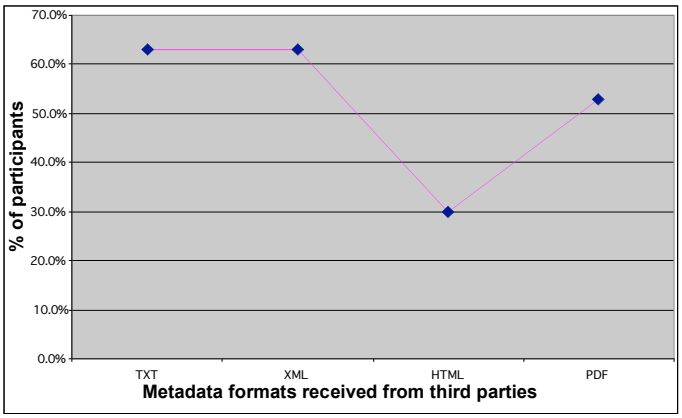
This part of the questionnaire was designed to assess the current status of metadata creation and updating in organizations to find those issues which should be addressed by the automation process.

The first question in this part was in regard to the percentage of data sets where their related metadata is created by manual input (without any automatic conversion). As a result of this, 43% of them indicated that metadata relevant to 100% of data sets are currently generated manually. Also three independent groups of participants (each around 6%) indicated that metadata relating to 5%, 20% and 90% respectively of spatial data sets are created manually.

Furthermore, the participants were asked to indicate the format(s) in which their organizations receive metadata from third parties. As shown in Figure 5, TXT and XML formatted metadata reached the highest percentage (63%) within the organizations. The PDF format (53%) was also popular in the organizations. The HTML format had the minimum amount (30%) among the organizations. These results show that the third parties are following no special policy to deliver spatial metadata; and therefore there is inconsistency between the metadata formats

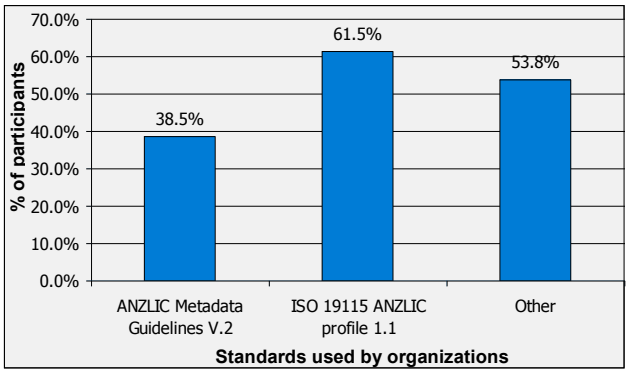
received by the organizations which may produce future problems when sharing these data.

Figure 5: Distribution of metadata formats received from third parties



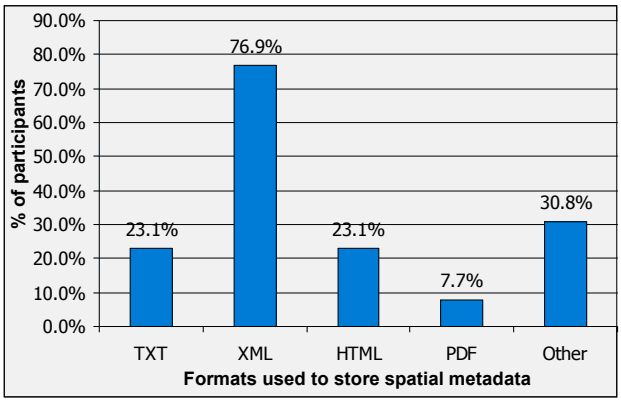
Moreover, information was gathered in relation to the metadata standards used by the organizations. The results are shown in Figure 6. The ISO 19115 ANZLIC profile 1.1 had the highest percentage of use (61.5%) among the organizations. The ANZLIC Metadata Guidelines V.2 was also used by 38.5% of organizations. The FGDC standard independently was used by none of the participants. Over half of the participants (53.8%) also noted that there are “other” standards used in their organizations which maybe the combination of other standards including in-house standard based on FGDC and ANZLIC Metadata Guidelines V.2, and ANZLIC Metadata Guidelines V.1. These results show that there is inconsistency between the metadata standards used by the organizations in Australia.

Figure 6: Distribution of metadata standards used by organizations



In addition, the participants reported the format(s) used to store spatial metadata (Figure 7). Among these formats, XML was reported by most of the participants (76.9%), TXT format was noted by 23.1% of them, and HTML and PDF formats were reported by 23.1% and 7.7% respectively. About 31% of participants also selected the “other” category and indicated RDBMS and Excel formats. The high percentage using XML format to store metadata illustrates a good potential for metadata interoperability in Australia.

Figure 7: Distribution of formats used to store spatial metadata



The participants were also asked if they create or update spatial metadata automatically in their organizations. Most of them (70%) reported that there is no automatic approach to create or update metadata. The remainder presented some in-house tools which automate a few parts of these approaches. These results show the significance of metadata automation research.

Additionally data was gathered on the general updating time interval for spatial data sets in the organizations. The results show that less than half of the organizations (46%) update metadata whenever spatial data are changed in the real world. About 8% of them update metadata files monthly and annually. These results show that there is no defined regime for metadata update within the organizations in Australia.

Furthermore, the participants were asked about the responsible party for updating spatial data. Over half of the organizations (54%) reported an internal team in the organization and about 16% of them reported an outside third party as responsible to update spatial data. Some of the participants which selected the “other” category also indicated the data custodians as the responsible party. These results show that updating spatial data internally is dominant within organizations in Australia.

Another significant question in this section was about the relation between time of updating metadata and spatial data sets. Over half of the participants (54%) reported that they update metadata and spatial data at different times and the remainder reported that they do the update process simultaneously. Thus there is a critical need to define a regime to update spatial data and metadata simultaneously within organizations in Australia.

Among the participants, 62% noted that the same team is responsible for updating spatial data and metadata and the remainder reported that the responsible teams differ. The separation of updating teams would make it longer to update metadata with any change in spatial data.

5.4. Metadata Tools

The main aim of this part of the questionnaire was to find the organizations’ approaches to apply the automatic or semiautomatic methods of metadata creation and updating.

Therefore, the detail of metadata entry tools used by the organizations was requested in this section. The participants reported some in-house tools, ANZMet

Lite metadata entry tool, GeoNetwork, Bluenet MEST, ESRI Geoportal Extension 9.3.1 and ESRI ArcCatalog. The wide variety of metadata entry tools used by the organizations is another issue arising from the survey questionnaire.

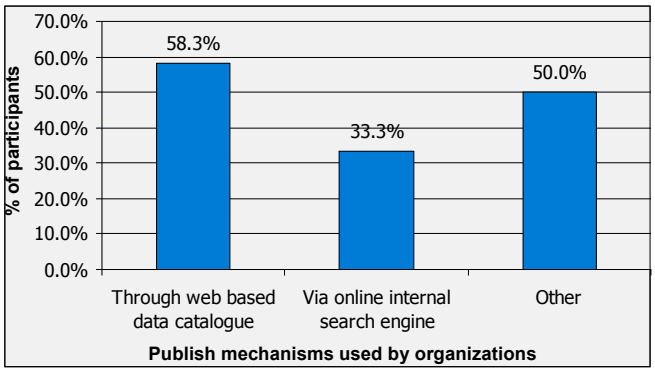
5.5. Sharing Metadata

This part of the questionnaire was to assess the platforms for metadata sharing and determine how the automation process can improve the sharing activities. In this section, the participants were asked some questions related to sharing spatial metadata with other parties. First, all of them reported that they share metadata of resources with other parties.

The data was collected on the percentage of metadata that currently can be accessed by other parties. The study shows that less than half of the organizations (41.6%) currently share 100% of their spatial data with others. 8.3% of them share 80%, 16.6% share 70%, 8.3% share 50%, 8.3% share 20% and 16.6% share no spatial data.

Secondly, the mechanism to publish metadata was the other item asked of the participants (Figure 8). About 58% of participants reported that they publish metadata through a web based data catalogue, 33.3% of them noted that they publish metadata through online internal search engine and half of them selected the “other” category as well which included other mechanisms such as packaged with data sets which are available via DVD or online download, Web Pages, ESRI Arc Catalog, a table of data with attributes and basic search tool, and through ASDD. The wide variety of publishing methods including non-Web solutions (e.g. DVDs) is another issue affecting interoperability and sharing.

Figure 8: Distribution of publishing mechanisms used by organizations

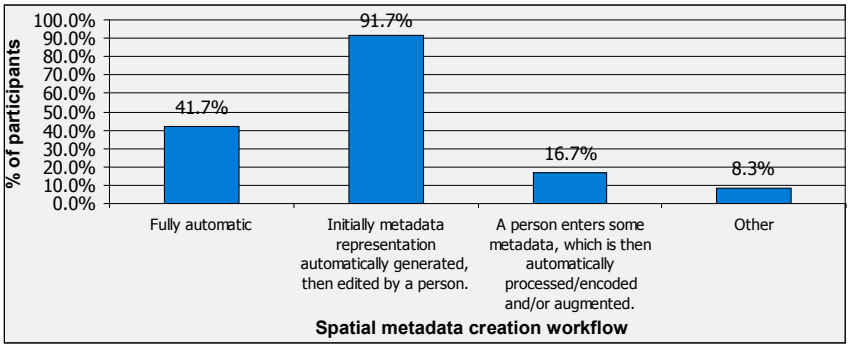


Moreover, the participants were asked to report the mechanism that their data consumers use to access the digital data resources. The results show that ordering required data via email and payment was reported by 67% of the participants. 58% of participants also noted on free online access and 50% of them reported a commercial contract with other organizations. A quarter of participants reported the online access after payment and 16% also stated that digital resources are accessed by free online access after registration on their website. These results show that most of the digital resources cannot be downloaded directly and the variety of access mechanisms illustrates that there is still a lack of governance policy for sharing data in Australia.

Finally, the participants were requested to indicate the metadata automation workflow they prefer (Figure 9). As a result, most of them (91.7%) agreed with the

option “initially metadata representation automatically generated, then edited by a person”. Less than half of them (41.7%) agreed that the automation process should be fully automatic. About 16% of them selected the option that a person enters some metadata, which is then automatically processed/encoded and/or augmented. All of the participants disagreed that a person creates all metadata. In addition, 8.3% of them selected the “other” category and noted that the creation process can be a combination of a tool to identify spatial data sets and manual entry so that completeness of metadata can be audited.

Figure 9: Distribution of preferred metadata automation workflow by organizations



6. FINDINGS AND FUTURE RESEARCH QUESTIONS

The result of the questionnaire responses analysis identified a number of areas for improvement where spatial metadata automation research should further investigate. These areas are here discussed.

6.1 The Issues and Challenges

The study found the main issues and challenges associated with generating and updating spatial metadata within organizations that can be listed as below:

- Inconsistency between the standards used to create metadata.
- Lack of appropriate policy to deliver metadata by third parties.
- Time consuming, monotonous and a labor-intensive process for generating and updating spatial metadata manually.
- Lack of appropriate regime for updating spatial data and metadata.
- Lack of a proper approach for updating spatial data and metadata simultaneously which creates inefficiency through more data redundancy.
- Multiple parties involved in the spatial data and metadata creation and update processes.
- Lack of internal policies and governance to share and publish metadata.
- Encouraging custodians on effectively maintaining their records for the organizations dealing with different spatial data sets.
- The variety of tools used for metadata creation and updating.

6.2 Metadata Standards and Formats

Automatic metadata generation can be facilitated when the structure of metadata is based on a selected standard (Olfat *et al.*, 2010). However, the study shows that organizations are still using various metadata standards such as in-house standards, regardless of recent efforts of ANZLIC.

As a result, considering a common standard with which most organizations agreed on might be one of the more fruitful areas in future metadata automation research.

On the other hand, the format of creating and storing spatial metadata is another key area considered for automation research in terms of technological trends. As the survey shows, XML is the most common format which the organizations are working on or they are moving toward.

6.3 The Spatial Data and Metadata Updating Time Interval and Responsibility

According to participants' responses, the updating time interval for spatial data sets cannot be exactly defined and it usually depends on data custodians and the nature of data sets. Based on this problem, the updating time interval for metadata would not follow any specific time frame.

Besides, the study shows that there is no specified organizational structure as a responsible party to update spatial data and metadata simultaneously within the organizations. This may prevent metadata from being current with spatial data set changes.

6.4 Spatial Metadata Entry Tools

Different organizations are seeking automated approaches to facilitate their efforts of creating and updating metadata for data sets. Based on the results of the questionnaire, most organizations are currently using the semiautomatic metadata creation approaches and they are applying different metadata entry tools which suit their needs. Within these tools, GeoNetwork, ANZMet Lite and ESRI products are the most common ones.

6.5 Sharing and Publishing Spatial Metadata

Despite the agreement of all organizations to share metadata, currently only less than half of them are sharing their whole metadata records with other parties. In addition, different platforms are used by the organizations to share and publish metadata, with ASDD being the most popular one.

6.6 Metadata Automation Workflow

The survey shows that the organizations do not prefer fully automatic metadata generation and they desire the direct operator auditing to be involved in the process. Indeed the preferred workflow is initially metadata representation automatically generated, then edited by an operator (particularly for the metadata elements that may not be automated). In this workflow the operator would audit the completeness of metadata and specify the next updating date as well.

6.7 Future Research Questions

As a result of the study, in addition to the findings of the metadata survey, some questions that would be considered in the future research were also derived from the participants' responses. These questions were almost all related to the feasibilities of metadata automation process.

Involving the capacity for object reuse (e.g. contact details), the ability to integrate virtually with spatial data tools, building a modeling tool to design or even discover

metadata requirements in advance, dynamic generating of any metadata element from a data set, storing and delivering metadata with the data set, simultaneous updating spatial data and its related metadata, consistency checks between metadata for similar data types and data sets, harvesting the percentage of attribute information that was complete for a spatial data set, addressing the discovery, access and licensing issues relating to both human and machine requirements (because of the movement to machine-to-machine interactions), designing data models to have metadata created as part of the update process at a feature level, creating metadata at the time of data collection either by the users involved in the data capture process or automatically by the instrument itself, and compliance with relevance metadata schemas especially ANZLIC and ISO 19115 were the research questions that have been identified. These questions are important and will be considered in future metadata automation research.

7. CONCLUSION

A variety of organizations at different jurisdictional levels in Australia have been improving the spatial metadata management process for several years to address the users' needs regarding spatial data discovery and access.

This paper has reviewed the main activities and initiatives and also the current status of spatial metadata management in Australia. Results show that there is a critical need to improve the existing approaches to create and update spatial metadata in an automated fashion, which illustrates the significance of the current Spatial Metadata Automation Research for spatial industry.

The study on organizational needs assessment around the automation research also showed that the organizations were dealing with a huge number of spatial data sets which might be shared with the other parties. So the first step would be publishing the up-to-date metadata records, which are now facing different challenges and issues. The major issues are regarding inconsistent standards, lack of appropriate user interfaces, the time and cost of manual metadata creation and updating, lack of a proper approach to create metadata automatically, the multiple parties involved in metadata creation and the updating process, lack of the ability to update spatial data and metadata simultaneously and automatically, and inconsistency of metadata structure within different organizations.

In addition to the existing issues which should be addressed via the automation research, several key areas related to the research were also found. Popularizing a common standard which most organizations agreed on, defining a regime to update spatial data and metadata, using Markup Languages (e.g. XML) to store and publish metadata, developing a semiautomatic approach to update spatial metadata aided by a direct operator auditing, integrating spatial data and metadata together to facilitate the simultaneous updating, and compatibility with ASDD and GeoNetwork environments could be some of these areas.

However, it should be emphasized that the spatial metadata automation research is still in its infancy and needs more concentration on users' needs, technological trends, existing tools and the feasibility of automatic approaches.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of the Australian Research Council Linkage Project on Metadata Automation and the members of the Centre for Spatial Data Infrastructures and Land Administration at the Department of Geomatics, 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

ANZLIC (2010a). *ANZMet Lite metadata entry tool (version 1.01)*, at http://www.anzlic.org.au/read/2469370794.html, [accessed 3 March 2010].

ANZLIC (2010b). *Metadata Standards*, at http://www.anzlic.org.au/infrastructure_metadata.html, [accessed 20 February 2010].

ASDD (2010). *Australian Spatial Data Directory (ASDD)*, at http://asdd.ga.gov.au/asdd/, [accessed 14 March 2010].

BlueNet (2010a). *About BlueNet*, at http://www.bluenet.org.au/About.html, [accessed 12 March 2010].

BlueNet (2010b). *The BlueNet Metadata Entry and Search Tool (MEST)*, at http://www.bluenet.org.au/MEST_intro.html, [accessed 20 March 2010].

CS2i (2010). *About CS2i*, at http://www.cs2i.nsw.gov.au/about_cs2i [accessed 21 March 2010].

CSDILA (2010). *Spatial Metadata Automation- Research Aims, at* http://www.csdila.unimelb.edu.au/projects/metauto/aims.html , [accessed 2 October 2010].

Göbel, S., and Lutze, K. (1998). "Development of Meta Databases for Geospatial Data", *in The WWW.* Paper presented at the 6th ACM International Symposium on Advances in Geographic Information Systems 1998, Washington DC, USA.

Interragator+ (2010). *What is Interragator+*, at https://www2.landgate.wa.gov.au/interragatorplus/help/index.html?what_is_interragator.htm, [accessed 16 March 2010].

Kalantari, M., Rajabifard, A., and Olfat, H. (2009). "Spatial metadata automation: a New Approach"*. Paper presented at the Surveying and Spatial Sciences Institute Biennial International Conference*, Adelaide, South Australia.

Limbach, T., Krawczyk, A., and Surowiec, G. (2004). "Metadata lifecycle management with GIS context", *Paper presented at the 10th EC GI & GIS Workshop*, ESDI State of the Art, Warsaw, Poland.

NISO. (2004). Understanding Metadata*, NISO Press Booklets,* at www.niso.org, [accessed 5 June 2009]

Olfat, H., Rajabifard, A., and Kalantari, M. (2010). Automatic Spatial Metadata Update: a New Approach. *Paper presented at the FIG 2010 Congress*, Sydney, Australia.

Pasca, M., Petriglia, L., Mattioni, F., Torchio, M., and Mariotti, C. (2009). "Experiences in the Creation and Updating of INSPIRE Compliant Metadata Catalogue". *Paper presented at the GSDI - 11 Conference*, at http://www.gsdi.org/gsdi11/papers/pdf/305.pdf, [Accessed 4 September 2009]

Rajabifard, A., Kalantari, M., and Binns, A. (2009). "SDI and Metadata Entry and Updating Tools", in B. v. Loenen, J. W. J. Besemer and J. A. Zevenbergen (Eds.), *Spatial Data Infrastructure Convergence. Research, Emerging Trends, and Critical Assessment*, pp 121-136.

Rajabifard, A., Kalantari, M., Binns, A., and Bennett, R. (2007). *Final Report on BlueNet- Anzlic ver 1.0*. Melbourne: The University of Melbourne.

Tsou, M.H. (2002) "An Operational Metadata Framework for Searching, Indexing, and Retrieving Distributed Geographic Information Services on the Internet", in Egenhofer M, Mark D (Eds.), *Geographic Information Science-Second International Conference GIScience 2002*, Lecture Notes in Computer Science 2478, Springer, Berlin, pp 312–333.

Williams, S., Pettit, C. J., Hunter, D. C., and Cherry, D. (2009). "A Prototype Metadata tool for land use change and impact models - A Case Study in Regional Victoria, Australia", in B. v. Loenen, J. W. J. Besemer and J. A. Zevenbergen (Eds.), *Spatial Data Infrastructure Convergence: Research, Emerging Trends, and Critical Assessment,* pp 137-150.

Author/s:
OLFAT, HAMED; Kalantari, Mohsen; RAJABIFARD, ABBAS; Williamson, Ian P.; Pettit, Christopher; Williams, Stephen

Title:
Exploring the key areas of spatial metadata automation research in Australia

Date:
2010

Citation:
Olfat, H., Kalantari, M., Rajabifard, A., Williamson, I. P., Pettit, C., & Williams, S. (2010). Exploring the key areas of spatial metadata automation research in Australia. In GSDI 12 World Conference: Realising Spatially Enabled Societies, Singapore.

Publication Status:
Published

Persistent Link:
http://hdl.handle.net/11343/28953

File Description:
Exploring the key areas of spatial metadata automation research in Australia