Automatic Spatial Metadata Updating and Enrichment

By

Hamed Olfat

January 2013

A thesis submitted to The University of Melbourne in fulfilment of the degree of Doctor of Philosophy

Centre for Spatial Data Infrastructures (SDIs) and Land Administration,
Department of Infrastructure Engineering, School of Engineering,
The University of Melbourne,
Victoria, Australia

ABSTRACT

Spatial information is necessary to make sound decisions at the local, regional and global levels. As a result, the amount of spatial datasets being created and exchanged between organisations or people over the networked environment is dramatically increasing. As more data and information is produced, it becomes more vital to manage and locate such resources. The role in which spatial metadata, as a summary document providing content, quality, type, creation, distribution and spatial information about a dataset, plays in the management and location of these resources has been widely acknowledged.

However, the current approaches cannot effectively manage metadata creation, updating, and improvement for an ever-growing amount of data created and shared in the Spatial Data Infrastructures (SDIs) and data sharing platforms. Among the available approaches, the manual approach has been considered monotonous, time-consuming, and a labour-intensive task by organisations. Also, the existing semi-automatic metadata approaches mainly focus on specific dataset formats to extract a limited number of metadata values (e.g. bounding box). Moreover, metadata is commonly collected and created in a separate process from the spatial data lifecycle, which requires the metadata author or responsible party to put extra effort into gathering necessary data for metadata creation and updating. In addition, dataset creation and editing are detached from metadata creation and editing procedures, necessitating diligent updating practices involving at a minimum, two separate applications. Metadata and related spatial data are often stored and maintained separately using a detached data model that results in avoiding automatic and simultaneous metadata updating when a dataset is modified. The spatial data end users are also disconnected from the metadata creation and improvement process.

Accordingly, this research investigated a framework and associated approaches and tools to facilitate and automate the spatial metadata creation, updating and enrichment processes. This framework consists of three complementary approaches namely ‘lifecycle-centric spatial metadata creation’, ‘automatic spatial metadata updating (synchronisation)’, and ‘automatic spatial metadata enrichment’ and a newly integrated data model for storing and exchanging spatial dataset and metadata jointly.

The lifecycle-centric spatial metadata creation approach aimed to create metadata in conjunction with the spatial data lifecycle steps. The automatic spatial metadata updating (synchronisation) approach was founded on a GML-based integrated data model to update metadata affected by the dataset modification concurrent with any change to the dataset, regardless of dataset format. The automatic spatial metadata enrichment approach was also design-rooted in Web 2.0 features

(tagging and folksonomy) to improve the content of spatial metadata keyword element through monitoring the end users’ interaction with the data discovery and retrieval process.

The proposed integrated data model and automatic spatial metadata updating and enrichment approaches were successfully implemented and tested via prototype systems. The prototype systems then were assessed against a number of requirements identified for the spatial metadata management and automation and effectively responded to those requirements.

DECLARATION

This is to certify that the thesis has not been submitted for a higher degree to any other university or institution. The text does not exceed 100,000 words.

Parts of this work were published in books, journals, refereed conference proceedings, professional magazines, and newsletters as listed in Appendix 1.

Hamed Olfat

January 2013

ACKNOWLEDGEMENTS

After all these years, I've got quite a list of people who have contributed in some way to this thesis, for which I would like to express thanks.

First and foremost I wish to acknowledge and thank my supervisors for their enduring support.

I would like to express my deep and most sincere gratitude to Professor Abbas Rajabifard, my supervisor and mentor. His extensive knowledge, enthusiastic support, inspiration, personal guidance and untiring help during my difficult moments have been of great value for me. I would have been lost without him.

I am also deeply grateful to my supervisor, Professor Ian Williamson for his detailed and constructive comments, and for his significant support throughout this work. His understanding, encouraging and sustained commitment to the ideas of my PhD have provided a good basis for the present thesis.

I would also like to especially thank my supervisor, Dr Mohsen Kalantari for his kind and incessant support and help to keep me on track during my PhD journey. He spared no effort to transfer his vast knowledge and experience to me as well as helping me to develop the ideas that came up to my mind in tackling the PhD challenges as they arose.

I warmly thank my supervisor, Associate Professor Chris Pettit who provided his expert advice, encouragement and support during my candidature. His detailed and key feedback, particularly during the thesis preparation, played a significant role in improving the quality of this thesis.

I also wish to thank the staff and postgraduate students from the Department of Infrastructure Engineering for their assistance and support. In particular, to my past and present fellow researchers in the Centre for Spatial Data Infrastructures and Land Administration (CSDILA), thank you for your friendship and support during my candidature; Professor Ian Bishop, Dr Rohan Bennett, Dr Kevin McDougall, Dr Payam Ghadirian, Dr Hossein Mohammadi, Dr Sheelan Vaez, Dr Ida Jazayeri, Dr Serryn Eagleson, Dr Malcolm Park, Ms Jude Wallace, Mr Brian Marwick, Ms Katie Potts, Ms Nilofer Tambuwala, Ms Serene Ho, Mr Ali Aien, Mr Davood Shojaei, Mr Sam Amirebrahimi, Mr Muyiwa Agunbiade, Mr Heri Sutanta, Mr Amir Nasr, Mr Farhad Laylavi, Mr Paul Box, and Mr Faisal El-Qureshi. I would like to especially thank Mr Hervé Senot for all his help in developing technical solutions during my PhD.

My special thanks goes to Professor Stephane Winter, Geomatics Discipline Leader in the Department of Infrastructute Engineering at the University of Melbourne, for all his valuable feedback and support throughout my PhD.

Thanks are also due to the visiting staff to the Department who provided their invaluable advice and research perspectives, especially Ms Liz Marchant (ANZLIC), Professor David J. Cowen (University of South Carolina), Professor Stéphane Roche (Université Laval), Dr Adam Iwaniak (Wroclaw University of Environmental and Life Sciences), Dr Joep Crompvoets (University of Leuven), Mr Mark Reichardt (OGC), Mr Steven Ramage (OGC), Dr Seyyed Abdolhadi Daneshpour (Iran University of Science and Technology), Dr Abolghasem Sadeghi-Niaraki (Inha University), and Dr Zaffar Sadiq Mohamed-Ghouse (Sinclair Knight Merz).

My research would of course not been possible without the support of the Australian Research Council (ARC) Linkage–Project on Spatial Metadata Automation and industry partners including the Victorian Departments of Primary Industries (DPI) and Sustainability and Environment (DSE); the Land and Property Management Authority (LPMA) – NSW; AusSoft Solutions; CubeWerx; and Logica Pty Ltd. I would like to especially thank Mr Stephen Williams and Mr David Hunter (DPI), Mr Gary Mahon (SII, DSE), Mr Greg Windsor (LPMA), Mr Jan Wandek (AusSoft), Mr Brad Spencer (CubeWerx Australia), Mr Phil Tattersall (Logica) and Mr Michael Dixon (PSMA).

I was also extraordinarily fortunate to have very good friends during the undertaking my PhD, in particular Dr Mehran Ektesabi, Mr Ali Orangi, Mrs Elham Karami, Mr Amir Orangi, Mr Mohammad Kalbasi and Dr Hootan Shah Hosseini, Thank you.

My special thanks goes to Ms Felicity Clissold, without her, this thesis would be unintelligible to read.

My deepest gratitude goes to my family for their unflagging love and support throughout my life; this dissertation is simply impossible without them.

I am indebted to my parents, Dr Mohammad Jafar Olfat and Mrs Haeideh Ghassemi for their everlasting care, support, advice, and love. My father spared no effort to provide the best possible environment for me to be successful in life. I cannot ask for more from my mother, as she is simply perfect. I have no suitable word that can fully describe her everlasting love to me. I also feel proud of my sisters, Haleh and Elham, my brother-in-law Mr Ahmad Seirafi, my niece Yalda, and my nephews Pedram and Aram for their care, support, and love during these years. I would also like to thank my father and mother in-law Mr and Mrs Shajiei and my

brothers-in-law, Khosro and Khashayar for their continuous support and love during these years.

And last but certainly not the least thanks goes to Nazanin, my beautiful and lovely wife, for her understanding, great patience, support and encouragement, and being the source of inspiration to keep me going. My name is written on the first page of this work, but her name is written on every page – because there was not one moment of this journey that she was not with me. Nazanin, it’s my turn now to provide a proper environment for you to pursue your academic education and make your wishes come true.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ABS	Australian Bureau of Statistics
AGIMO	Australian Government Information Management Office
AHP	Analytic Hierarchy Process
AJAX	Asynchronous JavaScript and XML
AMeGA	Automatic Metadata Generation Applications
ANP	Analytic Network Process
ANZLIC	Australia New Zealand Land Information Council
AODCJF	Australian Ocean Data Centre Joint Facility
APSIM	Agricultural Production Systems sIMulator
ARC	Australian Research Council
ASDD	Australian Spatial Data Directory
AURIN	Australian Urban Research Infrastructure Network
BPPT	Badan Pengkajian dan Penerapan Teknologi
CAD	Computer Aided Design
CGIAR	Consultative Group on International Agricultural Research
CGSB	Canadian General Standards Board
CSDGM	Content Standard for Digital Geospatial Metadata
CSDILA	Centre for Spatial Data Infrastructures (SDIs) and Land Administration
CSW	Catalogue Services for the Web
CUGIR	Cornell University Geospatial Information Repository
DBMS	Database Management Systems
DCMES	Dublin Core Metadata Element Set
DCMI	Dublin Core Metadata Initiative
DPI	Department of Primary Industries
DSE	Department of Sustainability and Environment
DTM	Digital Terrain Model
EPA	Environmental Protection Agency
ESA	European Space Agency
EUOSME	European Open Source Metadata Editor
FAO	Food and Agriculture Organisation
FGDC	Federal Geographic Data Committee
FOSS	Free and Open Source Software

GEMINI	GEo-spatial Metadata INteroperability Initiative
GIS	Geographic Information System
GML	Geography Markup Language
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSL	GeoSpatiumLab
GUI	Graphical User Interface
GWT	Google Web Toolkit
HMM	Hidden Markov Model
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
ICA	International Cartographic Association
IDE	Integrated Development Environment
ILS	Integrated Library System
INCITS	International Committee for Information Technology Standards Technical Committee
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organisation for Standardisation
IT	Information Technology
IVR	Interactive Voice Response
JSON	JavaScript Object Notation
KML	Keyhole Markup Language
LGPL	Lesser General Public License
LPMA	Land and Property Management Authority
MEF	Metadata Exchange Format
MET	Metadata Entry Tool
MGR	Metadata Generation Research
MIKE	Model Information Knowledge Environment
NAP	North American Profile
NGDA	National Geospatial Data Asset
NISO	National Information Standards Organisation
NSW	New South Wales
OAI-MHP	Open Archives Initiative Metadata Harvesting Protocol
OGC	Open Geospatial Consortium
OMB	Office of Management and Budget

OSDM	Office of Spatial Data Management
OSS	Open Source Software
OWL	Web Ontology Language
PCGIAP	Permanent Committee on GIS Infrastructure for Asia and the Pacific
PDO	PHP Data Objects
PNG	Portable Network Graphics
PSMA	Public Sector Mapping Agencies
PURL	Persistent Uniform Resource Identifier
QA	Quality Assurance
RDBMS	Relational Database Management System
RDF	Resource Description Framework
SDI	Spatial Data Infrastructure
SDTS	Spatial Data Transfer Standard
SKOS	Simple Knowledge Organisation System
SII	Spatial Information Infrastructure
SLIP	Shared Land Information Platform
SOAP	Simple Object Access Protocol
SPARQL	SPARQL Protocol and RDF Query Language
SUS	System Usability Scale
THREDDS	Thematic Real-time Environmental Distributed Data Services
UML	Unified Modelling Language
UN	United Nations
UNEP	UN Environmental Programme
UNOCHA	UN Office for the Coordination of Humanitarian Affairs
URL	Uniform Resource Locator
UUID	Universally Unique Identifier
VCCAP	Victorian Climate Change Adaptation Program
VGI	Volunteered Geographic Information
VIC	Victoria
WALIS	Western Australian Land Information System
WAR	Web Application Archive
WebDAV	Web Distributed Authoring and Versioning
WED	Western Ecology Division
WFS	Web Feature Service

WFS-T	Web Feature Service-Transaction
WMS	Web Map Service
WOS	Web Ontology Service
WPS	Web Processing Service
WWW	World Wide Web
XML	eXtensible Markup Language
XSL	EXtensible Stylesheet Language

CHAPTER 1

INTRODUCTION

1 INTRODUCTION

1.1 Research Background

In recent years, the amount of spatial datasets being created and exchanged between people or organisations over the networked environment has increased dramatically. As more data and information is produced, it becomes more vital to manage and locate such resources (Göbel and Lutze 1998). The role in which spatial metadata plays in the management and location of these resources, has been widely acknowledged (Tsou 2002, Limbach *et al.* 2004).

Metadata is commonly defined as 'data about data' and is the key to ensuring that digital resources will survive and continue to be accessible into the future (NISO 2004). Metadata plays a critical role in any Spatial Data Infrastructure (SDI) initiatives and spatial data sharing platforms (Ezigbalike and Rajabifard 2009, Rajabifard 2007) of which the aims are to facilitate data, and in a wider context, resource sharing, discovery and access. Meanwhile, 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 datasets, 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 datasets for use (Rajabifard *et al.* 2009).

However, long-standing issues and challenges regarding spatial metadata management (collection, creation, storage, improvement, maintenance, discovery, and access) have been acknowledged by the geospatial community (Batcheller *et al.* 2007, Batcheller *et al.* 2009, Kalantari *et al.* 2009, Manso *et al.* 2004, Manso-Callejo *et al.* 2009, Manso-Callejo *et al.* 2010, Morris *et al.* 2007, Taussi 2007, Zarazaga-Soria *et al.* 2003, Balfanz 2002).

Spatial metadata creation 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 spatial data lifecycle (Olfat *et al.* 2012c). Collecting and creating metadata later requires considerable effort and not all the information might be available (Timpf *et al.* 1996). The metadata gathered in this way is often missing or incomplete (Rajabifard *et al.* 2009).

The current spatial metadata creation approaches are highly dependent on the metadata author or responsible party's knowledge of the spatial data lifecycle. Among these approaches, the manual approach has been considered as monotonous, time consuming, and a labour-intensive task by the end users (West and Hess 2002, Guptill 1999). The spatial data industry also views the manual metadata approach as an overhead in terms of time and cost (Najar *et al.* 2007). The manual metadata approach effectively inserts a layer of human intervention into the Web-based search and retrieval process (Friesen *et al.* 2002) which affects the metadata quality and usefulness (Hatala and Forth 2003).

Furthermore, the semi-automatic metadata creation and updating approaches (implemented in metadata entry tools) that extract a limited number of metadata elements from the dataset are mainly restricted to dataset formats. For instance, CatMDEdit automates metadata generation only for Shapefile, DGN, ECW, FICC, GeoTIFF, GIF/GFW, JPG/JGW, and PNG/PGW formats (CatMDEdit 2011b).

Having explored the current manual and semi-automatic metadata approaches, it was also realised that spatial data and metadata storage mechanism are rooted in a detached data model (Kalantari *et al.* 2009). In this data model the spatial data and its associated metadata are stored separately in different files or databases. Separation of storage creates two independent datasets that must be managed and updated: spatial data and metadata (Rajabifard *et al.* 2009). This would result in prevention of metadata from being current with the spatial dataset changes and considerable delay in metadata updating. Thus, the most up-to-date metadata would not be accessible for the end users at the same time with dataset modification.

In addition to these challenges, Cooper *et al.* (2011a) discuss that users are not involved in the development of standards, such as assessing quality or documenting metadata. Kalantari *et al.* (2010) also argue that the users are disconnected from the spatial metadata creation and improvement process. They further discuss that the current data catalogue systems require more interaction with the end users to improve the content of metadata keyword element. According to Klien *et al.* (2004), in current standards-based catalogues, users can formulate queries using keywords which will be matched against the metadata fields. Therefore, the keywords embedded in the metadata should be comprehensive and address the probable queries made by the end users from diverse categories. A keyword should be popular meaning that majority of end users along with the metadata responsible party agree on that keyword.

Accordingly, the geospatial community requires new approaches to overcome the issues and challenges discussed in this section. Minimising human interference during metadata management should be the main consideration for designing any new approach or method. With this in mind, investigation of the feasible automatic approaches for creating, updating and improving the content of spatial metadata is central to this PhD research entitled ‘Automatic

Spatial Metadata Updating and Enrichment'. This research is part of an ARC[1] – Linkage Project entitled 'Spatial Metadata Automation' coordinated by the researchers from the Centre for Spatial Data Infrastructures (SDIs) and Land Administration, at the University of Melbourne in conjunction with some industry partners. The project's partners include the Victorian Departments of Primary Industries (DPI) and Sustainability and Environment (DSE), the Land and Property Management Authority (LPMA) – NSW, AusSoft, CubeWerx, and Logica Pty Ltd.

1.2 Research Problem

The number of spatial datasets is increasing considerably. Thus an effective, efficient and continual approach for managing the metadata of vastly growing spatial data is essential to facilitate sharing, discovery, and access of data.

However, the current spatial metadata management approaches deal with different issues and challenges. Metadata is commonly collected and created in a separate process from the spatial data lifecycle, which requires the metadata author or responsible party to put extra effort into gathering necessary data for metadata creation. Also, dataset creation and editing are detached from metadata creation and editing procedures, necessitating diligent updating practices involving at a minimum, two separate applications. Metadata and related spatial data are often stored and maintained separately using a detached data model. This issue results in avoiding the automatic and simultaneous metadata updating when a dataset is modified. In addition, the end users are disconnected from the metadata creation and improvement process. The current spatial data catalogues require more interaction with the end users, and in particular, to improve the content of the 'keyword' metadata element. This element is the key for discovering and finding desirable datasets over the Web.

Therefore, the research problem to be investigated in this thesis is as follows:
Current spatial metadata management approaches are:

a) not integrated with the spatial data lifecycle

b) not effective and efficient for automatic and continuous updating of the metadata content whenever the datasets are modified

c) not sufficiently interactive with the end users to involve them in creating and improving the content of metadata.

[1] Australian Research Council

This problem would result in the prevention of metadata from being current with spatial dataset changes and limitation of metadata easy accessibility; whilst increasing the risk of delivering incomplete, imprecise, and out-of-date metadata.

1.3 Research Questions

In considering the research problem a number of key research questions emerged, namely:

1. What is the current status of spatial metadata management and its requirements in terms of automation?
2. What is the relationship between spatial metadata elements and spatial data lifecycle?
3. Can an approach be designed to integrate spatial metadata creation with the spatial data lifecycle?
4. Can an automatic and dataset format agnostic approach be designed and implemented to overcome the ineffectiveness and inefficiency in updating metadata whenever the dataset is modified?
5. Can an automated approach be designed and implemented to improve the content of spatial metadata through more interaction with the end users seeking spatial data over the Web?

1.4 Research Aim

In recognising the research problem the central aim of the research is to design and develop a framework to facilitate and automate the spatial metadata management process, which:

a) integrates the metadata creation process with the spatial data lifecycle and reduces the burden of the metadata author’s intervention during metadata creation

b) effectively and efficiently supports the updating of metadata whenever the datasets are modified

c) enriches the content of metadata through the end users’ interaction in the data discovery process.

1.5 Research Objectives

The following objectives were formulated to achieve the research aim and address the research questions:

1. To investigate the underlying principles and components of spatial metadata automation
2. To study the current status of spatial metadata management in the geospatial community and identify the requirements of metadata automation and its related challenges
3. To design an approach for integrating metadata creation process with the spatial data lifecycle
4. To design approaches for automatic spatial metadata updating and enrichment
5. To implement prototype systems for automatic spatial metadata updating and enrichment
6. To evaluate the automatic spatial metadata updating and enrichment prototype systems and identify the areas which need improvement.

1.6 Research Approach

The research approach consists of 4 major phases, starting with the conceptualisation of research requirements, followed by the design (including the design of the lifecycle-centric spatial metadata creation approach and automatic spatial metadata updating and enrichment approaches), implementation (of prototype systems), and evaluation phases (Figure 1-1).

The first phase, *Conceptual Phase*, investigated the requirements of spatial metadata automation. In this phase, to establish the theoretical background of the research, an extensive literature review was undertaken on two main areas: metadata (non-spatial and spatial), and its automation approaches. Books, journals, organisation reports, conference proceedings, available metadata entry tools, visits and information published over the World Wide Web (WWW) were used to collate a range of information for reviewing in these two areas. In addition to the literature review, a case study was undertaken in Australia in order to identify the current status of spatial metadata management and the requirements for the spatial metadata automation. Also, a number of spatial metadata management tools were selected and assessed against a set of criteria developed for this research. Finally, the results achieved from the whole first phase were integrated and the main challenges regarding the spatial metadata management and automation were recognised. This phase was undertaken to address the first two objectives of the research.

After conceptualising the research requirements, in the second phase, *Design Phase*, a framework was designed and developed, which consists of three main approaches, namely 'lifecycle-centric spatial metadata creation', 'automatic spatial metadata updating' and 'automatic spatial metadata enrichment'. As a result, this phase was undertaken to address the research objectives 3 and 4.

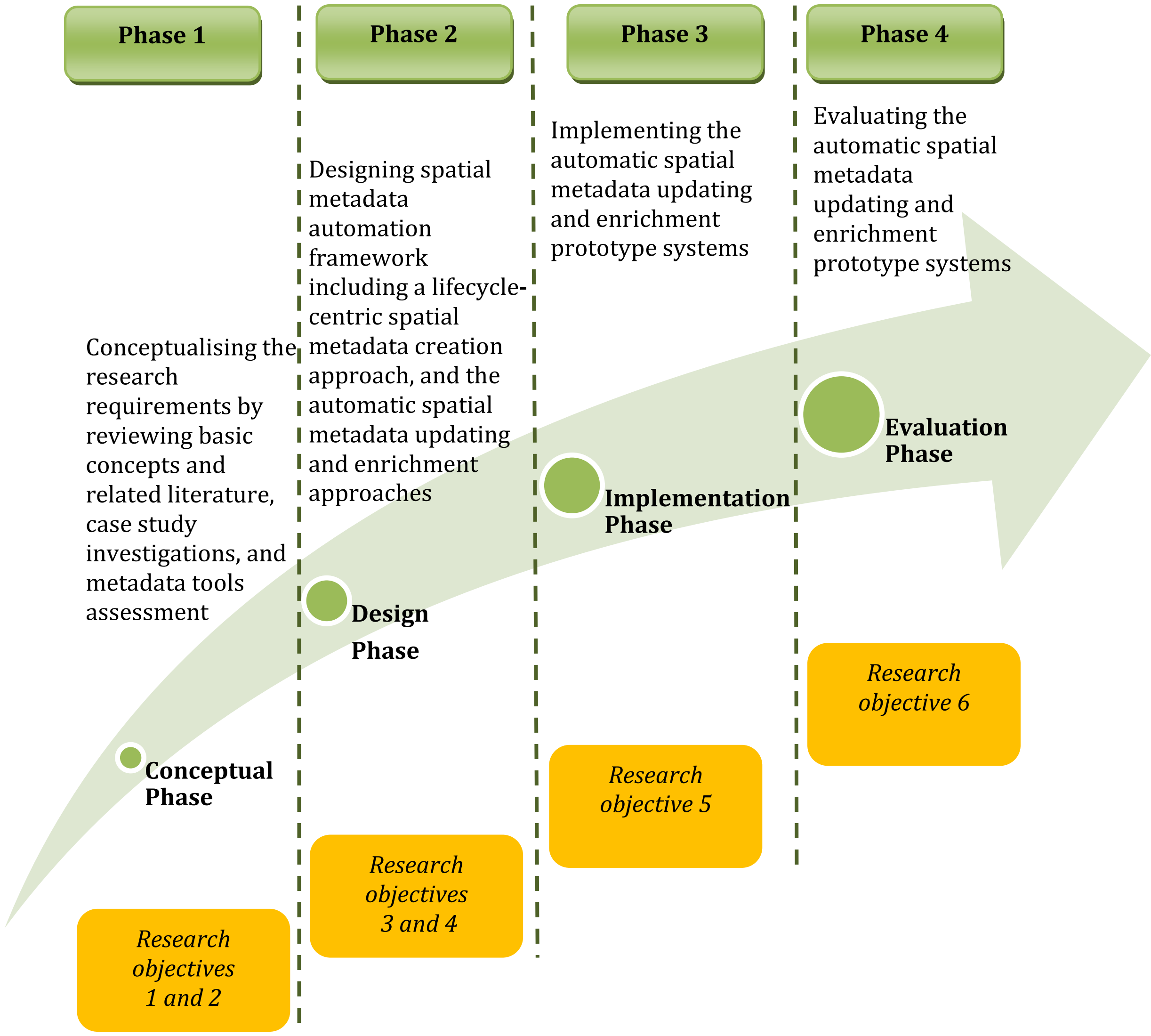


Figure 1-1: Research approach and connectivity with research objectives

Once the automatic spatial metadata updating and enrichment approaches were designed, in the next phase, *Implementation Phase*, two prototype systems were implemented to address the main challenges regarding spatial metadata management and automation identified during the research. The automatic spatial metadata updating prototype system was developed rooted in open source environments such as GeoNetwork, deegree, PostgreSQL/PostGIS, OpenLayers, and GeoExt. Also, the automatic spatial metadata enrichment prototype system was implemented within two different environments: GeoNetwork as an open source spatial data catalogue application and Model Information Knowledge Environment (MIKE) as an example of a data product – data modelling environment. MIKE was developed by the DPI–Victoria. MIKE was used in this phase for two reasons: first to use a real world system to prove the concept of automatic metadata enrichment, and secondly, to apply metadata enrichment approach to a non-spatial data catalogue (model catalogue) in addition to a spatial data catalogue (GeoNetwork). Accordingly, this phase was carried out to address the research objective 5.

In the final phase, *Evaluation Phase*, a set of criteria was developed for the assessment of the prototype systems. In this phase, two questionnaires were designed and distributed among the organisations that participated in the case study of the research in phase 1, as well as, some other interested parties. The results were then evaluated against the requirements of the spatial metadata management and automation and thus the areas which currently need improvement, were identified. Therefore, this phase was performed to address the research objective 6.

1.7 Delimitation of Scope

This research focuses on facilitating the process of spatial metadata creation, automating the process of metadata updating concurrent with dataset modification, and automating the enrichment of metadata ‘keyword’ content based on the end users’ interaction in the spatial data discovery and retrieval process.

In terms of automating the process of metadata updating, the focus of research is on vector datasets. Investigation into the metadata automation methods for raster datasets is not considered in the scope of this research. However, a number of researches on metadata automation for raster data have been explored in the literature review.

In addition, this research concentrates on addressing the problem of disconnection between the end users and spatial metadata creation and improvement through designing and developing an automatic spatial metadata enrichment approach that could also indirectly result in facilitating the data discovery and retrieval process. Accordingly, investigation into the approaches that can directly facilitate the spatial data discovery and retrieval process through improving the machine-processing capability of data discovery systems is out of the scope of this research.

1.8 Thesis Structure

The remaining chapters of this thesis have been formed as illustrated in Figure 1-2 to address the research objectives.

Chapter 2, *Spatial Metadata Underlying Principles and Related Literature*, starts with an introduction to the spatial metadata concepts and its significant role in any SDI platform. Then, different categories of spatial metadata, which are important for the automation purposes, are reviewed. It is followed by a comparison between different metadata data models that have a vital impact on metadata automation research. The chapter then presents and discusses different steps of the spatial metadata lifecycle and the relationship between these steps and the metadata automation. Finally, the importance of metadata standardisation for the purpose of automation is outlined.

Chapter 3, *Spatial Metadata Automation: Background and Technological Perspective*, reviews metadata automation related works in two main areas: digital library and information science community and the geospatial community. Different approaches and methods that facilitate the metadata collection, creation, updating and improvement are then explored in both communities. Based upon a metadata automation literature review in the geospatial community, this chapter identifies and summarises the main challenges that this community faces for metadata management and automation. Finally, the technologies focused in this thesis for the spatial metadata automation purpose are explored in detail.

Chapter 4, *Research Design and Methods*, outlines the research design and methods that were used to answer the research questions and achieve the research aim and objectives. This chapter first investigates the research conceptual design framework by reviewing the research problem and questions, and then explores the possible research methods available to answer these questions. The chosen research methods are then justified and the final research design including 'conceptual', 'design', 'implementation', and 'evaluation' phases presented. Finally, the ethical considerations and prototype systems communication relating to the undertaken surveys are described.

Chapter 5, *Spatial Metadata Management and Automation Current Status*, presents the results of the qualitative case study investigations of identifying the current status of spatial metadata management and also the metadata automation requirements in the context of Australia – the selected case study region. Moreover, the chapter explores a number of proprietary and open source spatial metadata management tools. These tools are examined against a set of criteria developed for this thesis. Finally, this chapter integrates the results of the case study investigations and tools assessment with the findings of the literature review chapters (Chapters 2 and 3) and summarises the main challenges the geospatial community faces regarding metadata collection, creation, updating and improvement.

Chapter 6, *Lifecycle-centric Creation, Automatic Updating and Enrichment of Spatial Metadata*, designs and develops a new framework to address the main challenges regarding spatial metadata management and automation identified in this research. The new framework includes three complementary approaches namely 'lifecycle-centric spatial metadata creation', 'automatic spatial metadata updating (synchronisation)', and 'automatic spatial metadata enrichment'. The lifecycle-centric approach focuses on creating metadata values in parallel to the spatial data lifecycle. The synchronisation approach is designed to update spatial data and metadata in real time and is rooted in a GML[2]-based integrated data model for spatial data and metadata storage and delivering. The automatic spatial metadata enrichment approach is also designed based on Web 2.0 features and aims at enriching the content of spatial metadata keyword element through interaction with the end users seeking spatial data over the Web.

Chapter 7, *Automatic Spatial Metadata Updating and Enrichment Prototype Systems Implementation*, first explores the implementation of automatic spatial metadata updating prototype system following the conceptual design of related approach discussed in Chapter 6. The prototype system is built upon the GeoNetwork opensource catalogue. This will result in an improvement of share-ability and accessibility of the prototype system within the geospatial community. Next, the chapter reviews the implementation of the automatic spatial metadata enrichment prototype system within two different environments, including the GeoNetwork and Model Information Knowledge Environment (MIKE).

Chapter 8, *Prototype Systems Evaluation*, explores the results of automatic spatial metadata updating and enrichment prototype systems evaluation. This chapter also summarises the recommended feedback from end users and experts to improve the proposed approaches for automating metadata and prototype systems designed and implemented to achieve this.

Chapter 9, *Conclusions and Recommendations*, examines the outcomes achieved during this research, reflects on the original research problem and suggests directions for future research efforts.

[2] Geography Markup Language

Objective 1: To investigate the underlying principles and components of spatial metadata automation
Spatial Metadata Concept, Role of Spatial Metadata within SDI Platforms, Spatial Metadata Data Models, Spatial Metadata Lifecycle, Development of Worldwide Spatial Metadata Standards, Worldwide Research and Development on Metadata Automation in Digital Library and Information Science Community and the Geospatial Community, Technological Perspective of Spatial Metadata Automation
Chapters 2 and 3

Objective 2: To study the current status of spatial metadata management in the geospatial community and identify the requirements of metadata automation and its related challenges
Research Design and Methods, Conceptual Design Framework, Case Study Framework, Prototype Systems Evaluation Criteria, Questionnaires Design and Distribution, Case Study Findings, Criteria for Assessing Spatial Metadata Management Tools, Findings of Spatial Metadata Management Tools Assessment, Summary of Spatial Metadata Management and Automation Main Challenges
Chapters 4 and 5

Objective 3: To design an approach for integrating metadata creation process with the spatial data lifecycle
A Framework for Addressing the Spatial Metadata Management and Automation Main Challenges, Designing a Generic Spatial Data Lifecycle, Mapping ISO 19115:2003 Metadata Elements against the Spatial Data Lifecycle Steps, Conceptual Design for Lifecycle-centric Spatial Metadata Creation Approach
Chapter 6

Objective 4: To design approaches for automatic spatial metadata updating and enrichment
Automatic Spatial Metadata Updating (Synchronisation) Conceptual Design, Technical Requirements for Implementation of Metadata Synchronisation, Automatic Spatial Metadata Enrichment Conceptual Design, Indirect Metadata Enrichment Model, Direct Metadata Enrichment Model
Chapter 6

Objective 5: To implement prototype systems for automatic spatial metadata updating and enrichment
Automatic Spatial Metadata Updating and Enrichment Implementation Architecture, Prototype Systems Capabilities Overview, Lessons Learned from Prototype Systems Implementation within the GeoNetwork opensource
Chapter 7

Objective 6: To evaluate the automatic spatial metadata updating and enrichment prototype systems and identify the areas which need improvement
Findings of the Automatic Spatial Metadata Updating and Enrichment Prototype Systems Evaluation, Discussion on Improvements
Chapter 8

Conclusions and Recommendations
Chapter 9

Figure 1-2: Thesis structure to address the objectives

1.9 Chapter Summary

This chapter has laid the foundations for the research and introduced the problem, questions, aim, objectives, approach, and structure of the research. The problem statement shows that the geospatial community now faces an increasing number of spatial datasets being created and exchanged over the networked environment; and therefore an effective, efficient and continuous approach for creating, updating and improving the content of metadata is essential to facilitate the spatial data sharing, discovery, and access. However, the current spatial metadata management approaches are, first of all, highly dependent on the metadata author's intervention and knowledge of the spatial data lifecycle for metadata creation, and, also, are not effective and efficient for automatic and continuous updating of metadata content whenever the datasets are modified, and, finally, do not interact properly with the end users to improve the content of spatial metadata.

To respond to the problem statement, the research sets out six objectives: firstly, to investigate the underlying concepts and components of spatial metadata automation; secondly, to study the current status of spatial metadata management in the geospatial community and indentify the requirements of metadata automation and its related issues and challenges; thirdly, to design an approach for integrating metadata creation process with the spatial data lifecycle; fourthly, to design the automatic spatial metadata updating and enrichment approaches; fifthly, to implement prototype systems for automatic spatial metadata updating and enrichment; and finally, to evaluate the prototype systems.

The next chapter provides a background to the spatial metadata including its definition, importance, objectives, categories, role in SDI, data models, lifecycle, and worldwide standards.

CHAPTER 2

SPATIAL METADATA UNDERLYING PRINCIPLES AND RELATED LITERATURE

2 SPATIAL METADATA UNDERLYING PRINCIPLES AND RELATED LITERATURE

2.1 Introduction

In order to address the first research objective, stated in Chapter 1, this chapter along with next chapter aims to review the underlying principles and related literature of spatial metadata automation. This chapter first describes the definition, importance and objectives of metadata in the geospatial community. Due to the significant role of metadata in any SDI platform for spatial data sharing and discovery, this role is then discussed in detail. Then, different categories of spatial metadata that are important for the automation purposes are reviewed. It is followed by a comparison between different metadata association models and data models that have a key impact on metadata automation research. The chapter then presents and discusses different steps of spatial metadata lifecycle including collection, creation, storage, publication, discovery, retrieval and access, and updating and the relationship between these steps with the metadata automation. Finally, the importance of metadata standardisation for the purpose of automation is outlined.

2.2 Spatial Metadata Definition

The term metadata, purportedly first used in 1969 (Howe 1996) is often called ‘data about data’ (ANZLIC 1996, Zarazaga-Soria *et al.* 2003) or ‘information about information’. The term ‘meta’ derives from the Greek word denoting a nature of a ‘higher order’ or more ‘fundamental kind’ (Taylor 2003), or ‘above’, ‘beyond’, and ‘of something in a different context’ (Litwin and Rossa 2011).

A more formal definition is:

> *‘metadata is data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics (Dempsey and Heery 1997).’*

A metadata record consists of a number of pre-defined elements representing specific attributes of a resource, and each element can have one or more values (Taylor 2003). These elements could be the content, quality, condition, origin, and other characteristics of data or other pieces of information which presents the best way to share information about the data without having to provide the actual data (Taussi 2007). In other words, metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource (Timpf *et al.* 1996). NISO (2004) also states that metadata is the key to ensuring that [digital] resources will survive and continue to be accessible into the future.

Similarly, the metadata concept has been defined by the geospatial community. It has been classified by Philips *et al.* (1998) into two different forms. The first, and oldest, form of metadata occurs within Geographic Information System (GIS), Computer Aided Design (CAD) packages and databases where it is '[an] underlying set of rules which tells a software program how to handle data' (Wilson 1998). The second form of metadata is a recent development. Metadata have become products in their own right, especially in the spatial data management field, where they are used to describe the characteristics of datasets. This form of metadata is a summary document providing content, quality, type, creation, and spatial information about a dataset (ESRI 2002b). The spatial metadata term used in this thesis is based on the definition of the second form of metadata.

This kind of metadata describes different aspects of any spatial data, including (FGDC 2000):

Identification: What is the name of the dataset? Who developed the dataset? What geographic area does it cover? What themes of information does it include? How current is the data? Do restrictions exist for accessing or using the data?

Data Quality: How good is the data? Is information available that allows a user to decide if the data is suitable for his or her purpose? What is the positional and attribute accuracy? Is the data complete? Was the consistency of the data verified?

Spatial Reference and Organisation Information: How is the data referenced to the real world (coordinate systems, datum)? How is the data organised (data models, topology)?

Entity and Attribute Information: What geographic information (roads, houses, elevation, temperature, etc.) is included? How is the information encoded? Were codes used? What do the codes mean?

Distribution: From whom can I obtain the data? What formats are available? What media is available? Is the data available online? What is the price of the data?

Having defined the spatial metadata concept, the significance of metadata in the geospatial community is discussed in the next section.

2.3 Importance of Metadata in the Geospatial Community

Spatial information is necessary to make sound decisions at the local, regional and global levels (Nebert 2004). Therefore, the amount of spatial datasets being created and exchanged between organisations or people is increasing considerably. According to a released study by Daratech (2011) for the period of 2004–2010, the overall growth of geospatial industry has increased by 11% in the areas of data, software and services. The report highlights that the spatial data is the fastest growing segment of the geospatial industry and is definitely becoming a major contributor to the overall growth of the industry. Nowadays, the volunteers (e.g. contributors to the OpenStreetMap initiative[3]) are producing a large amount of data. The social media (e.g. Facebook[4] and Twitter[5]) is also broadly used by smart phone users, for exchanging location-based data.

As more spatial data 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). Batcheller (2008) also agrees with this view and states that metadata is often employed by institutions to organise, maintain and document their spatial resources internally, and may also provide a vehicle for exposing marketable data assets externally when contributed to online geospatial exchange initiatives.

Besides, metadata has long been recognised as crucial to geospatial asset discovery (Batcheller *et al.* 2009) and selection. An important reason for creating descriptive metadata is to facilitate discovery of relevant information (NISO 2004) and reduce data duplication (ESRI 2002b). Data selection is the activity that aims to choose the most appropriate data for a specific application. It is strongly affected by the complexity of spatial data. Two main aspects are critical in the selection of spatial data (Albertoni *et al.* 2004):

- *Usually it is not possible to access spatial resources to have a look and to realise 'at a glance' the information that they contain since spatial data are resources that can be heavy in terms of volume or that can be not available for free.*
- *To compare different spatial resources requires strong efforts since data are available in a huge kind of variety. They differ in characteristics like Scale, Reference System, Geographic Extension, Themes, Quality, and Fees and so on.*

[3] http://www.openstreetmap.org/
[4] http://www.facebook.com/
[5] https://twitter.com/

Spatial metadata is adopted to overcome these issues: metadata gives a detailed description of the spatial data characteristics according with a specific standard. They provide a first level of data integration and allow comparing sources provided by different organisations. Moreover, they represent a means to choose spatial data without resource download. Therefore, they enable both professional and non-professional spatial users to find the most appropriate, applicable and accessible datasets for use (Rajabifard *et al.* 2009).

FGDC (2000) considers three major objectives for the creation of metadata. The first one is to organise and maintain an organisation's investment in data. As personnel change or time passes, information about an organisation's data will be lost. Later workers may have little understanding of the content and uses for a database and may find that they cannot trust results generated from these datasets. The second objective is to provide information to data catalogues and clearinghouses. By making metadata available through data catalogues and clearinghouses, organisations can find data to use, partners to share data collection and maintenance efforts, and customers for their data. Finally, the third objective of metadata is to provide information to aid data transfer. Metadata should accompany the transfer of a record. The metadata aids the organisation receiving the data process and interprets data, incorporates data into its holdings, and updates internal catalogues describing its data holdings.

Metadata has also been key in the development of spatial data sharing platforms (including Spatial Data Infrastructure (SDI)) since their first emergence (Batcheller *et al.* 2009). Metadata has been acknowledged as one the fundamental indicators for assessing any spatially enabling platform (Rajabifard 2007, Ezigbalike and Rajabifard 2009). The Infrastructure for Spatial Information in Europe (INSPIRE) Geo Portal (INSPIRE 2012), Australian Spatial Data Directory – ASDD (ASDD 2010), and Australian Urban Research Infrastructure Network (AURIN) Portal (AURIN 2012) could be good examples of using spatial metadata for sharing, discovering and accessing datasets in the context of SDI. The definition of SDI and the role of metadata in its context are discussed in further detail in the next section.

After all, further incentives for metadata use arise when the implications of neglecting them entirely are considered. Some claim that the cost of not creating metadata can outweigh that of authoring it, citing concerns associated with employee turnover, data redundancy, conflicts and inappropriate decision making (Batcheller 2008).

According to the significance and objectives of metadata, any approach or method that could facilitate and improve its management, are welcomed by the geospatial community. The aim of this thesis discussed in Chapter 1 confirms that the research is in-line with the current need of community.

2.4 The Role of Spatial Metadata within the SDI Platforms

An SDI provides a basis for spatial data discovery, evaluation, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general (Nebert 2004).

According to Coleman and Nebert (1998), the main components of an SDI should include data providers (sources of spatial data), databases and metadata, data networks, technologies (dealing with data collection, management, search and representation), institutional arrangements, policies and standards, and end users (Figure 2-1).

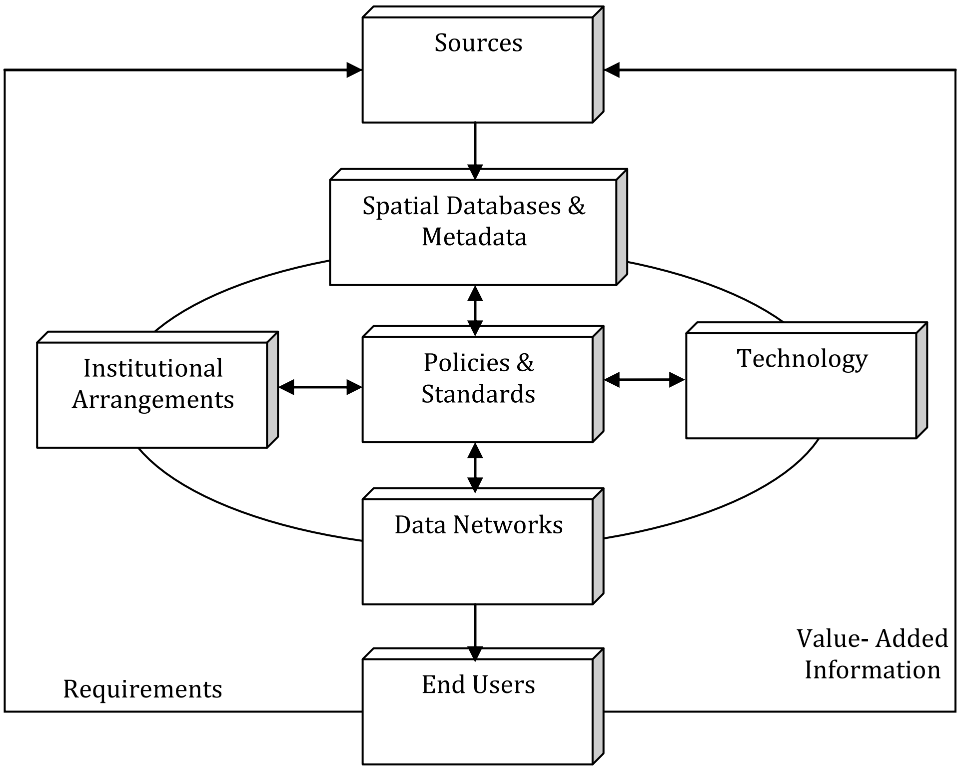


Figure 2-1: A system view of the spatial data infrastructure components (adopted from Coleman and Nebert (1998) and Nogueras-Iso *et al.* (2005))

Whilst there are many definitions of SDIs, a useful framework is the one put forward by Rajabifard *et al.* (2002), as shown in Figure 2-2, which places particular emphasis on the dynamic relationship between data, people, and a package that includes technology, policy and standards. The authors argue that the relationship between these categories is dynamic because changes of communities and society (people) and their needs, require access to different sets of data mediated by the ever-changing technology. The interactions among these components in turn put new demands on rights, restrictions and responsibilities enshrined in policy.

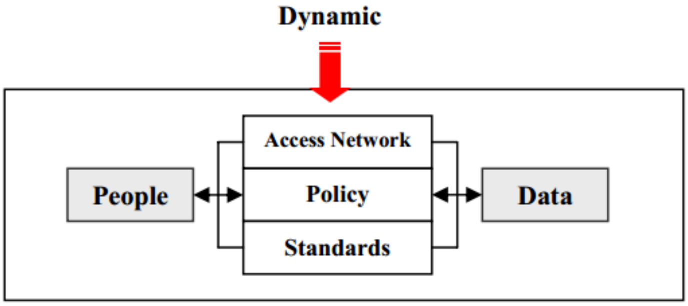


Figure 2-2: Nature and relations between SDI components (Rajabifard *et al.* 2002)

Coetzee (2012) investigates the generations of SDIs during the past decades. It is discussed that the SDIs have been evolving from a top-down product-based approach (e.g. National SDI of Australia) to a more process-based approach. This discussion is followed by the conclusion that today the SDIs need to accommodate the challenges of ubiquitous read-write access by millions of users from all kinds of devices and that the strict top-down approach of early SDIs is evolving into more dynamic bottom-up approaches. Similarly, Sadeghi-Niaraki *et al.* (2010) categorise the development of SDIs into three main groups: SDI 1.0 (data-centric), SDI 2.0 (service-centric), and SDI 3.0 (user-centric).

Within the SDI platform, metadata plays a key role to facilitate accessing up-to-date and high quality spatial data and services (Williamson *et al.* 2003) by allowing a potential user of a dataset to determine whether the dataset is useful to them or not (Phillips *et al.* 1998). Support of a discovery and access service for spatial information is known variously within the geospatial community as 'catalogue services' (Open Geospatial Consortium (OGC)), 'Spatial Data Directory' (Australian Spatial Data Infrastructure (ASDD)), and 'Clearinghouse' (U.S. Federal Geographic Data Committee (FGDC)). Although they have different names, the goals of discovering spatial data through the metadata properties they report are the same (Nebert 2000).

In this regard, Zarazaga-Soria *et al.* (2003) discuss that one of the main problems for launching an SDI is to have appropriate and well-defined contents for its catalogues, that is to say, metadata. Also, Nogueras-Iso *et al*. (2005) identify the spatial data catalogue as the most important part of an SDI. Similarly, Pasca *et al*. (2009) argue that the creation of metadata standards and a corresponding metadata catalogue is one of the first steps for the setting up of an SDI. Cooper *et al.* (2011b) also discuss that lack of metadata might provide a possibility for an SDI to fail.

Figure 2-3 shows the basic interactions of various individuals or organisations involved in the advertising and discovery of spatial data (Nebert 2000). In this Figure, the Catalog Gateway and its user interface allows a user to query distributed collections of spatial information through their metadata descriptions. The boxes are identifiable components of the distributed Catalog service; the lines that connect the boxes illustrate a specific set of interactions described by the words next to the line.

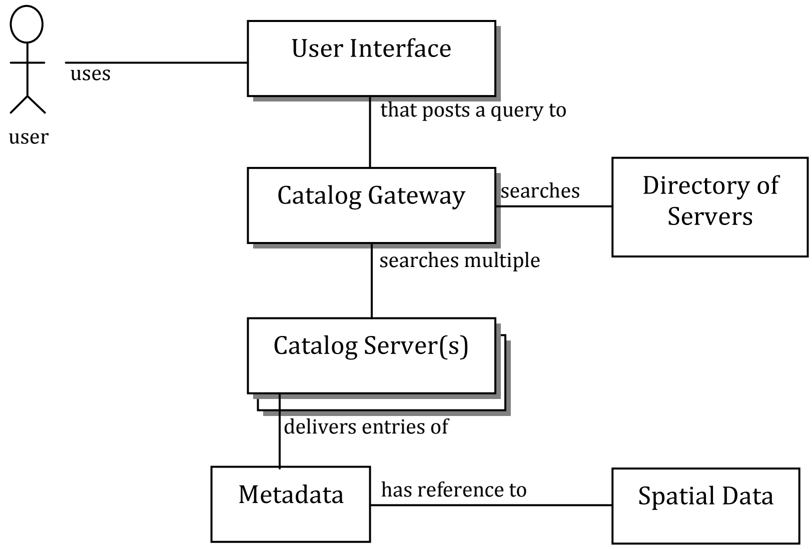


Figure 2-3: Interaction diagram showing basic usage of distributed catalogue services and related SDI elements from a user point of view (Nebert 2000)

A user interested in locating spatial information uses a search user interface, fills out a search form, specifying queries for data with certain properties. The search request is passed to the Catalog Gateway and poses the query of one or more registered Catalog Servers. Each Catalog Server manages a collection of metadata entries. Within the metadata entries, there are instructions on how to access the spatial data being described. There are a variety of user interfaces available in this type of Catalog search in various national and regional SDIs around the world. Interoperable searches across international Catalogs can be achieved through the use of a common descriptive vocabulary (metadata), a common search and retrieval protocol, and a registration system for servers of metadata collections.

The components of an SDI platform relating to the spatial metadata introduced by Nebert (2000) are used as the basis to design and implement the architecture of metadata automation prototype systems in this thesis. The next section categorises the spatial metadata into different groups following the roles they play in metadata automation.

2.5 Spatial Metadata Categories

So far, spatial metadata has been classified into different categories by the researchers.

Hill *et al.* (1999) categorise the spatial metadata in terms of collection into two groups:

- *Inherent Metadata* – that is, information that can be derived through the computer analysis of the contents of any collection, such as: temporal coverage (e.g. visualisation of the time periods covered or publication dates); types of items and the number of each; formats of items and the number of each; example of full metadata content; geospatial and image collections; number of thumbnail/browse images available; types of geospatial footprints (e.g. points, bounding boxes); geographic coverage of the information (map visualisation based on latitude and longitude coordinates of items in the collection)

- *Contextual Metadata* – that is, information supplied by the collection provider or collection maintainer that cannot otherwise be derived from the collection's contents, such as: title, responsible party, scope and purpose, type of collection (digital items, offline items, gazetteer, etc.), date (creation or latest update), update frequency, metadata schema(s), terms and conditions of use for the collection, contact(s) (name, position, email, etc.), special behaviours (e.g. search semantics) that the collection may exhibit or require in specific operational contexts (e.g. when accessed by a particular search engine).

Similarly, Jokela (2001) introduces the concepts of *implicit* and *explicit* metadata. When metadata are implicitly present in data and could be generated automatically by examining the data, metadata can be called implicit metadata. Explicit metadata requires manual work; it is typically descriptive information in the form of free text. When producing explicit metadata the author of the metadata must make explicit and perhaps subjective statements about the dataset. The majority of current spatial metadata are of the explicit kind. Examples of explicit metadata are information about the specific usage of the dataset; use and access constraints of the dataset; and specific file descriptions. Examples of implicit spatial metadata are information considering the geographic extent of the dataset, such as set of points defining the bounding polygon; the corner points; or the resolution of the dataset.

Moreover, according to Ahonen-Rainio (2005), metadata can be categorised into *objective* and *subjective* classes. Objective metadata are plain facts about the dataset at hand and can partially be produced automatically. Objective metadata are implicitly present in data, and therefore it is objective metadata that could be generated automatically by examining the dataset itself. Taussi (2007) specifically divides these objective metadata into:

- Objective parameters, which may be automatically produced out of the dataset. These objective parameters refer, for example, to the spatial and temporal extent of the dataset.

- Objective parameters, which may be produced out of already present and available metadata itself. This refers, for example, to information about the completeness of metadata in a particular dataset.

As these objective metadata often describe the status of the resource, which may change as time passes, they are usually dynamic. Subjective metadata is documentation about the dataset that is open to interpretation by the author of the metadata. Examples of subjective metadata are information about the source of the data and information about restrictions on the access and use of a resource. These metadata are usually static.

Accordingly, it can be realised that the researchers agree on classifying the spatial metadata into two main groups with the same purpose but different names. The first group includes the metadata that could be extracted automatically from the dataset (e.g. resource bounding box) namely ‘inherent’, ‘implicit’, or ‘objective’. The second group contains the metadata that should be generated by the author (e.g. resource restrictions) namely ‘contextual’, ‘explicit’, or ‘subjective’. This thesis has also focused on these two main groups for categorising spatial metadata for the automation purpose.

Following the current issues of spatial metadata in terms of storage and delivering discussed in Section 1.1 in Chapter 1, the next section explores different data models for storing spatial dataset and metadata.

2.6 Data Models and Spatial Metadata

There are various ways to associate metadata with resources, as discussed by Duval *et al.* (2002) and illustrated in Figure 2-4:

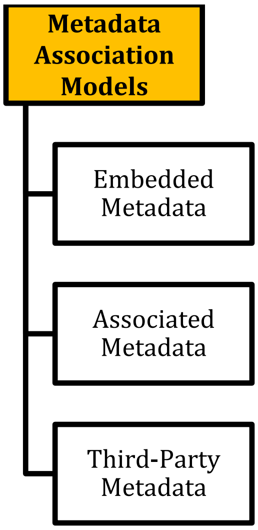


Figure 2-4: Metadata association models (adopted from Duval *et al.* (2002))

- *Embedded metadata* resides within the markup of the resource. This implies that the metadata is created at the time that the resource is created, often by the author. Experts differ concerning whether author-created metadata is best or whether it is better to have trained practitioners evaluate and describe resources. As a practical matter, resource description expertise is a scarce and costly commodity, and thus any investment by authors in the description of their intellectual products is likely to be of value. Embedded metadata can also be harvested, and the presumptive increase in visibility that might result is an incentive for creators to assign metadata.
- *Associated metadata* is maintained in files tightly coupled to the resources they describe. Such metadata may or may not be harvestable. The advantage of associated metadata derives from the relative ease of managing the metadata without altering the content of the resource itself, but this benefit is purchased at the cost of simplicity, necessitating the co-management of resource files and metadata files.
- *Third-Party metadata* is maintained in a separate repository by an organisation that may or may not have direct control over or access to the content of the resource. Typically such metadata is maintained in a database that is not accessible to harvesters, though the emerging Open Archives Initiative Metadata Harvesting Protocol (OAI-MHP) proposes a system that encourages the disclosure of metadata repositories among federated OAI servers.

Also, it was discussed by NISO (2004) that metadata can be *embedded* in a digital object or it can be stored *separately*. Metadata is often embedded in HyperText Markup Language (HTML) documents and in the headers of image files. Storing metadata with the object it describes ensures the metadata will not be lost, obviates problems of linking between data and metadata, and helps ensure that the metadata and object will be updated together. Using this type of metadata storage, which supports the automatic and concurrent metadata updating after any change to the dataset, has been focused for storing spatial metadata in this thesis.

Likewise, Taussi (2007) classifies the link between a metadata record and the resource it describes into two types: *embedded* or *external*. Metadata is embedded when stored and transmitted together with the data; this is typical in online services (Green and Bossomaier 2002) and used typically by Web search machines. Metadata is external when separated from the data they describe. External metadata is typical in cataloguing services, which is used to help the end user to find the appropriate datasets for their specified purposes.

As a result, it can be understood that metadata and its related resource can be generally stored either together or separately. Based on this, Kalantari *et al.* (2009) introduce the following two data models to store spatial dataset and metadata (Figure 2-5):

- Detached data model
- Integrated data model.

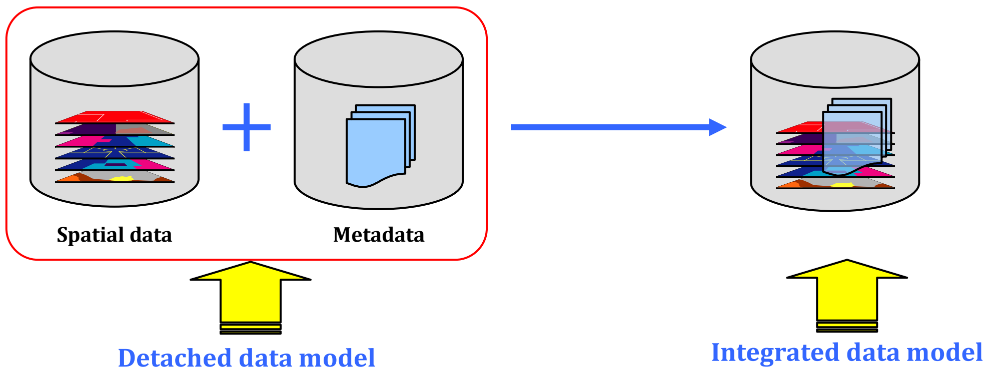


Figure 2-5: Data models for storing spatial data and metadata (adopted from Kalantari *et al.* (2009))

These data models are discussed in more detail below.

2.6.1 Detached Data Model

Traditionally, spatial metadata acquisition and management often play a less important role in many organisations. Metadata is usually acquired much after the spatial data has been captured and is stored in separate repositories. In a 'detached data model' the spatial data and its associated metadata are stored separately in different files or databases that makes them either without a relationship with each other, or to have only a common identifier (Olfat *et al.* 2012a). Consequently in this method of recording metadata there are two independent datasets to manage and update: spatial data and metadata; which are often redundant and inconsistent (Rajabifard *et al.* 2009).

Automation of spatial metadata stored based on this data model required synchronisation between two separate databases: metadata and spatial data databases. With the huge amount of spatial data in different formats, structure, and Database Management Systems (DBMS), automation of spatial metadata will be heavily associated with the issue of interoperability (Kalantari *et al.* 2009).

2.6.2 Integrated Data Model

To respond to the issue of the detached data model, recently researchers focus on the optimisation of metadata management by integrating metadata and spatial data in a common file or database (Giger and Najar 2003, Najar 2006). This common metadata-spatial dataset can be considered to be a 'comprehensive spatial data'. The concept of metadata-spatial data integration enables the spatial data to carry their own metadata description with them.

According to the results of the case study implemented by Najar (2006), the integration, managing and modelling existing metadata and spatial data commonly is realistic. The integration approach distinguishes between already existing spatial data models, which have to be extended and newly planned data models and sets, and can manage both spatial data and metadata together with commonality from the beginning (Najar *et al.* 2007).

An integrated model for metadata and spatial datasets will benefit the spatial data industry in general, as well as all industries that increasingly utilise spatial data in their day-to-day tasks. This will enable metadata to be maintained dynamically in a way that addresses the increased real-time requirements of people and organisations that use spatial data (Rajabifard *et al.* 2009).

When automating spatial metadata, the issue of interoperability between metadata and spatial databases can be addressed using an integrated database. This method in particular is in-line with the Synchronisation and Learning Object Methods of spatial metadata automation because of the structured and strict way of handling metadata imposed by the integrated data model (Kalantari *et al.* 2009). Therefore, this data model has been considered to be designed and developed as part of this thesis. It will be used particularly in the implementation of automatic spatial metadata updating approach discussed in Section 6.4, Chapter 6.

Having explored the above two data models, the next section reviews the spatial metadata lifecycle. In order for this thesis to focus on metadata automation requirements, there is a need to discover the steps involved in the metadata lifecycle.

2.7 Spatial Metadata Lifecycle

Reviewing the background of spatial metadata (Timpf *et al.* 1996, FGDC 2000, Nebert 2000, Taussi 2007, Greenberg 2003b, Nogueras-Iso *et al.* 2005, Moellering *et al.* 2005) confirms that the following steps form the spatial metadata lifecycle:

- Collection
- Creation
- Storage
- Publication
- Discovery
- Retrieval and Access
- Update.

The spatial metadata lifecycle is illustrated in Figure 2-6.

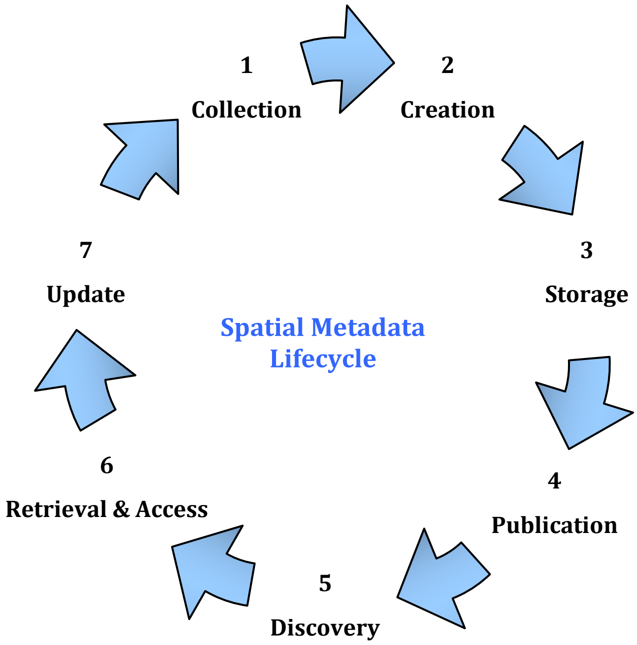


Figure 2-6: Spatial Metadata Lifecycle

These steps are here discussed in detail:

2.7.1 Spatial Metadata Collection

There are three main factors that should be considered for spatial metadata collection: (1) time of collection; (2) the necessary metadata values to be collected; and (3) the level the metadata should be collected.

The metadata is either gathered during the data collection process itself or at some later time. Generating metadata later requires considerable effort and not all the information might be available (Timpf *et al.* 1996). According to FGDC (2000), the best time to collect metadata is while the data is being developed, when the information needed for metadata is known. Waiting until after the data is developed risks less accurate information from being recorded and increases the cost caused by searching for information. The time period between the collection of spatial data and metadata can be crucial for quality of metadata (Najar 2006), as Beard (1996) also states that retroactive compilation of metadata is error-prone and sometimes even impossible.

Regarding the metadata values that should be collected during this step, the International Standards (e.g. ISO 19115: 2003) have attempted to enlighten the data custodians in which spatial metadata elements could be considered. The most widely accepted standards are explored in Section 2.8.

In terms of the level of metadata collection, according to Nebert (2000) metadata can be collected for spatial data at different levels. Metadata may exist at the collection level (e.g. satellite series), at a data product level (an image mosaic), at a data unit level (a vector dataset), at a group of features of a given type (certain roads), or even at a specific feature instance (a single road). Regardless of the level of abstraction, these associations of metadata to data objects should be maintained.

In practice, most metadata are currently collected at the dataset level, and a metadata entry in a catalogue refers the user to its location for access. Increasingly sophisticated providers of spatial data are including metadata at other levels of detail so as to preserve information richness. Metadata standards such as ISO 19115: 2003 allow different levels of metadata abstraction, and catalogue services will also need to accommodate this richness without confusing the user in its complexity. Public Sector Mapping Agencies (PSMA) Australia can be an illustration of the organisations interested in other levels of metadata by providing feature-level metadata for the 'Transport & Topography' dataset (PSMA 2012).

Once the required information of metadata is collected, the spatial metadata record should be created as part of metadata lifecycle.

2.7.2 Spatial Metadata Creation

The creation of spatial metadata can be separated into three approaches including automatic, semi-automatic and manual (Taussi 2007), as illustrated in Figure 2-7. Automatic approaches, such as automatic retrieval or searching and sorting of data are based on computerisation. Semi-automatic approaches combine automatic and manual approaches. Manual approaches are based on human reasoning and decision making.

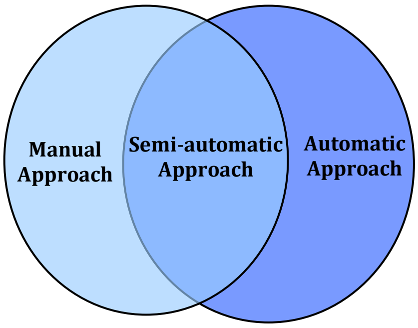


Figure 2-7: Spatial metadata creation approaches (adopted from Taussi (2007))

Among these approaches, many people view manual metadata generation as monotonous and time consuming, a labour-intensive process which is a major undertaking in itself (Guptill 1999, West and Hess 2002). This has resulted in a pervasive outlook which shuns metadata creation (Mathys 2004). Meanwhile, one of the main obstacles to the widespread adoption of systems that make intensive use of metadata, is the time and effort required to apply metadata to multiple resources and the inconsistencies and idiosyncrasies in interpretation that arise when this is a purely human activity (Hatala and Forth 2003). Moreover, it is commonly viewed by organisations as an overhead and extra cost (Najar *et al.* 2007). Finally, metadata for spatial datasets is often missing or incomplete and is acquired in heterogeneous ways (Rajabifard *et al.* 2009).

The use of automatic approach can, in turn, permit human resources to be directed to more intellectually challenging metadata creation and evaluation tasks. These factors underlie automatic metadata generation research efforts and the desire to build superior and robust automatic metadata generation applications (Greenberg *et al.* 2005). More importantly, the ability to automatically generate metadata relating to spatial data, and make it available through any SDI platform will have important benefits to all practitioners including spatial data producers, vendors, distributors and users. The benefits of automated metadata are savings in time and human resources, and in certain cases, an increased level of consistency when compared against human-created metadata (Baird and Jorum-Team 2006).

In addition to the importance of metadata creation methods, identifying the appropriate responsible party to create metadata is very critical within spatial metadata lifecycle.

In this regard, Greenberg (2003b) classifies the metadata responsible parties into four groups:

- professional metadata creators
- technical metadata creators
- content creators
- community or subject enthusiasts.

Professional metadata creators include cataloguers, indexers, Web masters and other persons who have had high-level training through a formal educational curriculum or an official on-the-job training program. Professional metadata creators have the intellectual capacity to make sophisticated interpretative metadata-related decisions and work with classificatory systems, complex schemas and other complex information standards. Technical metadata creators include para-professionals, data entry and other persons who have had metadata training – but much less extensive training than metadata professionals. Content creators are individuals responsible for the creation of the intellectual content, such as researchers that regularly produce abstracts, keywords and other types of metadata for their scientific and scholarly publications. Community or subject enthusiasts have not had any formal metadata creation training, but have special subject knowledge and want to assist with documentation. Relying on the community for metadata generation is becoming more widespread with Crowdsourcing and Volunteer Geographic Information (VGI) as part of Web 2.0 initiatives. The flickr[6] photo sharing and youtube[7] video sharing Websites are good examples of users' interaction in metadata provision for shared contents.

Similarly, Kalantari *et al.* (2010) categorise the responsible parties for the spatial metadata creation into three groups. Metadata experts form the first group, and this is the basis of most catalogues in SDIs. This often requires serious knowledge and background of spatial metadata management. The geospatial community in the metadata field has developed standards and schemes for cataloguing, categorisation and classification of spatial data. ISO 19115: 2003 and FGDC Content Standard for Digital Geospatial Metadata (CSDGM) are the examples of widely used standards, which are discussed in detail in Section 2.8.

While professionally created metadata are often considered to be of high quality, it is costly in terms of time and effort to produce. This makes it very difficult to scale and keep up with the vast amounts of new spatial data being produced and updated, especially with new technologies

[6] http://www.flickr.com/
[7] http://www.youtube.com/

like Global Navigation Satellite System (GNSS), Satellite imagery, automatic map creation methods and in particular mediums like the World Wide Web (WWW). An alternative is author created metadata, as the second group. Original producers of the spatial data provide metadata along with their creations. The Dublin Core Metadata Initiative has been used with some success in this area (Greenberg *et al.* 2001). Author created metadata may help with the scalability problems in comparison to professional metadata, but both approaches share a basic problem: the intended and unintended eventual users of the spatial information are disconnected from the process.

There is a third approach that can be utilised to create spatial metadata. Spatial data users will use their own language to describe the data and therefore they form the third group. This can go even deeper and they can express their comments not only about the data title but also on the other aspects of the metadata. This is a way of connecting users of spatial data to the process of creating and improving metadata. For instance, the capability of assigning information to user-generated spatial data within the wikimapia[8] platform (Figure 2-8) is based on the third metadata creation approach.

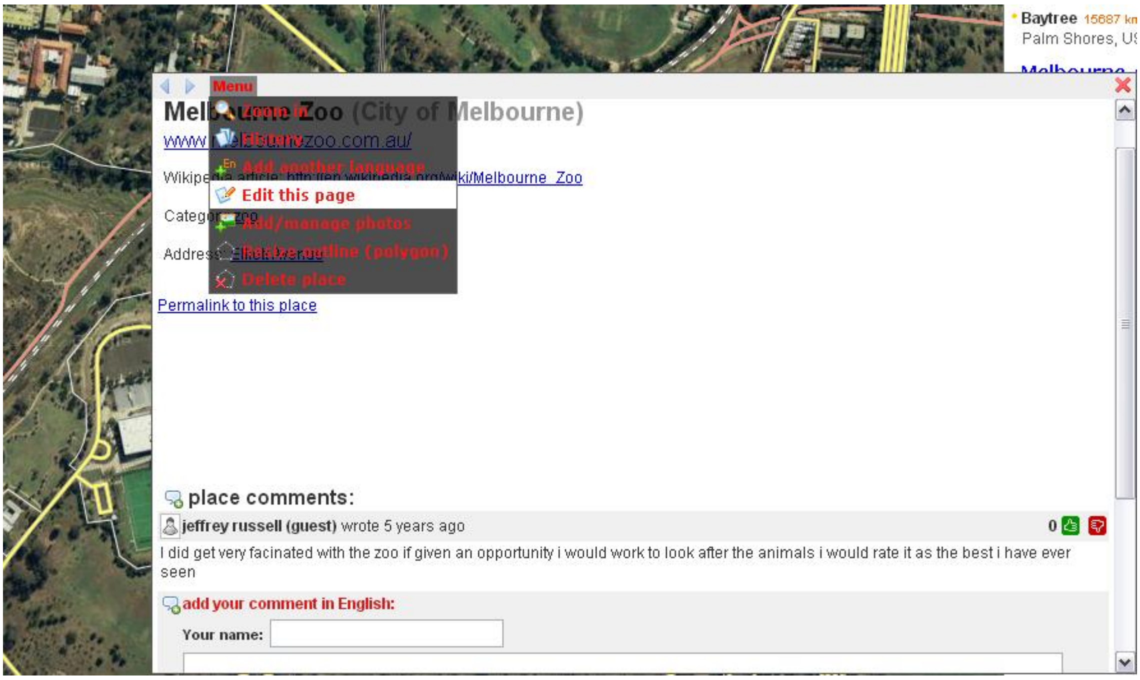


Figure 2-8: Provision of metadata for spatial data by users in wikimapia platform

Sharing spatial datasets in SDIs and allowing users to discover them and write notes about them, tag them and even share their notes with other users, opens another horizon for automation of spatial metadata and the efficient use of them. Designing and implementing the

[8] http://wikimapia.org/

automatic spatial metadata enrichment approach in this thesis, discussed in Section 6.5 in Chapter 6, has been undertaken based on this concept.

When the metadata is created, it should be stored in a structured manner. The next section reviews the storage step of metadata lifecycle.

2.7.3 Spatial Metadata Storage

Spatial metadata creation approaches have been formed and evolved based on the technological initiatives over time; thus the storage of spatial metadata has been influenced by these initiatives. For instance, after the PC Era and Internet initiative, the spatial metadata were created and stored in Markup Languages (e.g. HTML and eXtensible Markup Language (XML)) since the early 1990s (Olfat *et al.* 2010b). Figure 2-9 illustrates the spatial metadata creation approaches and different types of spatial metadata storage based on technological initiatives.

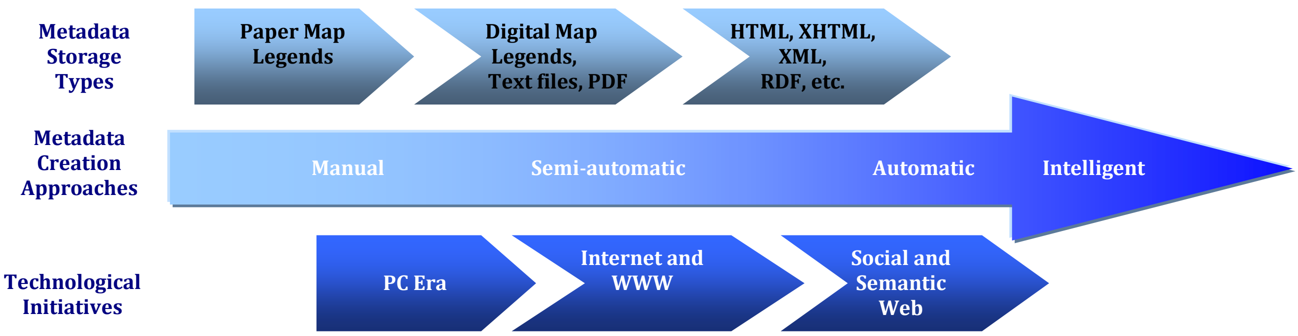


Figure 2-9: Spatial metadata creation approaches, storage types and technological initiatives (Olfat *et al.* 2010b)

Metadata publication for the purpose of spatial data sharing, discovery and access is considered as the next step of metadata lifecycle as explored in the following section.

2.7.4 Spatial Metadata Publication

Publishing metadata facilitates data sharing. Sharing data between organisations stimulates cooperation and a coordinated, integrated approach to spatially related policy issues (Land Information Council of Jamaica 2008).

Metadata records are usually published through catalogue systems, sometimes called directories or registries (Nogueras-Iso *et al.* 2005). Also, Catalogue Services for the Web (CSW) open standard by the Open Geospatial Consortium (OGC) supports the ability to publish and search collections of metadata for data, services, and related information objects. When collated into catalogues, metadata collections can be indexed for rapid query contributed to data clearinghouses or similar data exchange initiatives where they can be used to externally expose marketable data assets. It aids in the coordination of data procurement efforts by raising the awareness of extant datasets, thereby avoiding duplication of effort, redundant storage and obscuring search results (Batcheller 2008).

After publishing metadata, it would be discovered by the end users looking for required data. Thus, the next section explores the discovery step of metadata lifecycle.

2.7.5 Spatial Metadata Discovery

Because of spatial metadata's small size compared to the data it describes, metadata is more easily shareable (ESRI 2002b) and is considered as the surrogate of spatial datasets which is referenced to its related spatial dataset. Hence, in a networked environment, such spatial surrogates are discovered by the end users seeking required spatial datasets, through catalogue systems, Web services and user interfaces. The user interface usually supports making a variety of queries (via basic and advanced search) on spatial metadata records to retrieve the characteristics of the most appropriate datasets for the end users. Moellering *et al*. (2005) illustrate the spatial metadata discovery process as Figure 2-10.

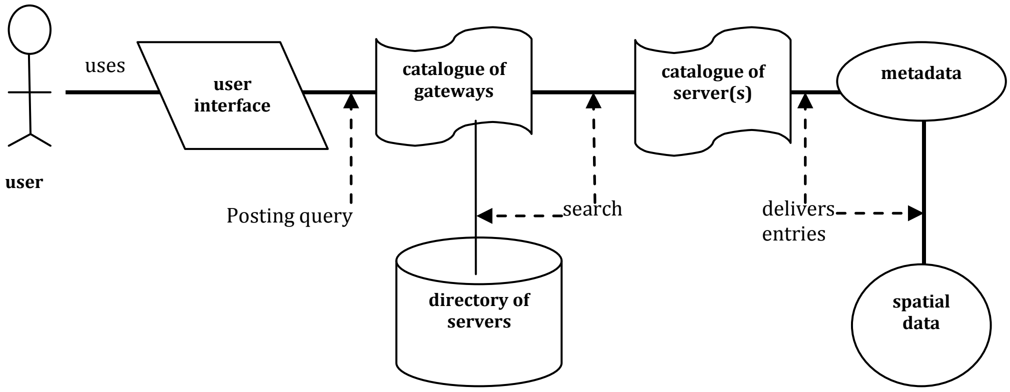


Figure 2-10: Finding spatial metadata through a catalogue of gateways and servers to find the spatial data itself (Moellering *et al.* 2005).

Once the results of metadata discovery are presented to the end users, the metadata records need to be retrieved and accessed by them. The next section gives a brief overview of the retrieval and access step of the metadata lifecycle.

2.7.6 Spatial Metadata Retrieval and Access

The result of spatial metadata discovery process is a list of metadata records corresponding to end users' queries. Here, the metadata are retrieved from a repository through a catalogue system and accessed by the end users in different formats (e.g. XML, HTML, and PDF) to help them make a decision on what spatial datasets suit their needs.

The last step of metadata lifecycle is metadata updating which should be undertaken in parallel with dataset maintenance process. The next section reviews this step of the metadata lifecycle.

2.7.7 Spatial Metadata Update

Spatial metadata should be updated according to any change in its related spatial dataset. For instance, when the geographic extent of spatial dataset is changed the metadata should be updated with the new information.

The metadata updating process can be either undertaken before metadata publication or after that. Updating published metadata would be a continuous activity that guarantees the consistency between metadata and their related spatial datasets.

Similar to the spatial metadata creation process, the updating process can be performed through manual, semi-automatic and automatic approaches. But here, the automatic spatial metadata updating is a process by which properties of a spatial dataset are read from the dataset and written into its spatial metadata. This automatic function will support the spatial metadata to be updated at the same time with their related spatial data update process. Therefore, it will benefit the organisations associated with spatial metadata to save time and effort and will also reduce the risk of inconsistency and redundancy in the spatial data and metadata. Following the predictable advantages of automatic spatial metadata updating, there has been increasing investigations by researchers (Balfanz 2002, Batcheller 2008, Batcheller *et al.* 2007, Batcheller *et al.* 2009, Kalantari *et al.* 2009, Manso-Callejo *et al.* 2009, Manso-Callejo *et al.* 2010).

However, the automatic updating is facing some challenges. The structure of spatial data and metadata data models is an important part of these limitations. Whereas, dataset creation and editing are detached from metadata creation and editing procedures, necessitating diligent updating practices involving at a minimum two separate applications (Batcheller 2008). Rajabifard *et al.* (2009) also discuss that separation of storage creates two independent datasets that must be managed and updated: spatial data and metadata. The authors highlight that an integrated data model can play an important role for handling spatial metadata by combining spatial data and metadata in a seamless approach. They further discuss that although using the integrated metadata data model facilitates metadata automation some elements of metadata obviously cannot be automatically updated and need human intervention (e.g. resource restrictions, abstract, and responsible party).

Furthermore, the spatial metadata entry tools designed to automate some part of the metadata updating process are restricted to dataset formats. For instance, CatMDEdit automates metadata generation only for Shapefile, DGN, ECW, FICC, GeoTIFF, GIF/GFW, JPG/JGW, and

PNG/PGW formats (CatMDEdit 2011b). This results in limited adoption of metadata entry tools due to the diversity of spatial data formats in the geospatial community (e.g. ESRI's Shapefile, file-based Geodatabase, and Enterprise Geodatabase, AutoCAD's DWG and DXF, MapInfo's TAB, Bentley's DGN, GeoJSON, BSB, PDS, PCIDSK, ISIS, ASCII Grid, GeoTIFF, PNG, etc.).

Therefore, there is a need in the geospatial community for an automated metadata updating approach that is not restricted to any dataset format and can overcome the detached metadata data model issues. Addressing this need has been considered in the fourth research objective, stated in Chapter 1.

Having explored the spatial metadata lifecycle, it should be emphasised that consideration of a metadata standard is critical for designing such a lifecycle. This standard will affect the whole lifecycle through defining the elements that should be gathered, the schema that metadata should be based on, and more importantly, the format that metadata should be stored in. Hatala and Forth (2003) discuss that metadata can fulfil its purpose only when it complies with some agreed upon standard. As the demand for standardised metadata increases, the geospatial community needs to identify automated metadata production methods that are more efficient and less costly than those practices involving manual production (Greenberg 2004). Automatic metadata generation can be facilitated when the structure of metadata is based on a selected standard. There are different standardisation methods in the metadata domain as discussed in the next section.

2.8 Development of Worldwide Spatial Metadata Standards

In order for metadata to be as useful and cost-effective as possible, it is essential that its structure, semantics and syntax conform to widely supported standards, so that it is effective for the widest possible consistency, maximises its longevity, and its processing can be automated as far as possible (Gill 2000). Taussi (2007) also discusses that metadata needs to conform to internationally acknowledged standards to make them readable, retrievable and interoperable for users all over the world. The metadata developed and applied using various and other than standardised rules, cannot be treated as the basis for comparison and objective assessment of quality of spatial data resources, even if the resource features are characterised as accurately as possible by these metadata (Litwin and Rossa 2011).

The advantages of metadata standardisation have been already summarised by Nebert (2004). One benefit of standards is that they have been developed through a consultative process (with other ‘experts’) and provide a basis from which to develop national or discipline-oriented profiles. As standards become adopted within the wider community, software programs will be developed to assist the industry in implementing the standard. Consistency in metadata content and style is recommended to ensure that comparisons can be made quickly by data users as to the suitability of data from different sources. According to Litwin and Rossa (2011), to ensure the functioning of metadata in computer networks (Internet/Intranet) and to achieve full interoperability between meta-information systems (metadata), it is necessary that both the metadata and these systems are based on clearly defined standards for spatial data. They further discuss that for the purposes of characterising metadata standards for geoinformation, they can be divided into two main groups. The first group includes standards that define rules for describing resources, i.e. for creating metadata. In the second group standards can be placed for the dedicated metadata services – that is, standards for metadata catalogues.

Different organisations and committees have been working on the development of standards for spatial metadata creation, such as the Federal Geographic Data Committee (FGDC) and the European Committee for Standardisation (CEN/Technical Committee 287) (Crompvoets *et al.* 2004). Also, the international organisations have been working for several years in order to achieve a common standard regarding metadata creation for spatial information. The International Organisation for Standardisation (ISO) through the ISO /TC211 Committee on Geographic Information/Geomatics published the world metadata standard (ISO 19115) in 2003, extensions for imagery and grid data (ISO 19115–2) in 2009, metadata for services (ISO 19119) in 2005 and the way of encoding ISO metadata in XML (ISO/TS 19139) in 2007 (Litwin and Rossa 2011). The Open Geospatial Consortium (OGC) also developed the OGC CSW (Catalogue Service for the Web) 2.0 and OGC CSW Application Profile for ISO 19115: 2003 and 19119: 2005. These standards relate to the capabilities of publishing and discovery of information describing spatial datasets and spatial data services. It should be noted that OGC specifications for CSW (metadata catalogue) could also be applied to different (non-ISO) profile such as: FDGC and Dublin Core.

In another effort, Moellering *et al.* (2005) represent the work undertaken over a period of eight years by the members of ICA[9] Spatial data Standards Commission who developed a meticulous set of scientific and technical characteristics, and then examined 22 national and international spatial metadata standards using these scientific and technical characteristics.

[9] International Cartographic Association

The following sections explore the most important metadata standards within the geospatial community including ISO 19115: 2003 (Kresse and Fadaie 2004) and FGDC along with Dublin Core as the most common standard for general metadata (Schindler and Diepenbroek 2008).

2.8.1 Dublin Core Metadata Standard

The Dublin Core Metadata Initiative (DCMI) created its first metadata standard, the Dublin Core Metadata Element Set (DCMES), to facilitate search and retrieval of Web-based resources. The American National Standards Institute approved DCMES in September 2001 as ANSI/NISO Z39.85 and, in February 2003, the ISO ratified the standard as ISO 15836 (McClelland 2003). DCMES is a standard for cross-domain information resource description. It provides a simple and standardised set of conventions for describing things online in ways that make them easier to find. Dublin Core is widely used to describe digital materials such as video, sound, image, text, and composite media like Web pages and can also be used for spatial data indexing (Kalantari *et al.* 2009).

The Dublin Core current standard version 1.1 includes 15 well-defined elements for describing ‘core’ information properties, ‘core’ because its elements are broad and generic, usable for describing a wide range of resources (DublinCore 2010). These metadata elements are illustrated in Table 2-1.

Table 2-1: Dublin Core metadata element set, version 1.1 (DublinCore 2010)

No.	Element name	Definition
1	Contributor	An entity responsible for making contributions to the resource.
2	Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
3	Creator	An entity primarily responsible for making the resource.
4	Date	A point or period of time associated with an event in the lifecycle of the resource.
5	Description	An account of the resource.
6	Format	The file format, physical medium, or dimensions of the resource.
7	Identifier	An unambiguous reference to the resource within a given context.
8	Language	A language of the resource.
9	Publisher	An entity responsible for making the resource available.
10	Relation	A related resource.
11	Rights	Information about rights held in and over the resource.
12	Source	A related resource from which the described resource is derived.
13	Subject	The topic of the resource.
14	Title	A name given to the resource.
15	Type	The nature or genre of the resource.

The Dublin Core metadata standard, version 1.1 in particular, has had a major impact on resource description in the context of the Web. Developed under the direction of the DCMI, the Dublin Core is an international and interdisciplinary metadata standard that has been adopted by an array of communities wanting to facilitate resource discovery and build an interoperable information environment (Greenberg 2004).

In the metadata community, standardisation of the Dublin Core has spurred development of a wide variety of metadata generation applications for creating metadata. Among such applications are a series of tools called metadata generators that almost exclusively rely on automatic processes (Greenberg 2003a).

Also, the studies undertaken by Batcheller *et al.* (2007) and Batcheller (2008) can be the illustrations of metadata automation for Dublin Core standard in the geospatial community. These studies are explored in more detail in Section 3.2.2, Chapter 3.

2.8.2 FGDC Content Standard for Digital Geospatial Metadata (CSDGM)

The first organisation to consider data about data was the United State Federal Geographic Data Committee (FGDC) with its Spatial Data Transfer Standard (SDTS). The goal of SDTS is to provide a common ground for data exchange by defining logical specifications across various data models and structures (Fegeas *et al.* 1992). This started the discussion on metadata and its organisation (Timpf *et al.* 1996). The Content Standard for Digital Geospatial Metadata (CSDGM) was developed by the FGDC to be used in the National Spatial Data Infrastructure (FGDC 2000).

During June 1992, the FGDC sponsored an 'Information Exchange Forum on Spatial Data' in which the need for a common definition of metadata in the community was identified. The FGDC developed a draft standard and sponsored a public review of this draft from October 1992 to April 1993. Finally, the standard was approved on June 8, 1994. In April 1997, the FGDC circulated, for public review, and second version of the CSDGM. In June 1998, the FGDC endorsed the second version of the FGDC Content Standard for Digital Geospatial Metadata (FGDC-STD-001-1998).

There are seven main sections and three supporting sections in this standard that describe different aspects of data that potential users might need to know (FGDC 2000, Westbrooks 2004a):

- Section 0 – Metadata
- Section 1 – Identification
- Section 2 – Data Quality
- Section 3 – Spatial Data Organisation
- Section 4 – Spatial Reference

- Section 5 – Entity and Attribute
- Section 7 – Distribution
- Section 7 – Metadata Reference
- Section 8, 9, 10 – Citation, Time Period, Contact.

Of these areas only Identification Information (basic information about the file such as originator, abstract, and purpose) and Metadata Reference Information (information about the production of the metadata) are defined as being mandatory for all records. All other areas of the standard are mandatory if applicable. Within each section are sub-fields that can be defined as mandatory, mandatory if applicable, or optional. This flexibility allows metadata creators to determine the level of detail that they can provide or support based on perceived user needs. It also guarantees that at least basic metadata will be recorded about each dataset.

The ESRI ArcCatalog product can be an illustration of spatial metadata entry tools provided by the geospatial industry that automates some part of FGDC metadata creation and updating. This tool is reviewed in more detail in Section 5.3.1, Chapter 5.

2.8.3 ISO 19115: 2003 Metadata Standard

The ISO 19115: 2003 metadata standard is one of the most comprehensive (but also the most complicated) metadata schemes for distributed geographic information services (Tsou 2002). The framework of geospatial metadata includes three conceptual levels: a data level, an application level, and a meta-model level. Each level highlights different aspects of the metadata model and its relationship to spatial datasets (Figure 2-11).

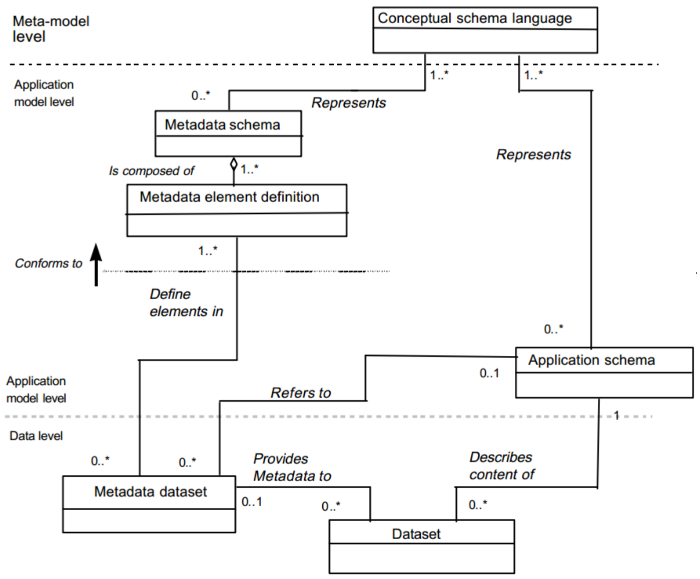


Figure 2-11: ISO/TC 211 metadata relationship (ISO/TC211/WG1: 1998)

Metadata for spatial data is presented in UML[10] packages in the standard. Each package contains one or more entities (UML classes), which can be specified (sub-classed) or generalised (super-classed). These packages include metadata entity set, identification, constraints, data quality, maintenance, spatial representation, reference system, content, portrayal catalogue, distribution, metadata extension, application schema, extent and citation and responsible party information. Figure 2-12 illustrates the layout of the packages.

The ISO19115: 2003 Standard defines an extensive set of metadata elements; typically only a subset of the full number of elements is used. However, it is essential that a basic minimum number of metadata elements be maintained for a dataset (ISO19115: 2003). Thus, this standard includes 22 core elements consisting of Dataset title, Dataset reference date, Dataset responsible party, Geographic location of the dataset (by four coordinates or by geographic identifier), Dataset language, Dataset character set, Dataset topic category, Spatial resolution of the dataset, Abstract describing the dataset, Distribution format, Spatial representation type, Reference system, Lineage, Online resource, Metadata file identifier, Metadata standard name, Metadata standard version, Metadata language, Metadata character set, Metadata point of contact, Metadata date stamp and Additional extent information for the dataset (vertical and temporal).

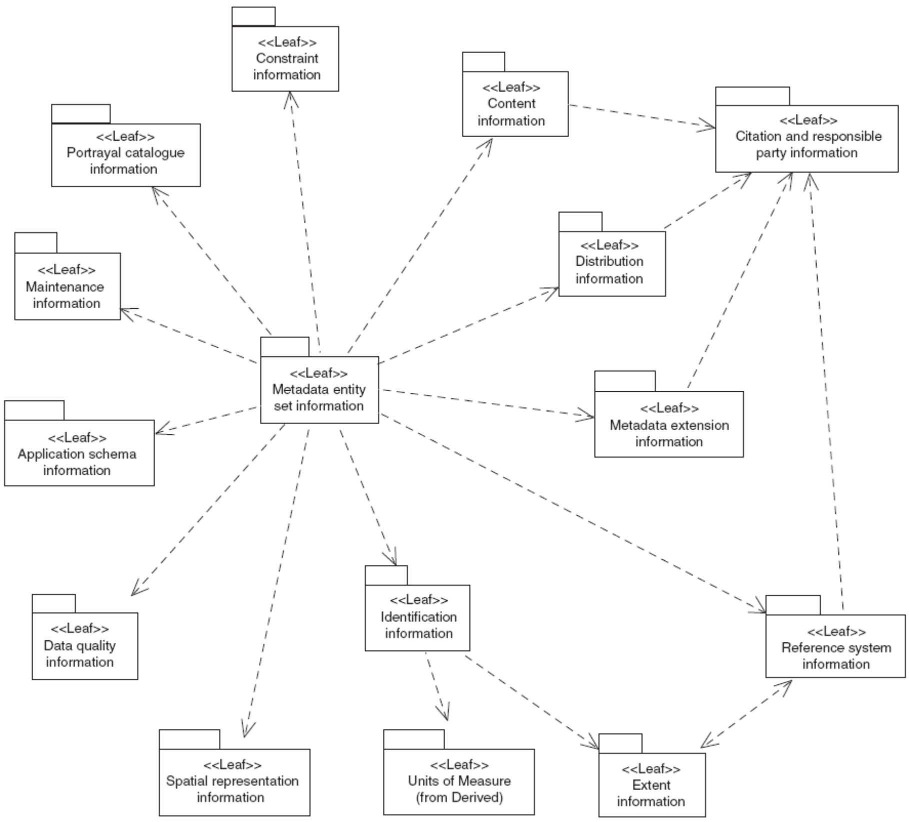


Figure 2-12: ISO 19115: 2003 metadata packages (ISO19115: 2003)

10 Unified Modelling Language

Since the ISO 19115: 2003 does not provide any encoding, the actual implementation of geographic information metadata could vary based on the interpretation of metadata producers. In an attempt to facilitate the standardisation of implementations, another international standard entitled ISO/TS 19139: 2007 as a comprehensive metadata implementation specification was published by the ISO/TC 211 which provides a definitive, rule-based encoding for applying the ISO 19115: 2003 (ISO/TS19139: 2007). This technical specification provides XML schemas that enhance interoperability by providing a common specification for describing, validating and exchanging metadata about spatial datasets, dataset series, individual spatial features, feature attributes, feature types, feature properties, etc.

According to the broad use of ISO 19115: 2003 and ISO/TS 19139: 2007 within the geospatial community, in particular for the automation purpose (Manso *et al.* 2004, Manso-Callejo *et al.* 2009, Taussi 2007, Zarazaga-Soria *et al.* 2003), the metadata automation approaches were designed based upon these standards in this thesis. The profiles of ISO 19115: 2003 discussed in the next section, confirms the popularity of this standard among different countries.

- **Profiles of ISO 19115: 2003**

Metadata extensions and profiles are discussed in Annex C of the ISO 19115: 2003 Standard (ISO19115: 2003). It describes the possibility of creating specific profiles or extension of the standard. This might be interesting for fulfilling specialised purposes for certain information, as well as, user community or nation (Najar 2006).

Figure 2-13 illustrates the relationship between the core metadata components, the comprehensive metadata application profile and national, regional, domain specific or organisational profiles.

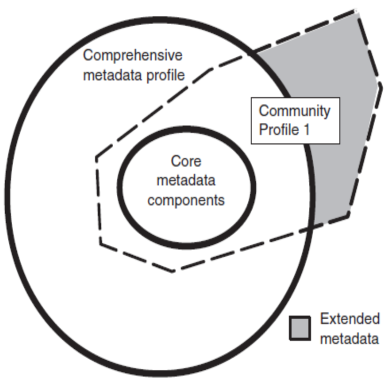


Figure 2-13: Metadata community profile (ISO19115: 2003)

The inner circle contains the core metadata components. The comprehensive metadata includes the core metadata components. A community profile shall contain the core metadata components, but not necessarily all the other metadata components. Additionally it may contain metadata extensions (shaded area) which shall be defined following the metadata extension rules in Annex C of the ISO 19115: 2003 Standard.

Accordingly, following the popularity of the ISO 19115: 2003 Standard in recent years some organisations have decided to adopt profiles of this standard. For instance, the Australia New Zealand Land Information Council (ANZLIC) released an Australian/New Zealand profile of AS/NZS ISO 19115:2005, Geographic information-Metadata (implemented using ISO/TS 19139: 2007, Geographic information-Metadata-XML schema implementation) in August 2007 (ANZLIC 2009).

Also, through efforts seeking collaboration, US and Canadian scientific volunteers from the International Committee for Information Technology Standards Technical Committee L1 (INCITS/L1) and the Canadian General Standards Board Committee on Geomatics (CGSB-COG) developed the INCITS 453-2009, the North American Profile (NAP) of ISO 19115: 2003, Geographic Information – Metadata, to meet the requirements of both countries in July 2009 (FGDC 2009).

The Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP) under review of United Nations (UN) has also released a draft version of Asia-Pacific Geospatial Metadata Profile based on ISO 19115: 2003 core metadata elements and other regional profiles as well as the survey results of participating countries (UN 2009).

In addition to the above profiles, the INSPIRE (Infrastructure for Spatial Information in Europe) initiative has published the implementation rules for metadata based on the ISO 19115: 2003 (INSPIRE 2009). INSPIRE is a proposed European Directive establishing the legal framework for setting up and operating an Infrastructure for Spatial Information in Europe based on infrastructures for spatial information established and operated by the Member States (Craglia and Annoni 2007), which focuses on data interoperability (Rautenbach *et al.* 2012).

As the demand for standardised metadata increases, researchers need to identify metadata production methods that are more efficient and less costly than practices involving humans. Automatic metadata generation can help address this need (Greenberg 2004).

2.9 Chapter Summary

This chapter reviewed the underlying concepts of spatial metadata. The significant role of metadata in the geospatial community and in particular the SDI platforms to facilitate accessing up-to-date and high quality spatial data and services was discussed in detail. The chapter also explored two data models to store spatial data and metadata: detached and integrated data models. The spatial metadata lifecycle consisting of collection, creation, storage, publication, discovery, retrieval and access, and update steps was then reviewed in the chapter. Finally, the chapter highlighted the commonly used worldwide standards including ISO 19115: 2003, FGDC and Dublin Core that make metadata readable, retrievable and interoperable for users all over the world.

The next chapter will review the background of spatial metadata automation research and development to provide necessary information (in addition to the information discussed in this chapter) to address the first research objective, stated in Chapter 1. It will also explore the main technologies involved in designing and implementing the spatial metadata automation approaches in this thesis to address the fourth and fifth research objectives.

CHAPTER 3

SPATIAL METADATA AUTOMATION: BACKGROUND AND TECHNOLOGICAL PERSPECTIVE

3 SPATIAL METADATA AUTOMATION: BACKGROUND AND TECHNOLOGICAL PERSPECTIVE

3.1 Introduction

In order to address the first research objective, along with the information provided in the previous chapter, this chapter reviews metadata automation related works in two main areas: digital library and information science community and the geospatial community. Different approaches and methods that facilitate the metadata collection, creation, updating and improvement are then explored in both communities. According to the metadata automation literature review in the geospatial community, the chapter identifies and summarises the main challenges that this community faces for metadata management and automation. Finally, the technologies focused in this thesis for the spatial metadata automation purpose are reviewed in detail. These technologies will be used in designing and developing prototype systems to address the fourth and fifth objectives of this research.

3.2 Metadata Automation Research and Development Background

According to Batcheller (2008), to explore the background of spatial metadata automation the related work can be reviewed in two main areas: digital library and information science community, and the geospatial community. In this regard, this section first explores the main research activities undertaken in digital library and information science community which have similarities with the studies undertaken in the geospatial community. Next, it reviews the literature of metadata automation research and development in the geospatial community.

3.2.1 Digital Library and Information Science Community

Given the rapid and continual growth of accessible digital resources observed since the advent of the World Wide Web (WWW), it is unsurprising that efforts to facilitate effective information location, navigation and retrieval through resource documentation have followed. A number of research initiatives investigating automated metadata generation have arisen from the digital library and Internet cataloguing arenas, motivated by the view that it is ‘unrealistic to depend on traditional humanly generated metadata approaches' when considering the volumes of resources involved (Greenberg *et al.* 2006).

In this regard, as one of the early and significant research initiatives in this area, Greenberg (2003b) elaborated a framework for metadata generation for online content, noting the part standards play in guiding metadata authoring in addition to the roles of human and computing resources (Figure 3-1). Automated practices therein were categorised into those which employ resource content indexing i.e. were not predicated on the presence of recognised metadata elements, and those employed by commercial search engines, whether using pre-formed metadata or that produced at run-time.

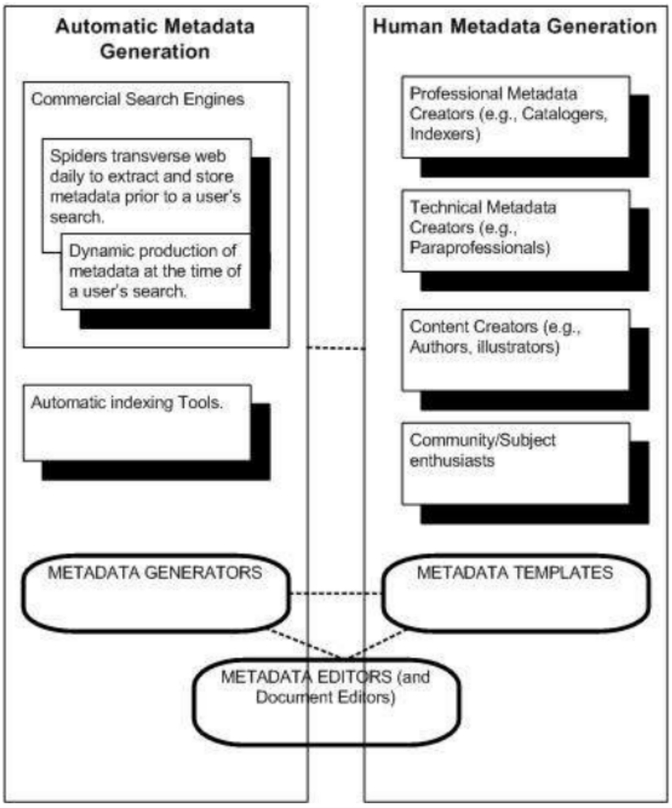


Figure 3-1: Metadata generation framework proposed by Greenberg (2003b)

Within this framework, metadata generators were considered to support automatic metadata production. In the context of the Web, generators first required the submission of a Uniform Resource Locator (URL), a Persistent Uniform Resource Identifier (PURL) or another Web address in order to locate the object. An algorithm was then used to comb an object's content, including its source code, and automatically assign metadata.

In addition, Han *et al.* (2003) discussed several methods that could be used for automatic metadata extraction including regular expressions, rule-based parsers, and machine learning. From their point of view, machine-learning methods were robust and adaptable and, theoretically, could be used on any document set. They classified the machine learning techniques for information extraction into symbolic learning, inductive logic programming, grammar induction, Support Vector Machines, Hidden Markov Models (HMMs), and statistical methods. They also highlighted that Hidden Markov Models were the most widely used generative learning method for representing and extracting information from sequential data, regardless of being based on the assumption that features of the model they represented were not independent from each other. Finally, they focused on using the Support Vector Machine classification-based method for metadata extraction from the header part of research papers (e.g. authors' names, affiliations, addresses, and the title of the paper) and showed that this proposed method outperformed other machine learning methods on the same task.

As part of a study conducted by the Metadata Generation Research (MGR) project at the University of North Carolina, Greenberg (2004) identified and analysed the effectiveness of two different methods for automatic metadata generation applicable to digital resources (based on Dublin Core Standard): metadata *extraction* and *harvesting*. To undertake this study, two Dublin Core automatic metadata generation applications were explored: Klarity, which emphasised the extraction method; and DC.dot, which emphasised the harvesting method. It was discussed that metadata extraction occured when an algorithm automatically extracted metadata from a resource's content displayed via a Web browser similar to what happened in many commercial search engines in response to a search. Also, metadata harvesting was defined as the other key automatic metadata generation method. It occured when metadata was automatically collected from META tags found in the 'header' source code of an HTML resource or encoding from another resource format. It was also found that the harvesting process relied on the metadata produced by humans or by full or semi-automatic processes supported by software. The outcomes of this study indicated that extraction-processing algorithms could contribute to useful automatic metadata generation; and harvesting metadata from META tags created by humans could have a positive impact on automatic metadata generation.

Metadata *extraction* and *harvesting* was also acknowledged by the researchers in the geospatial community as two key methods for automating spatial metadata creation and updating (Manso *et al.* 2004, Batcheller 2008, Batcheller *et al.* 2007, Batcheller *et al.* 2009, Manso-Callejo *et al.* 2010, Borjas *et al.* 2011). The related studies are explored in more detail in the next section. In addition, the spatial metadata management tools use these methods for automating metadata generation and updating (e.g. GeoNetwork opensource harvests metadata from external nodes, and ESRI ArcCatalog product extracts metadata out of Shapefiles). These tools will be discussed in Section 5.3, Chapter 5. Extraction and harvesting methods have also been considered in this thesis for designing the new automated metadata approaches.

Having focused on indexing and extraction methods, Cardinaels *et al.* (2005) developed and presented a framework for automatic metadata generation for learning objects in a learning management system, as illustrated in Figure 3-2. In this framework the learning object metadata could be derived from two different types of sources: the learning object itself and the context in which the learning object was used. Metadata derived from the object itself was obtained by content analysis, such as keyword extraction, language classification and so on. The contexts typically were learning management systems in which the learning objects were deployed. As a result, the framework consisted of two major groups of classes that generated the metadata, namely Object-based indexers and Context-based indexers. The object-based indexers generated metadata based on the learning object, isolated from any other learning object or learning management system. The second class of indexers used a context to generate metadata. The framework was easily extensible for new learning object types and new contexts. To be complete, the framework also had some Extractors that for example extracted the text and properties from a PowerPoint-file, and a MetadataMerger that could combine the results of the different indexers into one set of metadata.

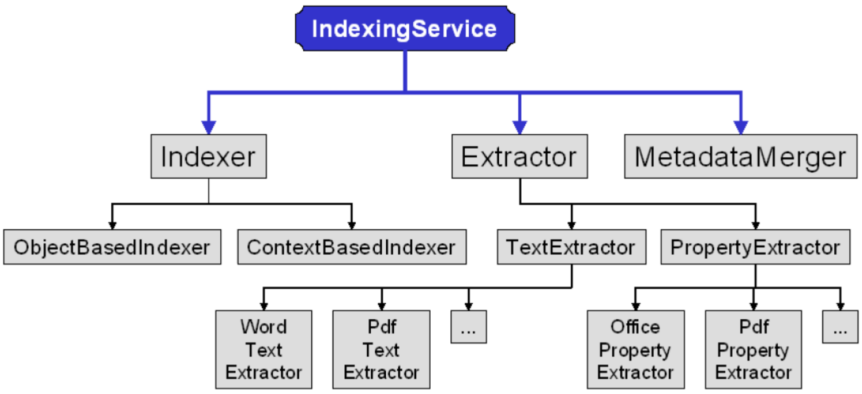


Figure 3-2: Overall structure of the automatic indexing framework proposed by Cardinaels *et al.* (2005)

In another effort in digital library community, Greenberg *et al*. (2005) evaluated the current automatic metadata generation functionalities supported by content creation software and automatic metadata generation applications; and reviewed the automatic metadata generation functionalities supported by Integrated Library Systems (ILSs). As a result of this study, they classified the research in the area of automatic metadata generation into experimental and application areas. The experimental research focused on information retrieval techniques and digital resource content, and applications research focused on the development of content creation software and metadata generation tools used in the operational setting. As the main finding, they raised that there was a disconnection between experimental research and application development and most likely metadata generation applications could be vastly improved by integrating experimental research findings.

Greenberg *et al*. (2006) undertook a survey as part of the Automatic Metadata Generation Applications (AMeGA) project to identify the functionalities that metadata experts desire in automatic metadata generation applications. The goal of the AMeGA project was to identify and recommend functionalities for applications supporting automatic metadata generation for digital resources, and ultimately, to improve the state-of-the-art for metadata applications. Results of this survey indicated that metadata experts were interested in using automatic metadata generation, particularly for metadata that could be created accurately and efficiently. However, participants were generally not in favour of eliminating human evaluation or production for the more intellectually demanding metadata (e.g. subject metadata). The majority of participants agreed that automatic processes should be employed to aid humans creating metadata – including metadata requiring intellectual discretion.

According to the results of the case study investigations undertaken for addressing the second research objective, it was realised that the participating organisations from the case study region – Australia also prefer a semi-automatic spatial metadata creation and updating approach where the human is able to evaluate and approve the automated metadata values. The results of this case study will be discussed in detail in Section 5.2.1, Chapter 5.

With regard to the role of the human in evaluating and improving the automated metadata, Guy *et al.* (2004) proposed a workflow based on an automatic/semi-automated/manual creation of metadata (Figure 3-3). This workflow foresaw the following tasks: a) automated, b) automated improved by the author of data, c) automated improved by an author and by an information specialist, d) created manually by an author and improved by an information specialist or e) created by an information specialist.

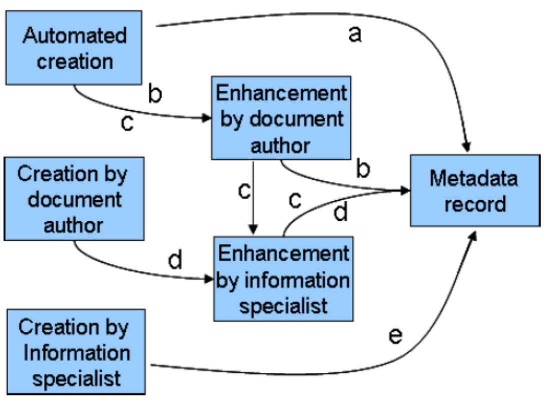


Figure 3-3: Proposed tasks for the generation of metadata according to the human interaction in the creation of metadata (adopted from Guy *et al.* (2004) *and* Manso-Callejo *et al.* (2010))

As a result, it can be realised that enhancement of metadata values by the author once they are automatically generated plays a key role in metadata management. This consideration has been noted in designing and implementing the automatic spatial metadata approaches in this thesis.

In addition to the studies discussed above for metadata automation, Al-Khalifa and Davis (2007b, 2007a) presented another way of automatic metadata creation for the Web resources rooted in folksonomy in which its concept will be discussed in Section 3.3.2. They explored the value of the folksonomy tags as a potential source of keyword metadata by examining the relationship between folksonomies, community produced annotations, and keywords extracted by machines. The results of this experiment showed that the folksonomy tags agreed more closely with the human generated keywords than those automatically generated. Following this study, the current thesis has also built up a new automatic approach for enriching the content of spatial metadata keyword element rooted in folksonomy and tagging to address the fourth research objective. The conceptual design of this approach will be discussed in Section 6.5, Chapter 6.

The next section explores metadata automation background in the geospatial community.

3.2.2 Geospatial Community

The very nature of spatial data dictates of a somewhat different approach for metadata automation than the ones in the digital library and information science community (Batcheller 2008). In this regard, different studies have been undertaken in the geospatial community to automate the generation and updating of metadata describing spatial data resources.

As one of the earliest researches in this area, Beard (1996) proposed five methods for the generation of metadata considering the possible automation methods. The proposed methods included: a) manually (with a keyboard); b) by extending the stored information with values obtained via consultations (i.e. the geographic identifier based on the geographic extension in a gazetteer); c) automated measurements and observations; d) extracted and calculated; and finally e) inferred from other elements.

In a later study, Balfanz (2002) classified the metadata for the automatic generation purpose into three groups. The first group included metadata that was explicitly part of spatial data (e.g. name of data format). The second group contained metadata that was implicitly part of spatial data (e.g. spatial extent). The third group consisted of metadata that was not included in the spatial data in any way (e.g. contact information). It was discussed that because of the third group of metadata a combination of compilation methods were required for automating metadata generation. The metadata compilation methods were categorised into five groups including a) key in, b) look-up, c) measured, d) computed, and e) inferred. Then, a repository-based approach was adopted for automatic spatial metadata generation. The repository contained a set of modules that could use different input sources as extracting information from spatial data, inferring from previously extracted information, using templates or simply demanding user input in the worst case.

Following this study, a combination of metadata generation methods investigated by the researchers have been used in designing a new spatial metadata creation approach in this thesis to address the third research objective, explained in Chapter 1. However, the main focus of this new approach namely 'lifecycle-centric' is on identifying the metadata elements that should be created in any step of spatial data lifecycle. This approach will be discussed in detail in Section 6.3, Chapter 6.

As another research in spatial metadata automation, Zarazaga-Soria *et al.* (2003) presented a tool for creating spatial metadata based on ISO/DIS 19115 Standard developed at the University of Zaragoza. Apart from the basic metadata creation functionality, this tool was enhanced with a set of tools to improve the quality of metadata: a thesaurus management tool and an automatic metadata generation tool. The first tool enabled metadata originators to use look-up tables in order to fill some metadata elements with predefined lists of controlled keywords. The second tool was an automatic metadata generation tool able to derive metadata from the data sources (Shapefiles in this study) by means of interconnection with commercial GIS tools or proprietary software. Examples of derived metadata were information about bounding box; number and type of geometric objects; URL of the Shapefile; and the relational structure of the associated attribute information.

Following this research, Manso *et al.* (2004) investigated the ISO 19115: 2003 metadata elements that might be automatically extracted from its original file storage (for raster, digital terrain model (raster grid), and vector datasets). Tables 3-I and 3-II in **Appendix 2** summarise the result of this study and present different types of geographic information (raster-based and vector-based) along with the ISO 19115: 2003 metadata elements that can be populated automatically through the extraction techniques. This provides information that underpins the design of an automatic spatial metadata updating approach, which will be addressed in the fourth research objective. The automatic spatial metadata updating approach will be discussed in Section 6.4, Chapter 6.

In another effort, Westbrooks (2004a) reviewed the CUGIR metadata management model including the automatic metadata creation and synchronisation function. The Cornell University Geospatial Information Repository (CUGIR) was a Web-based repository providing free access to spatial data and metadata (based on FGDC standard) for New York State, which was created by the Albert R. Mann Library in 1998. In this study, the automatic metadata synchronisation capability provided by the ESRI ArcCatalog as part of the software suite to develop the CUGIR was studied. This tool automatically created metadata (including coordinate system, entity, and attribute information) for datasets stored in the geodatabase if none existed. ArcCatalog automatically updated or synchronised dataset properties with its most current values every time the metadata librarian viewed the metadata. However, regardless of the benefits resulted from the ArcCatalog it was argued that this tool brought forth a host of problems associated with archiving and bibliographic control, due to its inability to differentiate a version of a metadata record from an edition or update.

This identified problem along with the limitation of ArcCatalog synchronisation process to some dataset formats (e.g. Shapefile), and the need for cataloguers to manually run the synchronisation process have been considered in designing a new automatic metadata updating approach in this thesis. The synchronisation process provided by the ArcCatalog will be also discussed in detail in Section 5.3.1, Chapter 5.

Design and development of a prototype system based on ESRI's ArcGIS, Microsoft's .NET, personal geodatabase, and a Qualified Dublin Core profile (including 23 metadata elements as illustrated in Table 3-III, **Appendix 2**) by Batcheller *et al.* (2007) and Batcheller (2008) was another significant progress in the spatial metadata management and automation area. The prototype was designed to integrate a systematic data management model with data initialisation and documentation processes, the aim being to conflate the component workflows whilst facilitating the automatic creation of appropriate metadata.

The operations the prototype performed included (Batcheller 2008):

1. harvesting pre-existing metadata elements generated by ArcCatalog
2. extracting file hierarchy, data and dataset properties and attributes for use as metadata elements
3. harvesting user-prepared metadata templates
4. guiding the visual inspection, modification and completion of metadata records through the structured presentation of record fields on an editing form
5. enabling the importation from and exportation to other standards through the use of a basic metadata crosswalk.

A flow diagram of the above operations is illustrated in Figure 3-4.

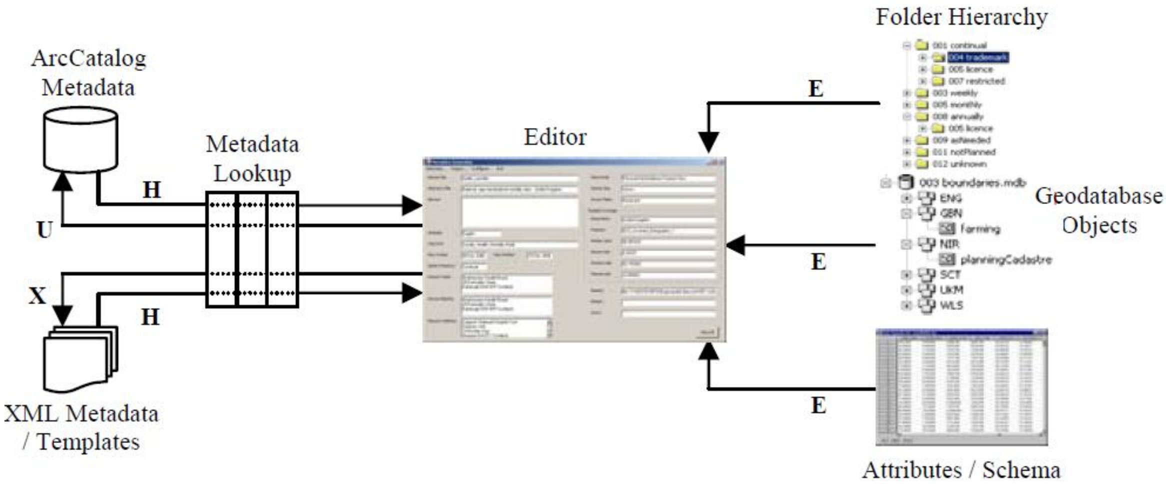


Figure 3-4: Flow diagram of the operations performed by the prototype, generating elements from a number of separate sources for review on the main editing form. H: Harvesting; E: Extraction; X: Export; U: Update proposed by Batcheller *et al*. (2007) and Batcheller (2008).

According to Figure 3-4, two different routines were implemented within this prototype: initial harvesting and extraction. Initial harvesting took advantage of the inherent metadata already collected from registered datasets by ArcCatalog. Harvesting routines were run using XPath expressions, defining from where to retrieve pre-formed metadata from both internal, ArcCatalog-specific XML (stored alongside the dataset in question) as well as, from external user-defined XML templates. In the case of the latter, system variables read from the underlying operating system (such as workstation domain and username) could be used to determine the appropriate templates to query. XPath expressions were encoded in a look-up table interpreted by the tool and which might be readily adapted to a variety of metadata conventions, as well as used as a fundamental crosswalk for metadata output. In addition, custom routines were used to extract further information from the dataset, its data content, as well as the dataset's location within a refined folder hierarchy. Extraction and harvesting routines could either be performed automatically or individually executed once the form has loaded; similarly, elements could be interactively deselected to prevent from being overwritten.

Accordingly, of the total twenty-three metadata standard elements were considered, twenty were completed using the proposed approach (Table 3-1); the compound element 'keyword' comprising of four sub-elements retrieved through the various extraction methods.

Table 3-1: Breakdown of metadata items retrieved and the corresponding routines used. Qualified Dublin Core elements not completed include Abstract, Relation and Source (Batcheller *et al*. 2007).

Routine	Element (abridged)
Harvesting – pre-formed metadata	Title; Language; Date Created; Format; Dataset Type; Projection; Spatial Box Coordinates; Identifier – 11 in total
Harvesting – external templates	Creator; Publisher; Contributor – 3 in total
Element Extraction – hierarchy	Date Period; Access Rights; Spatial Box Name; Keyword (x2); - 3 ½ in total
Element Extraction – Dataset	Alternative Title – 1 in total
Element Extraction – Data	Date Modified; Keyword (x2) 1 ½ in total

The authors further discussed that the proposed approach was not so much the automatic generation of metadata but the transfer of effort from metadata authoring to data preparation and management. Being limited to ArcCatalog as a commercial application, using the Dublin Core standard, which is not widely used in the geospatial community comparing to the ISO 19115: 2003, and finally being limited to personal geodatabase are some of the main factors that might prevent broadly adoption of the proposed approach.

As a result of this study and referring to the research initiatives in the digital library and information science community, metadata extraction and harvesting are the cooperative methods used in both non-spatial and spatial metadata automation efforts.

Morris *et al.* (2007) also proposed a workflow for the generation of spatial metadata based on a combination of manual and automatic methods. This workflow started with definition of a template for the agency or organisation, and after this template was personalised for a given collection of geospatial datasets. Once the template was defined for a set of data, and if metadata existed, the data was processed in order to adapt it to the template, and it was created if the data did not exist. Next, all the lineage information was added and finally, a process of synchronisation of metadata was applied with a tool of commercial extraction of metadata, as shown in Figure 3-5.

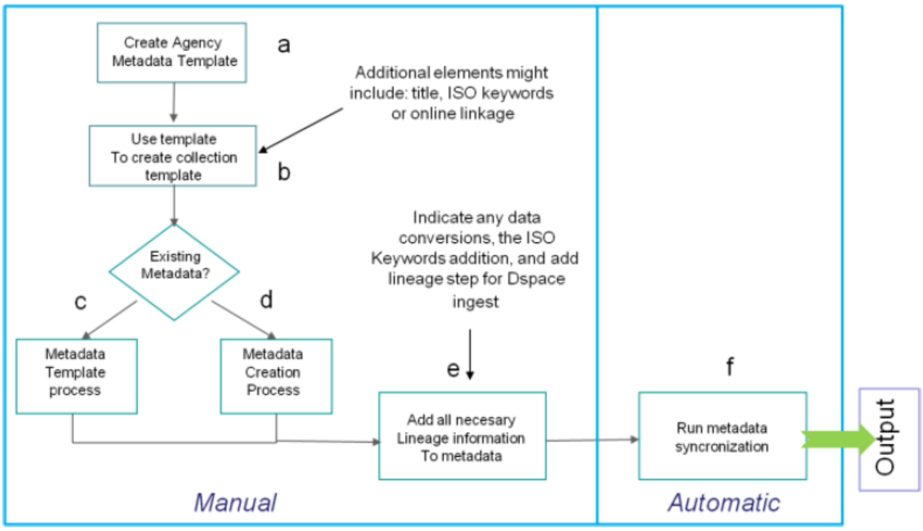


Figure 3-5: A workflow process for the generation of metadata (Adopted from Morris *et al.* (2007) and Manso-Callejo *et al.* (2010)).

Adding lineage information to metadata proposed by this workflow has been adopted by the current thesis. In this regard, an automated function has been designed and implemented as part of the automatic spatial metadata updating approach to update the dataset lineage after any modification to dataset. This function will be discussed in detail in Section 7.2.1, Chapter 7.

In another effort in the geospatial community, Bradford (2007) presented the project of the spatial data catalogue implementation for the US Environmental Protection Agency (EPA)'s Western Ecology Division (WED). Automating the metadata creation process as much as possible was recognised as one of the core requirements of this project. In this regard, two custom EPA tools were designed for ArcCatalog: the EPA Synchroniser (the hook which created a metadata file for a single dataset), and the EPA Metadata Editors (a custom ArcCatalog toolbar tool which simply executed the EPA Synchroniser program on multiple selected datasets) using Microsoft Visual Studio 2005 and ESRI ArcObjects library. This was undertaken through customising the code behind ArcCatalog's 'Create/Update Metadata' button, which included a 'hook' or customisable subroutine that allowed the software developer to insert custom code. This code enabled one to add or change any metadata values that were being written by ArcCatalog.

As a result of this study, the EPA Synchroniser, the primary piece of software developed during this project, allowed ArcCatalog to create largely EPA-compliant metadata (including geographic extent, datum, projection, and on-disk file location) with the push of a button, and only the Title, Abstract, Purpose, and Supplemental Info fields were left blank, as they obviously required human intervention.

This project along with the developments by other researchers (Batcheller 2008, Batcheller *et al.* 2007, Westbrooks 2004a) confirms that ArcCatalog metadata synchronisation capability, which is also expandable, has played a key role in progressing the spatial metadata automation. However, focusing on a commercial tool and its supported data formats by the researchers might result in limiting the wide adoption of their proposed metadata automation approaches only to organisations that work with such a tool. Developing the spatial metadata automation approaches on an open source application, which is easily accessible for the whole community, has been investigated in this thesis when addressing the fifth research objective. The results of the prototype systems evaluation survey discussed in Chapter 8 also confirm the significance of this kind of development.

Following the research and development activities discussed above, Taussi (2007) investigated the standardised and non-standardised (but, useful) spatial metadata elements that could be produced automatically through examining the current standards on spatial metadata and spatial data quality measures. In this research three approaches for extracting metadata out of geographic datasets were investigated to:

- find out the possibility of automatically extracting standardised metadata (ISO 19115: 2003) out of geographic datasets
- find out which standardised quality measures can be extracted automatically from geographic datasets
- find out if some additional knowledge can be extracted by data mining methods based on an overall study of different data mining methods.

The overall process of extracting different metadata elements was presented in this research as shown in Figure 3-6.

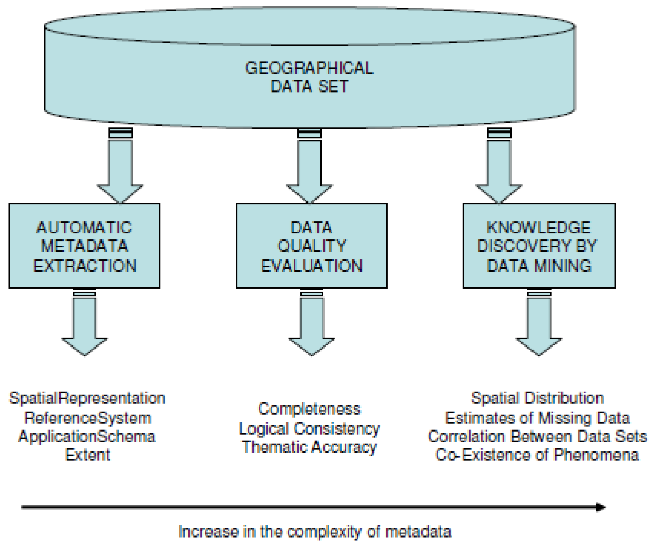


Figure 3-6: Process of extracting different metadata out of geographic datasets proposed by Taussi (2007)

As a result of this research, first the standardised metadata elements for vector and raster data that could be generated automatically by simple automatic analysis were identified. These elements included the information about spatial representation, reference system, application schema and spatial extent of dataset. Whether this information could be generated automatically depended on the overall format of dataset and the application that was used for creating and maintaining the dataset and its features. This research also argued that the ISO 19115: 2003 mandatory core elements are subjective by nature, and therefore not to be generated automatically. All the elements that could be created automatically were considered optional by the standard, except geographic location, which was considered mandatory under certain conditions.

In addition, the possibility of extracting the dataset quality information (including completeness, logical consistency and thematic accuracy) was examined in this research through evaluating different simple methods including 'Brute Force', 'Comparing against Reference', and 'Stochastic'. Moreover, some meta-knowledge information about the datasets that were not included in existing standards and could be revealed by using the spatial mining methods was identified in the research. This information consisted of the type of spatial or spatio-temporal distribution of specific phenomenon in the dataset, estimation of missing data in the dataset, the existence of spatio-temporal correlation and the strength of this correlation, and the co-existence of multiple spatial variables in datasets. This information was strongly related to the semantics of the datasets and thus could be valuable for the users.

The results of this research provide information for designing a lifecycle-centric spatial metadata creation approach in this thesis to address the third research objective.

As an extension to the original approach already proposed by Batcheller (2008), Batcheller *et al*. (2009) developed a prototype system embedded within the ESRI ArcCatalog for the automation of metadata authoring process steps. Figure 3-7 illustrates an overview of the routines employed in the prototype.

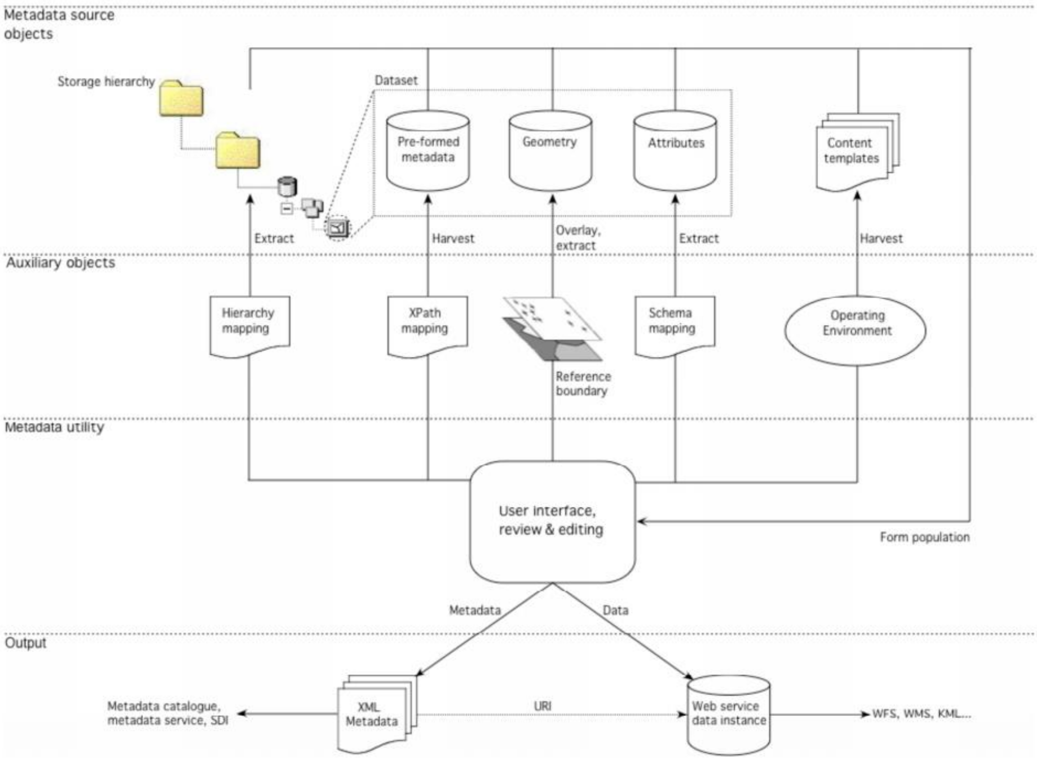


Figure 3-7: An overview of prototype's routines proposed by Batcheller *et al.* (2009)

Batcheller *et al*. (2009) discussed that using the prototype out of the seventeen mandatory elements detailed by the UK GEMINI (GEo-spatial Metadata INteroperability Initiative, a.k.a. GEMINI) convention (outlined in Table 3-IV, **Appendix 2**); fifteen were automatically populated, the exceptions being the Abstract (partial) and Dataset reference date. Twelve further optional elements were generated (two arising from default values hard-coded into the utility) leaving three entries requiring manual completion – these being the elements associated with supplemental dataset description and portrayal: Online resource, Browse graphic and Additional information source.

However, using a detached data model for storing spatial data and metadata, consideration of metadata generation outside the spatial data lifecycle, and dependency of tool on a specific dataset format can be recognised as the main issues of the proposed approach. The metadata generated in XML is only linked to its related spatial data using a Unified Resource Identifier (URI). This model of metadata storage brings forth a set of problems already discussed in Section 2.6.1, Chapter 2. Also, only the dataset itself as the final output of spatial data lifecycle is employed in metadata creation process. Therefore, some of the required elements need to be authored manually after dataset is fully created such as abstract, and reference date. Accessing this information may not be easy and needs extra effort by the metadata author. Moreover, the proposed approach is limited to ArcGIS data formats for creating metadata.

The significance of spatial data and metadata integration was also emphasised by Rajabifard *et al.* (2009). In this research, a set of criteria was developed to evaluate a selection of Metadata Entry Tools (METs) such as the CatMDEdit, and GeoNetwork. This set of criteria was categorised into four main groups including technical requirement, compliance with International Standards, user-friendly interface and finally availability of necessary functions for handling metadata records. The capability to automatically update the metadata when any change occurred to the spatial data was considered as one of the necessary function for handling metadata. As a result of this study, none of the assessed tools could support the automatic metadata updating capability. These tools along with some other widely used metadata management tools will be reviewed in Section 5.3 in Chapter 5 to address the second research objective.

Another study by Manso-Callejo *et al.* (2009) investigated the ISO 19115: 2003 metadata elements that supported dynamic interoperability of the SDI platform and could be automatically produced by extraction, computation and inference. It was discussed that dynamic interoperability would be achieved when systems were able to detect state changes and to take advantage of those changes (Turnitsa and Tolk 2006) or when systems were able to locate resources for their use based on the existence of standardised metadata (Shanzhen *et al.* 1999). As a result of this study, of the 151 metadata elements providing dynamic interoperability, 54 elements might be automatically produced (35%). Depending on the type of spatial data storage, its nature and the possibility of determining the content type, these values might also diminish or increase markedly. Table 3-V in **Appendix 2** illustrates the output of this research including the metadata elements that might be generated automatically, dataset type and the explanation of the metadata elements. This provides information for identifying the metadata elements affected by dataset modification that is used in designing the automatic spatial metadata updating approach in this thesis to address the fourth research objective.

A framework was also presented by Kalantari *et al.* (2010) for spatial metadata automation including three complementary approaches: automatic metadata creation, updating and enrichment. In their point of view, when there is no metadata for spatial dataset there is a need for methods to extract metadata from a dataset automatically. On the other hand, when metadata exists for a dataset there is a need for methods that can automatically update metadata at the same time with any change to the dataset. The authors highlighted the fact that end users are disconnected from the spatial metadata creation and improvement process and an automatic method is required to involve the users to create and improve the content of spatial metadata.

Manso-Callejo *et al.* (2010) also proposed a new workflow for the automated metadata generation for spatial datasets. This workflow consisted of a set of tasks that were grouped into four main functions aimed at the automated generation of metadata:

- Discovery: enabling users to find/locate spatial datasets
- Use: enabling users to explore spatial datasets
- Evaluate: enabling users to explore whether a spatial dataset suits their needs
- Retrieval: enabling users to know how spatial datasets can be obtained.

Figure 3-8 illustrates the synchronisation of the proposed tasks in this approach.

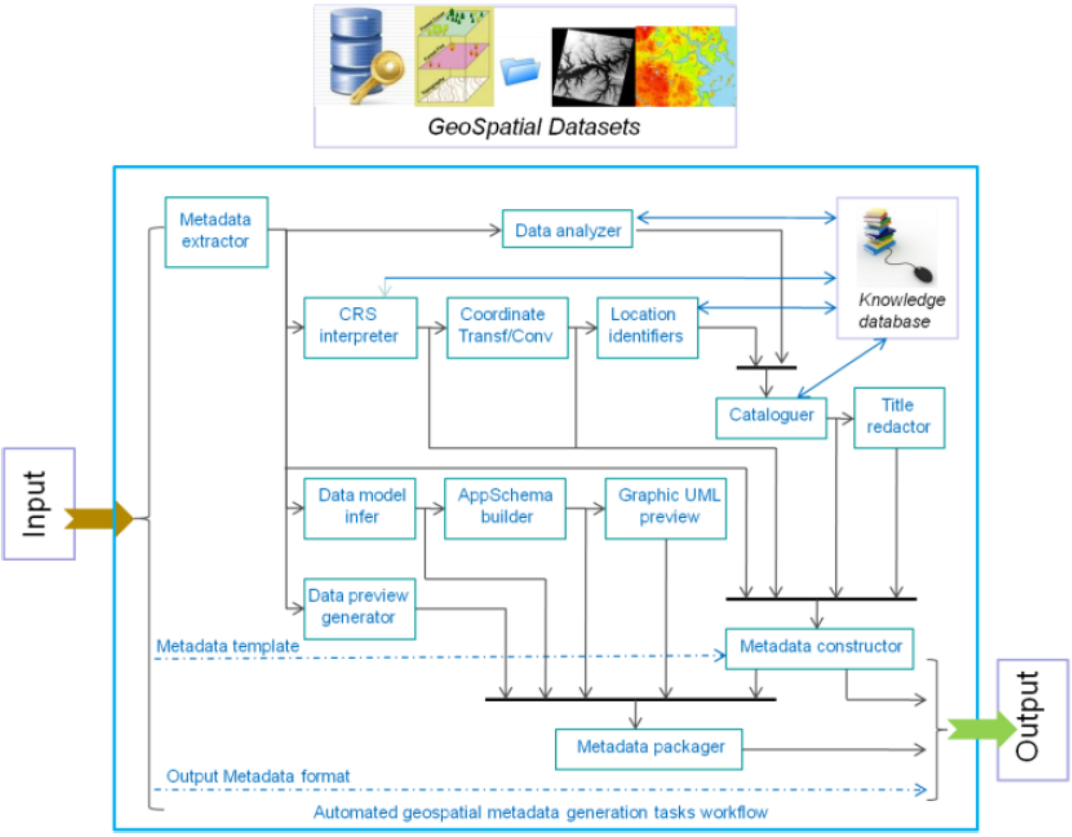


Figure 3-8: Overview of the task synchronisation of the workflow proposed by Manso-Callejo *et al.* (2010)

Although this workflow aimed at automating the spatial metadata generation, the cataloguer's knowledge of the spatial data lifecycle was still required for creating some of the metadata values. Employing a knowledge database in the workflow confirmed this need. Consideration of a separate process for metadata creation from the spatial data creation lifecycle has resulted in requiring the cataloguer's intervention to make the workflow valid.

As a result of the study, Table 3-VI in **Appendix 2** illustrates the metadata elements that could be generated automatically using this workflow. The first column namely 'MD_Metadata: Packed' shows the class where the metadata belongs (e.g. MD_Metadata); while the second one, namely 'Metadata element' identifies the element. The third column classifies the items as 'G', which stands for 'generated' (extracted, calculated or inferred); 'C' with cardinality dependent on the dataset; and finally 'I', when a fixed value can be set by agreement. The fourth column identifies those elements that could only be applied to certain types of geospatial datasets ('R' raster data, 'D' DTM[11] and 'V' vector data). The fifth column of the table identifies the function performed by the metadata element ('D' discovery, 'E' evaluation, 'R' retrieve, and 'U' use).

This study concluded that the proposed workflow for the automated generation of metadata could compile 83 metadata elements for images, 69 for vector data and 68 for DTM. It also emphasised that the final number of elements would depend on the used format and the information it contained. The number of elements could be increased if the digital repository contained more than one band or a layer with 21 more elements identified for the three classifications (R, V and D), with cardinality higher than or equal to one.

In another endeavour for automating the spatial metadata, Borjas *et al.* (2011) proposed an architecture to carry out an automatic characterisation of Geoportal Web resources (Figure 3-9). The system consumed a document describing a Web resource, detected automatically the model of the input metadata, discovered type and format of the described resource, extracted necessary information (metadata structures), and generated enhanced metadata according to the input model. The architecture allowed the application of different logic for extracting metadata structures, and also to use different logic to improve the existing description (from simple mapping to the analysis of the related resources). Based on this architecture, a prototype was dedicated for generation OGC CSW records that describe Geoportals Web resources.

[11] Digital Terrain Model

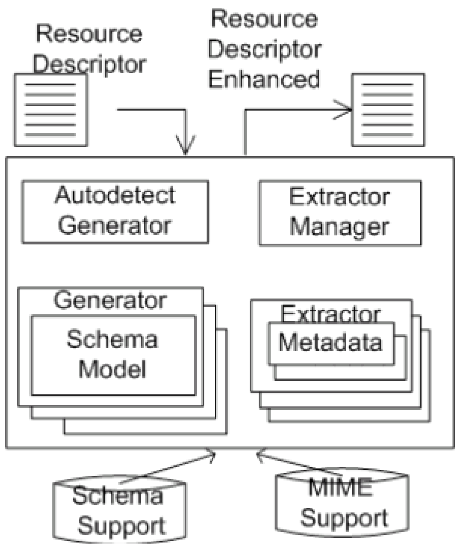


Figure 3-9: The architecture for automatic metadata generator for Geportal Web resources proposed by Borjas *et al.* (2011)

A set of resources consisted of Web resources of the geoportals forming the Spanish SDI regional level were tested in this study. Following the first experiment, the main problem was the extraction of geographic extent information (coverage) because the pages did not contain the geographic metadata. For this reason, the second experiment was devoted to this problem, and the system was extended with natural language processing analysis algorithms to extract place names. The place names were translated into the geographical extent that they represented. It was concluded that the coverage element was completed in most cases and human review confirmed that 67% corresponded to the reality.

Similarly, Florczyk *et al.* (2012) proposed architecture of a system dedicated to the automatic creation of geospatial metadata of Web resources. They also proposed a heuristic-based method for geographic scope estimation of Web pages. Based on this study, a prototype capable of generating a geospatial metadata (in Dublin Core (DC) profile) of a HTML Web page was developed and tested on Web pages that belonged to the Websites that publish Geospatial Web resources. The experiment showed that the proposed metadata automation approach might also be useful to improve existing SDI resource metadata; for example, the metadata offered by an OGC Web Service getCapabilities response.

Having explored the background of spatial metadata automation research and development, it could be purported that the geospatial community has been mainly investigating the methods to automate the generation and updating of spatial metadata within the last two decades. The research also showed that metadata extraction from datasets, harvesting, inference, and computation have been the key methods used in most studies. However, the research confirmed that the geospatial community still faces a number of challenges, as summarised here:

- Metadata is commonly created in a separate process from the spatial data lifecycle that results in highly dependency of metadata creation approaches on metadata author's knowledge of the dataset.
- A detached data model is typically used for storing and maintaining the spatial metadata and data that prevents the support for the real-time dataset and metadata updating.
- The automatic metadata generation and updating approaches are dependent on specific data formats that limit the wide adoption of these approaches.
- Spatial data end users are disconnected from the metadata creation and improvement process.

The next section discusses the technologies involved in design and implementation of spatial metadata automation approaches proposed in this thesis to address the fourth and fifth research objectives, which are technical in nature.

3.3 Spatial Metadata Automation Technological Perspective

As Goodchild (2007) has discussed, much progress has been made in the past two decades, and increasingly since the popularising of the Internet and the advent of the World Wide Web (WWW), in exploiting new technologies in support of the dissemination of spatial data. Data warehouses, spatial data libraries, and geoportals have proliferated, and today's users of spatial data have a wealth of potential sources that can be searched for suitable datasets.

WWW, as the main platform of the new technologies, offers the potential benefits of flexibility, ubiquity, and reduced costs and risks of obsolescence and isolation which result in helping the organisations to: (1) maximise productivity and efficiency; (2) protect critical information infrastructure; and (3) overcome problems related to data sharing, security and data maintenance, as well as software special requirements and steep learning curves (Anderson and Moreno-Sanchez 2003).

As a result of recent years development of Web technologies, spatial data can now be easily shared, discovered and downloaded over the Web (Devillers *et al.* 2002) – e.g. the Shared Land Information Platform enabling framework, SLIP Enabler, is the core technological infrastructure that allows users to access Western Australia's significant land and geographic information resources over the Web and has been in operation since 2007 (Landgate 2012).

By increasing distribution, sharing and access of spatial data over the Web, the demand for identifying and employing the state-of-the-art technologies, which can facilitate the automation of metadata creation, updating and improvement in the networked environment, is growing. Accordingly, in order to identify such related technologies – particularly to address the fourth and fifth research objectives stated in Chapter 1, which are technical in nature – the technologies involved in the following three main areas have been explored in this thesis:

- Web generations (Semantic Web and Web 2.0)
- Software applications
- Open standards.

Web generation has been moving from Web 1.0 towards Social (Web 2.0), Semantic (Web 3.0), and Intelligent Web (Web 4.0) progressively, according to Saraf (2008). As illustrated in Figure 3-10, during each generation different technologies are involved in information connections as well as, people connections. For instance, during Web 2.0 generation information are connected through Java, JavaScript, Simple Object Access Protocol (SOAP), HTML, Hypertext Transfer Protocol (HTTP), Resource Description Framework (RDF) and XML technologies and people are connected via Websites, Directory portals, social networking, Wikis, social media networking, etc.

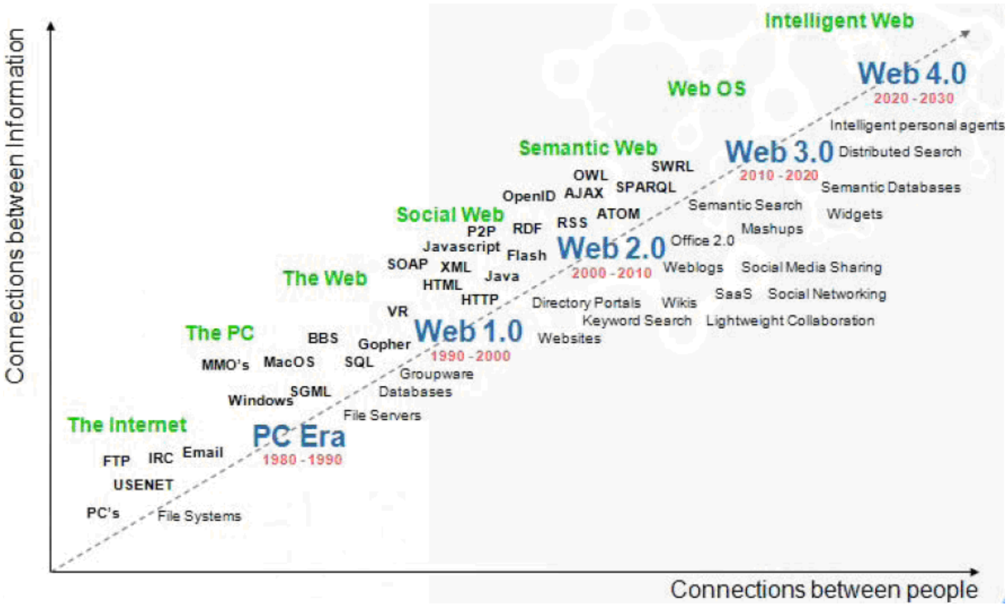


Figure 3-10: Generations of the Web (Saraf 2008)

Among the Web generations, the Semantic Web and Web 2.0 are often presented as competing visions for the future of the Web. However, there is growing realisation that the two ideas complement each other and that in fact both communities need elements from the other's technologies to overcome their own limitations (Ankolekar *et al.* 2008). For instance, the study undertaken by Xu *et al.* (2006) confirms that the integration of technologies involved in both Semantic Web (e.g. ontology) and Web 2.0 (e.g. tagging and folksonomy) would improve the information retrieval over the Web. As a result, both Web 2.0 and Semantic Web were selected for exploration in more detail for the purpose of spatial metadata automation in this thesis.

For the second group, software applications, the two main Open Source and proprietary software were explored for the purpose of spatial metadata automation. The study shows that the debate over the benefits of Open Source versus proprietary software continues among software users and practitioners (Boulanger 2005, Bonaccorsi and Rossi 2003). However, Anderson and Moreno-Sanchez (2003) discuss that when organisations try to use the Web as a platform to deliver geographic data and provide geo-processing functionality to their end users, they commonly find that commercial Web-GIS software raises the following issues: (1) it does not currently offer out of the box geo-processing functionality to perform many of the analyses demanded by their end users; (2) it is expensive; (3) it has a steep learning curve; (4) it requires that some of their IT personnel become specialists in the software operation and maintenance; and (5) it is difficult to integrate with existing IT infrastructure (personnel skills, software and applications). This discussion is followed by the fact that Open Source Software (OSS) offers the potential to overcome the above mentioned issues and facilitate the deployment of geographic data and geo-processing functionality on the WWW.

Therefore, this research has mainly focused on designing and developing the spatial metadata automation prototype systems based on OSS, which will be discussed in more detail in this section. However, one of the prototype systems has also been developed within a real-world system rooted in the proprietary software.

According to the widely adoption of open standards in implementing the technical components of worldwide SDI initiatives such as INSPIRE (Bray and Ramage 2012), the third group of technologies explored in this thesis included the open standards. Pountain (2003) defines open standard as ‘a standard that is independent of any single institution or manufacturer, and to which users may propose amendments.’ Open standards can help assure interoperability. Interoperability has many meanings, including the facets of communication, exchange, cooperation, and sharing of information between systems. In fact, the essence of interoperability is the relationships between systems, where each relationship is a form of communication, exchange, cooperation and sharing (Carney *et al.* 2005).

The following sections review the main groups of technologies mentioned above in more detail.

3.3.1 Web 3.0 (Semantic Web)

W3C (http://www.w3.org) defines the third generation of Web, Semantic Web, as a Web of Data – of dates and titles and part numbers and chemical properties and any other data one might conceive of. The collection of Semantic Web technologies (Resource Description Framework (RDF), Web Ontology Language (OWL), Simple Knowledge Organisation System (SKOS), etc.) provides an environment where application can query that data, drawing inferences using vocabularies, etc.

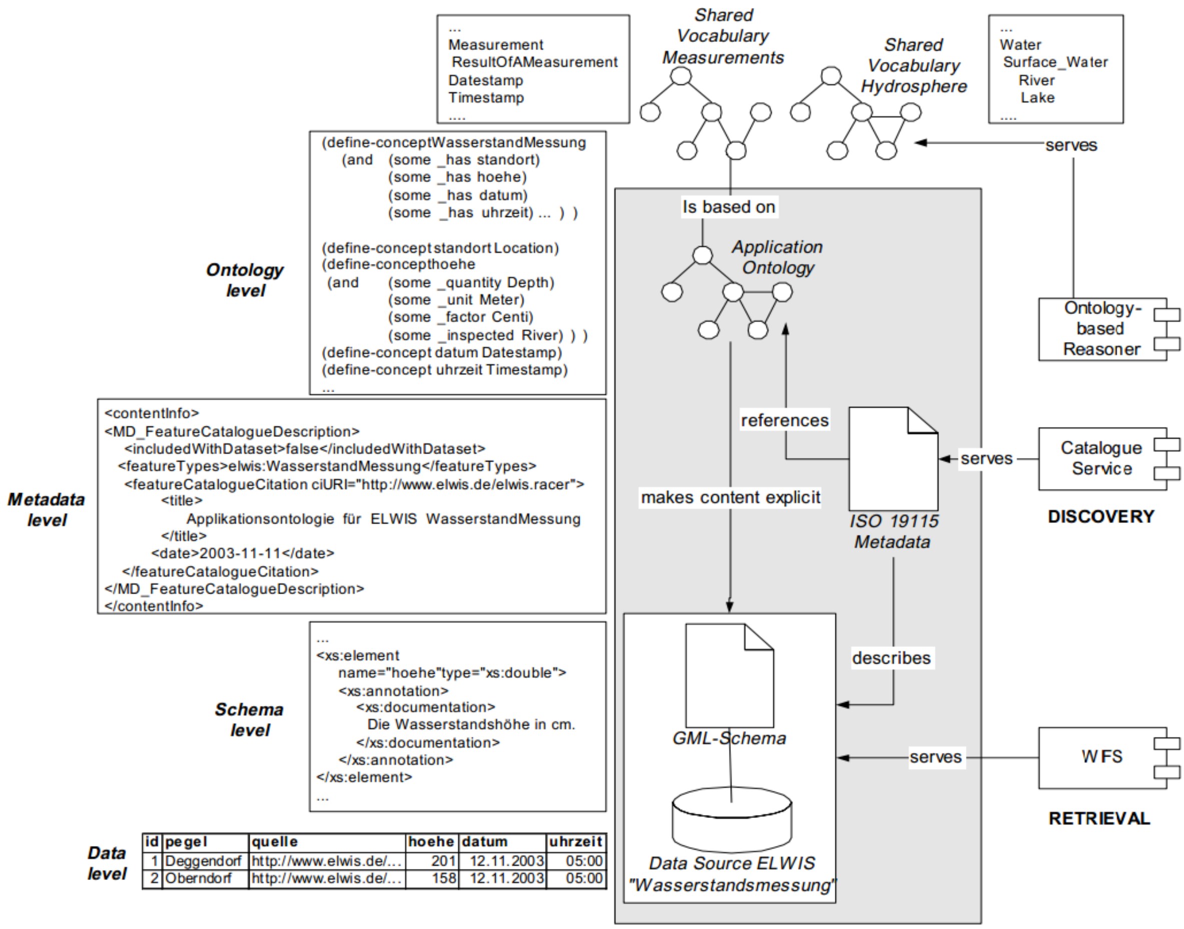
Berners-Lee *et al.* (2001) compare the XML and RDF as two important technologies for developing the Semantic Web. They discuss that XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean; however, meaning is expressed by RDF, which encodes it in sets of triples – each triplet being rather like the subject, verb and object of an elementary sentence. Steinacker *et al.* (2001) also discuss that RDF offers a way to publish human-readable and machine-processable vocabularies designed to encourage the re-use and extension of metadata semantics among disparate information communities.

According to Maedche and Staab (2001), Semantic Web relies heavily on the formal ontologies that structure underlying data for the purpose of comprehensive and transportable machine understanding. They introduce the ontologies as (meta)data schemas, providing a controlled vocabulary of concepts, each with an explicitly defined and machine-processable semantics. By defining shared and common domain theories, ontologies help both people and machines to communicate concisely, supporting the exchange of semantics and not only syntax.

In comparison with the earlier Web generations, the advent of Semantic Web promises better retrieval methods by incorporating the data’s semantics and exploiting the semantics during the search process (Egenhofer 2002). In this regard, Lacasta *et al.* (2007) discuss the potential benefits that a Web Ontology Service (WOS), compliant with the OGC Web Services Architecture specification, may provide the discovery components of an SDI. Based on this study, the first benefit of using WOS service would be to facilitate the creation of metadata content, since it provides access to terminological ontologies (concepts, properties, definitions and relations between concepts and other ontologies) recommended by metadata standards. Also, the second benefit would be the easy integration of a WOS service within an information retrieval system that facilitates the construction of user queries and improves the recall of such systems.

Lutz (2005) also explores the typical data discovery techniques in SDI platforms and discusses that the available keyword-based techniques are inherently restricted by the ambiguities of natural language; even though natural language processing techniques can increase the semantic relevance of search results with respect to the search request (Richardson and Smeaton 1995). Bernstein and Klein (2002) also discuss that keyword-based search can have low recall if different terminology is used and/or low precision if terms are homonymous or because of limited possibilities to express complex queries.

Klien *et al.* (2004) present an approach and architecture (Figure 3-11) for ontology-based discovery and retrieval of spatial information in an SDI platform that can contribute to solving existing problems of semantic heterogeneity, which has not been tackled by the OGC-compliant catalogues. The architecture presented in this study offers two functionalities to enhance the usability of existing spatial information: using defined concept queries to overcome semantic heterogeneity problems during information discovery, and providing interpretation support for feature types and properties during information retrieval.



id	pegel	quelle	hoehe	datum	uhrzeit
1	Deggendorf	http://www.elwis.de/...	201	12.11.2003	05:00
2	Oberndorf	http://www.elwis.de/...	158	12.11.2003	05:00

Figure 3-11: Architecture for geographic information ontology-based discovery and retrieval proposed by Klien *et al.* (2004)

Iwaniak *et al.* (2011) emphasise on the power of Web 3.0 tools to provide the ontologies that create the register for logical models of objects and their features, thesauri that group concepts according to semantic links between them, and finally graphs of logical connections according to the scheme of subject → predicate → object (RDF). Kubik and Iwaniak (2010) also discuss that metadata repositories and intelligent catalogues and registries built upon ontology tools and semantic Web technologies (e.g. RDF, OWL and SPARQL[12] (query language for RDF)) could facilitate and improve the spatial data discovery and retrieval process.

Having reviewed the role that Semantic Web technologies played in SDI platforms, it can be deduced that the geospatial community needs to employ such technologies when implementing the spatial data catalogue systems for the purpose of data discovery and retrieval improvement. However, following the scope of the current research already discussed in Chapter 1, facilitating the information retrieval has not been seen as the main focus of this thesis. Instead, designing and developing an automated approach to improve the content of spatial metadata through more interaction with the end users has been one of the areas of interest for this thesis, which indirectly improves the data discovery and retrieval as well. Therefore, Web 2.0 technology was mainly considered and used due to its power in building interaction with the end users of Web applications. The next section explores Web 2.0 features.

3.3.2 Web 2.0 (Social Web)

According to Goodchild (2007), Web 1.0 was primarily one-directional, allowing a large number of users to view the contents of a comparatively small number of sites; however, the Web 2.0 is a bi-directional collaboration in which users are able to interact with and provide information to central sites, and to see that information collated and made available to others.

Alexander (2006) also discusses that Web 2.0 Services respond more deeply to users than Web 1.0 Services. A leading form of this is a new form of metadata, the folksonomy. Whereas traditional metadata is usually hierarchical (topics nested within topics), structured (e.g. the fields within Dublin Core), and predetermined by content authorities, folksonomic metadata consists of words that users generate and attach to content, which are well known as tags.

Many examples already exist of Web 2.0 Services designed to acquire, assemble, and publish geographic information. Wikimapia, introduced in Section 2.7.2 in Chapter 2, is an example of Web 2.0 which provides for extensive loosely structured metadata and its users are able to access the complete history of any entry, including all previous versions and edits.

[12] SPARQL Protocol and RDF Query Language

Tagging and folksonomy as the main features of Web 2.0, which have been used in this thesis to address the fourth research objective relating to an automatic spatial metadata enrichment approach, are detailed below.

- **Tagging and Folksonomy**

Tags are keywords added to the resources on the Web and the process of creating tags is known as tagging (Murugesan 2007). While one of the main ideas of Web 2.0 is to let users play an important role in the process of creating content, tagging goes a step further by letting them control the way they organise it (Passant and Laublet 2008). By adding simple keywords, or tags, to their data or to data they browse online, they can decide themselves which metadata must be related to any content. Voss (2007) argues that tagging can be referred to with several names: collaborative tagging, social classification, social indexing, folksonomy etc. The basic principle is that end users do subject indexing instead of experts only, and the assigned tags are being shown immediately on the Web. The number of Websites that support tagging has rapidly increased since 2004. Websites like Delicious.com allow users to save the URL for a Website; provide an annotation if desired; supply a number of tags that will help to retrieve it again; and group similar URLs together. These tags can be whatever or how many associations the user wishes to make with the URL. In this way, the URL can be associated with many concepts at once.

The ability to use tags to bring out different aspects of a resource is a major advantage of tagging over formal systems of classification and taxonomies – controlled vocabularies (Thomas *et al.* 2009), which can be adopted for spatial metadata automation (Kalantari *et al.* 2010). Users do not have to be experts and learn complex rules and specialised vocabulary in order to tag as they do when applying a controlled vocabulary (Mathes 2004). Sinclair and Cardew-Hall (2008) also argue that using an uncontrolled vocabulary generated through tagging service is less costly than employing an expert to perform a domain analysis to determine an appropriate classification scheme. Allowing authors to classify their own work also avoids the ongoing cost of employing an expert to classify resources added to the repository.

According to Shrinky (2005), the free associations made by taggers are the only appropriate way to organise resources on systems as large and chaotic as the Web for three reasons: (1) classification fails to allow more than one place for an item; (2) it is impossible to keep a classification system stable over time; and (3) it is also impossible for an expert to truly predict how a user will search for something. This cannot be entirely relevant to SDIs but to an extent in spatial arena with an increasing number of spatial data layers, users and applications (Kalantari *et al.* 2010). First, it is difficult to stabilise a classification of information, secondly (and more importantly), it is impossible to predict how increasing numbers of interested users and applications might name or interpret a spatial dataset.

The tagging approach provides more freedom for users, because when tagging, the user does not have to make a decision and restrict the resource to just one or two formal terms from a controlled keywords they may or may not be familiar with. Instead the users supply their own terms which are meaningful to them (Shirky 2005). For instance, spatial data users can select different tags to describe the same item. Items related to scale may be tagged '500', '1:500', '1/500'. This flexibility allows users to classify their collections of items in a way that they find useful and users also have to decide whether each tagged item is actually relevant to what they are looking for.

The tags are usually illustrated in a weighted list, known popularly as a 'tag cloud' (Sinclair and Cardew-Hall 2008), which has appeared on many popular folksonomy-based Websites such as Websites for photo sharing (e.g. flickr), bookmark sharing (e.g. del.icio.us), and blog searching (e.g. Technorati). Bateman *et al.* (2008) discuss that the tag clouds represent variables of interest (such as popularity) in the visual appearance of the keywords themselves – using text properties such as font size, weight, or colour. For instance, amazon.com tag cloud represents more frequently used tags as larger and more recently used tags as darker (Figure 3-12).

Figure 3-12: Example tag cloud from amazon.com/tags

The emergent dataset arising from all the participants' tags and resources is commonly referred to as a 'folksonomy' (Sinclair and Cardew-Hall 2008). The folksonomy concept was first introduced by Thomas Vander Wal in 2004. He stated: 'Folksonomy is the result of personal free tagging of information and objects for one's own retrieval. The tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information. The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object. People are not so much categorising, as providing a means to connect items (placing hooks) to provide their meaning in their own understanding.'

Vander Wal adds, 'Folksonomy is tagging that works. This is still a strong belief the three tenets of a folksonomy: 1) tag; 2) object being tagged; and 3) identity, are core to disambiguation of tag terms and provide for a rich understanding of the object being tagged.'

Having explored the benefits of Web 2.0 tagging and folksonomy features in non-geospatial community, it could be shown that expanding these features to the geospatial community could result in improving the content of spatial metadata through monitoring and recording the tags which are assigned to datasets by end users. This kind of spatial metadata can help to describe a dataset and allows it to be found again simply using the spatial data catalogue. Tags for spatial data can be identified automatically by monitoring the most popular search words used for finding datasets or assigned by spatial data responsible parties or end users directly. On a spatial data directory if many users are allowed to tag many spatial datasets, this collection of tags can become a spatial folksonomy: a method that can collaboratively create and manage metadata to annotate and categorise spatial data. According to the potential benefits of tagging and folksonomy in improving the content of spatial metadata through the interaction with end users, these features have been used in this thesis to design an automatic spatial metadata enrichment approach for addressing the fourth research objective.

The next section explores the Open Source Software employed in this thesis.

3.3.3 Open Source Software (OSS)

Stallman, as reported by Bonaccorsi and Rossi (2003), proposed a revolutionary idea in 1984 with the Free Software Foundation, subsequently confirmed in 1997 in the Open Source Definition[13]. The key concept is that there should be unrestricted access to computer programming codes: anyone should be able to use them, modify them and circulate such modifications without having to pay anything.

Fuggetta (2003) introduces three factors motivating the increasing interest in OSS: (1) the success of products such as Linux and Apache, which are gaining increasing shares in their own markets (operating systems and http servers); (2) the uneasiness about the Microsoft monopoly in the software industry; and (3) the increasingly strong opinion that 'classical' approaches to software development are failing to provide a satisfactory answer to the increasing demand for effective and reliable software applications.

Over the past few years several open source projects have been established in the geospatial community (Neteler and Mitasova 2002). The examples of these projects can be found at the

[13] Open Source Definition, http://www.opensource.org

'Open Source Geospatial Foundation (OSGeo)' Website[14]. OSGeo is a not-for-profit organisation whose mission is to support the collaborative development of open source geospatial software, and promote its widespread use.

The current OSGeo projects are categorised into five groups, as following:

- Web Mapping (including deegree, geomajas, GeoMoose, GeoServer, Mapbender, MapBuilder, MapFish, MapGuide Open Source, MapServer, OpenLayers, and Zoo)
- Desktop Applications (including GRASS GIS, Quantum GIS, gvSIG, and Opticks)
- Geospatial Libraries (including FDO, GDAL/OGR, GEOS, GeoTools, MetaCRS, OSSIM, PostGIS, and rasdaman)
- Metadata Catalog (including GeoNetwork)
- Other Projects (including Public Geospatial Data and Education and Curriculum).

In order to address the fifth research objective regarding the implementation of metadata automation prototype systems, GeoNetwork, deegree, GeoServer, OpenLayers, and PostGIS were used from the OSGeo projects. In addition, a number of other geospatial open source projects consisting of OpenGeo, PostgreSQL, and GeoExt were employed in this thesis.

The next section reviews the open standards in the context of this thesis.

3.3.4 Open Standards

It is important to distinguish between open standards and open source in order to be aware of differences in the meanings of these terms, as they are sometimes used interchangeably (Hall and Leahy 2008). An open standard (i.e. a standard that is publicly available to use) can be implemented by open source (i.e. the principles and methodologies to promote open access to design and production) software, as well as, commercial or proprietary solutions, in much the same manner. Similarly, Krechmer (2006) discusses that the purpose of open source is to support continuous software improvement, while the purpose of open standards is to support common agreements that enable an interchange available to all. Krechmer (2006) also summarises the minimal characteristics that a specification and its attendant documents must have in order to be considered as an open standard:

[14] http://www.osgeo.org

- The standard is adopted and will be maintained by a not-for-profit organisation, and its ongoing development occurs on the basis of an open decision-making procedure available to all interested parties (consensus or majority decision, etc.).

- The standard has been published and the standard specification document is available either freely or at a nominal charge. It must be permissible to all to copy, distribute and use it for no fee or at a nominal fee.

- The intellectual property-i.e. patents possibly present-of (parts of) the standard is made irrevocably available on a royalty-free basis.

- There are no constraints on the reuse of the standard.

Open standard is also considered as one of the two main categories of standards classified by Yeung and Hall (2007). As illustrated in Figure 3-13, the other category includes the proprietary standard. The authors discuss that the concept of open standards often means different things to different people. Some people perceive an open standard as one that is documented, generally available for use, and free of charge. Others put less restriction on the definition of 'open' by including those standards that are generally available but require a licence fee to use.

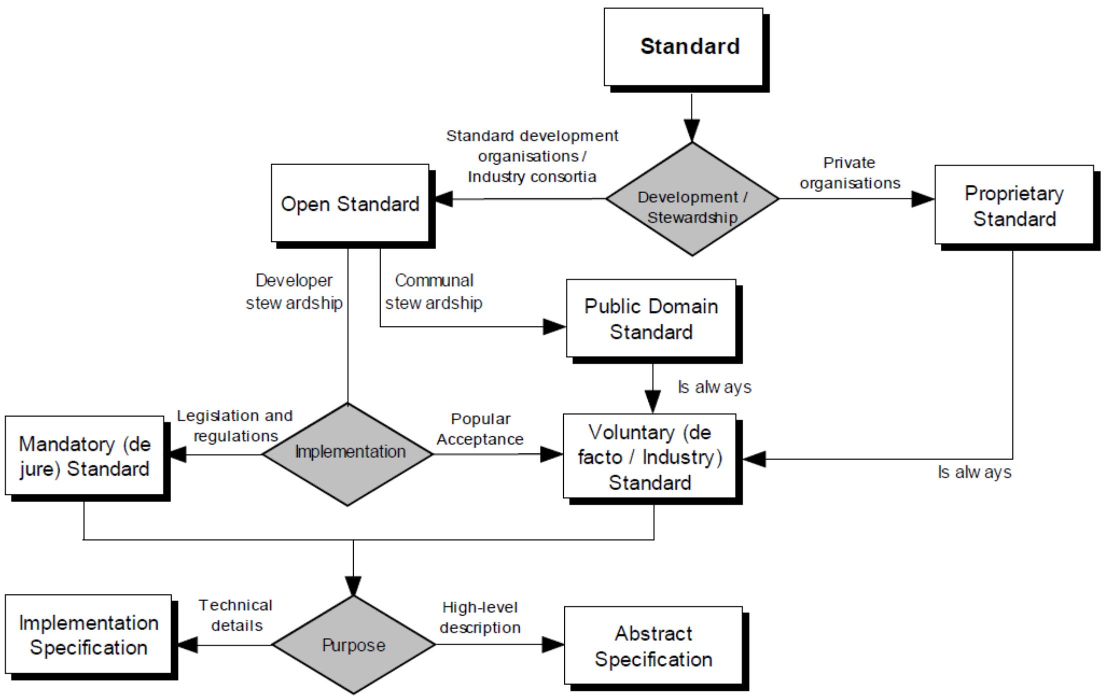


Figure 3-13: Classification of standards (Yeung and Hall 2007)

Within the geospatial community, the Open Geospatial Consortium (OGC) has specified several open standards for integrating spatial data and geo-processing resources into mainstream computing for enhancing interoperable spatial information sharing (OGC 2009). The most important open standards by the OGC which have been used in this research in particular to address the research objective 5, relating to implementation of a prototype system for automatic spatial metadata updating, are as follows:

- Catalogue Services for the Web (CSW): supports the ability to publish and search collections of descriptive information (metadata) about geospatial data, services and related resources.

- Geography Markup Language (GML): an XML grammar for expressing geographical features – GML serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet.

- Web Feature Service (WFS): defines interfaces for data access and manipulation operations on geographic features using HTTP as the distributed computing platform.

- Web Map Service (WMS): provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases.

The way in which the above standards have been used in this thesis will be discussed in Chapters 6 and 7.

3.4 Chapter Summary

This chapter reviewed the research and development activities regarding metadata automation in two key areas: digital library and information science community and the geospatial community. Within the methods investigated and proposed in digital library and information science community extraction, harvesting, machine learning, classification, indexing, and tagging have been at the forefront of metadata automation methods. Also, these methods have formed the foundation for emerging spatial metadata automation approaches.

The chapter also investigated the major research and development activities on vector and raster dataset metadata automation in the geospatial community. Out of the available automation methods, metadata extraction from dataset, harvesting, inference, and computation have been the key methods used in most studies. As a result of spatial metadata automation literature review, a number of main challenges relating to metadata management and automation were also identified and discussed in this chapter. Finally, the technological aspect of spatial metadata automation was explored. The technologies considered in this thesis were categorised and reviewed in three main groups: Web generations, Open Source Software, and open standards.

The next chapter will discuss the research design and methodology to answer the research questions discussed in Chapter 1.

CHAPTER 4

RESEARCH DESIGN AND METHODS

4 RESEARCH DESIGN AND METHODS

4.1 Introduction

This chapter outlines the research design and methods that were used to answer the research questions and achieve the research aim and objectives, stated earlier in Chapter 1. The first part of the chapter investigates the conceptual design framework by reviewing the research problem and questions, and then explores the possible research methods available to answer these questions. The chosen research methods are then justified and the final research design including ‘conceptual’, ‘design’, ‘implementation’, and ‘evaluation’ phases presented. Finally, the ethical considerations and prototype systems communication relating to the undertaken surveys are described.

4.2 Conceptual Design Framework

The underlying concepts of spatial metadata and the literature of spatial metadata automation in the areas of research and development were reviewed in Chapters 2 and 3.

As a result, a number of challenges were identified in the research, including separation of metadata creation process from the spatial data lifecycle; dependency of automatic metadata generation and updating approaches on dataset format; delay in updating metadata when the corresponding dataset changes; storing dataset and metadata using a detached data model; and lack of support for interaction between the spatial data catalogues and end users to improve the content of metadata and facilitate the data discovery process.

As stated in Chapter 1 and supported by the literature review, the key research questions were clarified as follows:

1. What is the current status of spatial metadata management and its requirements in terms of automation?
2. What is the relationship between spatial metadata elements and spatial data lifecycle?
3. Can an approach be designed to integrate spatial metadata creation with the spatial data lifecycle?
4. Can an automatic and dataset format agnostic approach be designed and implemented to overcome the ineffectiveness and inefficiency in updating metadata whenever the dataset is modified?

5. Can an automated approach be designed and implemented to improve the content of spatial metadata through more interaction with the end users seeking spatial data over the Web?

The conceptual design framework shown in Figure 4-1 illustrates the relationship between research context, problem, questions, aim and objectives, methodologies and outcomes.

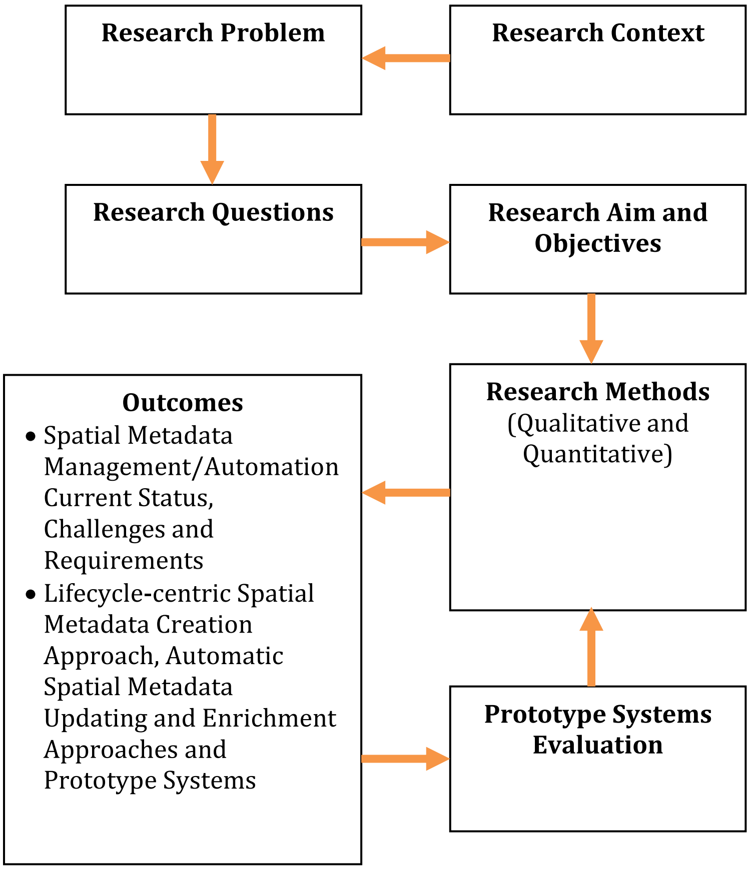


Figure 4-1: Conceptual design framework

The first research question is primarily qualitative in nature and aims to identify the current status of spatial metadata management and its requirements in terms of automation. The second and third research questions are also qualitative in nature and seek to recognise the relationship between spatial metadata elements and the spatial data lifecycle and integrate the metadata creation process with the data lifecycle. However, to address the last two research questions the blending of qualitative and quantitative methods are required. The prototyping technique is needed to design and implement the prototype systems. Also, qualitative and quantitative methods are required to test the prototype systems to answer the relating research questions.

The next section discusses the selection of research approach in more details.

4.3 Selection of Research Approach

At an early stage of the research design it became evident that both qualitative and quantitative methods along with the prototyping technique should be examined. On one hand, there was a need to explore the current status of spatial metadata management, identify the spatial metadata automation requirements, and recognise the relationship between spatial metadata elements and spatial data lifecycle. On the other hand, it was required to prototype the spatial metadata automation approaches and evaluate the developed prototype systems in terms of functionality and usability. Therefore, a mixed method is proposed for answering the research questions. This section reviews the research methods considered in the research design.

4.3.1 Qualitative Research Methods

According to Myers (1997), qualitative research methods were developed in the social sciences to enable researchers to study social and cultural phenomena. Examples of qualitative methods are action research, case study research and ethnography. Qualitative data sources include observation and participant observation (fieldwork), interviews and questionnaires, documents and texts, and the researcher's impressions and reactions.

The strengths of qualitative research methods lie in their usefulness for understanding the meaning and context of the phenomena studied, and the particular events and processes that make up these phenomena over time, in real-life, natural settings (Maxwell 1996). Mack *et al.* (2005) also summarise two main advantages of qualitative research methods. The first advantage is using open-ended questions, which give participants the opportunity to respond in their own words. Open-ended questions have the ability to evoke responses that are meaningful and culturally salient to the participant, unanticipated by the researcher, and rich and explanatory in nature. Another advantage of qualitative research methods is that they allow the researcher the flexibility to probe initial participant responses – that is, to ask 'why' or 'how'.

Within the qualitative research methods, it was decided that a case study strategy would be an appropriate approach in the context of this research for investigation of current status of spatial metadata management and its automation requirements within the geospatial community. The characteristics of case study strategy and justification for this choice are outlined below.

- **Case Study Strategy and Justification**

The cases can be classified into three categories (Gillham 2000). A case can be an *individual*: it can be a group – such as a family, or an office; it can be an *institution* – such as a school or a factory; it can be a large-scale *community* – a town, an industry, a profession. This classification is then followed by the definition of the case study as one which investigates the cases to answer specific research questions and which seeks a range of different kinds of evidence, evidence which is there in the case setting, and which has to be abstracted and collated to get the best possible answers to the research question.

The case study strategy was selected for investigating the current status of spatial metadata management and automation requirements for a number of reasons:

- Case studies allow generalisations either about an instance or from an instance to a class (Adelman *et al.* 1980) which could help to identify the current status of metadata management in the whole geospatial community.
- A case study is particularly suited to research questions which require detailed understanding of organisational processes because of the rich data collected in context (Hartley 2004), such as the first research question.
- The current spatial metadata management approaches could be studied in their natural settings (Kaplan and Maxwell 1994).
- The case study enabled the ‘how’ and ‘why’ research questions (Yin 1994), which are significant for exploring the current approaches used in spatial metadata management and recognising the automation requirements.

Because the case study aimed to investigate the current status of spatial metadata management and automation needs within the whole geospatial community, out of three categories of a case outlined at the beginning of this section a large-scale community was selected as the appropriate category for this research. The case needs to be a representative for the whole geospatial community. Selection of the case study region will be discussed in Section 4.4.1.

Within the broad case study strategy a number of methods may be used – either qualitative, quantitative or both. Case studies generally include multiple methods because of the research issues, which can be best addressed through this strategy. Participant observation, direct observation, ethnography, interviews, focus groups, documentary analysis and questionnaires may be used, or in combination (Hartley 2004). For the purpose of this research, using the questionnaire to undertake the case study was adopted among the data gathering methods due to wide distribution of participating organisations.

4.3.2 Quantitative Research Methods

Quantitative research uses numbers and statistical methods to explain and validate phenomena. Quantitative methods focus on ‘measurements and amounts (more or less, larger or smaller, often or seldom, similar or different) of the characteristics displayed by people and events that the researcher studies (Thomas 2003).

Mack *et al.* (2005) introduce the flexibility of quantitative and qualitative methods as the key difference between them. They further discuss that quantitative methods are fairly inflexible. With quantitative methods such as surveys and questionnaires, for example, researchers ask all participants identical questions in the same order. The response categories from which participants may choose are 'closed-ended' or fixed. The advantage of this inflexibility is that it allows for meaningful comparison of responses across participants and study sites. However, it requires a thorough understanding of the important questions to ask, the best way to ask them, and the range of possible responses.

In the context of this research, the quantitative research method was adopted for examining the prototype systems usability, as part of prototype systems evaluation phase.

ISO 9241-11:1998 (Guidance on usability) defines usability as:

> *'Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use (ISO9241-11: 1998).'*

Jokela *et al.* (2006) also discuss that the above definition means that usability requirements are based on measures of users performing tasks with the product to be developed:

- An example of an effectiveness measure is the percentage of users who can successfully complete a task.
- Efficiency can be measured by the mean time needed to successfully complete a task.
- User satisfaction can be measured with a questionnaire.

In order to evaluate the prototype systems, it was decided that the System Usability Scale (SUS) would be a suitable survey. The SUS definition and justification for this choice are detailed below.

• System Usability Scale (SUS) and Justification

The System Usability Scale (SUS) was developed by Brooke (1996) as a 'quick and dirty' survey scale that would allow the usability practitioner to quickly and easily assess the usability of a given product or service. In fact, SUS is a ten-item Likert scale giving a global view of subjective assessments of usability (Brooke 1996). The SUS statements are shown in Figure 4-2. It can be seen that these statements actually cover a variety of aspects of system usability, such as the need for support, training, and complexity, and thus have a high level of face validity for measuring usability of a system.

	Strongly Disagree				Strongly Agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Figure 4-2: SUS original statements (Brooke 1996)

Respondents should be asked to record their immediate response to each statement, rather than thinking about them for a long time. If a respondent feels that they cannot respond to a particular item, they should mark the centre point of the scale.

System Usability Scale (SUS) yields a single number representing a composite measure of the overall usability of the system being studied. The scores for individual items are not meaningful on their own. To calculate the SUS score, first the score contributions from each item are sum. Each item's score contribution will range from 0 to 4. For items 1, 3, 5, 7 and 9 the score contribution is the scale position minus 1. For items 2, 4, 6, 8 and 10, the contribution is 5 minus the scale position. Then, the sum of the scores is multiplied by 2.5 to obtain the overall value of system usability. SUS scores have a range of 0 to 100.

Bangor *et al.* (2008) investigated nearly 10 year's worth of System Usability Scale (SUS) data collected on numerous products in all phases of the development lifecycle. In this study, the authors provided details of what constitutes an acceptable SUS score. In this regard, they proposed a rule-of-thumb standard based on the typical grading scale used in most schools, which became known as the 'university grade analog'. This standard proposes an adjective rating scale for a product including the 'worst imaginable', 'poor', 'ok', 'good', 'excellent', and 'best imaginable' ratings based on the SUS score. Also, the standard considers a rating scale including the 'not acceptable', 'low marginal', 'high marginal', and 'acceptable' ranges for the acceptability of a product based on its SUS score. According to this standard, for instance a product with the SUS score of 65 out of 100 can achieve an adjective rating between 'OK' and 'Good' and a 'high marginal' score of acceptability. In addition, they discussed that the SUS survey has some advantages against a number of other excellent alternatives, as following:

- The survey is technology agnostic, making it flexible enough to assess a wide range of interface technologies, from Interactive Voice Response (IVR) systems and novel hardware platforms to the more traditional computer interfaces and Websites.
- The survey is relatively quick and easy to use by both study participants and administrators.
- The survey provides a single score on a scale that is easily understood by the wide range of people (from project managers to computer programmers) who are typically involved in the development of products and services and who may have little or no experience in human factors and usability.
- The survey is non-proprietary, making it a cost effective tool as well.

The above advantages, although general in context, provided the basis for justifying the SUS as a suitable survey method in this research for evaluation of the prototype systems usability. In fact, using the SUS survey enables the spatial metadata automation prototype systems developed in this research to be assessed in terms of user-friendliness. Based on the results of case study investigations, which will be discussed in the next chapter, lack of user-friendly tools for cataloguing spatial metadata has been identified as one of the main challenges of the geospatial community. In addition, using the SUS survey the prototype systems will be examined in terms of end users' willingness and confidence to utilise these systems, which can show the effectiveness of designed approaches in this research for addressing the identified main challenges regarding spatial metadata management and automation. The main challenges will be summarised in the next chapter.

However, in the context of this research the prototype systems were required to be evaluated, not only in terms of usability but also in terms of developed functionalities. Assessing the system functionalities is qualitative in nature. Kaplan and Maxwell (1994) discuss that the qualitative research methods are used increasingly in evaluation studies, including evaluations of computer systems and information technology. Therefore, there was a need to adopt qualitative research methods in addition to quantitative methods for the evaluation purpose.

According to the research methods adopted for both phases of data collection in this research (first phase for exploring the current status of spatial metadata management through a case study strategy; and second phase for evaluating the prototype systems), a mixed-methods research approach was selected. This kind of research is explored in the following section.

4.3.3 Mixed-Methods Research Approach

The debate over the benefits of qualitative versus quantitative methods continues, with the proponents in each camp vigorously defending the benefits and rigour of each approach (Tashakkori and Teddlie 2003). According to Carey (1993), quantitative and qualitative methods are merely tools; integrating them allows us to answer questions of substantial importance. Also, Teddlie and Tashakkori (2003) identify three reasons where the utility of mixed methods research may be superior to single method approach:

1. Mixed methods research can answer research questions that other methodologies cannot.
2. Mixed methods research provides better (stronger) inferences.
3. Mixed methods provide the opportunity for presenting a greater diversity of divergent views.

The above reasons, although general in context, provided the basis for justifying the mixed-methods approach as a suitable research approach in this thesis. The research questions already identified could not be addressed by a single research method. In order to study the current status of spatial metadata management and identify the automation requirements, there was a need to undertake a case study to answer the 'why' and 'how' questions. For instance, 'how' the organisations currently create and update spatial metadata. Also, identifying the relationship between metadata elements and spatial data lifecycle was qualitative in nature, which needed response to questions such as 'how' a metadata element can be mapped against the spatial data lifecycle. Moreover, in order to understand the effectiveness and usability of the developed prototype systems based on the recognised requirements there was a need to evaluate these systems. Both qualitative and quantitative methods were involved in this evaluation. The evaluation results help to understand how the requirements have been addressed by this research.

Creswell *et al.* (2003) design a matrix, as shown in Figure 4-3, which illustrates that four decisions go into selecting a mixed-methods strategy:

1. What is the implementations sequence of the quantitative and qualitative data collection in the proposed study?
2. What priority will be given to the quantitative and qualitative data collection and analysis?
3. At what stage in the research project will the quantitative and qualitative data and findings be integrated?
4. Will an overall theoretical perspective (e.g. gender, race/ethnicity, lifestyle, class) be used in the study?

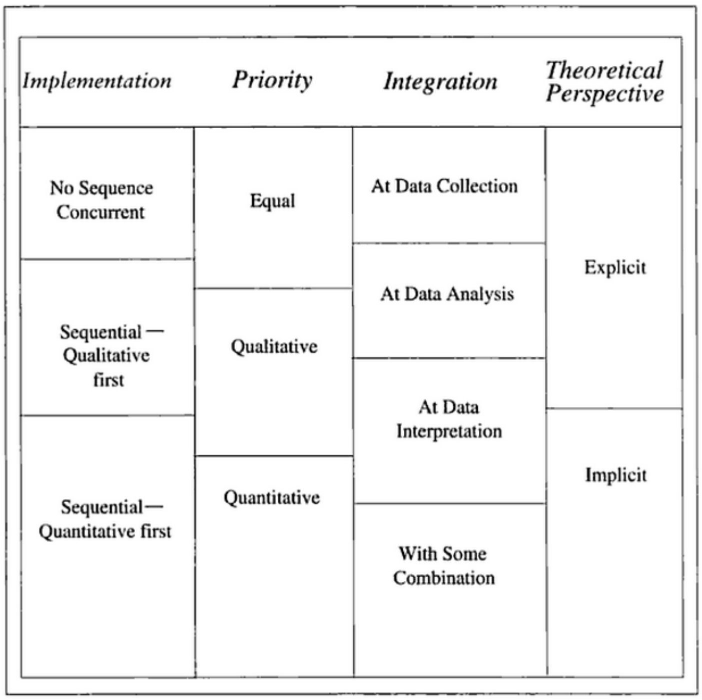


Figure 4-3: Decision matrix for determining a mixed methods design (Creswell *et al.* 2003)

This research followed the second implementation strategy, namely 'Sequential – Qualitative first'. The qualitative method was first used to investigate the spatial metadata management current status and automation requirements and the relationship between metadata and spatial data lifecycle. Then, once the prototype systems were designed and developed these systems were evaluated through qualitative and quantitative methods. Finally, the results were integrated to see if the research aim is addressed or not.

4.4 Research Design

The research design consisted of four overall phases including *conceptual*, *design*, *implementation*, and *evaluation* phases. Figure 4-4 illustrates these phases. Also, this Figure shows the relationship between the research phases, objectives and thesis chapters.

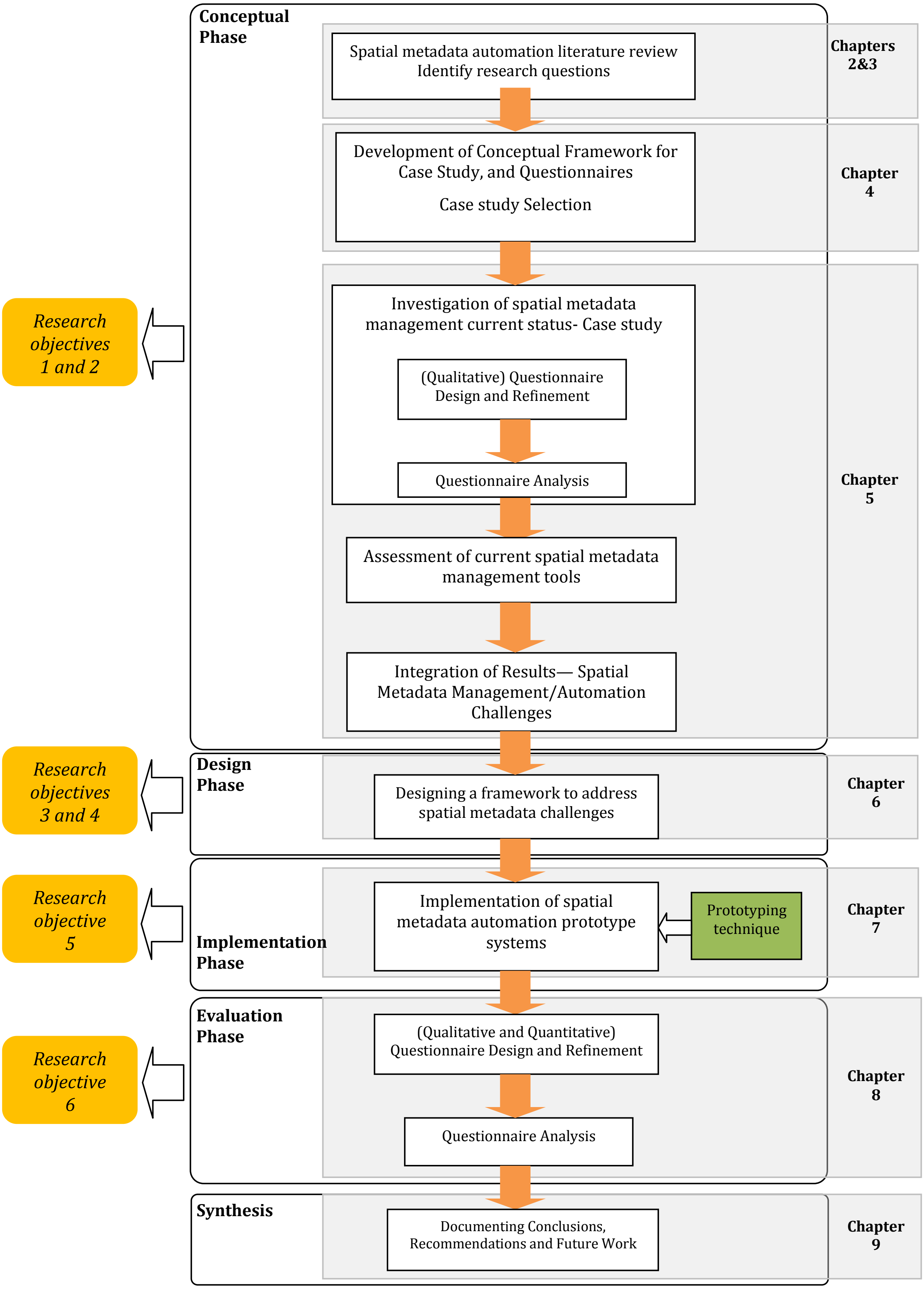


Figure 4-4: The relationship between research design, objectives and chapters

The research phases together with the research methods employed are detailed as following.

4.4.1 Conceptual Phase

- **Literature Review**

In this phase, an extensive literature review has been undertaken on two main areas: metadata (spatial and non-spatial), and its automation approaches. Different resources are used to collect a range of information for reviewing these two areas. These resources include books, journals, international standards, organisation reports, conference proceedings, and information published over the World Wide Web (WWW). Following the results of literature review the research questions, as stated in Chapter 1, were identified. The results of the literature review were discussed in previous two chapters.

- **Case Study Selection**

The case study strategy was selected for investigating the current status of spatial metadata management and automation requirements within the geospatial community. Australia was selected as the case study region for undertaking this investigation for a number of reasons:

- Australia has been contributing to design and development of spatial metadata standards for many years. Releasing the ANZLIC Metadata Guidelines Version 2: February 2001, ANZLIC Metadata Profile Version 1.1: August 2007, and ANZLIC Metadata Profile Guidelines Version 1.2: July 2011 by ANZLIC confirms this key contribution.
- Australia has planned to standardise the spatial metadata in the organisations through early adoption of the ISO Metadata Standards and developing the ANZLIC Metadata Profile version 1.1 in August 2007.
- Australia has been a key player in revising the ISO 19115 Metadata Standard (ANZLIC has participated in the review process through Standards Australia's IT-004 Geographic Information Committee).
- Australia has been heavily involved in developing and improving one of the world's commonly used data catalogues- GeoNetwork opensource platform (through the activities on BlueNetMEST/ANZ-MEST).
- In addition to the activities based on the GeoNetwork, Australia has recently designed and developed some spatial metadata management tools based on state-of-the-art technologies to facilitate spatial metadata management (e.g. ANZMet Lite, xMet Client).
- Australia has provided the users looking for data with different data catalogues using spatial metadata. Australian Spatial Data Directory (ASDD), Australian Ocean Data

Network Portal, and Australian Urban Research Infrastructure Network (AURIN) Portal are some of these catalogues.

Once the case study region was selected, the survey was designed to gather the required data.

- **Survey Design and Implementation**

Wyatt (2000) discusses that a survey needs to be designed by defining two major features: how participants are identified and the data capture technique.

➢ **Selecting Participants**

How participants are identified determines how much control the investigators exert over their selection. For complete control, eligible participants are selected according to specific characteristics or pre-defined criteria – 'purposive' sampling (Mack *et al.* 2005). If, instead, only a modest degree of control is considered, copies of the survey can be distributed to a group known to the investigators and ask them in turn to send it out to their contacts – the 'snowball' technique (Gold *et al.* 2000). With this method there is much less control over who receives a copy of the survey, and it makes sending reminders or follow-up surveys more difficult, too. However, snowball sampling has been widely used in qualitative research (Biernacki and Waldorf 1981).

In the context of this research, the survey was designed based on snowball sampling. The Victorian Departments of Primary Industries (DPI) and Sustainability and Environment (DSE) known to the researcher, were asked to nominate their contacts whom deal with spatial data and metadata as their main business.

As a result, 12 organisations including DPI and DSE were selected for the survey. Table 4-1 illustrates the information of these organisations.

Table 4-1: Selected organisations to participate in the first survey of the research within the case study region (Australia)

No.	Organisation Name	Organisation Website
1	Office of Spatial Data management (Former Australian Government Office of Spatial Data Management (OSDM))	http://spatial.gov.au/
2	Victorian Department of Primary Industries	http://www.dpi.vic.gov.au/
3	Victorian Department of Sustainability and Environment	http://www.dse.vic.gov.au/
4	ACT Planning and Land Authority	http://www.actpla.act.gov.au/
5	PSMA Australia Limited	http://www.psma.com.au/
6	VicRoads	http://www.vicroads.vic.gov.au/
7	Bureau of Meteorology	http://www.bom.gov.au/
8	Tasmanian Department of Primary Industries, Parks, Water and Environment	http://www.dpiw.tas.gov.au/
9	Australian Hydrographic Service	http://www.hydro.gov.au/
10	South Australian Department for Transport, Energy and Infrastructure	http://www.dpti.sa.gov.au/
11	Western Australian Land Information System (WALIS)	http://www.walis.wa.gov.au/
12	Sinclair Knight Merz Pty Ltd.	http://www.skmconsulting.com/

➢ Selecting Data Collection Method

Once the participants were selected for the survey, different data gathering techniques were investigated. Among the available techniques e.g. face-to-face interview (Greenfield *et al.* 2000), paper questionnaire (Velikova *et al.* 1999), phone survey (McCarty *et al.* 2006) and Web-based questionnaire (Wyatt 2000), Web-based questionnaire was selected for a number of reasons:

- The selected organisations were distributed widely in Australia.
- Web-based survey was cheap to carry out (Wyatt 2000).
- It was easy to monitor the progress of survey and see how many of invitees participated in different stages of research.
- The data could be captured directly in electronic format, making analysis faster and cheaper.
- Using multimedia to design a user-friendly questionnaire was supported.

➢ Designing Questionnaire

The questionnaire aimed at gathering information about the current status of spatial metadata management and discovering the requirements of metadata automation within the case study region. This questionnaire was designed in 8 main parts, as following:

Part 1) Introduction: aimed to give an introduction to both research and questionnaire.

Part 2) Participant's Information: designed to gather participant's name, email address and organisation name.

Part 3) Organisational Spatial Data and Metadata Characteristics: focused on collecting data about the number of spatial datasets; activities regarding metadata for spatial datasets; and activities regarding spatial metadata lifecycle in the participating organisations.

Part 4) Spatial Metadata Applications: aimed to gather data on the main applications of spatial metadata in participating organisations, in addition to the well known applications including data discovery, exploration and exploitation.

Part 5) Spatial Metadata Creation and Updating: focused on collecting data regarding spatial metadata creation and updating methods; the amount and format of metadata created by external teams; standards used for metadata creation; the formats of metadata storage; existence of any automatic metadata creation application; the mechanism for updating spatial data; the updating time interval for spatial metadata; and finally the responsible party for updating spatial metadata within the participating organisations.

Part 6) Spatial Metadata Management Tools: aimed to gather data on the tools used in metadata management.

Part 7) Spatial Metadata Sharing: designed for collecting data on the possibility of sharing metadata with other parties; the amount of metadata accessed by other parties; and the mechanisms used to publish spatial metadata within the participating organisations.

Part 8) Spatial Metadata Management Issues and Challenges/Desirable Automation Workflow: aimed to gather data on current issues and challenges regarding metadata management; desirable metadata automation workflow; and further metadata automation requirements.

A copy of the questionnaire is provided in **Appendix 3**.

- **Questionnaire Refinement**

The questionnaire was tested internally and externally. A draft questionnaire was developed in digital format and distributed to the members of the Centre for SDIs and Land Administration (CSDILA) within the Department of Infrastructure Engineering (former Department of Geomatics) at the University of Melbourne to check for the understanding of the questions being asked. Also, a contact person from the DSE was invited to test the questionnaire to check for terminology and understanding. The feedback was collected and incorporated into the questionnaire. The Web-based questionnaire was then prepared and tested internally to ensure that the URL was accessible and also that responses were being recorded at the Web server.

➢ Questionnaire Distribution and Collection

In order to distribute the questionnaire, first a contact person was selected from each organisation. In this regard, the management team of each selected organisation was provided with the introduction to the questionnaire and its detail and formally invited to nominate a contact person who would be the best choice for responding the questions. Then, the contact people were formally invited to participate in the survey. The survey duration was two months to collect the participants' responses. The data from the questionnaires were automatically collected into an excel spreadsheet via the Web server.

➢ Ethics Considerations

Appropriate ethical approval to conduct the human research was gained through the University Ethics Committee. The information from the case study and questionnaire remained confidential and was utilised for research purposes only.

• Assessment of Spatial Metadata Management Tools

In the conceptual phase a number of commonly used spatial metadata management tools were also assessed to understand the main functionalities and automation capabilities supported by these tools.

Following the main challenges of spatial metadata management and automation identified during the previous stages of the conceptual phase (literature review and case study); in this stage, a set of criteria was suggested to assess the current status of spatial metadata management tools in terms of support for:

- integration of metadata creation with spatial data lifecycle
- integrated data model for storing metadata and dataset
- automatic metadata creation
- automatic metadata updating whenever dataset changes
- interaction with the end users to improve the content of metadata.

Once a set of criteria was proposed, a number of commonly used spatial metadata management tools were selected for the assessment purpose. In order for this selection, an investigation was undertaken to find out the broadly used tools in the geospatial community that could be adapted to suit the spatial metadata needs of the case study region – Australia. The results from the case study investigations were also used for the selection purpose.

According to Rajabifard *et al.* (2007), CatMDEdit, and GeoNetwork have been recognised as the international metadata management tools that could be adapted to suit the needs of ANZLIC, as the peak intergovernmental organisation for the collection, management and use of spatial information in Australia and New Zealand. Also, according to findings of case study investigations, ANZMet Lite, GeoNetwork, BlueNetMEST, and ESRI ArcCatalog were frequently used by the participating organisations in Australia.

Therefore, all the above tools were selected to be assessed against the criteria proposed in this research. In addition to these tools, two other international tools, which have been recently used by different countries, were also chosen to be involved in the assessment process. These tools included GeoNode, and European Open Source Metadata Editor (EUOSME). Moreover, another tool, which has been recently developed in Australia namely 'xMet Client', was considered for the evaluation purpose.

As a result of this investigation, the spatial metadata management tools selected to be reviewed and examined against the criteria, were categorised in two main groups, as following:

- International tools
    - GeoNetwork opensource
    - GeoNode
    - ESRI ArcCatalog
    - CatMDEdit
    - EUOSME
- Australian tools
    - BlueNetMEST
    - ANZMet Lite
    - xMet Client

The selected tools were then examined against the proposed set of criteria and the results were documented. These results are discussed in the next chapter.

• Integration of Results

As the final stage of the conceptual phase, the results achieved from the previous stages of this phase (including literature review, case study, and metadata tools assessment) were integrated and the main challenges regarding the spatial metadata management and automation were identified. These results are discussed in the next chapter.

4.4.2 Design Phase

In this phase, following the findings of the conceptual phase, a framework was developed to address the main challenges identified in the previous phase and respond to the research objectives 3, 4, 5 and 6. This framework consisted of three complementary approaches namely lifecycle-centric spatial metadata creation, automatic spatial metadata updating (synchronisation), and automatic spatial metadata enrichment. This framework will be presented in Section 6.2, Chapter 6.

- **Lifecycle-centric Spatial Metadata Creation Approach**

The lifecycle-centric spatial metadata creation approach was designed to address the issues of current metadata creation approaches investigated in the conceptual phase of the research. The current approaches mainly focus on dataset to generate metadata, while the lifecycle-centric approach aims at creating metadata along with different steps of the spatial data lifecycle. In order for designing such an approach, first there was a need for designing the generic spatial data lifecycle. Because the prototype systems development and evaluation occurred in Australia, the case study region, the ‘information lifecycle’ recommended by the Australian Government Information Interoperability Framework (AGIMO 2006) was adopted in this research to identify the steps involved in the generic spatial data lifecycle. Once the lifecycle was designed, the spatial metadata elements recommended by the ISO19115: 2003 Standard were mapped against the steps of the spatial data lifecycle. The ISO19115: 2003 Standard was chosen for this purpose due its popularity within the geospatial community; in particular, the case study region – Australia, according to the results of case study investigations. Finally, following the results arisen from the mapping of metadata elements against the lifecycle a new approach was designed and proposed for spatial metadata creation in parallel with the spatial data lifecycle.

- **Automatic Spatial Metadata Updating (synchronisation) Approach**

The automatic spatial metadata updating (synchronisation) approach aimed at updating vector spatial dataset and metadata in real time. According to the findings of the research conceptual phase (including literature review, case study investigations and metadata tools assessment), lack of support for real-time metadata and dataset updating was identified as one of the main challenges that the geospatial community faces. In order to design an approach to address this challenge, first an integrated data model to store both spatial data and metadata in a middleware was designed and developed. Thereafter, the other technical requirements for synchronising metadata and dataset modification were identified. Also, the relationships between the technical requirements were defined.

- **Automatic Spatial Metadata Enrichment Approach**

The findings of the research conceptual phase showed that the current spatial metadata management and discovery systems do not interact properly with the end users to improve the content of spatial metadata and the end user is disconnected from the spatial metadata creation and enrichment process. Therefore, the automatic spatial metadata enrichment approach was designed to address the mentioned challenge and engage the end users in improving the content of spatial metadata (keyword metadata element). In this regard, two complementary models namely 'indirect' and 'direct' were designed based on the features of Web 2.0, already discussed in Section 3.3.2, Chapter 3.

4.4.3 Implementation Phase

In order to prove the concepts of automatic spatial metadata updating and enrichment approaches designed in the 'design' phase, this phase aimed at implementing two prototype systems.

O'Leary (1988) discusses that '*prototyping'* the system provides a *'proof of concept'*. This discussion is followed by the fact that prototyping is a well-established tradition in engineering as a means of testing a product in order to determine if a product can be built and to ascertain if there are any operating anomalies in the design. Also, Eagleson (2002) introduces the prototyping as a technique for the development of software of which its aim is to produce software quickly using simple methods and tools. In this instance, the prototype allows the concept of the automatic spatial metadata updating and enrichment approaches to be tested in a real situation, using real data and real constraints.

In the context of this research, the stages of prototyping were defined based on the '*system development research methodology'* introduced by Nunamaker *et al*. (1990-91), as illustrated in Figure 4-5.

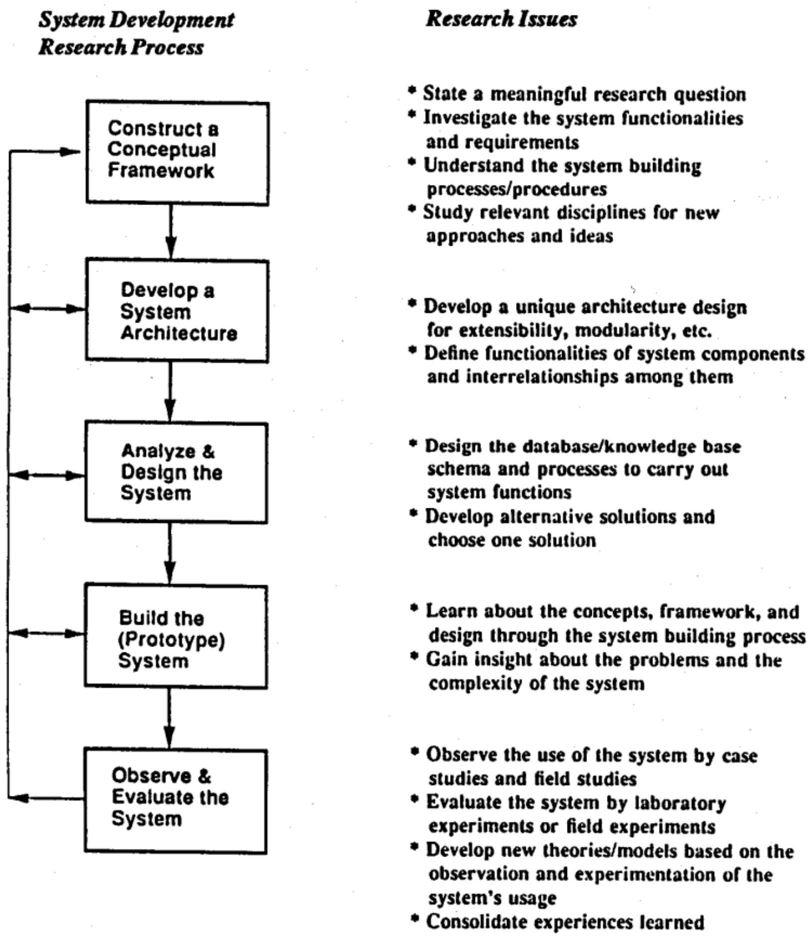


Figure 4-5: A process for systems development research (Nunamaker *et al.* 1990-91)

These stages are detailed below.

1) Construct a Conceptual Framework

The first stage of systems development research is constructing a conceptual framework that leads to theory building. In this stage, clear and meaningful research questions should be stated, the requirements of system should be investigated, the processes for building the system should be identified and relevant disciplines should be investigated for new approaches and ideas. Accordingly, the results of 'conceptual phase' of the current research formed the conceptual framework required for the automatic spatial metadata updating and enrichment prototype systems.

2) Develop a System Architecture

A system architecture provides a roadmap for the system building process. It puts the system components into perspective, specifies the system functionalities, and defines the structural relationships and dynamic interactions among system components. In this regard, the conceptual designs of both automatic spatial metadata updating, and enrichment approaches, as the outputs of ‘design phase’, were used to develop the detailed architectures for both prototype systems.

3) Analyse and Design the System

Design, one of the most important parts of a systems development process, is rooted in engineering (Denning *et al*. 1989). It involves the understanding of the studied domain, the application of relevant scientific and technical knowledge, the creation of various alternatives, and synthesis and evaluation of proposed alternative solutions. In this stage, the data structures, technologies involved in the integrated data model for storing both spatial data and metadata, and databases used in the prototype systems were determined. Also, the programming languages for both front and back ends were specified.

As already described in Section 3.3 in Chapter 3, this research focused on using the open source environments to build the prototype systems. For instance, to build the automatic spatial metadata updating prototype system, the GeoNetwork opensource was selected as the base platform. Also, open source Web server (deegree), database (PostgreSQL and PostGIS) and programming languages (PHP, HTML, JavaScript, OpenLayers, GeoExt, ExtJS) were chosen for developing this prototype system.

4) Build the Prototype System

Implementation of a system is used to demonstrate the feasibility of the design and the usability of the functionalities of a system. In this stage of the implementation phase, both the automatic spatial metadata updating and enrichment prototype systems were built according to their conceptual frameworks, system architectures and system designs.

In order to build both prototype systems the GeoNetwork opensource catalogue was chosen as the main development environment for a number of reasons:

1. The results of the case study investigations confirmed that the GeoNetwork opensource catalogue is commonly used within the organisations in the case study region-Australia.
2. The GeoNetwork has been used in Australia as the main platform for building up spatial metadata tools such as BlueNet MEST/ANZ-MEST, which will be discussed in the next chapter.
3. Since the GeoNetwork is an open source software, the contributions to this software can be shared among the GeoNetwork community of users and developers, so that others can use and build upon the work provided by this thesis.

Also, the automatic spatial metadata enrichment prototype system was decided to be built within the Model Information Knowledge Environment (MIKE) developed by the Victorian Department of Primary Industries (DPI) in addition to the GeoNetwork opensource catalogue to:

- use a real world system to prove the concept of automatic metadata enrichment
- apply enrichment approach to a non-spatial data catalogue (an example of data product-data modelling environment) in addition to a spatial data catalogue (GeoNetwork opensource).

5) Observe and Evaluate the System

Once the system is built, its performance and usability should be tested and its impacts on individuals, groups or organisations should be observed. The test results should be interpreted and evaluated based on the conceptual framework and the requirements of the system defined at the earlier stages. The evaluation of the automatic spatial metadata updating and enrichment prototype systems in this research was undertaken in the 'evaluation phase', which is discussed in the following section.

4.4.4 Evaluation Phase

In this phase, the automatic spatial metadata updating and enrichment prototype systems were evaluated. Consequently, the areas that required improvement were identified. Since the automatic spatial metadata enrichment approach was implemented within two different environments – GeoNetwork and MIKE – one of these environments was selected for evaluating the prototype system. Among both prototype environments, the GeoNetwork opensource was selected as the test environment for a number of reasons:

1. The GeoNetwork has been used and developed by the geospatial community for a long time (since 2001), so that more people were familiar with this catalogue in comparison with the MIKE, which has a smaller community; thus better feedback could be provided by comparing the GeoNetwork original functionality with the new functionality proposed by the prototype system.
2. The MIKE system was still under development by the DPI, Victoria in this phase of the research, so that it could not be accessed by the Australian and international participants for the evaluation purpose.
3. The GeoNetwork was used by different countries and for this reason, the number of international participants in the prototype system evaluation process could likely increase.

4. The prototype system implemented within the GeoNetwork included all the functionalities developed for the prototype system within the MIKE; therefore, the feedback provided for the GeoNetwork could be applied to the MIKE system as well.

In order to evaluate the automatic spatial metadata updating and enrichment prototype systems, a survey was designed and undertaken which is detailed below.

- **Survey Design and Implementation**

As discussed earlier, in designing a survey two major features should be defined: how participants are identified and the data capture technique. These two features for designing a survey to evaluate the prototype systems are discussed here.

➢ **Selecting Participants**

In order to select the participants for this survey, the snowball sampling technique was used. However, this time the organisations, which had participated in the initial survey of the research, were formally invited to evaluate the prototype systems. They were also asked to distribute the survey to their contacts. Moreover, a URL of the survey was shared within the professional social media (Victoria Geo Message[15], and geomatics-related groups in Linkedin[16]) for other interested participants from Australia or other countries. Victoria Geo Message is prepared monthly and distributed by the Surveying and Spatial Sciences Institute (SSSI), Australia and shares the important information and news regarding the spatial activities and events within Victoria. LinkedIn is also the world's largest professional network with over 175 million members and growing rapidly (Linkedin 2012).

➢ **Selecting Data Collection Method**

A Web-based questionnaire was selected as the most suitable data collection method for this survey according to the reasons already discussed in designing the initial survey of the research, discussed in Section 4.4.1.

➢ **Designing Questionnaires**

Two questionnaires were designed for this survey detailed below.

---

15 http://www.sssi.org.au/index/region/6.html
16 https://www.linkedin.com/

A) Questionnaire for the Evaluation of Automatic Spatial Metadata Updating Prototype System

This questionnaire was designed to evaluate the automatic spatial metadata updating prototype system. In order to design this questionnaire, there was a need for a set of criteria to be proposed for the evaluation purpose.

This set of criteria was suggested to make sure that the automatic spatial metadata updating prototype system had successfully addressed: the challenges of current spatial metadata management approaches regarding the use of detached metadata data model; lack of support for an effective and efficient real-time spatial data and metadata updating; and dependency of metadata automation methods on dataset format. The evaluation criteria is summarised in Table 4-2. As it can be seen, the set of criteria consists of three main categories including system functionality, usability, and efficiency. Also, each category is classified into some important areas in which the system should be evaluated against them.

Table 4-2: Evaluation criteria for automatic spatial metadata updating prototype system

Category	Criteria
System functionality	Willingness to apply the developed integrated data model into the existing datasets and metadata in the organisations
	Overall effectiveness of integrated data model in facilitating datasets publication and related interoperability issues
	Usefulness of system in updating the dataset and metadata in real time
	Overall effectiveness of system in improving the dataset discoverability
	Confidence of end users to maintain metadata records using the system
	Satisfaction of end users with metadata elements supported by the system
System usability	Willingness to use the system among end users
	System ease of use (user-friendliness)
	Confidence of end users to use the system
	Need for learning a lot of things before using the system
	Need for technical support to work with the system
	Quickness of learning the system
	The quality of functionalities integration within the system
	Sufficiency of system functionalities
System efficiency	The system efficiency in terms of saving time
	The system efficiency in terms of saving cost

Once the set of criteria was proposed, the questionnaire was designed in the following 5 main parts:

Part 1) Introduction: aimed to give an introduction to the prototype system and the questionnaire.

Part 2) Participant's Information: aimed to collect the participant's information such as name, email, organisation, unit or division and position title.

Part 3) Evaluation of System Functionality: designed to give a view of the functionalities of the prototype system and then collect required data to address the criteria. Therefore, this part first introduced the new integrated data model designed for storing spatial data and metadata, as well as, the whole functionality of the prototype system in terms of updating metadata and dataset in real time. Then, the related questions were asked. For the majority of questions the responses were measured on a five point Likert scale in order to standardise and categorise the responses. The debate over the consideration of Likert scale data as either ordinal or interval data continues (Newman 1994); however, it has been argued successfully that Likert scales can validly be considered as interval data, particularly if five or more scale points are utilised (Jaccard and Wan 1996).

Part 4) Evaluation of System Usability: described earlier in this chapter, the System Usability Scale (SUS) method was selected for evaluating the prototype systems usability. However, according to the criteria developed for evaluating the usability of automatic spatial metadata updating prototype system 7 key statements (1, 3, 4, 5, 7, 9, and 10) already introduced in Section 4.3.2 were selected from the 10 original SUS statements and then customised. In addition to these 7 statements, another statement was defined and considered necessary for assessing the system usability that was compatible with the SUS scoring for odd number statements (already discussed in Section 4.3.2). Accordingly, to calculate the SUS score out of 100 the sum of scores for these 8 statements multiplied by 3.125. Table 4-3 illustrates the statements used in this section of the questionnaire.

Table 4-3: Statements used for evaluation of automatic spatial metadata updating prototype system usability

Original SUS Statement No.	Original SUS Statement	Customised Statement
1	I think that I would like to use this system frequently.	I would like to use this system more frequently than the existing metadata management system in my organisation.
3	I thought the system was easy to use.	I perceive that the system is user friendly.
4	I think that I would need the support of a technical person to be able to use this system.	I need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.	Various functions in this system are well integrated.
7	I would imagine that most people would learn to use this system very quickly.	I perceive that people responsible for metadata could learn to use this system very quickly.
9	I felt very confident using the system.	I feel confident using the system.
10	I needed to learn a lot of things before I could get going with this system.	I need to learn a lot of things before I can get going with this system.
N/A	N/A	There is enough functionality in the system to make it a useful tool.

Part 5) Evaluation of System Efficiency: to address the criteria, this part of the questionnaire included questions regarding the time and cost of metadata updating.

A copy of the questionnaire is provided in **Appendix 4**.

B) Questionnaire for the Evaluation of Automatic Spatial Metadata Enrichment Prototype System

The second questionnaire was designed to evaluate the automatic spatial metadata enrichment prototype system. Similar to the first questionnaire there was a need to propose a set of criteria for the evaluation purpose.

This set of criteria was suggested to make sure that the automatic spatial metadata enrichment prototype system had successfully addressed the challenge of current spatial data discovery systems regarding the limited interaction with the end users for metadata creation and improvement. The evaluation criteria is summarised in Table 4-4. As it can be shown, the set of criteria consists of two main categories including system functionality and usability. Also, each category is classified into some important areas that the system should be evaluated against them.

Table 4-4: Evaluation criteria for automatic spatial metadata enrichment prototype system

Category	Criteria
System functionality	Preferences regarding spatial data discovery methods
	Overall effectiveness of system in improving data discovery
	Overall effectiveness of designed search word weighting system
	Overall effectiveness of user-generated search words in data discovery against the original search words embedded in metadata
	Need for an approval process for user-generated search words
	Overall effectiveness of tag cloud in facilitating dataset search and discovery
System usability	Willingness to use the system among end users
	System ease of use (user-friendliness)
	Need for technical support to work with the system
	The quality of add-ons integration within the system

Once the set of criteria was proposed, the questionnaire was designed in the following 4 main parts:

Part 1) Introduction: aimed to give an introduction to the prototype system and the questionnaire.

Part 2) Participant's Information: aimed to collect participant's information such as name, email, organisation, unit or division and position title.

Part 3) Evaluation of System Functionality: designed to give a view on the functionalities of the prototype system and then collect required data to address the criteria. Therefore, this part first introduced the add-ons and weighting system developed for the automatic enrichment of spatial metadata. Then, the related questions were asked. Similar to the first questionnaire, for the majority of questions the responses were measured on a five point Likert scale.

Part 4) Evaluation of System Usability: to design the questions for evaluating the prototype system usability, 4 critical statements that addressed the developed criteria were selected from 10 original System Usability Scale (SUS) statements already introduced in Section 4.3.2. Statements 1, 3, 4, and 5 were chosen and customised for this purpose. Accordingly, to calculate the SUS score out of 100 the sum of scores for these 4 statements multiplied by 6.25. Table 4-5 illustrates the statements used in this section of the questionnaire.

Table 4-5: Statements used for evaluation of automatic spatial metadata enrichment prototype system usability

Original SUS Statement No.	Original SUS Statement	Customised Statement
1	I think that I would like to use this system frequently.	I think that I would like to use this system for searching and discovering datasets more frequently than the current system in my organisation.
3	I thought the system was easy to use.	I perceive that the system is user friendly.
4	I think that I would need the support of a technical person to be able to use this system.	I think that I need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.	I think that the various functions (add-ons) in this system are well integrated.

A copy of the questionnaire is provided in **Appendix 5**.

➢ Questionnaires Refinement

Both questionnaires were tested internally and externally. A draft version of the questionnaires was developed in digital format and distributed to the members of the Centre for SDIs and Land Administration (CSDILA) at the University of Melbourne to check for understanding. Also, a contact person from the Australian Urban Research Infrastructure Network (AURIN) technical team was invited to test the questionnaires to check for terminology and understanding of the questions being asked. The feedback was collected and incorporated into the questionnaires. The Web-based questionnaires were then prepared and tested internally to ensure that the URLs were accessible and also that responses were being recorded at the Web server.

➢ Questionnaires Distribution and Collection

The participants of the initial survey of the research were formally invited to participate in this survey. As mentioned earlier, in order to prompt and initiate more participants, the URLs of both questionnaires were also shared in professional social networks. The survey duration took one month to collect the participants' responses. The data from the questionnaires was automatically collected into an excel spreadsheet via the Web server.

There were 13 organisations who participated in the evaluation of automatic spatial metadata updating prototype system. Table 4-6 illustrates the detail of these organisations. However, out of these participating organisations only five of them first listed in this Table were involved in the initial survey of the research.

Table 4-6: Organisations participated in the evaluation survey of automatic spatial metadata updating prototype system

No.	Organisation Name	Organisation Website
1	Victorian Department of Primary Industries (DPI), Australia	http://www.dpi.vic.gov.au/
2	Victorian Department of Sustainability and Environment (DSE), Australia	http://www.dse.vic.gov.au/
3	PSMA Australia Limited	http://www.psma.com.au/
4	Tasmanian Department of Primary Industries, Parks, Water and Environment, Australia	http://www.dpiw.tas.gov.au/
5	South Australian Department for Transport, Energy and Infrastructure	http://www.dpti.sa.gov.au/
6	Victorian Department of Planning and Community Development, Australia	http://www.dpcd.vic.gov.au/
7	Land and Property Information, NSW, Australia	http://www.lpi.nsw.gov.au/
8	City of Ryde Council- NSW, Australia	http://www.ryde.nsw.gov.au/
9	Office of Environment and Heritage - NSW, Australia	http://www.environment.nsw.gov.au/
10	GroundTruth Ltd. , Australia	http://groundtruth.com.au/
11	Ministry of Lands, Housing and Human Settlement Development, Tanzania	http://www.ardhi.go.tz/
12	The Arizona Geological Survey, US	http://www.azgs.az.gov/
13	Survey of Pakistan	http://www.surveyofpakistan.gov.pk/

There were 26 organisations who also participated in the evaluation of automatic spatial metadata enrichment prototype system. Table 4-7 illustrates the detail of these organisations. For this survey, out of these participating organisations only six of them first listed in Table 4-7 were engaged in the initial survey of the research.

Table 4-7: Organisations participated in the evaluation survey of automatic spatial metadata enrichment prototype system

No.	Organisation Name	Organisation Website
1	Victorian Department of Primary Industries (DPI), Australia	http://www.dpi.vic.gov.au/
2	Victorian Department of Sustainability and Environment (DSE), Australia	http://www.dse.vic.gov.au/
3	PSMA Australia Limited, Australia	http://www.psma.com.au/
4	Tasmanian Department of Primary Industries, Parks, Water and Environment, Australia	http://www.dpiw.tas.gov.au/
5	Department for Transport, Energy and Infrastructure, South Australia	http://www.dpti.sa.gov.au/
6	VicRoads, Australia	http://www.vicroads.vic.gov.au
7	Victorian Department of Planning and Community Development, Australia	http://www.dpcd.vic.gov.au/
8	Land and Property Information - NSW, Australia	http://www.lpi.nsw.gov.au/
9	City of Ryde Council- NSW, Australia	http://www.ryde.nsw.gov.au/
10	Office of Environment and Heritage - NSW, Australia	http://www.environment.nsw.gov.au/
11	GroundTruth, Australia	http://groundtruth.com.au/
12	Ministry of Lands, Housing and Human Settlement Development, Tanzania	http://www.ardhi.go.tz/
13	The Arizona Geological Survey, US	http://www.azgs.az.gov/
14	Survey of Pakistan, Pakistan	http://www.surveyofpakistan.gov.pk/
15	Esri, US	http://www.esri.com/
16	GEOCAT, US	http://geocat.net/
17	Teck Australia	http://www.teck.com
18	National Oceanic and Atmospheric Administration (NOAA)- US Department of Commerce	http://www.noaa.gov/
19	GeoMaxim, US	http://www.geomaxim.com/
20	Geobeyond Srl, Italy	http://www.geobeyond.it/
21	E.T.S. Topography, Geodesy and Cartography, Universidad Politecnica de Madrid, Spain	www.upm.es
22	Badan Pengkajian dan Penerapan Teknologi (BPPT), Indonesia	http://www.bppt.go.id/
23	Bureau of Land Management, US Department of Interior	http://www.blm.gov
24	Australian Government Bureau of Meteorology	http://www.bom.gov.au/
25	National Snow and Data Centre, US	http://nsidc.org/
26	University of Colorado, US	http://www.colorado.edu/

➢ Ethics Considerations

Similar to the initial survey of the research, appropriate ethical approval to conduct the human research was gained through the University Ethics Committee. The information from the questionnaires remained confidential and was utilised for research purposes only.

➢ Prototype Systems Communication

The automatic spatial metadata updating and enrichment prototype systems were built up on a desktop computer within the University's IT network. Therefore, these prototype systems could only be accessible from inside this network. However, in order to expand the evaluation to other users outside the University's network, a detailed video was captured from each prototype system and its functionalities. The videos were also enriched with voice narration and graphical comments. Both videos were published on youtube[17] and embedded in the Web-based questionnaires (Figure 4-6).

Figure 4-6: A screenshot of the Web-based questionnaire enriched with the video for evaluating the automatic spatial metadata enrichment prototype system

Using the video technology for presenting the prototype systems resulted in increasing the number of participants and involving more people from different countries in the evaluation process. As a result, to evaluate the automatic spatial metadata updating prototype system, other than with the directly invited participants, some other people participated in the evaluation process from Australia, United States, Tanzania and Pakistan. Similarly, in addition to directly invited participants, a number of other people from different countries including Australia, United States, Italy, Spain, Indonesia, Pakistan, and Tanzania participated in the evaluation of automatic spatial metadata enrichment prototype system. Accordingly, the prototype systems representing the spatial metadata automation approaches designed in this thesis could be

[17] www.youtube.com

evaluated, not only in the case study region – Australia, but also internationally. The results of participants' responses analysis will be discussed in Chapter 8.

4.5 Chapter Summary

This chapter charted the development of the research design and presented the methods utilised for this research. In this regard, the chapter described the conceptual design framework for the research by investigating the challenges identified during the literature review and the research questions. Then, the possible research methods available to answer the research questions were explored. Also, the research methods (qualitative and quantitative) were examined and thus the most suitable approach, the mixed-methods approach, was selected and justified to answer the research questions. As part of this approach, the case study strategy and Australia as the case study region were chosen to study the current status of spatial metadata management and identify the metadata automation requirements. Also, the System Usability Scale (SUS) survey was selected to evaluate the prototype systems usability.

In addition, the chapter developed the research design including four main phases: conceptual, design, implementation, and evaluation. The research methods utilised in each phase were presented. In addition, the prototyping technique was selected and justified for proving the concepts designed in this research. Moreover, two surveys were designed and implemented to gather required data for the research. The first survey was undertaken to gather the case study data based on a Web-based questionnaire. The second survey was executed to evaluate the developed prototype systems through two Web-based questionnaires. The ethics consideration and prototype systems communication to undertake the surveys were also reviewed.

Next chapter will investigate the results of the case study investigations and presents the current status of spatial metadata management in the case study region – Australia. Also, this chapter will explore the results of assessing the selected spatial metadata management tools against the criteria suggested in the research design.

CHAPTER 5

SPATIAL METADATA MANAGEMENT AND AUTOMATION CURRENT STATUS

5 SPATIAL METADATA MANAGEMENT AND AUTOMATION CURRENT STATUS

5.1 Introduction

This chapter presents the results of the qualitative case study investigations for identifying the current status of spatial metadata management and also the metadata automation requirements in the context of Australia – the selected case study region. Moreover, the chapter explores a number of proprietary and open source spatial metadata management tools. These tools are examined against a set of criteria developed for this thesis. Finally, this chapter integrates the results of the case study investigations and tools assessment with the findings of the literature review chapters (Chapters 2 and 3) and summaries the main challenges which the geospatial community faces regarding metadata collection, creation, storage, updating and improvement.

5.2 Spatial Metadata Current Status – Australian Case Study

In order to address the second research objective, a qualitative case study was designed and implemented according to the research design described in Section 4.4, Chapter 4. This case study aimed at understanding the current status of spatial metadata management and metadata automation requirements within the case study region. Australia was selected as the case study region. Within this region, a number of organisations were identified to participate in the survey. In order to undertake the survey, a Web-based questionnaire (see **Appendix 3**) was designed in eight parts. The structure of the questionnaire is illustrated in Table 5-1.

Table 5-1: Structure of research questionnaire designed for case study investigations

Part	Topics Covered	Sub-topics Covered
1	Introduction	• Introduction to research and questionnaire
2	Participant's information	• Name, email address and organisation name
3	Organisational Spatial Data and Metadata Characteristics	• The number of spatial datasets • Activities regarding metadata for spatial datasets • Activities regarding spatial metadata lifecycle
4	Spatial Metadata Applications	• The main applications complementing data discovery, exploration and exploitation
5	Spatial Metadata Creation and Updating	• Metadata creation/updating methods • The amount and format of metadata created by external teams • Standards used for metadata creation • The formats of metadata storage • Existence of any automatic metadata creation application • The mechanism for updating spatial data • The updating time interval for spatial metadata • The responsible body for updating spatial metadata
6	Spatial Metadata Management Tools	• The tools used in metadata management
7	Spatial Metadata Sharing	• The possibility of sharing metadata with other parties • The amount of metadata accessed by other parties • The mechanisms used to publish spatial metadata
8	Spatial Metadata management Issues and Challenges/ Desirable Automation Workflow	• Current issues and challenges regarding metadata management • Desirable metadata automation workflow • Further metadata automation requirements

In total, 12 main identified organisations participated in the survey and responded to the questionnaire. The analysis of the survey results is provided in the next section.

5.2.1 Results of Case Study Investigations

The results of case study investigations are categorised and presented in the following six main categories. These categories have been formed based on the last six parts of the research questionnaire already illustrated in Table 5-1.

- **Organisational Spatial Data and Metadata Characteristics**

The survey instrument first gathered data on general characteristics of spatial data and metadata existing in the participating organisations. The study found that the number of spatial datasets maintained in the organisations varied from a few to millions (e.g. datasets maintained by the

Victorian Department of Sustainability and Environment (DSE)). The issue thus arose that in some organisations the datasets are too various to quantify; hence, managing these data and related metadata sets posed a real challenge. In this regard, maintaining the spatial data and its related metadata together, and in an automated manner, would potentially address this challenge.

Data was also gathered on activities regarding spatial metadata undertaken by participating organisations. As a result, Figure 5-1 illustrates the distribution of activities regarding metadata for spatial datasets in the participating organisations.

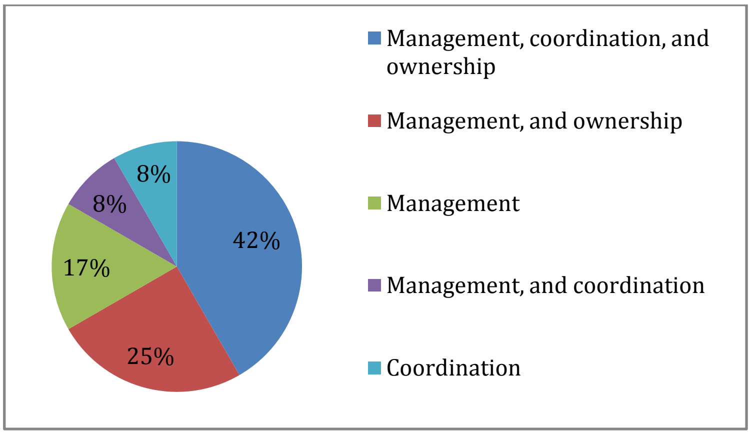


Figure 5-1: Distribution of activities regarding spatial metadata undertaken by participating organisations

As it can be seen in Figure 5-1, less than half of the participating organisations (42%) were dealing with spatial metadata management, coordination, and ownership; 25% were engaged in spatial metadata management and ownership activities. About 17% were only involved in management of spatial metadata; 8% only coordinated metadata; the other 8% managed and coordinated spatial metadata. From the analysis, it can be shown that management of spatial metadata is the common activity undertaken by most participating organisations (92%) internally.

In addition, data was collected on organisational activities regarding spatial metadata lifecycle (Figure 5-2). This lifecycle was already illustrated in Figure 2-6, Chapter 2. The results show that half of the participating organisations were involved in similar activities related to spatial metadata in terms of creation, updating, storage in clearinghouse and sharing with other organisations. It also showed 25% were dealing with metadata creation, updating and sharing; 9% were involved in metadata creation, updating and storing in clearinghouse and 8% only shared their metadata with others. The remainder (8%) did not answer the related question.

Although most of the participating organisations already indicated that metadata management was undertaken internally there were still multiple parties involved in creating, updating, storing and sharing metadata for the organisations. Coordination of the multiple parties to deliver the most up-to-date metadata when a dataset is created or updated would be a challenge.

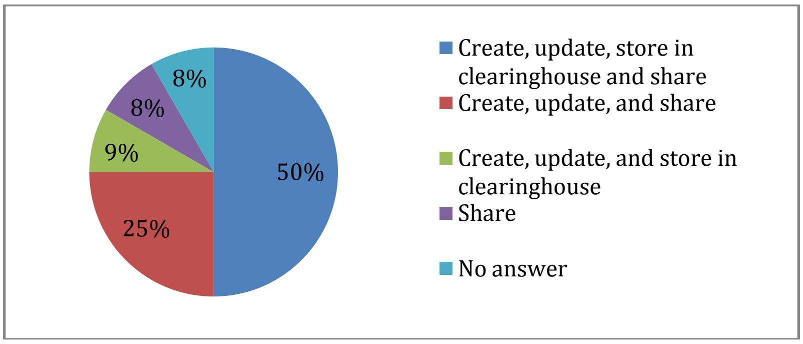


Figure 5-2: Distribution of organisational activities regarding spatial metadata lifecycle

- ## Spatial Metadata Applications

The participating organisations agreed that data 'discovery', 'exploration' and 'exploitation', acknowledged by Nebert (2004), are the main applications of spatial metadata according to their current practice. They also highlighted that metadata is expected to assist the spatial data management process through providing current information about data quality assurance and data maintenance (last and next update date/time). The participants reported that providing reliable and up-to-date metadata would be very critical for their organisations in order to limit their liability against the spatial data end users. Accordingly, focusing on keeping metadata current with any change to the dataset was realised as one of the main requirements of metadata automation research.

- ## Spatial Metadata Creation and Updating

The creation and updating approaches of spatial metadata were sought in the survey. As illustrated in Figure 5-3, no fully automatic approach was used by the participating organisations. Instead, 25% used the mixed manual/automatic (semi-automatic) approach. More importantly, most of them (75%) used the fully manual approach. This paucity of automation processes and tools points to the need for further research and development in the area of metadata automation.

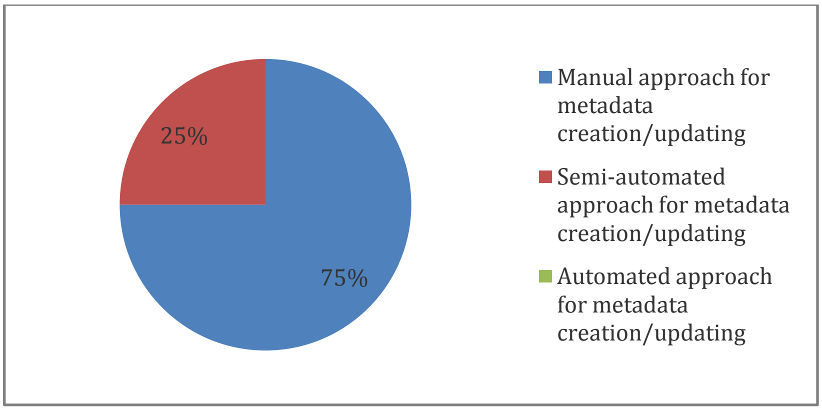


Figure 5-3: Distribution of metadata creation/updating approaches used by participating organisations

Furthermore, the participating organisations were asked to indicate the format(s) in which they receive metadata from external teams. As shown in Figure 5-4, 25% of organisations received metadata in both TXT and XML formats from external teams and 17% received metadata in TXT and PDF. Obtaining metadata in XML and PDF was reported by 8% of them; 8% indicated that they received metadata in PDF format. The figure also shows 8% collected metadata only in TXT format. Only 9% indicated that they received metadata in XML (the ISO/TS 19139: 2007 recommended metadata format). The remainder (25%) did not answer the related question.

These results showed that there was limited consistency across the industry to deliver spatial metadata and that it may produce future problems when sharing these data through semi-automated or automated workflows and business processes.

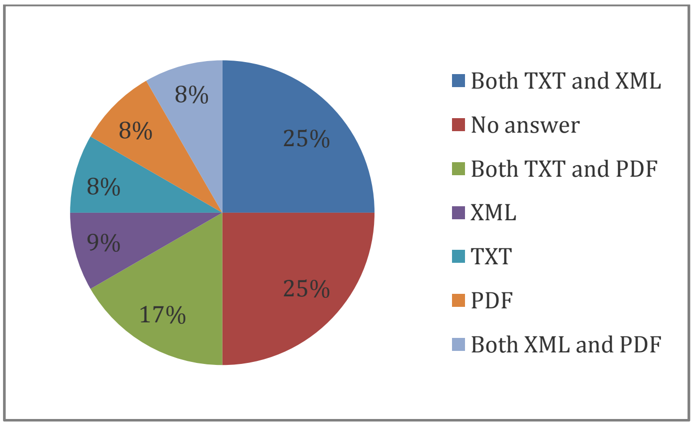


Figure 5-4: Distribution of metadata formats received by participating organisations from external teams

Moreover, data was gathered in relation to the metadata standards used by the organisations. The results are shown in Figure 5-5. Half of the organisations created metadata based on the ANZLIC Profile version 1.1:2007. The figure shows, 25% used both the ANZLIC Metadata

Guidelines V.2: 2001 and the ANZLIC Profile version 1.1: 2007. Both the ANZLIC Metadata Guidelines V.2: 2001 and the ISO 19115: 2003 were used by 9%. Also, 8% only used the ANZLIC Metadata Guidelines V.2: 2001. The remainder (8%) used an in-house standard based on FGDC and ANZLIC Metadata Guidelines V.2: 2001.

Although the spatial metadata standards have some considerable overlaps (e.g. page 0 of the ANZLIC Metadata Guidelines V.2: 2001 and core elements of the ANZLIC Profile version 1.1: 2007), the results confirmed that different standards are used by the participating organisations in Australia for the purpose of metadata creation regardless of recent efforts of ANZLIC. As discussed in Section 2.7.7 in Chapter 2, automatic metadata generation can be facilitated when the structure of metadata is based on a selected standard. As a result, considering a common standard in which most organisations agree on might be one of the more fruitful areas in the metadata automation research.

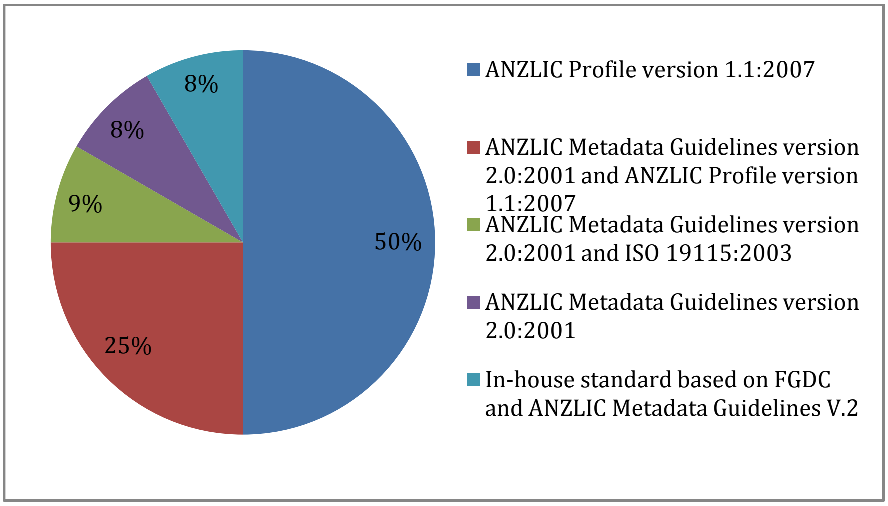


Figure 5-5: Distribution of metadata standards used by participating organisations

The participating organisations also reported the format(s) used to store spatial metadata (Figure 5-6). Among these formats, XML was reported by 42% as the main format for storing metadata. The combination of TXT, HTML, and XML for storing metadata was reported by 17% of them. The combination of TXT and XML was also reported by 9%. HTML and XML were the main formats used by 8% and another 8% noted that they used MS Excel for storing metadata. A further 8% used RDBMS for metadata storage and the combination of TXT, PDF, and XML for storing metadata was reported by 8% of them.

The format of spatial metadata creation and storage is a key area considered for metadata automation research in terms of technological aspects. According to Figure 5-6, it can be understood that a wide range of spatial metadata storage formats was used by the participating organisations. Among these data formats a high percentage of organisations using XML illustrated a good potential for metadata interoperability and compliance with standards in Australia. On the other hand, following the results of designing and developing an integrated data model for spatial data and metadata storage in this thesis, which will be discussed in Section 6.4.2 in Chapter 6, storing spatial metadata in a relational database and then exchanging them in XML (using the CSW service) could result in real-time spatial data and metadata updating. Therefore, a relational database to store the ISO 19115: 2003 metadata elements has been identified as one of the requirements for automating the spatial metadata updating (see Figure 6-9 in Chapter 6).

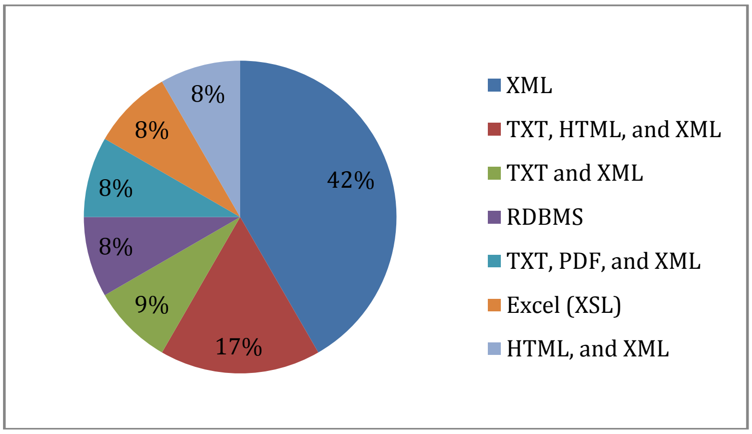


Figure 5-6: Distribution of formats used by participating organisations to store spatial metadata

Furthermore, the participating organisations were asked about the responsible party for updating spatial data. As shown in Figure 5-7, 59% of them reported that an internal team in the organisation is responsible for datasets updating; 17% indicated that an internal team together with an external team were in charge of datasets updating; and 8% mentioned that an external team updates datasets. Also, 8% reported that no role for updating datasets was defined in the organisation. The remainder (8%) did not answer the related question. The results showed that updating spatial data internally is dominant within the organisations surveyed in Australia.

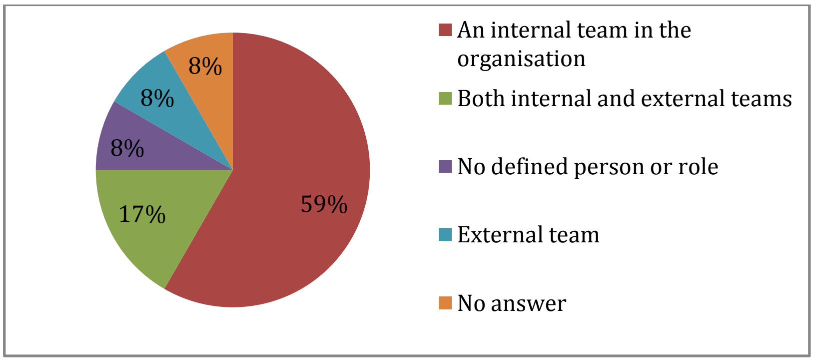


Figure 5-7: Distribution of responsible parties for dataset updating in the participating organisations

Another significant finding in this survey was about the relationship between time of updating spatial datasets and metadata. As illustrated in Figure 5-8, half of the participating organisations reported that they update metadata and spatial data at different times and half reported that they carry out the dataset and metadata updating processes simultaneously.

As discussed in Section 2.7.1 in Chapter 2, updating metadata at a later time than the actual dataset requires considerable effort and not all the necessary information might be available. Also, the end users need to access the most up-to-date, reliable and precise metadata for the data discovery purpose. This further points to the importance to define an approach to update spatial data and metadata simultaneously within organisations. Defining this approach has been considered as one of the requirements of spatial metadata automation research.

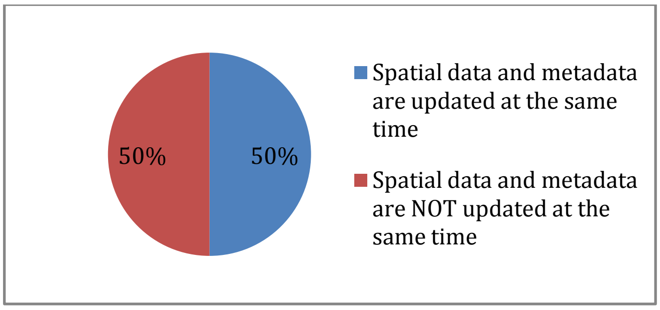


Figure 5-8: Distribution of participating organisations regarding real-time spatial data and metadata updating

Also, according to Figure 5-9, 67% of participants noted that the same team was responsible for updating spatial data and metadata and the remainder (33%) reported that the responsible teams differ. The separation of updating teams would make it longer to update metadata with any change in spatial data. However, if there was an automated approach for updating metadata at the same time with dataset modification, the metadata could be updated by the same responsible party in charge of dataset maintenance.

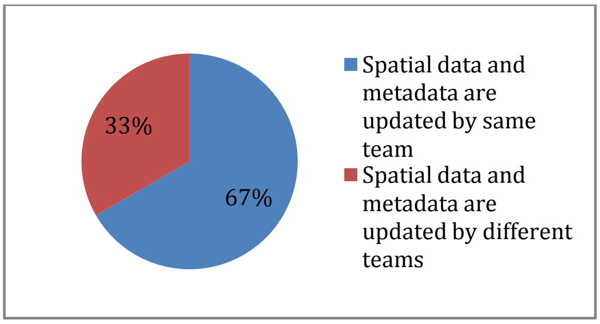


Figure 5-9: Distribution of organisations regarding the responsible parties for spatial data and metadata updating

- **Spatial Metadata Management Tools**

The detail of spatial metadata management tools used by the participating organisations was also requested in the survey. The tools reported by them are illustrated in Figure 5-10. The results showed that 34% of participating organisations were using their in-house applications for metadata management. The examples of these applications include MetaShare, and MS Excel Macros. The figure shows 17% reported that they used both the GeoNetwork and ANZMet Lite for metadata management purpose and 8% of organisations indicated that they only use the GeoNetwork. Also, 8% only used the ANZMet Lite and 8% reported using the BlueNetMEST. A further 8% used only the ESRI products (e.g. ArcCatalag and Geoportal Extension). Using both the ESRI products and the ANZMet Lite was also reported by 9%. The remainder (8%) did not answer the related question. The results show that there were a variety of tools used by different organisations surveyed in Australia and confirmed the paucity of a comprehensive spatial metadata management tool which can meet the needs of different organisations.

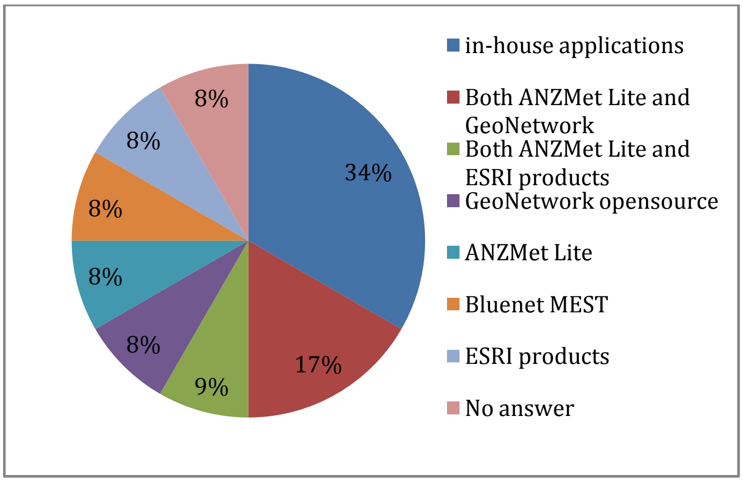


Figure 5-10: Distribution of metadata management tools used by the organisations

- **Sharing Spatial Metadata**

The participating organisations were asked questions related to sharing spatial metadata with other parties. All of them agreed that they share metadata of resources with other parties.

The data was also collected on the percentage of metadata that currently can be accessed by other parties. As illustrated in Figure 5-11, the survey showed that despite all the participating organisations' agreement on metadata sharing, only 42% currently share their whole metadata with others; 8% of them share 80%; 17% share 70%; 8% share 50%; 8% share 20%; and 17% share no metadata.

Referring to the main objectives of spatial metadata discussed in Section 2.3 in Chapter 2, it can be understood that metadata needs to be available online in order for users to discover and access the existing datasets. However, according to the results of the survey the participating organisations in Australia still needed automated approaches as part of the metadata management process to speed up the publication of metadata for their spatial datasets.

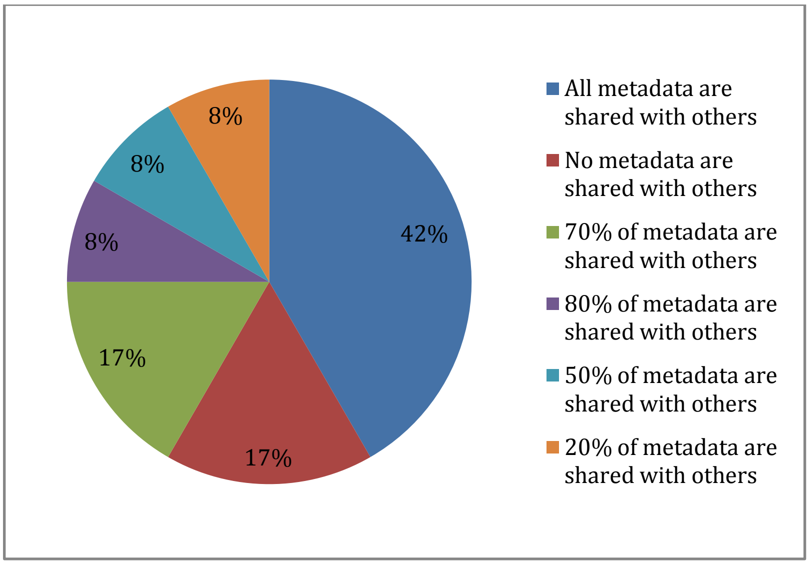


Figure 5-11: Percentage of metadata which is shared with other organisations

Once data was collected on the amount of shareable metadata, the mechanism to publish metadata was asked of the participating organisations. As shown in Figure 5-12, half of them reported that they publish metadata through a Web-based data catalogue (Australian Spatial Data Directory (ASDD) was introduced as this catalogue) and 25% of them indicated that they publish metadata through their organisations' Web pages. A further 17% reported using both Web-based data catalogues and Web pages. The remainder (8%) mentioned they use the ESRI products.

From the analysis, it can be surmised that most of the organisations are moving toward sharing spatial metadata on a national network using the ASDD platform. In order to publish spatial metadata to the ASDD, the organisations need to create metadata compliant with the ANZLIC metadata profile (based on ISO 19115: 2003 International Standard). Returning to the role that a selected standard can play in spatial metadata automation, discussed in Section 2.8 in Chapter 2, it can be realised that using a national platform for metadata publication would facilitate selecting the common standard within a community for metadata automation purpose.

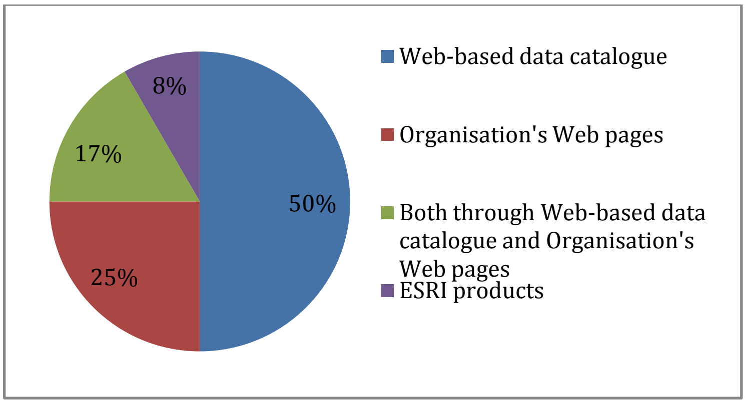


Figure 5-12: Distribution of metadata publication mechanisms used by participating organisations

- ## Current Spatial Metadata Management Issues and Challenges/ Desirable Automation Workflow

The participating organisations indicated main issues and challenges associated with managing spatial metadata. According to Figure 5-13, most of the participating organisations (92%) agreed on three main challenges including:

- Metadata authors are heavily involved in gathering information about metadata.
- Current metadata creation and updating are time consuming and labour-intensive.
- Lack of user-friendly tools available for cataloguing metadata.

Also, over half of them (59%) agreed on the following main challenges:

- complication of current metadata standards and data models
- ineffectiveness of existing semi-automatic metadata creation approaches in addressing the end users' needs
- lack of rigorous approaches for updating spatial data and metadata simultaneously.

Less than half of them (42%) also noted that involvement of multiple parties in the metadata creation and updating process would be a main challenge, which results in delay in updating metadata.

Also, 25% of them mentioned that generation and updating spatial metadata was expensive.

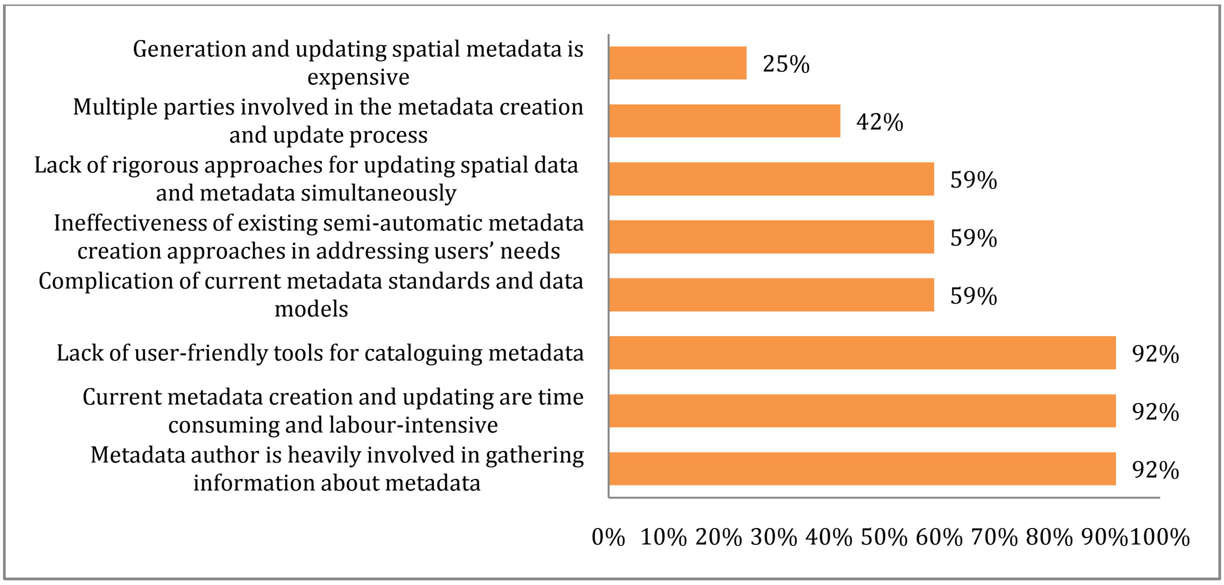


Figure 5-13: Spatial metadata management issues and challenges shared by the participating organisations

The participating organisations were requested to indicate the metadata automation workflow they prefer. As a result, according to Figure 5-14 most of them (92%) agreed that metadata should be generated automatically and then edited by a person. Only 8% preferred the fully automatic metadata automation workflow. This finding confirmed that the participating organisations preferred to assess the quality and completeness of metadata before storing the metadata values generated automatically. Therefore, in order to design and develop the spatial metadata automation approaches in this thesis, the ability to monitor and approve the metadata values created in an automated manner has been considered for metadata responsible parties.

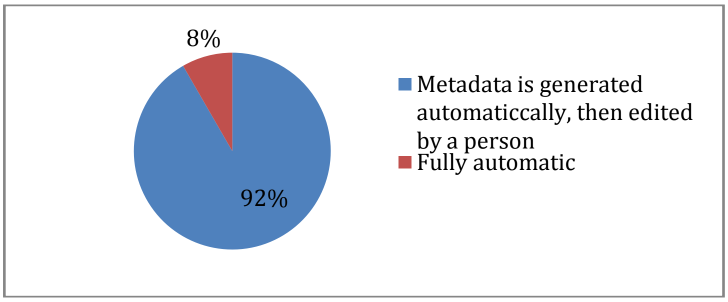


Figure 5-14: Distribution of preferred metadata automation workflow by participating organisations

Finally, the participating organisations shared their views regarding the requirements for spatial metadata automation approach. These requirements are here described:

- Creating metadata at the time of spatial data collection: would emphasise the need for accessing the metadata values that may not be accessible after the spatial dataset is created and published over the Web.

- Dynamic generation of metadata elements from a dataset (e.g. dataset extent): would not only be required for creation of non-existing metadata but also essential for updating the metadata values affected by the dataset modification.
- Storing and delivering metadata with the dataset: would emphasise the role of an integrated data model for spatial data and metadata storage and exchange. Through this data model, datasets and metadata should be stored and delivered together and also updated in real time.
- Simultaneous updating of spatial data and its related metadata: would require the metadata values influenced by the dataset modification to be synchronised with any change to the dataset.
- Advancing the data discovery process through automatic improvement of metadata content: since the value of metadata keyword element is mainly used for spatial data discovery in catalogues; this requirement in particular would emphasise the need for enriching the content of metadata keyword element over the time.
- Involving the capacity for object reuse (e.g. contact details): would illustrate the important role of predefined values for the metadata elements that cannot be automated.
- The ability to integrate with spatial data tools (e.g. Web-based GIS and data catalogue systems): would show the importance of adopting popular international standards and technologies for designing and developing spatial metadata automation tools.
- Compliance with relevant metadata schemas, especially ANZLIC and ISO/TS 191139: 2007 would increase the adoption rate of proposed spatial metadata automation approaches by different organisations within a community.

Following the research design (see Section 4.4 in Chapter 4), in the conceptual phase of the research, a number of spatial metadata management tools were selected to examine in addition to the case study investigations. Exploring these tools would provide proper information about the current spatial metadata management and automation capabilities. Moreover, this assessment would be a good basis for exploring how different tools address the challenges of spatial metadata management identified during the literature review (Chapter 2 and 3) and the case study (current chapter).

5.3 Assessment of Spatial Metadata Management Tools

Along with the worldwide research and development initiatives regarding spatial metadata automation, several proprietary and open source spatial metadata management tools have been designed and developed by the geospatial community. The main aim of these tools is to aid the metadata authors in metadata management through decreasing and minimising the human intervention and associated errors, and therefore increasing the quality of metadata.

According to the research design, a number of these tools were selected to examine against a set of criteria suggested for this thesis, in order to address the second research objective. A set of criteria was proposed to assess the selected tools in terms of five main aspects already described in Section 4.4.1 in Chapter 4. Table 5-2 outlines the criteria in more detail.

Table 5-2: Set of criteria proposed for examining the selected spatial metadata management tools

Criteria	Description
Support for the integration of metadata creation with spatial data lifecycle	Whether the tool provides the organisations with the facilities to generate metadata values in parallel with their spatial data lifecycle. This integration could result in reducing the burden of metadata creation for metadata authors through associating the generation of each metadata element to its relevant responsible party. Also, the integration could have the potential to overcome the problem of missing or incomplete metadata through recognising the stage to generate and update metadata within the data lifecycle.
Support for the integrated data model	Whether the tool provides an integrated data model for storing spatial datasets and metadata. This data model should allow datasets and metadata to be coupled in a common environment, which could result in managing and maintaining them together. Integrated data model was already discussed in Section 2.6.2, Chapter 2.
Support for automatic metadata creation	Whether the tool supports the automatic generation of spatial metadata values using different sources such as dataset file and pre-defined metadata. This could result in saving required time and resources for creating metadata and also increase the quality of metadata by reducing the human error.
Support for automatic metadata updating when dataset changes	Whether the tool automatically synchronises the metadata with any changes to the dataset in real time. This could result in being the spatial metadata always up-to-date.
Support for interaction with end users to improve the content of metadata	Whether the tool engages end users to improve the content of metadata during the data discovery process. For instance, the interaction might include the functionalities for tagging datasets with new search words or commenting on the datasets. Interaction with end users could result in connecting the end users to metadata creation and maintenance process, facilitating the discovery process by involving end users' knowledge of datasets, and making the metadata management tool more user-friendly. The end user here means people seeking spatial datasets and not the spatial data cataloguers.

The spatial metadata tools selected to be reviewed and examined against the criteria are categorised in two main groups, international and Australian.

5.3.1 International Tools

The results of examining the selected international metadata tools including GeoNetwork opensource, GeoNode, ESRI ArcCatalog, CatMDEdit, and EUOSME are discussed in this section.

- **GeoNetwork opensource**

GeoNetwork opensource has been developed following the principles of a Free and Open Source Software (FOSS) and based on International and Open Standards for services and protocols, such as the ISO/TC211 and the OGC specifications. GeoNetwork is a catalogue application to manage spatially referenced resources through the Web, designed to enable access to geo-referenced databases, cartographic products and related metadata from a variety of sources. It is currently used in numerous SDI initiatives across the world (GeoNetwork 2011c) and it is one the most commonly used metadata entry and discovery tools in Australia (Olfat *et al.* 2010a).

The prototype of the GeoNetwork catalogue was developed by the Food and Agriculture Organisation of the United Nations (FAO) in 2001 to systematically archive and publish the geographic datasets produced within the organisation (GeoNetwork 2012b). However, the GeoNetwork is the result of the collaborative development of many contributors. These include among others the FAO, the UN Office for the Coordination of Humanitarian Affairs (UNOCHA), the Consultative Group on International Agricultural Research (CSI-CGIAR), the UN Environmental Programme (UNEP), the European Space Agency (ESA) and many others.

The GeoNetwork architecture is largely compatible with the OGC Portal Reference Architecture, i.e. the OGC guide for implementing standardised geospatial portals. Indeed the structure relies on the same three main modules identified by the OGC Portal Reference Architecture that are focused on spatial data, metadata and interactive map visualisation. The system is also fully compliant with the OGC specifications for querying and retrieving information from Web catalogues (CSW). It supports the most common standards to specifically describe geographic data (ISO/TS 19139: 2007 and FGDC) and the International Standard for general documents (Dublin Core).

Figure 5-15 illustrates the user interface of the GeoNetwork. The main features of this catalogue are as follows:

- instant search on local and distributed geospatial catalogues

- uploading and downloading of data, documents, PDF's and any other content
- an interactive Web map viewer that combines Web Map Services from distributed servers around the world
- online map layout generation and export in PDF format
- online editing of metadata with a powerful template system
- scheduled harvesting and synchronisation of metadata between distributed catalogues
- group and user management
- fine-grained access control.

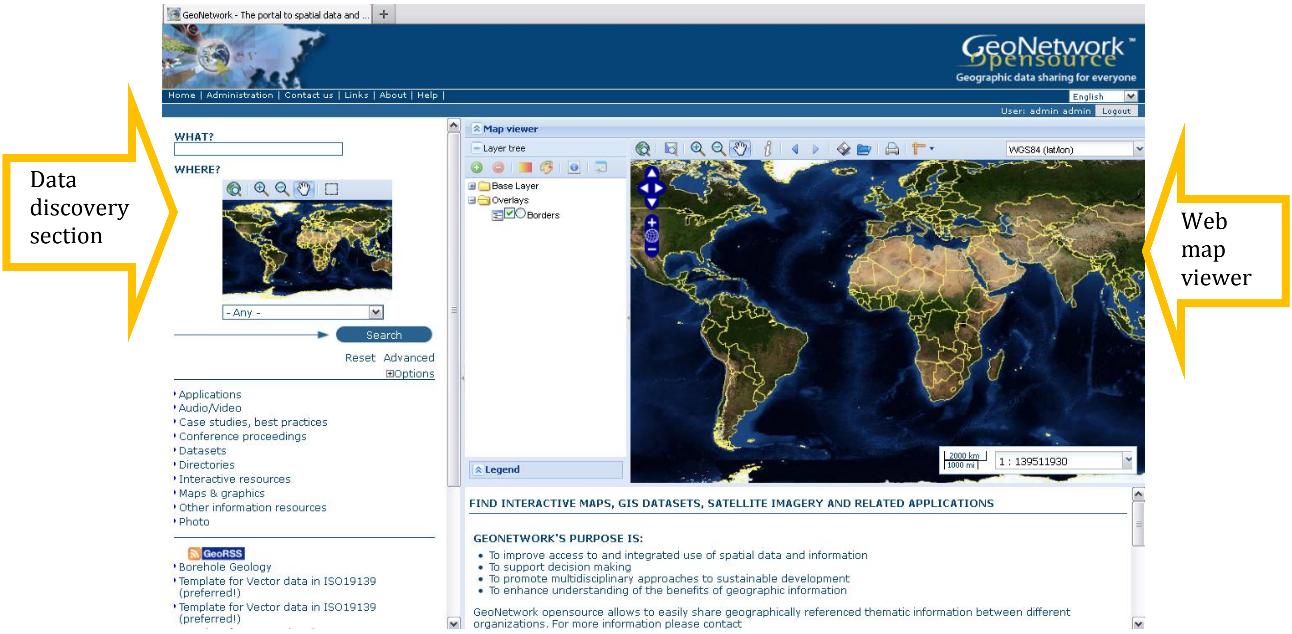


Figure 5-15: GeoNetwork opensource user interface

The results of examining GeoNetwork against the criteria outlined in Table 5-2 are discussed as following:

➢ Support for integration with spatial data lifecycle

The distribution of ISO 19115: 2003 metadata elements in GeoNetwork is based on ISO packages (metadata, identification, maintenance, constraints, spatial information, reference system, distribution, data quality, application schema, catalogue, content information, and extent information). GeoNetwork does not provide the spatial data cataloguers with any capability to distribute metadata elements according to different steps of the spatial data lifecycle. Therefore, the integration of metadata creation with the data lifecycle is not possible using this tool.

➢ Support for integrated data model

GeoNetwork installs with an internal DBMS[18] server, the McKoi SQL database. However, it has the capability to connect to other databases including Oracle, PostgreSQL, and MySQL. GeoNetwork stores the metadata records in XML format. An entire metadata XML string is stored in a single database column. In terms of relating the spatial dataset to metadata, GeoNetwork allows the spatial data cataloguers to provide required information for accessing the dataset when creating metadata. This information includes the URL, protocol (e.g. ArcIMS, WFS, KML[19]), and necessary description. It also enables uploading the dataset file for further downloads by end users. However, datasets and metadata are stored in separate environments and cannot be coupled in a middleware.

In addition, the GeoNetwork works with the Metadata Exchange Format (MEF), which is a special designed file format whose purpose is to allow metadata exchange between different platforms. A metadata exported into this format can be imported by any platform that is able to understand it. An MEF file is simply a ZIP file which contains the following files (GeoNetwork 2012a):

1. *metadata.xml*: this file contains the metadata itself, in XML format. The text encoding of the metadata is that one specified into the XML declaration.

2. *info.xml*: this is a special XML file which contains information related to the metadata but that cannot be stored into it. Examples of such information are the creation date, the last change date, and privileges on the metadata.

3. *public*: this is a directory used to store the metadata thumbnails and other public files. There are no restrictions on the images' format but it is strongly recommended to use the portable network graphics (PNG), the JPEG or the GIF formats.

4. *private*: this is a directory used to store all data (maps, shape files etc.) associated to the metadata. Files in this directory are private in the sense that an authorisation is required to access them. There are no restrictions on the file types that can be stored into this directory.

[18] Database Management System
[19] Keyhole Markup Language

Using this data format, although the spatial data and metadata are stored jointly within a ZIP file these two data cannot be managed and maintained concurrently.

As a result, the GeoNetwork follows the detached data model for storing dataset and metadata.

- **Support for automatic metadata creation**

The process of creating new records in the GeoNetwork can be undertaken through either the online metadata editor or the advanced metadata insert tool, based on XML documents (GeoNetwork 2011b). Metadata creation within the GeoNetwork is a manual process (except for computing bounding box from keywords, if the keyword exists in the thesaurus and has an extent) and metadata author is responsible to gather as much information as possible to create metadata.

However, GeoNetwork enables harvesting metadata records that results in the automatic creation of metadata from a remote node. In fact, harvesting is the process of collecting remote metadata and storing them locally for a faster access. This is a periodic process to do, for example, once a week. Harvesting is not a simple import: local and remote metadata are kept aligned. GeoNetwork node is capable of discovering metadata that has been added, removed or updated in the remote node. The following sources are used by the GeoNetwork harvesting mechanism to create metadata automatically:

- another GeoNetwork node
- WebDAV[20] server
- CSW 2.0.1 or 2.0.2 catalogue server
- OAI-PMH server
- OGC service using its GetCapabilities document (including WMS, WFS, WPS[21] and WCS[22] services)
- ArcSDE server
- THREDDS[23] catalogue
- OGC WFS using a GetFeature query
- Z3950 server.

For instance, when a WFS GetCapabilities document is harvested, some of the ISO 19115: 2003 compliant metadata elements including title, abstract, keywords, bounding box, metadata date

[20] Web Distributed Authoring and Versioning
[21] Web Processing Service
[22] Web Coverage Service
[23] Thematic Real-time Environmental Distributed Data Services

stamp, online resource URL, protocol and description can be generated automatically for the corresponding metadata records.

➢ Support for automatic metadata updating

The process of updating a metadata record is supported by the online metadata editor and the CSW ISO Profile test (test interface for the CSW ISO Profile Catalogue Interface, in version 2.6.4). According to Rajabifard *et al.* (2009), GeoNetwork lacks support for an automatic method to update metadata records in conjunction with the dataset modification. However, when metadata are harvested using this tool the new values of harvested elements (e.g. abstract, bounding box, and keywords for a WFS node) can be updated automatically in the database.

➢ Support for interaction with end users

The end users are able to discover, retrieve, and access metadata records using the GeoNetwork catalogue. There is no capability provided by this tool for more interaction with the end users to create, update, or improve the content of metadata. Only the spatial data cataloguers (authorised users) are allowed to manipulate or improve metadata content.

• GeoNode

GeoNode is an open source platform that facilitates the creation, sharing, and collaborative use of spatial data. It allows the upload vector and raster data in their original projections using a Web form. Vector data is uploaded in ESRI Shapefile format and satellite imagery and other kinds of raster data are uploaded as Geotiffs. After the upload is finished, the user is presented with a form to fill in the metadata based on ISO 19115: 2003 and it is made available using a CSW interface. GeoNode also provides a Web-based styler, that lets the user change how the data looks and preview the changes in real time. Once the data has been uploaded, GeoNode lets the user search for it geographically or via keywords and create maps. When the maps are saved, it is possible to embed them in any Web page or get a PDF version for printing.

The results of examining GeoNode against the criteria outlined in Table 5-2 are discussed as follows:

➢ Support for integration with spatial data lifecycle

GeoNode does not provide any facilities to integrate metadata creation with the spatial data lifecycle. As already mentioned, metadata creation option would be available once a dataset is uploaded. This means that GeoNode does not deal with the steps in the spatial data lifecycle and only creates metadata based on ready-to-use datasets uploaded.

➢ Support for integrated data model

Figure 5-16 illustrates the architecture of the GeoNode.

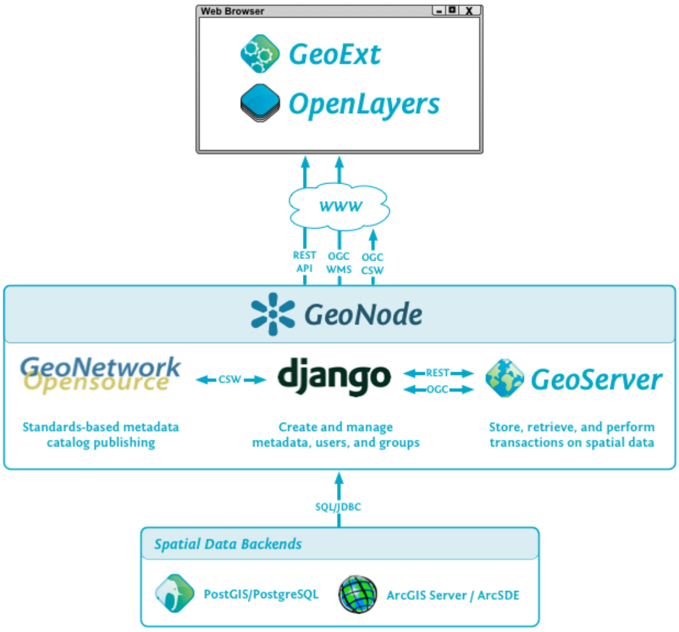


Figure 5-16: The architecture of the GeoNode (GeoNode 2012a)

Based on this architecture, GeoNode stores and queries information from both the GeoNetwork and the GeoServer, and will act as the integration layer between the GeoServer datasets and their corresponding metadata records in the GeoNetwork. Although, spatial data and its corresponding metadata are associated to each other using the GeoNode, they cannot be published in a middleware and maintained together. Therefore, this tool is rooted in a detached data model for storing metadata and datasets.

➢ Support for automatic metadata creation

GeoNode allows the creation of metadata elements based on the ISO 19115: 2003 Standard including Title, Date, Date Type, Edition, Abstract, Purpose, Maintenance Frequency, Keywords, Keywords Region, Constraints Use, Other Constraints, Spatial Representation Type, Language, Temporal Extent End, Geographic Bounding Box, Supplemental Information, Distribution URL, Data Quality Statement, Point of Contact, and Metadata Author. Of all these elements, the GeoNode creates date (date/time of upload), Bounding Box (calculated using the GeoServer), distribution URL, and contact information (using the user's profile) automatically once the dataset is uploaded.

➢ Support for automatic metadata updating

When the dataset is modified, its new version can be uploaded to the GeoNode. Once the file is uploaded, similar to the creation of metadata the user is presented with a Web form to update the metadata values. However, this time all the values need be updated manually if there is any update available. Therefore, at this stage there is no support for automatic metadata updating in the GeoNode when a dataset is modified. According to the GeoNode's Wiki (2012b), updating the bounding box metadata element automatically, whenever a dataset is modified, is considered as the future plan for the GeoNode development.

➢ Support for interaction with end users

GeoNode allows the end users to discover the existing datasets. Once any dataset is retrieved within the discovery process, the end user can see its corresponding metadata record. However, there is no support for the end users to interact with the discovery process to refine metadata.

• ESRI ArcCatalog

The ArcCatalog application is an ESRI product that provides a catalogue window to organise and manage various types of geographic information for ArcGIS Desktop. The kinds of information that can be organised and managed in ArcCatalog include but are not limited to:

- Geodatabases:
    - File geodatabase – a folder full of files on disk
    - Personal geodatabase – a Microsoft Access database file (.mdb); and
    - DBMS such as Oracle, SQL Server, Informix, DB2, or PostgreSQL.
- Raster files
- Map documents, globe documents, 3D scene documents, and layer files
- Geoprocessing toolboxes, models, and Python scripts
- GIS services published using ArcGIS Server
- Standards-based metadata for above GIS information items.

The metadata can be published through ArcCatalog using the ESRI ArcGIS Server Geoportal extension Publish Client. In fact, the Geoportal extension is a fully functioning Website with a set of application components that allow the publication, discovery and use of standards-based metadata documents based upon known XML metadata standards. The standards recognised and supported by default include the ISO/TS 19139: 2007, ISO 19119: 2005, Dublin Core, and FGDC (ESRI 2009).

The results of examining ArcCatalog against the criteria outlined in Table 5-2 are discussed as follows:

- **Support for integration with spatial data lifecycle**

ArcCatalog does not provide spatial data cataloguers with any facilities to integrate metadata creation with the data lifecycle. In terms of metadata management, this tool mainly aims to synchronise metadata with the latest status of existing 'dataset' as the final output of spatial data lifecycle. Accordingly, the metadata elements, which are not synchronised automatically (e.g. resource quality, point of contact, and abstract), need to be created and maintained manually by metadata author or responsible party.

- **Support for integrated data model**

Metadata is part of an ArcGIS item (ESRI 2012). When the item is copied in ArcGIS, its metadata is also copied. When the item is imported into a geodatabase, its metadata is also imported. Metadata is stored in the same location as the item's data in a manner that is appropriate for its data type. For instance, metadata is stored in the same location on disk as a shapefile's data in an accompanying XML file (file format is '.shp.xml'). For geodatabase items, metadata is stored in the geodatabase system tables. For an enterprise geodatabase, metadata is stored in the Documentation column in the GDB_Items table. Although metadata and datasets are stored in the same place (folder or database) through the ArcCatalog their integration and presentation in a common environment, is not supported by this tool.

- **Support for automatic metadata creation**

ArcCatalog special objects named metadata synchronisers are used to write metadata when synchronisation occurs. Synchronisation is the process by which properties of a dataset are read from the dataset and written into its metadata. The purpose of synchronisation is twofold:

- To complete as much of the metadata as possible automatically, thereby minimising the amount of work the user has to do to completely document a dataset
- To make sure the metadata is kept up-to-date with changes to the dataset.

Within ArcCatalog, synchronisation can be triggered by any of the following:

- When a user browses to a dataset in ArcCatalog with the Metadata tab selected and that dataset does not have metadata, synchronisation is triggered, and metadata is created for the dataset.

- When a user browses to a dataset in ArcCatalog with the Metadata tab selected and that dataset has metadata, synchronisation is triggered, and metadata is updated for the dataset.

- When a user clicks the Create/Update metadata button in ArcCatalog, synchronisation takes place and the metadata for the selected dataset is created (if metadata does not exist), or updated (if it exists).

In this regard, ArcCatalog comes shipped with three metadata synchronisers: the FGDCSynchroniser CoClass, which writes metadata according to the FGDC standard; the GNSynchroniser CoClass, which writes metadata according to the ISO standard; and the GNSynchroniser CoClass, which is used to synchronise metadata for the Geography NetworkSM. Figure 5-17 shows how the synchronisation manager works (ESRI 2002a).

Table 5-I in **Appendix 6** also summarises the objects that are passed into the Update method by the synchronisation manager. In this Table, the 'Value' column lists the type of the object passed in. The 'itemDesc' column lists the string identifier passed in. When working with the Update method, the string in the 'itemDesc' column is used to identify the object in the 'Value' column. Examples of synchronised metadata are information about dataset location, name, size, language, extent, and metadata date.

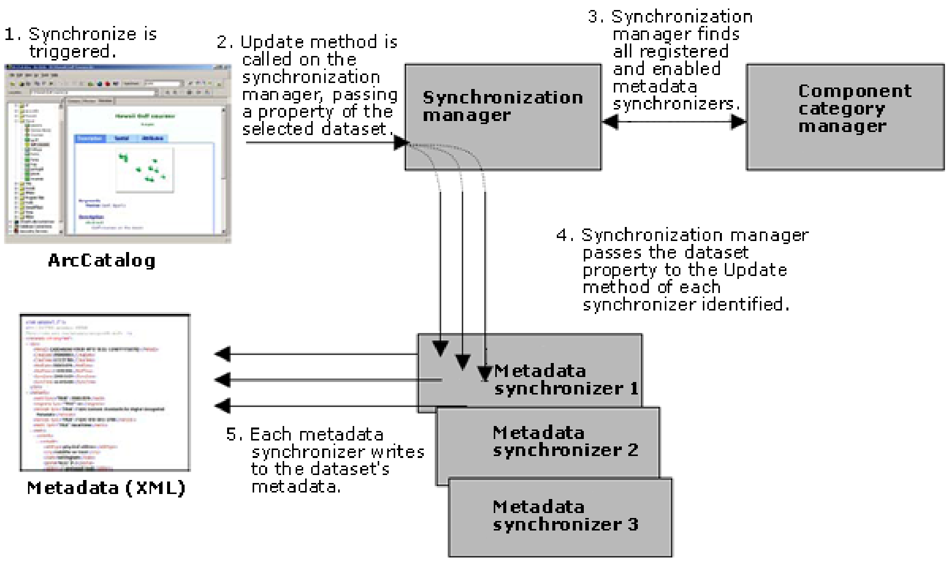


Figure 5-17: Workflow of ArcCatalog synchronisation manager (ESRI 2002a)

➢ Support for automatic metadata updating

As already mentioned, the synchronisation process works for both creation and updating of metadata. By default, metadata is automatically updated when anyone who has write access to the ArcGIS item views its metadata using ArcCatalog. In other words, synchronisation always occurs in combination with viewing metadata (ESRI 2010) which requires human intervention after dataset modification. As a result, if someone edits an ArcGIS item's data or changes its properties and does not view the metadata, the properties recorded in the metadata are out of date with the actual properties of the item. If this item's metadata was published in this state the information in the metadata would not be current.

➢ Support for interaction with end users

ArcCatalog is a desktop application and its end users would be the spatial data cataloguers and not the public users of a data discovery system. Therefore, no interaction with the users looking for data could be expected from this tool.

• CatMDEdit

CatMDEdit is a metadata editor tool that facilitates the documentation of resources, with special focus on the description of geographic information resources. It is an initiative of the National Geographic Institute of Spain, which is the result of the scientific and technical collaboration between IGN and the Advanced Information Systems Group (IAAA) of the University of Zaragoza with the technical support of GeoSpatiumLab (GSL) (CatMDEdit 2011a).

CatMDEdit supports metadata creation and edition in conformance with the ISO 19115: 2003 Standard. The editor also supports separate edition of subsets of the ISO 19115: 2003 metadata elements that correspond with the ISO19115, ISO19115 NEM, ISO19115 SDIGER/WFD, ISO19115 SDIGER/INSPIRE and the Dublin Core Metadata Element Set.

Iso *et al.* (2008) present the high-level architecture of CatMDEdit components. This architecture, as shown in Figure 5-18, includes three layers. The upper-layer contains the three main functional components of the application: a Resource Browser, a Metadata Editor, and a Resource Viewer. The second layer contains the middleware software libraries that support the development of the functional components. And the data layer includes the different repositories needed for the configuration of the tool and the storage of data and metadata, which are managed or produced by the application.

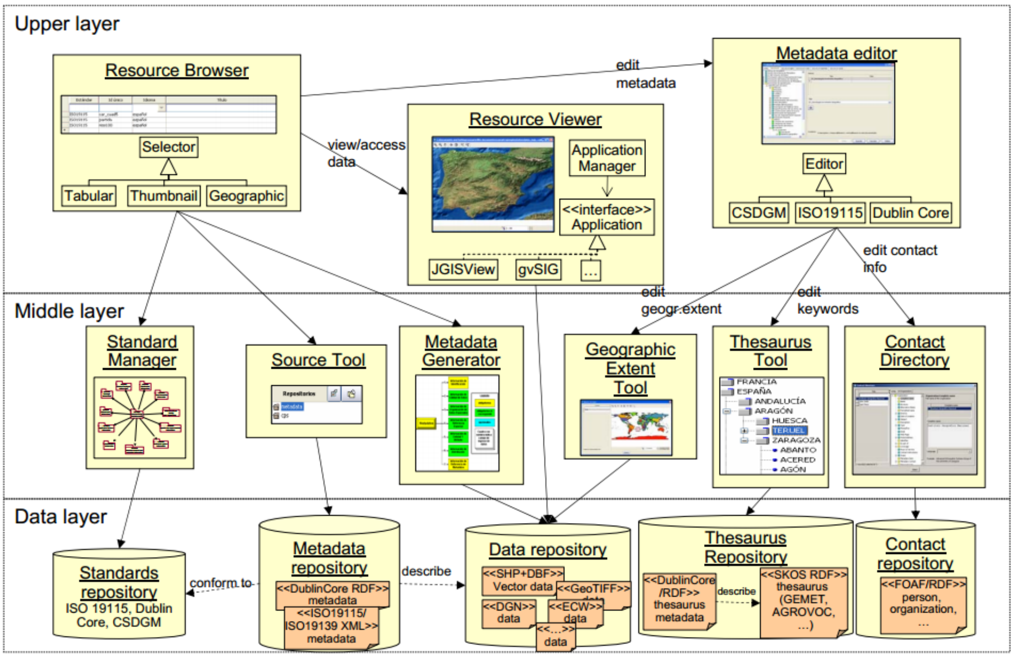


Figure 5-18: High-level architecture of CatMDEdit components (Iso *et al.* 2008)

In order for unifying data and metadata management, CatMDEdit is connected with the open source tool gvSIG for the presentation of geographic information; gvSIG is an Open Source GIS tool that arose in 2004 as a GIS client tool to cover the needs of spatial data users working at Generalitat Valenciana (government of the autonomous community of Valencia, Spain). Once metadata of a resource is generated (manually or via existent semi-automatic mechanism for some specific spatial data formats), CatMDEdit auto-generates the configuration file of a gvSIG project and directly opens this tool to have access to the resource with the adequate viewing parameters (e.g. coordinate reference system and geographic extent).

The results of examining the CatMDEdit against the criteria outlined in Table 5-2 are discussed as follows:

➢ Support for integration with spatial data lifecycle

The CatMDEdit does not provide spatial data cataloguers with any facilities to integrate metadata creation with the data lifecycle. This tool focuses on the dataset as the output of data lifecycle to create metadata in a semi-automatic way.

➢ Support for integrated data model

Following the CatMDEdit architecture already illustrated in Figure 5-18, dataset and metadata are stored in separate repositories in a data layer. Therefore, this tool is based on a detached data model for storing dataset and its associated metadata record and there is no possibility to integrate and maintain metadata and dataset in a common environment using this tool.

➢ Support for automatic metadata creation

The CatMDEdit automates metadata generation for some data file formats including: Shapefile, DGN, ECW, FICC, GeoTiff, GIF/GFW, JPG/JGW, PNG/PGW. In addition, it allows the automatic creation of metadata for collections of related resources, in particular spatial series arisen as a result of the fragmentation of geometric resources into datasets of manageable size and similar scale (e.g. mosaics of orthoimages and tiles 1:50,000). Moreover, it enables automatic metadata generation from the 'getCapabilities' operation supported by a service that complies with the OGC Specifications (WMS, CSW, WFS, WCS or WPS). According to the specific format, the tool generates the metadata record fields automatically as illustrated in Table 5-II in **Appendix 6**. Examples of derived metadata are information about spatial reference systems, number and type of geographic features, extension covered by a dataset, or information about the entities and attributes of alphanumerical related data (Ballari *et al.* 2006).

➢ Support for automatic metadata updating

According to Rajabifard *et al.* (2009), CatMDEdit does not support automated updating of metadata whenever its associated spatial dataset is modified.

➢ Support for interaction with end users

The CatMDEdit end users are the spatial data cataloguers. Therefore, no interaction with public users seeking data over the Web could be expected from this tool.

• The European Open Source Metadata Editor (EUOSME)

The EUOSME is a Web application written in Java and based on Google Web Toolkit (GWT) libraries. Its main purpose is to help create metadata compliant with the INSPIRE Directive (2007/2/EC) and the INSPIRE Metadata Regulation (1205/2008). More specifically, this implementation allows describing a spatial dataset, a spatial dataset series or a spatial data service compliant with the standards ISO 19115: 2003 (corrigendum 2003/Cor.1:2006) and ISO 19119: 2005. It is therefore an implementation of the INSPIRE Metadata Technical Guidelines based on these two ISO Standards, and published on the INSPIRE Website. This editor builds on the experience acquired in the development of the INSPIRE Metadata Implementing Rules, and includes the INSPIRE Metadata Validator Service available from the INSPIRE EU Geoportal[24] (Grasso and Craglia 2010). Figure 5-19 illustrates the user interface of this metadata editor.

[24] http://inspire-geoportal.ec.europa.eu/

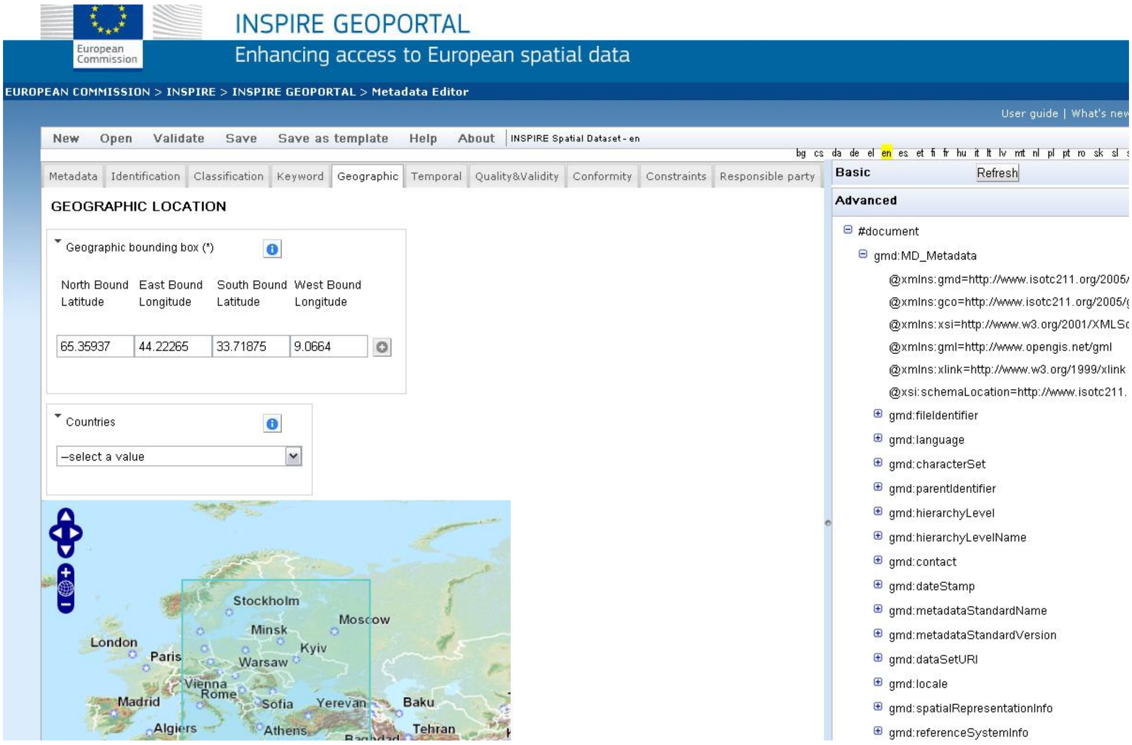


Figure 5-19: The EUOSME user interface

The results of examining the EUOSME against the criteria outlined in Table 5-2 are discussed as follows:

➢ Support for integration with spatial data lifecycle

The EUOSME is a Web form for creating and editing metadata records. The distribution of metadata elements within the form is based on the ISO 19115: 2003 metadata packages. This tool does not provide spatial data cataloguers with any facilities to integrate metadata creation with the data lifecycle.

➢ Support for integrated data model

The EUOSME provides metadata records in XML. The end user is able to store the XML metadata flat file on his personal computer once it is created or updated. Thus, the metadata record should be stored and maintained in a separate environment from its associated dataset without any defined relationship. This means that the EUOSME is rooted in a detached data model.

➢ Support for automatic metadata creation

There is no capability in the EUOSME for browsing and selecting or associating the actual spatial dataset related to metadata record. Therefore, there is no support for automatic creation of metadata values using this tool.

➢ Support for automatic metadata updating

The EUOSME does not support automatic metadata updating when a change occurs to the dataset.

➢ Support for interaction with end users

The EUOSME is part of the INSPIRE Geoportal. With the deadline of November 2011 for Member States to provide discovery and view services according to the INSPIRE Regulation on Network Services, a first release of the INSPIRE Geoportal was published (INSPIRE 2012). This Geoportal allows end users to search, discover and access geographic information provided by European governmental, commercial, and non-commercial organisations.

In addition to the basic and advanced search functions that normally exist in spatial data catalogues, the INSPIRE Geoportal provides end users with other two search functionalities: suggesting list of search words, and tag cloud. Figure 5-20 shows these two functionalities. The tag cloud here is a way of displaying the whole content within the 'Origin', 'Spatial Data Theme' and 'Topic Category' search criteria of the INSPIRE Geoportal. These criteria are also directly accessible within the data discovery system (see Figure 5-20). The number of results found for each tag is also assigned to it as its weight. For this reason, 'Poland' (included in the 'Origin' criteria) with 198192 found results, is shown bigger than 'Cadastral Parcels' (included in the 'Spatial Data Theme' criteria) with 21730 found results. The suggestion list functionality also lists the matching search word when the user starts typing in the search box. As shown in Figure 5-20, the number of results would be illustrated next to each search word as well.

Although, the tag cloud within the INSPIRE Geoportal facilitates the data discovery process, the approach and objectives of generating such a tag cloud is different from what is already discussed in Section 3.3.2 in Chapter 3 regarding the new potential application of tag clouds in the geospatial community. The tags within the INSPIRE Geoportal tag cloud are generated based on the content of metadata already created by the data responsible parties rather than the user-generated content. In fact, the end users are not allowed to improve the content of metadata records or tag them with the terms they desire.

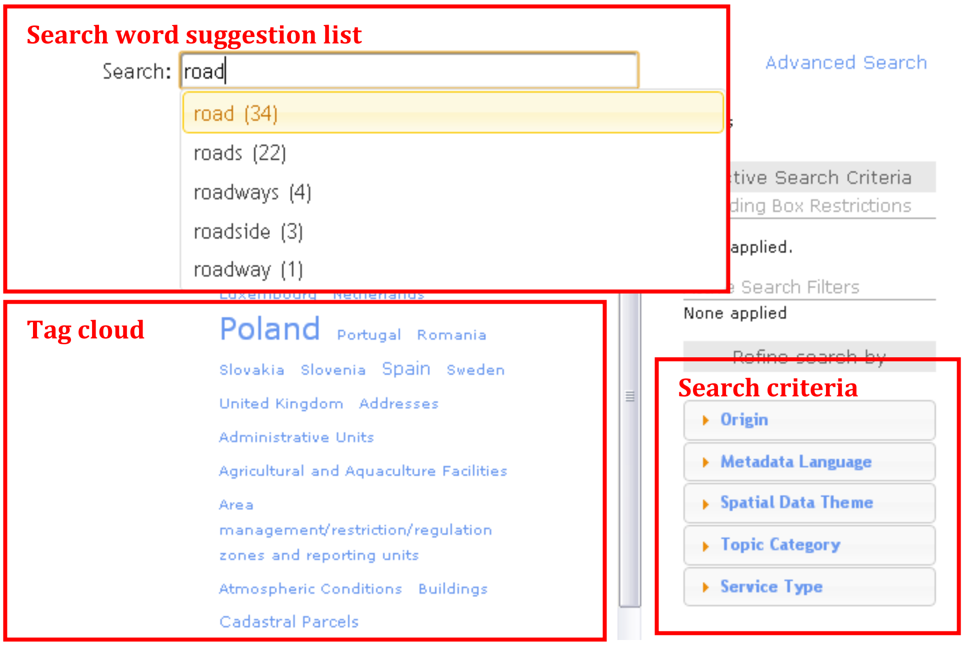


Figure 5-20: The INSPIRE Geoportal suggestion list and tag cloud functionalities

5.3.2 Australian Tools

This section examines the selected spatial metadata tools provided in Australia including the BlueNetMEST, ANZMet Lite and xMet Client against the criteria proposed for this thesis.

- **BlueNetMEST (Currently Known as ANZ-MEST)**

The University of Tasmania is the leading institute for a consortium that has been granted from the Australian Government to establish BlueNet – the Australian Marine Science Data Network. The project aims to provide a virtual data centre to support long-term duration and management of data for Australia's marine science researchers. BlueNet links vast data repositories and marine resources that currently reside in academic and government institutions both in Australia and overseas. It is an extension of the nation's first online, virtual facility – the Australian Ocean Data Centre Joint Facility (AODCJF).

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 BlueNetMEST is a branch of GeoNetwork opensource, which was already explained as part of international tools. BlueNetMEST allows end users to search for data and assess its quality through a metadata description (BlueNet 2010). The BlueNetMEST has been known as ANZ-MEST in Australia since January 2011 (ANZ-MEST 2012). Figure 5-21 illustrates the BlueNetMEST user interface.

Figure 5-21: The BlueNetMEST user interface

The results of examining the BlueNetMEST against the criteria outlined in Table 5-2 are discussed below:

➢ Support for integration with spatial data lifecycle

Similar to the GeoNetwork, the BlueNetMEST does not provide spatial data cataloguers with any facilities to integrate metadata creation with the data lifecycle.

➢ Support for integrated data model

The BlueNetMEST is supported by four types of RDBMS drivers including McKoi, Oracle, MySQL and PostgerSQL (Rajabifard *et al.* 2007). Within the database, the BlueNetMEST stores the metadata records in XML format. An entire metadata XML string is stored in a single database column. Using this tool the dataset and metadata cannot be coupled and managed in a common environment. This means that the integrated data model is not supported in the BlueNetMEST.

➢ Support for automatic metadata creation

Similar to GeoNetwork, metadata creation within the BlueNetMEST is a manual process and spatial data cataloguer is responsible to gather as much information as possible to create metadata. However, if metadata are harvested from remote nodes some of the metadata elements already discussed for the GeoNetwork can be created automatically.

➢ Support for automatic metadata updating

According to the BlueNetMEST user manual (2008), the end users can update metadata records by finding the record via a search and selecting the 'edit' button. At this point only the Revision

Date stamp within the metadata record will change to the current date and time. Therefore, it can be understood that the metadata in the BlueNetMEST should be updated manually, except for the Revision Date stamp element.

For the manually inserted metadata, due to lack of support for the association of actual dataset to metadata with any change to the dataset, its corresponding metadata cannot be updated automatically and at the same time. On the other hand, if the metadata is harvested from other nodes whenever the node updates its published information, the harvested metadata records would be replaced and updated automatically in the next round of harvesting.

➢ Support for interaction with end users

The end users are able to discover, retrieve, and access metadata records using the BlueNetMEST. Only the spatial data cataloguers are able to update metadata records. There is no capability provided in this tool for more interaction with the end users to update or improve the content of metadata.

• ANZMET Lite

The ANZLIC Spatial Resources Discovery and Access Toolkit (ANZMET Lite) has been developed by the Office of Spatial Data Management (OSDM) on behalf of ANZLIC Spatial Resources Discovery and Access Program Steering Committee to support the implementation of the ANZLIC Metadata Profile (OSDM 2009c).

ANZMET Lite is a wizard-based metadata editor written in Microsoft VB.NET for Windows. Unlike the GeoNetwork and ANZ-MEST, this tool is solely for editing metadata. By comparison, GeoNetwork and ANZ-MEST are intended as operational Web-based catalogues, which also allow/manage data delivery and data visualisation, metadata harvesting from other catalogues (or remote searching of them), visualisation of others' data, and so on. According to ANZ-MEST (2012), the key rationale for development of the ANZMET Lite was that the editing interface of the GeoNetwork/ANZ-MEST is not as user-friendly as is needed for 'novice' metadata creators, when using a complex metadata standard such as the ISO 19115: 2003. Therefore, the ANZMET Lite is intended as an editing tool that seamlessly interfaces with the GeoNetwork/ANZ-MEST.

Figure 5-22 illustrates the ANZMet Lite user interface.

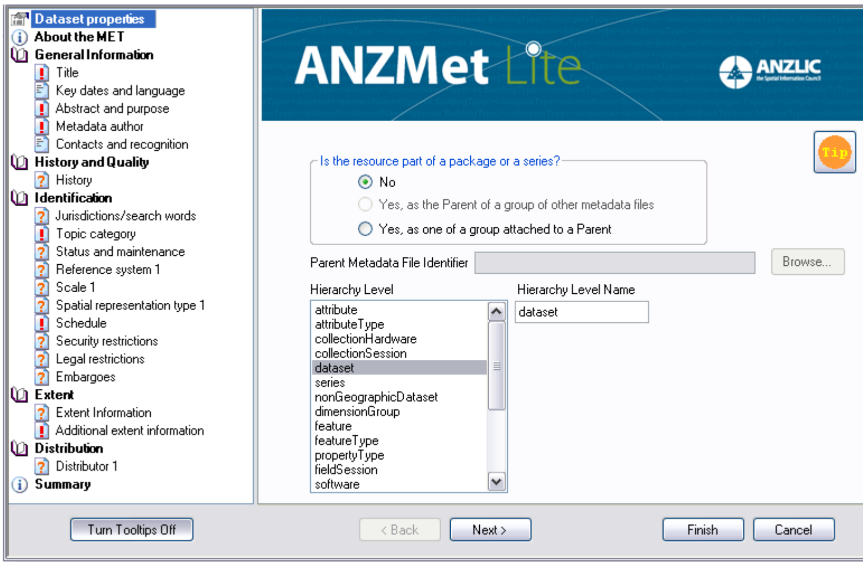


Figure 5-22: The ANZMET Lite user interface

The results of examining the ANZMET Lite against the criteria outlined in Table 5-2 are discussed here:

➢ Support for integration with spatial data lifecycle

The ANZMet Lite does not provide any facilities for integrating metadata creation with the data lifecycle.

➢ Support for integrated data model

According to the OSDM (2009b), the ANZMet Lite supports different ways to create metadata such as creating new parent, unlinked and linked metadata records. Creating linked metadata allows the metadata record to be created in the same directory as the resource and will remain linked to the resource. However, the spatial data and its associated metadata cannot be integrated in a common environment and published jointly. Therefore, the ANZMET Lite supports the detached data model for storing dataset and metadata.

➢ Support for automatic metadata creation

According to the OSDM (2009a), the elements described in Table 5-III in **Appendix 6** can be generated semi-automatically or fully-automatically via the ANZMet Lite metadata entry tool. Through the method of linked metadata creation, some of the metadata elements would be automatically derived from the dataset (such as dataset title and date of last update). The file name extensions for which the tool allows entry of metadata include ASCII, CSV, DBF, DGN, DWG, DXF, E00, ECW, IMG, SHP, BMP, DOC, JPG, JPEG, PDF, TIF, TIFF, TXT, and XLS. Some of the other elements can also be automated through accessing the pre-set values such as metadata standard name, standard version, and metadata contact details. The remaining elements such as resource quality and restrictions have to be populated manually.

- **Support for automatic metadata updating**

Although, the ANZMet Lite associates the resource to metadata file, there is no capability supported by this tool to update metadata automatically when the dataset is edited.

- **Support for interaction with end users**

The ANZMet Lite is a desktop metadata entry tool and there is no search and discovery functionality provided by this tool. Its end users are the spatial data cataloguers and not public users looking for data over the Web. Thus, no support for interaction with the end users for enriching the content of metadata could be expected from this tool. However, valid metadata records provided by the ANZMet Lite can be published to the Australian Spatial Data Directory (ASDD) to be used for data search and discovery purpose. The ASDD also does not support any facilities for the end users searching for data to improve the content of metadata record.

- **xMet Client**

The extensible Metadata Editing Tool (xMET) is a Java application developed by the OSDM-Australia that allows users to manage the ANZLIC Metadata Profile version 1.1 XML metadata (xMET 2011). Its purpose is to provide a platform for users to effectively create and manage metadata records based on the actual geographic data.

The premise behind xMET's development is that most metadata entry tools are very specific to the profile or schema of the metadata. They lack the extendibility that lets users or administrators to design or change the entry and editing interfaces without the tool being changed by developers. This tool is built in a way to facilitate designing metadata entry interfaces of any profiles without requiring any programming to be undertaken. The tool allows for the creation, editing and management of XML metadata records that are on the local file system and also on the GeoNetwork (CSW). Figure 5-23 illustrates the xMet Client user interface.

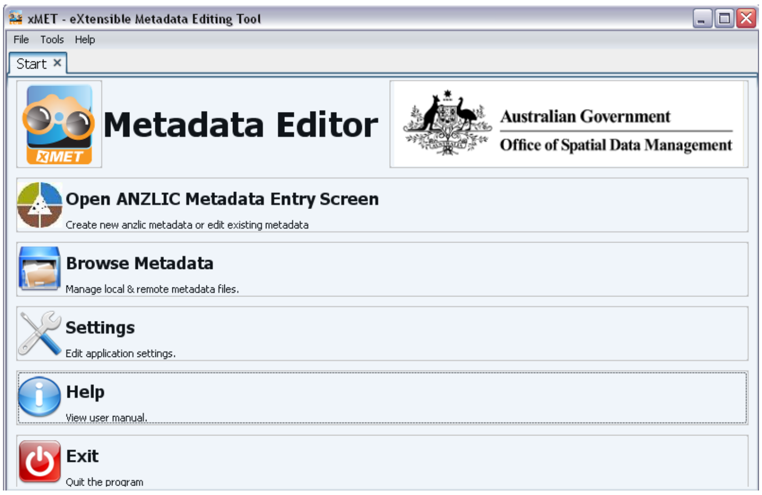


Figure 5-23: The xMET Client user interface

The results of examining the xMet Client against the criteria outlined in Table 5-2 are discussed below:

- **Support for integration with spatial data lifecycle**

The xMet does not provide any capabilities for spatial data cataloguers to integrate metadata creation process with the data lifecycle.

- **Support for integrated data model**

The xMet stores metadata locally or in the GeoNetwork database (through the CSW service) in XML. However, the relevant spatial datasets should be managed in a separate environment and no relationship between datasets and metadata can be defined using the xMet. Therefore, this tool supports a detached model for storing metadata.

- **Support for automatic metadata creation**

Automatic production of metadata from the actual datasets is not supported in the xMet.

- **Support for automatic metadata updating**

Whenever a dataset is modified, its associated metadata record should be updated manually using the xMet and no automation can assist this process.

- **Support for interaction with end users**

Similar to the ANZMET Lite, the xMet is a metadata management tool that does not provide the public users with metadata search and discovery functionalities. Therefore, no support for interaction with the end users for improving the content of metadata could be expected from this tool.

The next section summarises the main results arisen from the selected spatial metadata tools assessment.

5.3.3 Summary of Assessment Results

The assessment of selected spatial metadata management tools against the criteria proposed for this research has confirmed that the main challenges identified during the literature review (summarised in Section 3.2.2 in Chapter 3) and the case study investigations, have not been addressed by the metadata management tools. Table 5-3 summarises the results of assessing the tools against the criteria.

Table 5-3: Summary of results of examining selected metadata management tools against the criteria

Spatial metadata management tool		Criteria				
		Support for integration with spatial data lifecycle	Support for integrated data model	Support for automatic metadata creation	Support for automatic metadata updating	Support for interaction with end users
International	GeoNetwork opensource	NO	NO	YES– Only when metadata are harvested.	YES– Only when metadata are harvested; but not in real time.	NO
	GeoNode	NO	NO – only metadata and spatial data are associated to each other.	YES – only for a limited number of elements (e.g. date/time, Bounding Box, distribution URL, and contact information).	NO	NO
	ESRI ArcCatalog	NO	NO – only metadata and spatial data are stored in the same place.	YES– depending on the input dataset formats for the elements outlined in Table 5-I, Appendix 6.	YES– depending on the input dataset formats for the elements outlined in Table 5-I, Appendix 6; but not in real time.	N/A

Table 5-3: Summary of results of examining selected metadata management tools against the criteria

Spatial metadata management tool		Criteria				
		Support for integration with spatial data lifecycle	Support for integrated data model	Support for automatic metadata creation	Support for automatic metadata updating	Support for interaction with end users
	CatMDEdit	NO	NO	YES– depending on the input dataset formats for the elements outlined in Table 5-II, Appendix 6.	NO	N/A
	EUOSME	NO	NO	NO	NO	N/A
Australian	BlueNetMEST	NO	NO	YES– Only when metadata are harvested.	YES– Only when metadata are harvested; but not in real time.	NO
Australian	ANZMet Lite	NO	NO – Only 'linked' metadata is stored in the same directory which the dataset is stored.	YES– depending on the input dataset formats for the elements outlined in Table 5-III, Appendix 6.	NO	N/A
Australian	xMet Client	NO	NO	NO	NO	N/A

In the final stage of the conceptual phase of this research, the results of the selected tools assessment have been integrated with the findings of the case study investigations in the context of Australia and the results arisen from the literature review. Accordingly, a number of main challenges regarding spatial metadata management and automation have been identified which are summarised in the next section.

5.4 Integration of Results – Summary of Current Spatial Metadata Challenges

The integration of results achieved during the literature review, case study investigations and tools assessment resulted in identifying five main challenges which need to be addressed in the current thesis, as discussed below.

5.4.1 Consideration of Metadata Management outside the Spatial Data Lifecycle

Metadata describes different aspects of the dataset such as identification, quality, citation, extent, constraints, etc (ISO19115: 2003). Therefore, ideally metadata should be part of a spatial dataset and its values should be generated and updated with any change to the dataset from the very first stages of the data lifecycle (Olfat *et al.* 2012c). Producing metadata afterwards is difficult and may be a laborious task (Taussi 2007). However, the results of the case study, literature review and tools assessment reveal that 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 data lifecycle. Collecting metadata later requires considerable effort and not all the information might be available (Timpf *et al.* 1996) and the metadata gathered in this way is often missing or incomplete (Rajabifard *et al.* 2009).

5.4.2 Using Detached Metadata Data Model

Following the investigation of past and recent research and development activities on spatial metadata management, as well as, the selected metadata tools (such as the GeoNetwork opensource, ANZMET Lite, CatMDEdit, and ESRI), it can be deduced that the current metadata generation/updating approach is rooted in a detached data model (Figure 5-24). Moreover, this approach is entirely dependent on the knowledge of the metadata author or responsible party about the dataset (Olfat *et al.* 2012c).

As discussed in Section 2.6.1 in Chapter 2, in a 'detached data model' the spatial data and its associated metadata are stored separately in different files or databases which make them either without a relationship with each other or to have only a common identifier. In contrast, in an 'integrated data model' spatial data and metadata can be mapped to and stored in a middleware, so that with any change in the data the metadata can be updated at the same time (Kalantari *et al.* 2009).

As a result, designing and implementing an integrated metadata data model that can be used for automating the metadata updating process at the same time as dataset's modification, will benefit organisations.

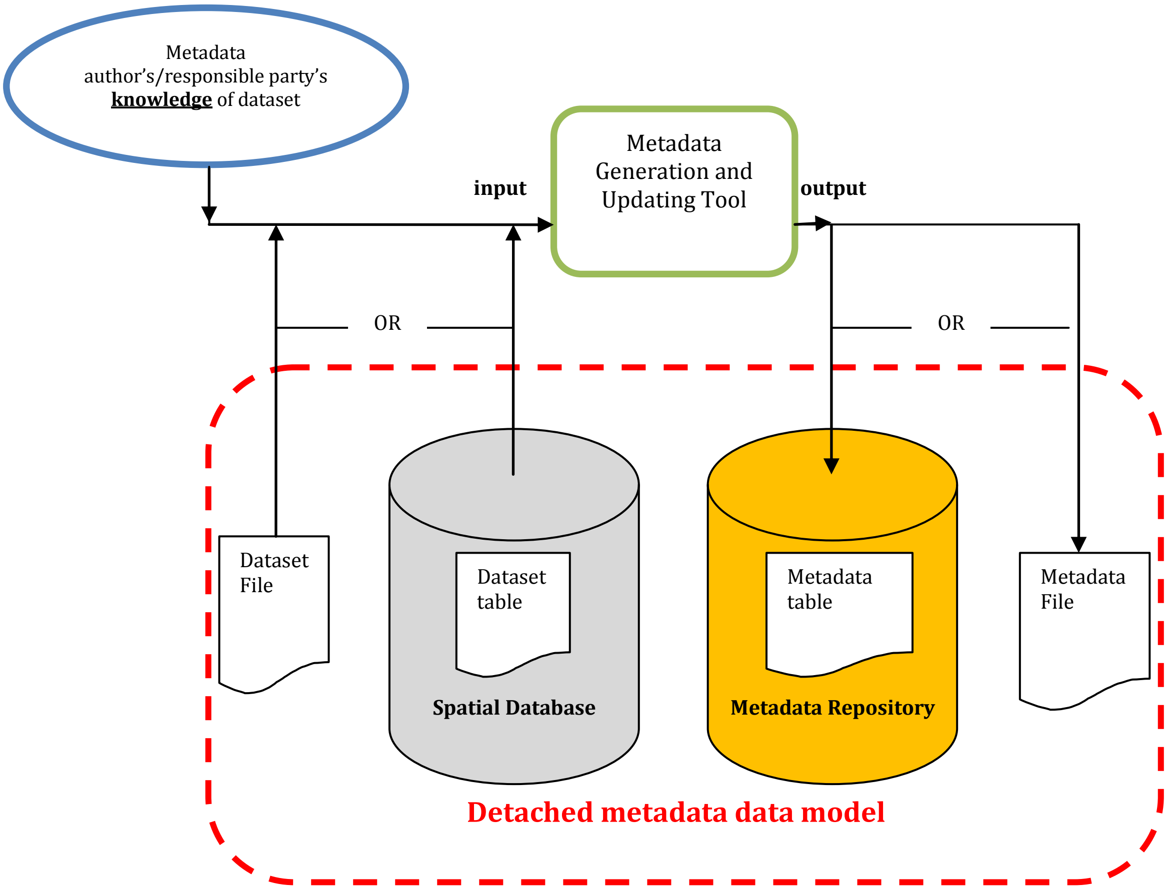


Figure 5-24: Current approach of metadata generation/updating (Olfat *et al.* 2012c)

5.4.3 Lack of Support for Real-time Spatial Data and Metadata Updating

The results of the survey run in the context of Australia showed that a large portion of spatial data custodians and creators still update metadata through a separate activity from the dataset modification. It was also found that separate teams have the responsibility to update dataset and metadata for the same organisation. This results in a delay between dataset and metadata updating time and therefore avoids the metadata from being always up-to-date, reliable, and precise. Following this current approach, the organisations need extra resources in terms of budget and time to undertake further effort to update metadata after any change to the dataset.

In addition, the results of spatial metadata tools assessment showed that these tools lack the support for real-time spatial data and metadata updating. Within the assessed tools, the ESRI ArcCatalog, which has also been commonly used by different researchers (Batcheller 2008, Batcheller *et al.* 2007, Westbrooks 2004a), synchronises metadata with the latest status of spatial data. However, the synchronisation process requires human intervention to run (ESRI 2010). Therefore, metadata and spatial data updating would not be real-time in nature.

5.4.4 Dependency of Metadata Automation Methods on Dataset Format

As a result of systematically reviewing the metadata automation research and development activities and selected spatial metadata tools, it can be surmised that the existing automation tools are highly restricted to dataset formats to extract metadata values. For instance, CatMDEdit automates metadata generation only for Shapefile, DGN, ECW, FICC, GeoTIFF, GIF/GFW, JPG/JGW, and PNG/PGW formats (CatMDEdit 2011b), or GeoNode automatically generates a few metadata values for the Shapefiles. Therefore, an agnostic dataset format automatic approach to create and update metadata will play an important role to address this issue.

5.4.5 Limited Interaction with End Users for Metadata Creation and Improvement

The results of assessing the selected spatial metadata tools, along with the case study investigations in the context of Australia, indicated that the current tools are not sufficiently user-friendly. The end users are also disconnected from the spatial metadata creation and improvement process. These tools need more interaction with the users to improve the content of metadata; especially 'keyword' metadata element (Kalantari *et al.* 2010), which is the main gateway for discovering and finding datasets over the Web. Kalantari *et al.* (2010) also argue 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 that 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 the users from diverse categories. Moreover, a keyword should be popular meaning that most of the users and spatial data responsible party agree on that.

5.5 Chapter Summary

This chapter explored the current status of spatial metadata management and the automation requirements through a qualitative case study research in the context of Australia. The results showed that there has been a critical need to improve the existing approaches to create and update spatial metadata in an automated fashion, which illustrates the significance of this thesis for the geospatial community. The chapter also proposed a set of criteria and examined a number of spatial metadata management tools against the criteria to understand the automation capabilities supported by these tools. Finally, the results of the literature review together with the case study investigations and tools assessment were integrated and the main challenges that need to be addressed in this thesis were discussed. These challenges include the consideration of metadata generation/updating outside the spatial dataset lifecycle, lack of support for real-time spatial data and metadata updating, dependency of automatic metadata creation and updating methods on dataset format, and limited interaction of metadata systems with the end users.

The next chapter develops a framework based on the research design to address the main challenges identified in this chapter.

CHAPTER 6

LIFECYCLE-CENTRIC CREATION, AUTOMATIC UPDATING AND ENRICHMENT OF SPATIAL METADATA

6 LIFECYCLE-CENTRIC CREATION, AUTOMATIC UPDATING AND ENRICHMENT OF SPATIAL METADATA

6.1 Introduction

This chapter discusses the design and development of a framework to address the main challenges regarding spatial metadata management and automation identified in Section 5.4, Chapter 5. In this regard and in order to address the third research objective, stated in Chapter 1, regarding the possibility of integrating metadata creation process with the spatial data lifecycle, the chapter designs a lifecycle-centric approach as part of the developed framework. This approach focuses on creating metadata values in parallel to the spatial data lifecycle. Also, the chapter designs an automatic spatial metadata updating (synchronisation) approach as another component of the developed framework to address the fourth research objective regarding an automatic solution for concurrent spatial data and metadata updating. The synchronisation approach aims to update spatial data and metadata in real time and is rooted in an integrated data model for spatial data and metadata storage. Finally, this chapter designs an automatic spatial metadata enrichment approach as part of the developed framework to address the fourth research objective relating to an automatic approach for engaging the end users to improve the content of spatial metadata. This approach is built up on Web 2.0 features and aims to enrich the content of keyword metadata element through interaction with the end users seeking spatial data over the Web.

6.2 A Framework to Address Current Spatial Metadata Management and Automation Challenges

According to the integration of results arisen from literature review, case study investigations and selected tools assessment, the previous chapter identified and summarised the main challenges regarding spatial metadata management and automation that need to be addressed in this thesis. As a result, a framework has been developed in this section to overcome the identified challenges (Figure 6-1).

According to Figure 6-1, the core of framework contains the ‘spatial data lifecycle’. Following the challenges discussed in Section 5.4 in Chapter 5, the metadata management has been mainly considered outside the spatial data lifecycle. The framework through the 'Lifecycle-centric Spatial Metadata Creation' approach aims to integrate the metadata creation with the steps involved in the spatial data lifecycle.

Also, lack of support for real-time spatial data and metadata updating and dependency of current metadata automation methods on the dataset formats had been recognised as other challenges summarised in Section 5.4, Chapter 5. The developed framework aims to address these challenges through the 'Automatic Spatial Metadata Updating' approach, which would be dataset format agnostic and integrated with the spatial data lifecycle. Also, in order to address the current challenges of detached data model for spatial data and metadata storage the framework develops an integrated data model in which the spatial dataset and its related metadata can be managed and updated together.

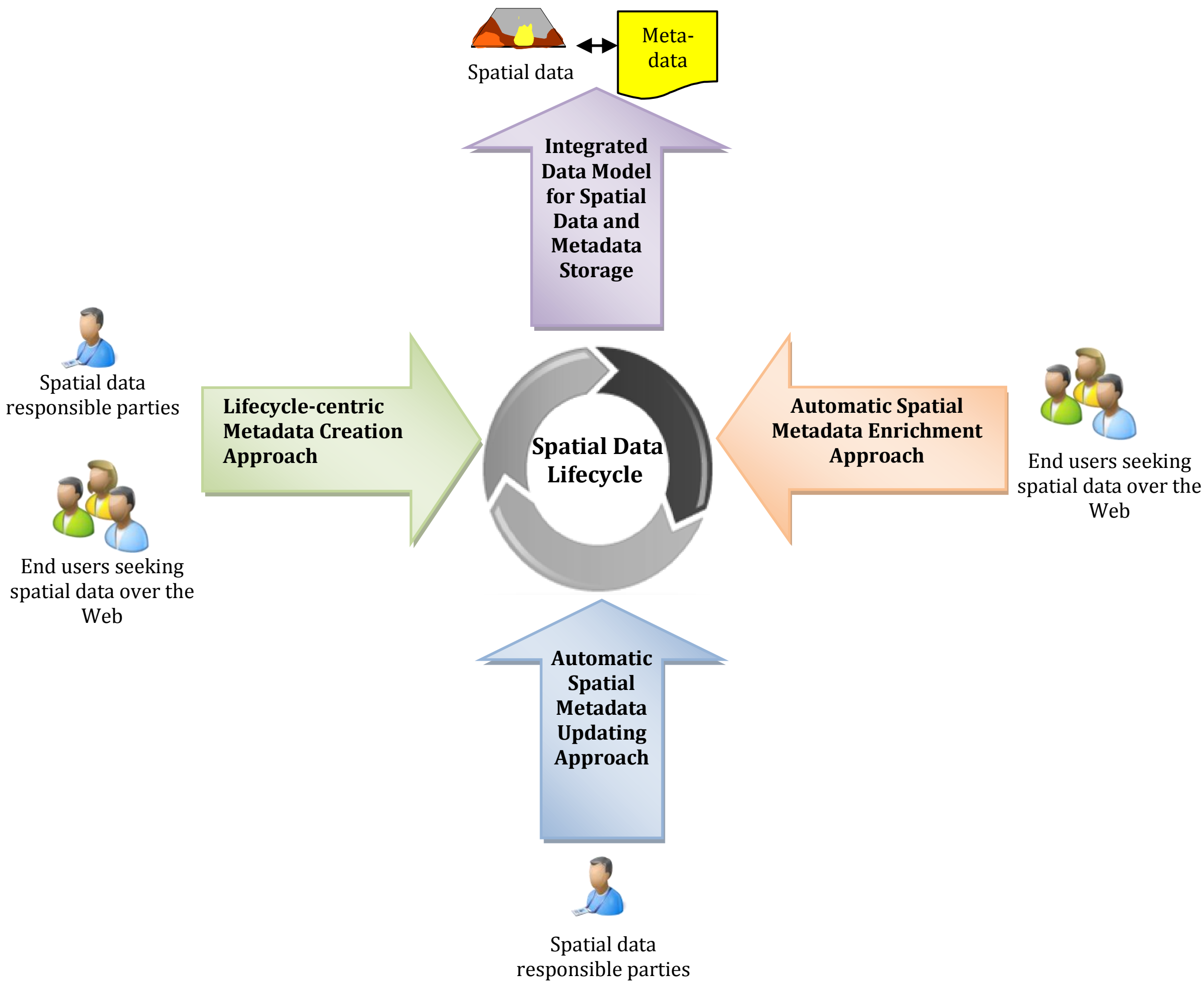


Figure 6-1: A framework to address the identified challenges regarding spatial metadata management and automation

Finally, the developed framework focuses on engaging the end users to improve the content of metadata to address the challenge of end users' disconnection from metadata creation and improvement process. This would be undertaken through the 'Automatic Spatial Metadata Enrichment' approach, which is also integrated with the spatial data lifecycle.

The next sections explore the components of the developed framework in detail.

6.3 Lifecycle-centric Spatial Metadata Creation Approach

Following the findings of literature review discussed in Section 3.2.2 in Chapter 3 and the results of selected metadata tools assessment discussed in Section 5.3 in Chapter 5, the recent research initiatives (Manso *et al.* 2004, Manso-Callejo *et al.* 2009, Batcheller 2008, Batcheller *et al.* 2007, Taussi 2007) and current metadata management tools (e.g. ArcCatalog, and CatMDEdit) mainly focus on developing methods to generate metadata from an existing dataset. Also, the findings of the case study, discussed in Section 5.2.1 in Chapter 5, confirmed that a process separate from the spatial data lifecycle has typically been considered in the organisations of the case study region – Australia to create the metadata records.

In this thesis, the current state of play in metadata creation approach is named as a 'dataset-centric approach' because it mainly focuses on generating metadata from the dataset itself. Figure 6-2 illustrates the dataset-centric metadata creation approach.

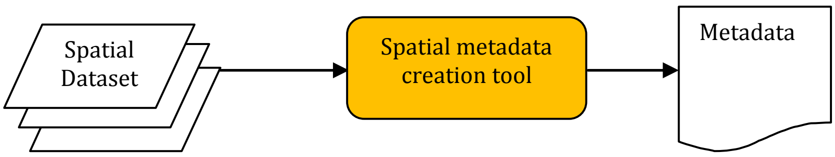


Figure 6-2: Dataset-centric metadata creation approach

Although the dataset-centric approach might be valuable for creating metadata for datasets of which associated metadata have not been incorporated in the data creation process and need to be retrospectively added, it would not be a holistic approach for metadata creation for forthcoming spatial datasets. The reason is that the dataset-centric metadata creation approach is capable of generating a limited portion of worldwide standards recommended metadata values (such as bounding box and coordinate reference system), and the remaining portion should be generated by the metadata responsible party, or author, in an organisation. This person should be aware of the whole data lifecycle to generate a complete, precise and up-to-date metadata. Also, as personnel change or time passes, information about an organisation's data will be lost (FGDC 2000). Moreover, from the organisations' point of view, this approach is highly labour-intensive, tedious, time-consuming, and costly (Guptill 1999, West and Hess 2002).

Accordingly, this thesis investigates an alternative approach to address the third research objective, outlined in Chapter 1, regarding the integration of metadata creation process with the spatial data lifecycle. However, designing such an approach requires the understanding of the relationship between spatial metadata elements and spatial data lifecycle.

In this regard, this thesis designed a generic spatial data lifecycle and investigated the relationship between metadata elements and the steps involved in the lifecycle. The following sections explore the outcomes of this investigation.

6.3.1 Design of a Generic Spatial Data Lifecycle

Niemann (2011) discusses that the data lifecycle is a term coined to represent the entire process of data management. It starts with concept study and data collection, but importantly has no end, as data is continually repurposed, creating new data products that may be processed, distributed, discovered, analysed and archived. Fully supporting the different steps in the lifecycle puts demands on metadata, standards, tools, and people.

The United States Office of Management and Budget (OMB) also developed a geospatial data lifecycle in 2010 that federal agencies should use when developing, managing and reporting on National Geospatial Data Asset (NGDA) Datasets under the auspices of OMB Circular A–16 (OBM 2010). The Circular provides direction for federal agencies that produce, maintain, or use spatial data either directly or indirectly in the fulfilment of their mission and provides for improvements in the coordination and use of spatial data. The Circular establishes a coordinated approach to electronically develop the National SDI and establishes the Federal Geographic Data Committee (FGDC). As illustrated in Figure 6-3, the lifecycle recommended by the OMB (2010) consists of 7 main stages, as following:

- Define – characterisation of data requirements based upon business-driven user needs
- Inventory/Evaluate – creation and publication of a detailed list of data assets and data gaps (both internal and external) as they relate to business-driven user needs
- Obtain – collection, purchase, conversion, transformation, sharing, exchanging, or creation of geospatial data that were selected to meet the business needs is identified
- Access – production of data known and retrievable to the community through documentation and discovery mechanisms so the users can meet their business requirements
- Maintain – ongoing processes and procedures to ensure that the data meet business requirements

- Use/Evaluate – ongoing assessment, validation, and potential enhancement of data to meet user needs and business requirements
- Archive – required retention of data and the data's retirement into long-term storage.

Figure 6-3: The geospatial data lifecycle recommended by the OBM (2010)

The Australian Government Information Interoperability Framework also developed an 'information lifecycle' in 2006 (AGIMO 2006). As illustrated in Figure 6-4, this lifecycle includes different steps consisting of planning, creation/collection, organisation/storage, access, using, maintenance, re-using and sharing the information. The steps in the information lifecycle are presented as if in sequence, but it should be understood that steps may be undertaken simultaneously, iteratively, partially or in different orders.

According to AGIMO (2006), 'planning' includes the identification of the information requirements relevant to any work activity. Information is 'created', 'collected', 'captured' or 'accessed' in a variety of ways from a variety of sources as part of a business need. Once created or collected, information needs to be 'organised' and 'stored' to enable easy location/access/retrieval to support business processes. Information may be 'accessed' and 'used' in a range of ways. These may involve using information in its original state, manipulating it in some way, integrating information from a number of sources, and re-using information. The information lifecycle includes the effective 'maintenance' of information, and in some circumstances, its disposal.

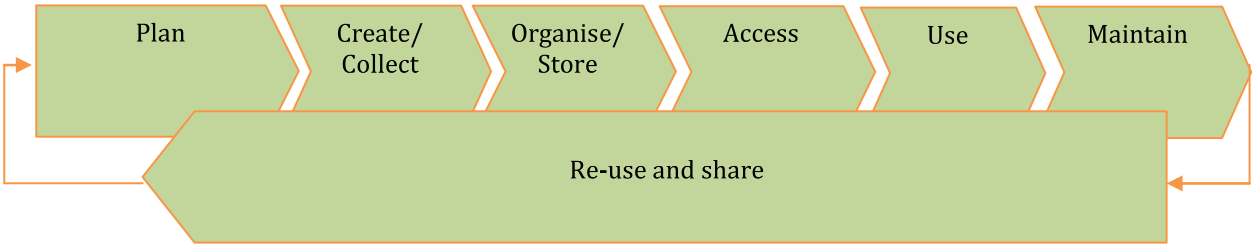


Figure 6-4: The information lifecycle (AGIMO 2006)

Because the prototype systems development and evaluation for this research occurred in Australia – the case study region, the 'information lifecycle' recommended by the AGIMO (2006) was adopted to identify the steps involved in the generic spatial data lifecycle. However, the information lifecycle was modified and expanded to include the detailed steps involved in the generic spatial data lifecycle. For instance, data discovery was considered as a separate step, or policy making was included in the planning step. Figure 6-5 illustrates the generic spatial data lifecycle informed by the AGIMO lifecycle, which consists of planning and policy making, spatial and non-spatial data collection, dataset creation, storage, publication, discovery and access, utilisation, and maintenance steps (Olfat *et al.* 2012c). Similar to the information lifecycle, these steps may be undertaken simultaneously, iteratively, or in different orders.

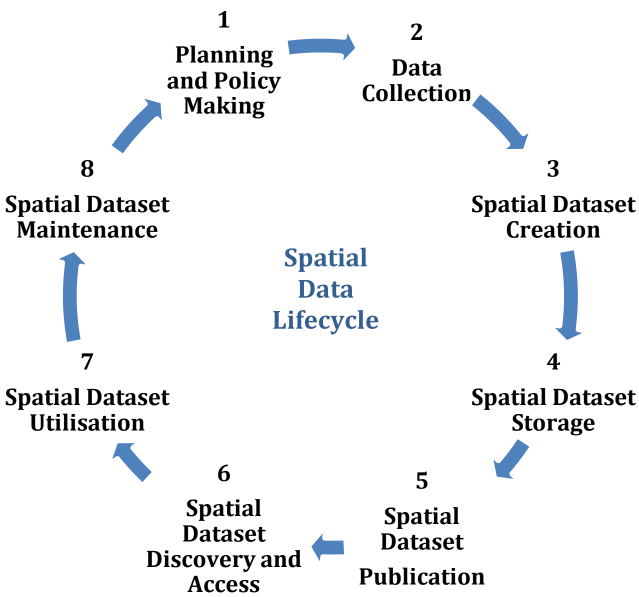


Figure 6-5: The generic spatial data lifecycle

The steps of the generic spatial data lifecycle illustrated in Figure 6-5 are here explained (Olfat *et al.* 2012c):

- **Planning and Policy Making**

Planning and policy making, as the first step of the spatial data lifecycle, has a tight similarity to the 'planning' step in the information lifecycle by AGIMO (2006). Because different policies are required in any step of the spatial data lifecycle, this step can also be considered as an ongoing process parallel to other steps. Different decisions regarding the spatial dataset/metadata creation, storage, distribution, responsibilities, rights, restrictions, standards, languages, extensions, etc. would be made.

- **Data Collection**

In the spatial data lifecycle, the next step would be collecting the spatial and non-spatial data to create the dataset. In fact, it is the first part of the 'collect and create' step in the information lifecycle that is separated. The related data to create spatial datasets can be collected through different methods (e.g. using GPS[25], aerial photographs, satellite images, etc.) depending on the planning decisions in terms of users' needs, the purpose of collection, the required quality, scale, extent, etc.

- **Spatial Dataset Creation**

After the data collection, in this step different activities would be undertaken to create the spatial dataset. Indeed it is the second part of the 'collect and create' step in the information lifecycle. First, the collected data would be standardised; then, different data would be aggregated to create the spatial dataset. Finally, to make sure that the dataset meets the required quality, the Quality Assurance (QA) process would be undertaken.

- **Spatial Dataset Storage**

Once the spatial dataset is created, it would be stored in the database. The database management system (DBMS) is usually designed before this step commences. It is based on the concept of storage in the 'organise and store' step in the information lifecycle.

- **Spatial Dataset Publication**

After the spatial dataset is stored, based on the concept of organisation in the 'organise and store' step in the information lifecycle, the dataset would be published to a networked environment in form of data or service (Web Services such as WFS, WMS, etc.) for sharing among the end users.

[25] Global Positioning System

- ### Spatial Dataset Discovery and Access

In a networked environment, the end users would be able to discover the existing and shared spatial datasets to find the most appropriate ones for their needs. In this step, the users would be able to discover the datasets through the data catalogue system and provide them with different search methods such as basic (based on keywords) and advanced (based on extent, dataset node, topic, etc.). This step is added to the information lifecycle. After discovering and finding the required datasets, the end user would be able to access and retrieve the required datasets to be utilised. Accessing the datasets is usually dependant on some policies such as dataset rights and restrictions. It follows the access part of the 'access and use' step in the information lifecycle.

- ### Spatial Dataset Utilisation

When the dataset is retrieved by the end user, it would be utilised in relevant spatial applications such as research, analysis, map creation, reporting, and value adding to the existing datasets. It is based on the 'use' part of the 'access and use' step in the information lifecycle.

- ### Spatial Dataset Maintenance

According to the huge amount of changes to the datasets, which usually occur in short time frames, an effective maintenance process is considered in the generic spatial data lifecycle; otherwise, the datasets would be out-of-dated and therefore unusable for the end users. The maintenance step is based on the concept of the 'maintain' step in the information lifecycle.

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

On one hand, the generic spatial data lifecycle confirms that there are different steps involved in any spatial data life-time of which their characteristics and events should be documented and published to the end users as 'metadata'. On the other hand, a wide variety of metadata elements recommended by the International Standards (e.g. ISO 19115: 2003) emphasises the documentation of metadata values during the spatial data lifecycle. Therefore, to have a comprehensive view of metadata creation, there is a need to investigate the relationship between metadata elements and the generic spatial data lifecycle. Because the ISO 19115: 2003 Standard was adopted in this thesis for designing and developing the new approaches to address the research objectives, an investigation was undertaken to realise the relationship between the generic lifecycle designed in this section and the metadata elements recommended by the ISO 19115: 2003. The following section discusses the results of this investigation.

6.3.2 Mapping Metadata Elements against the Generic Spatial Data Lifecycle

In this investigation, the elements recommended by the ISO 19115: 2003 were reviewed systematically and mapped against the steps of the generic spatial data lifecycle (8 steps). Table 6-I in **Appendix 7** illustrates the outcomes of this investigation. The first column of this table namely 'ISO element No.' describes the element number in the ISO 19115: 2003. The second column namely 'Element' includes the metadata element and its path within the ISO 19115: 2003 UML packages and the third column namely 'Proposed Spatial Data Lifecycle Related Step(s)' proposes the relevant step(s) within the data lifecycle for creating the metadata element. The names of the steps are shortened in some cases and include 'planning and policy making', 'collection', 'creation', 'storage', 'publication', 'discovery and access', 'utilisation', and 'maintenance'. The fourth column namely 'Proposed Creation Method' proposes the metadata creation methods (manual, semi-automatic, and automatic) derived from the findings of spatial metadata automation literature review (discussed in Section 3.2.2 in Chapter 3) and the capabilities of the selected metadata management tools discussed in Section 5.3, Chapter 5.

Besides, based on the findings outlined in Table 6-I in **Appendix 7**, the elements that could be used in designing and implementing the automatic approaches to address the fourth research objective were identified. Accordingly, the ISO metadata elements, which might be updated along with the dataset modification (within the 'maintenance' step), were identified to be employed in designing and developing an automatic spatial metadata updating approach. Table 6-1 outlines some of these elements for a vector dataset. In addition, the ISO metadata element number 53, 'Identification information: MD_Keywords: keyword', was recognised as the only metadata element which could be improved by the interaction of end users seeking spatial datasets over the Web (within the 'discovery and access' step). Therefore, this element was used in designing and developing an automatic spatial metadata enrichment approach.

Table 6-1: Some of the main ISO 19115: 2003 metadata elements affected by the vector dataset modification

ISO Element No.	Metadata Element	Description
187	Reference system	name of the reference system
190	Projection	identity of the projection used
394	Date of last update (revision)	reference date for the selected resource
28	Status	status of the resource(s)
144	Resource date of next update	next scheduled revision for the resource
60	Equivalent Scale	level of detail expressed as the scale of a comparable hardcopy map or chart
349	Geographic location of the resource (by description)	identifier used to represent a geographic area
344	West longitude	western-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east)
345	East longitude	eastern-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east)
346	South latitude	southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north)
347	North latitude	northern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north)
351	Temporal extent – time period	range of dates and times for the content of the dataset
355	Vertical extent – minimum value	lowest vertical extent contained in the dataset
356	Vertical extent – maximum value	highest vertical extent contained in the dataset
83	Lineage statement	general explanation of the data producer's knowledge about the lineage (or history) of a resource

Figure 6-6 illustrates the result of mapping the ISO metadata elements against the generic spatial data lifecycle steps. According to this Figure, the highest number of metadata elements should be created within the spatial dataset creation step. Planning and policy making, dataset maintenance, publication, data collection, dataset storage, utilisation, and discovery and access are respectively the next steps with the highest number of elements.

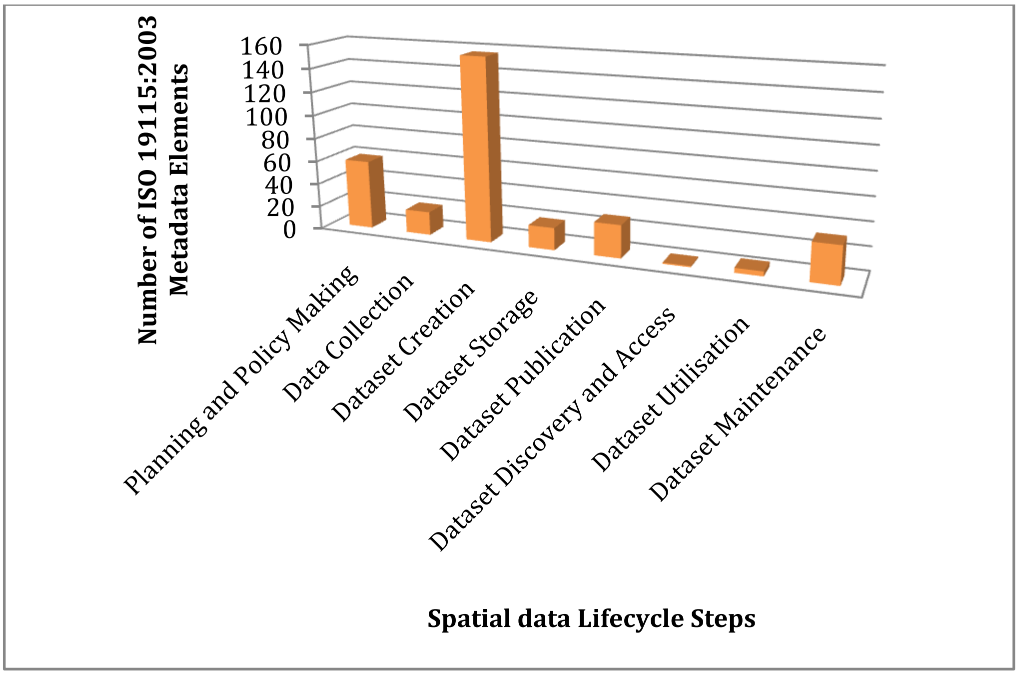


Figure 6-6: Mapping the ISO 19115: 2003 metadata elements against the generic spatial data lifecycle steps

To better conceptualise the relationship between the ISO metadata elements and the spatial data lifecycle, an example is described in Table 6-2. This Table shows the related steps in the data lifecycle for generating metadata values for a sample dataset entitled ‘Borehole Geology (Boresg)’. This dataset has been provided by the Victorian Department of Sustainability and Environment (DSE) and used in developing the automatic spatial metadata updating prototype system, which will be discussed in the next chapter.

Table 6-2 Example of the relationship between metadata elements of 'Borehole Geology (Boresg)' dataset and the spatial data lifecycle steps

Step(s) in Spatial Data Lifecycle	ISO 19115: 2003 Metadata Element	Metadata Value
Planning and policy making	Metadata file identifier	ANZVI0803002677
Creation/ Storage	Dataset title	Borehole Geology (from Earth Resource's database) (BORESG/)
Creation	Abstract	The dataset is a subset of Minerals and Petroleum's BORES dataset and contains geological data from Minerals and Petroleum's boreholes RDBMS.
Creation/ Storage/ Maintenance	Geographic location of the dataset	34.0 S 141.0 E 150.2 E 39.2 S
Creation/ Maintenance	Temporal Extent	Begin date: 01JAN1870 End date: Current
Collection/ Creation/ Maintenance	Lineage statement	Dataset Source: From Minerals and Petroleum's INGRES Database
Creation	Positional accuracy	Varies from 10 metres to 1km – accuracy is specified for each location
Storage/ Publication	Distribution format	ESRI Shapefile
Creation/ Maintenance	Maintenance and update frequency	Irregular
Publication/ Maintenance/ Discovery and access	Keyword	• ENERGY Coal none • ENERGY Petroleum none • GEOSCIENCES Exploration • GEOSCIENCES Geology Inventory
Planning and policy making	Dataset point of contact	Department of Primary Industries Administration Officer GPO Box 4440 Melbourne Vic 3001 Australia Tel: (03) 9658 4563

According to the results illustrated in Table 6-I in **Appendix 7** and Figure 6-6, it can be understood that the values regarding metadata elements (more than 300 elements recommended by the ISO 19115: 2003) cannot be completely achieved unless they are created within different steps of the spatial data lifecycle. The findings of the literature review, discussed in Section 3.2.2 in Chapter 3, along with the findings of the case study investigations and selected tools assessment, discussed in Sections 5.2 and 5.3 in Chapter 5, showed that the current metadata management approaches do not support the metadata to be created within the data lifecycle. Therefore, the next section designs a new approach for integrating metadata creation with the steps of spatial data lifecycle to address the third research objective.

6.3.3 Integration of Metadata Creation with the Generic Spatial Data Lifecycle

In this section, a new approach namely 'lifecycle-centric' has been designed to create metadata in parallel to the spatial data lifecycle. Figure 6-7 illustrates this new approach.

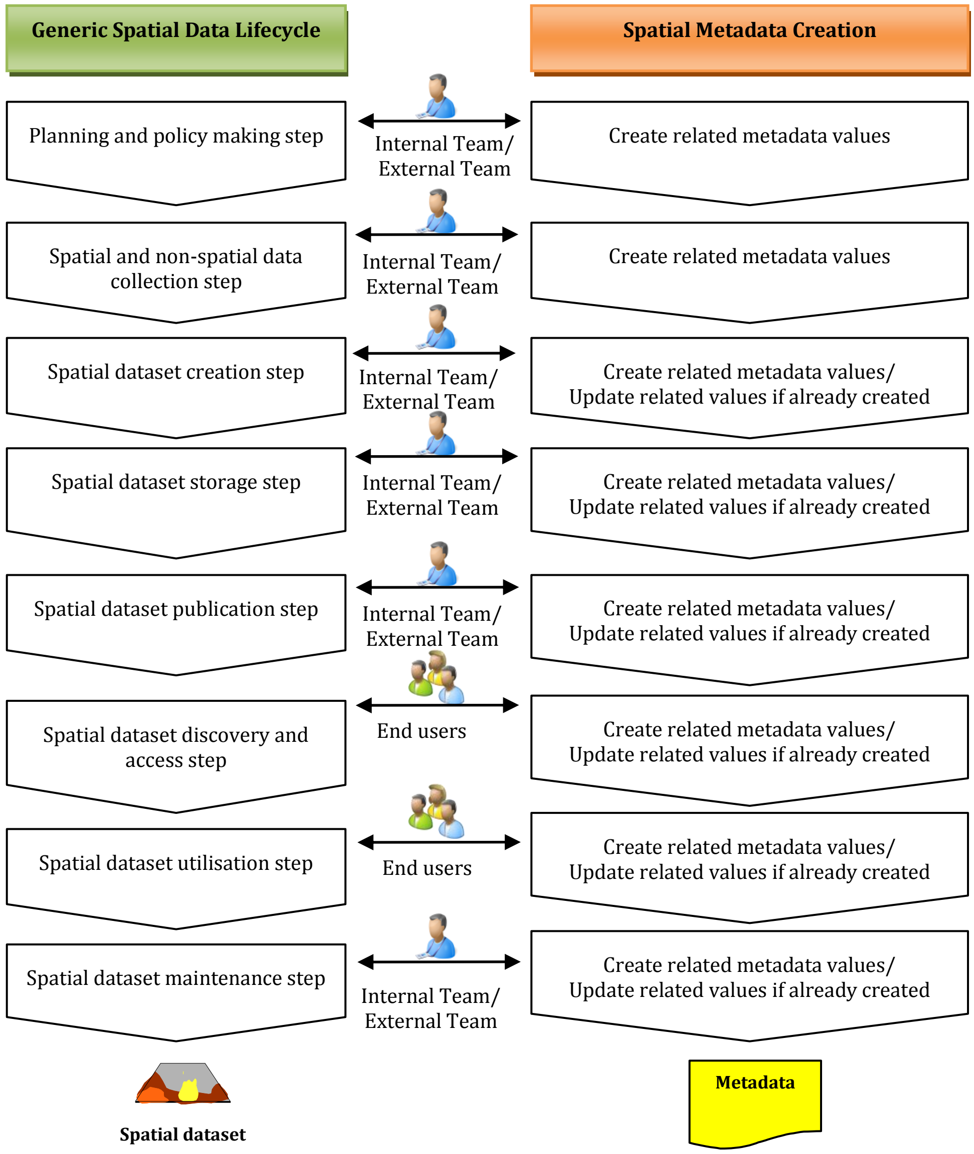


Figure 6-7: Lifecycle-centric spatial metadata creation approach

As illustrated in Figure 6-7, the responsible party for each step of the spatial data lifecycle would be in charge of creating (and maintaining) related metadata values within those steps. According to the results of the case study investigations discussed in Section 5.2.1 in Chapter 5, the responsible parties for creating and maintaining spatial data can be categorised into two main groups: internal team within the organisation and external team. Also, following the definition of 'discovery and access' and 'utilisation' given earlier in Section 6.3.1 the end users are mainly involved in these two steps of the lifecycle. Therefore, according to Figure 6-7 the end users would be engaged in creating (and improving) the content of metadata within these two mentioned steps.

However, according to Table 6-I in **Appendix 7** most of the metadata elements need to be created in one step of the lifecycle and then updated within the others (except the elements for 'planning and policy making' and 'data collection'). For instance, 'status' metadata element needs to be generated when dataset is 'created' and needs to be updated when dataset is 'maintained'. For this reason, in the lifecycle-centric spatial metadata creation approach (see Figure 6-7) 'updating' the content of metadata is also considered along with 'creation' of metadata values.

As a result of the lifecycle-centric approach, the metadata will be completed over time in conjunction with the spatial data lifecycle and therefore, it is more likely to be accurate and up-to-date. Also, in order to facilitate the development of this approach the existing dataset-centric metadata creation methods (e.g. computation, extraction and harvesting), discussed extensively in Section 3.2.2 in Chapter 3, need to be employed. As already mentioned, the last column of Table 6-I in **Appendix 7** includes the information about the available methods that could be utilised for developing the lifecycle-centric approach for metadata creation.

The proposed lifecycle-centric spatial metadata creation approach brings forth some advantages for the organisations. This approach will support the generation and updating of a wide range of the ISO metadata elements. The approach also has the potential to overcome the problem of missing or incomplete metadata through recognising the specific step to generate and update metadata within the data lifecycle. Moreover, it most likely reduces the burden of metadata creation for metadata authors by involving the spatial data responsible parties and interacting with the end users in creating and updating metadata values. However, in order for the lifecycle-centric approach to work properly a metadata entry/edit tool needs to be designed and developed to provide a combination of available manual/semi-automatic/automatic metadata creation methods and an appropriate level of access for the responsible parties involved in the spatial data lifecycle.

According to Figure 6-1, an automatic approach for updating metadata in conjunction with the dataset modification and an integrated data model for storing metadata and dataset are other components of the framework developed for addressing the spatial metadata challenges. The next section explores these two components.

6.4 Automatic Spatial Metadata Updating (Synchronisation) Approach

As discussed in Chapter 1, long-standing issues and challenges regarding metadata updating have been acknowledged by the geospatial community. Some of the major challenges have been identified following the results of the literature review and case study investigations which were summarised in Section 5.4 in Chapter 5; including lack of support for real-time spatial data and metadata updating, dependency of metadata automation methods on dataset format and using a detached data model for spatial data and metadata storage.

In order to address the above challenges and respond to the research objective 4, this section aims to design an automatic approach namely '*metadata synchronisation*' to update metadata concurrent with any modification of the dataset during its lifecycle (see Figure 6-1). Following the generic spatial data lifecycle, discussed in Section 6.3.1, datasets would be modified in the 'dataset maintenance' step. Therefore, the metadata synchronisation approach needs to be adopted in this step of the spatial data lifecycle. The approach also needs to be designed in a way to be dataset format agnostic and built upon an integrated data model for metadata and dataset storage.

The following sections explore the conceptual design and technical requirements for implementing the metadata synchronisation approach.

6.4.1 Metadata Synchronisation Conceptual Design

Real-time spatial metadata and dataset updating plays a significant role in accessing the most up-to-date and precise metadata in any sharing platform. The metadata synchronisation approach, as an automatic process by which properties of a spatial dataset are read from both back end (where dataset is stored) and front end (where the modification environment is up and running) and written into its spatial metadata at the same time as any modification of the dataset (Olfat *et al.* 2010b, Kalantari *et al.* 2010), is designed to address this need.

According to the aim of metadata synchronisation approach, the prerequisite for taking this approach into account would be designing and building an integrated data model for storing metadata and the dataset related to each other. This new kind of data model has already been introduced to the geospatial community (Giger and Najar 2003, Najar 2006). According to Najar (2006), the integration, managing and modelling existing metadata and spatial data commonly is realistic.

Through the integrated data model, each dataset would be related to its metadata record. Having the relationship between these two sources and accommodating dataset geometries, attributes and metadata values into a middleware would result in a comprehensive dataset, which can also be exchanged over the Web between different spatial systems as well as end users. Comprehensive datasets are datasets that are associated by three fundamental components: geometries (and topologies), attributes and metadata. By transferring this comprehensive dataset to a user interface through the middleware the users (spatial data responsible parties) would be able to represent and edit dataset and metadata concurrently. While modifying a dataset, the responsible parties should be able to see the modification reflection on corresponding metadata values simultaneously and automatically. Some of the metadata values affected by the dataset modification should be updated at the front end (e.g. date of revision and lineage) and the others should be updated at the back end (e.g. bounding box) via synchronisation scripts. Those elements that are updated at the front end would be transferred to the back end (metadata table) through the middleware and those updated at the back end would be directly replaced on the metadata table in the database.

By having the synchronisation approach in place, after any dataset modification the new values for geometry, attributes and metadata would be transferred to and stored at the back end and are shown at the same time on the user interface. The conceptual design for metadata synchronisation approach is illustrated in Figure 6-8 (Olfat *et al.* 2012a).

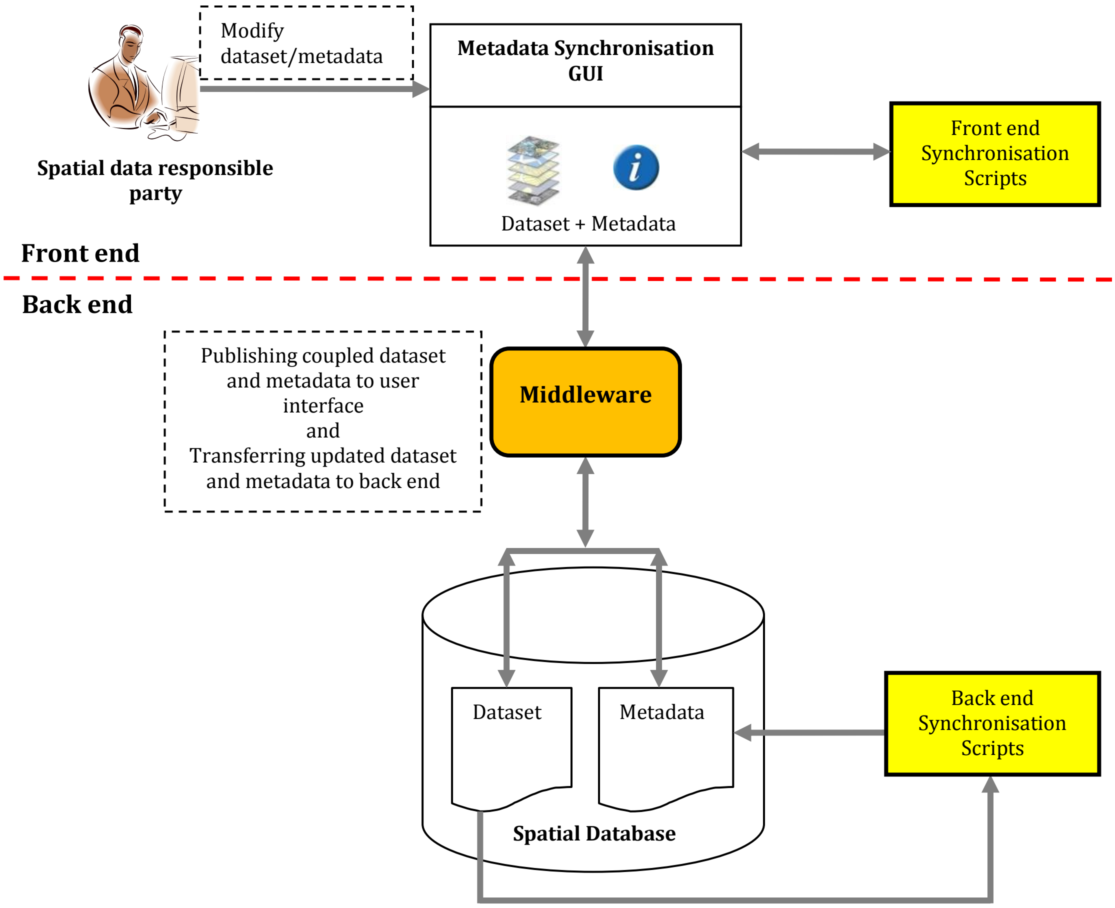


Figure 6-8: The conceptual design for metadata synchronisation approach (Olfat *et al.* 2012a), proposed to be adopted in 'dataset modification' step within the spatial data lifecycle

The technical requirements identified to implement this conceptual design (see Figure 6-8) are described in the next section.

6.4.2 Identification of Technical Requirements for Implementing Metadata Synchronisation

In order to design and implement a system based on the metadata synchronisation concept, the major technical requirements have been identified as (Olfat *et al*. 2012a):

- integrated data model for storing spatial metadata and dataset
- mapping software application to generate the integrated data model and support dataset and metadata updating
- user-friendly interface to view dataset and metadata from the integrated data model and modify the dataset
- identification of the metadata elements which might be updated after any dataset

modification

- synchronisation scripts to update metadata based on dataset changes.

The above requirements are detailed as following.

- **Integrated Data Model for Spatial Metadata and Dataset Storage (Requirement 1)**

To address the first requirement and design and form a data model to accommodate both spatial data and metadata values, the following four components are considered (Olfat *et al.* 2012a):

1. spatial database to store both dataset and metadata
2. dataset and metadata tables
3. defined relationship between dataset and metadata tables
4. middleware to accommodate dataset and metadata values and include their relationship.

According to Figure 6-9, the first component of the integrated data model would be a spatially enabled database that is capable of storing spatial and non-spatial tables including the dataset (and attributes) and metadata. The second component would be the tables to store the dataset and metadata. The dataset table would contain the attributes and geometry columns while the metadata table(s) would embrace the ISO 19115: 2003 recommended elements (such as the file identifier, title, abstract, identification, quality, and distribution information). The third component would be a relationship between the dataset table and the metadata table(s). This relationship indicates which metadata record is related to which dataset in the spatial database.

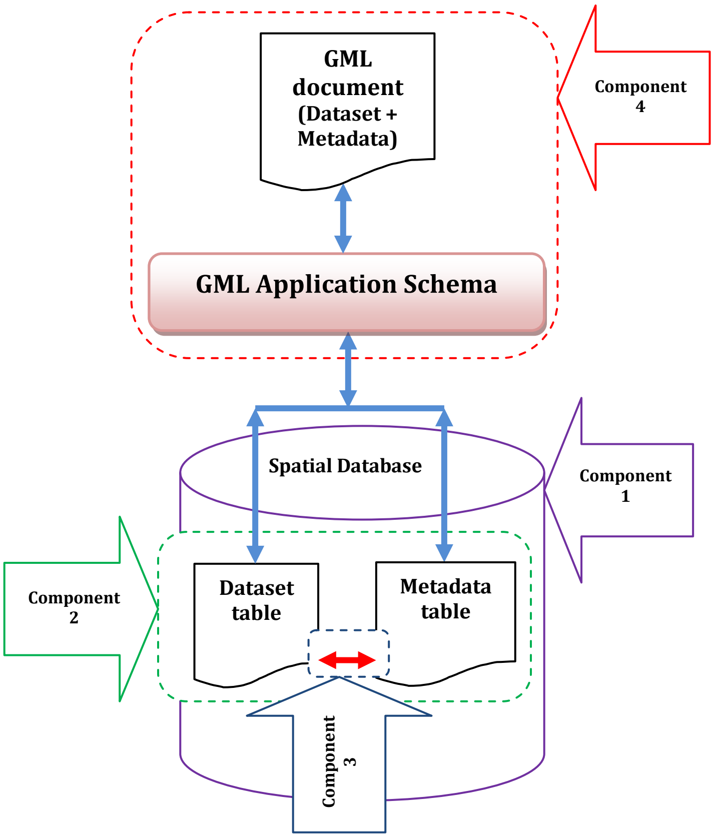


Figure 6-9: The components of the integrated data model for spatial metadata and dataset storage (Olfat *et al.* 2012a)

The fourth component of the integrated data model would be a middleware that is able to store both dataset and metadata values and their relationship. The middleware should also be involved in returning the changes (updated values) from the user interface to the required tables at the back end. To identify such an environment, two of the available metadata exchange formats including Metadata Exchange Format (MEF) and Geography Markup Language (GML) were investigated.

As discussed during the exploration of GeoNetwork opensource in Section 5.3.1 in Chapter 5, MEF is an especially designed file format whose purpose is to allow metadata exchange between different platforms that understand this format. An MEF file is a ZIP file which contains the metadata file in XML, a special XML file which embraces information related to the metadata, application schema, a directory to store the metadata thumbnails and other public files, and another directory to store all data (Maps, Shapefiles, etc.) associated to the metadata (GeoNetwork 2012a). This exchange format does not support the spatial metadata and dataset to be stored in the database, as needed for developing the integrated data model (see Figure 6-9).

On the other hand, the GML International Standard by the OGC is an XML grammar and vocabulary for expressing geographical features and serves as a modelling language for geographic systems and also as an open interchange format for geographic transactions on the Internet (OGC 2007). The GML has the capability to support encoding, transferring and storage of spatial data characteristics, as well as, metadata values that are stored in different tables in the same database or distributed databases. It also helps address many of the concerns relating to the data interoperability and communication (Zhang *et al*. 2003) and provides an open dialect for data transfer not bound to specific software offerings (Peng 2005). According to the requirements of the middleware in the integrated metadata data model such as communicating with the database and different tables to view and update the dataset and metadata, the GML was selected to play the role of middleware.

Because the GML is a Markup Language, it means that the GML document has to follow certain rules in order to be a valid GML document. This set of rules is defined in a GML Application Schema which is also an extension of the 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).

Following the integrated data model, the output that consists of values from several related tables (dataset and metadata) is known as a feature collection. A feature collection is a special category that represents a collection of features that have common metadata and formal relationships (OGC 2003). Collections possess all the characteristics of a feature, that is, they are complex features. From a technical view, complex features contain properties that can include further nested properties to arbitrary depth and are used to represent information not as an XML view of a single table, but as a collection of related objects of different types (GeoServer 2011b). As a result, the capability of a mapping software application to work with the complex features was considered as another significant requirement for implementing the metadata synchronisation approach.

- **Mapping Software Application (Requirement 2)**

The second requirement for developing the metadata synchronisation approach is a mapping software application. This application would be connected to the spatial database and support the integrated metadata data model by forming the middleware (GML document) to contain the dataset and metadata and deliver it to the users (data responsible parties). Also, the mapping application provides the transaction capabilities in order to insert, delete and update the dataset and metadata. For instance, an OGC compliant mapping software application that supports the Web Feature Service-Transaction (WFS-T) capability for the complex features can play this role.

- **Graphical User Interface (GUI- Requirement 3)**

The third requirement for setting up the metadata synchronisation approach is designing and implementing a user-friendly interface to interact with the users who are responsible for modifying the spatial datasets in 'dataset maintenance' step of the spatial data lifecycle. This user interface represents the extracted geometry and metadata from the GML document (as the output of the integrated data model) in different sections. This interface also includes an 'Edit Toolbar' and enables the responsible parties to modify the dataset.

To store the changes resulting from the dataset modification, there is also a need to have a 'Save' button which sends a transaction request to the mapping software application to update the dataset and metadata. Figure 6-10 shows the different components of the GUI.

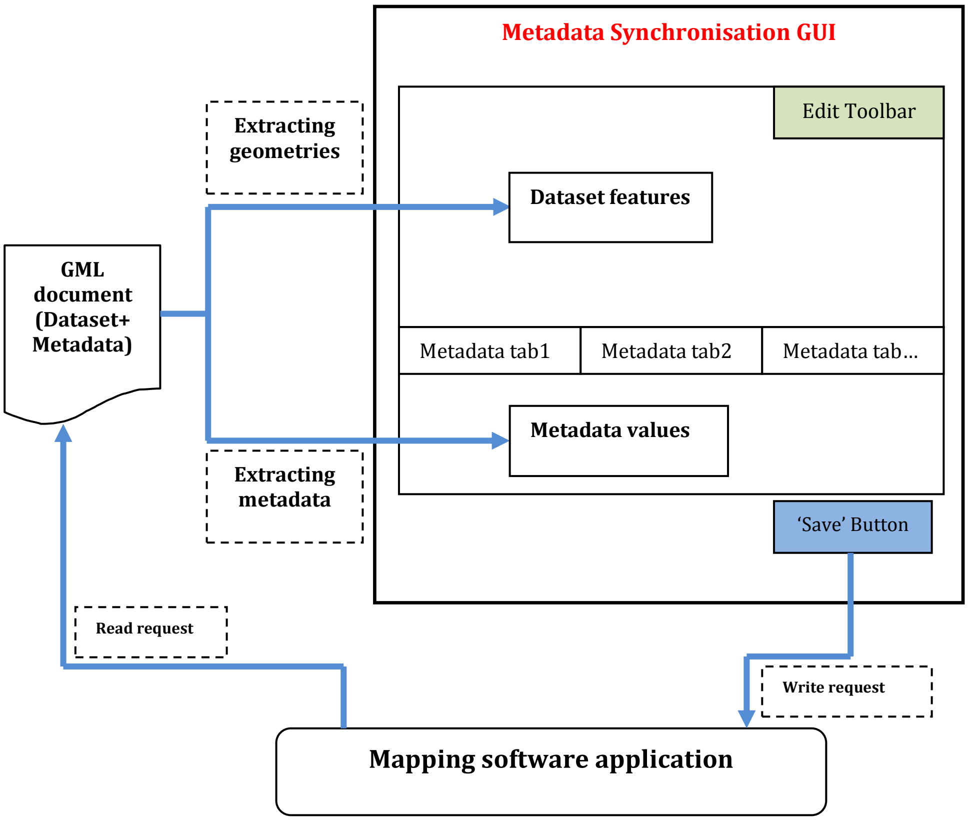


Figure 6-10: The components of metadata synchronisation GUI and its relation to the mapping software application

This interface should be able to read the WFS/WFS-T response in GML format. In addition, the interface through its 'Edit Toolbar' should enable the users (data responsible parties) to modify the dataset and save the changes by sending the WFS-T request to the Web Server. The technologies used for the purpose of GUI development are discussed in detail in the next chapter.

- ## Identification of Metadata Elements (Requirement 4)

In order to design and implement the metadata synchronisation approach, another requirement is to identify the metadata elements that might be updated after any dataset modification, and to replace the old values with new ones in the metadata table within the spatial database.

In this regard, the main metadata elements that might be affected by the dataset modification should be recognised during the design of the lifecycle-centric spatial metadata creation approach. Table 6-1 already illustrated some of the most significant metadata elements that might need to be synchronised after any modification to a vector dataset. Keeping these elements always up-to-date plays an important role in sharing any dataset, so that data is robust and reusable. The metadata elements employed in implementing a prototype system for metadata synchronisation will be introduced in the next chapter.

- **Synchronisation Scripts (Requirement 5)**

Once the metadata elements that might be updated concurrently with the dataset are identified, there is a need to develop scripts for generating the new values for these metadata elements based on the changes applied to the dataset. The synchronisation scripts reflect the changes of metadata based on the user interaction (the front end; e.g. updating of date and time of dataset revision) as well as, the system interaction (the back end; e.g. updating of the dataset bounding box). The scripts developed for implementing the metadata synchronisation prototype system will be reviewed in the next chapter.

Based on the above technical requirements and the methodology suggested and discussed in Section 4.4.3 in Chapter 4, the automatic spatial metadata updating (synchronisation) prototype system is designed and implemented within the GeoNetwork opensource catalogue. The architecture used to implement this system, as well as, its capabilities and challenges will be discussed in the next chapter.

According to Figure 6-1, the last component of the developed framework to address the identified challenges of spatial metadata would be an automatic approach to engage the end users seeking spatial data over the Web to enrich the content of metadata. This approach is discussed in the next section.

6.5 Automatic Spatial Metadata Enrichment Approach

In order to design an automatic approach to address the fourth research objective, there was a need to define the metadata element(s) of which their value(s) could be improved by the end users seeking spatial data over the Web. According to the results of mapping the ISO 19115: 2003 metadata elements against the generic spatial data lifecycle steps discussed in Section 6.3.2, 'descriptive keyword' was the only metadata element in which its value could be improved by the end users during the 'dataset discovery and access' step of the spatial data lifecycle. The descriptive keyword element is one of the mandatory elements recommended by the ISO 19115: 2003 that should be embedded into each spatial metadata record and is defined as ‘commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject’ (ISO19115: 2003).

The spatial data discovery systems usually support making a variety of queries via basic and advanced search modes on the spatial metadata records to retrieve the characteristics of the most appropriate datasets for the end users. As illustrated in Figure 6-11, using a search word or a library of search words (e.g. ANZLIC search word library) is common between both basic and advanced search modes. The data discovery system matches the search words with the values of metadata elements such as title, abstract, and more importantly the descriptive keyword to retrieve the relevant metadata records.

Accordingly, identifying the appropriate keywords to describe the spatial datasets is fundamental within any sharing platform, but on the other hand has become increasingly problematic (Chi 2009). The appropriate keyword for any spatial dataset means the keyword that is consistent with the content of the dataset and can reveal its essence and applications. In addition, a good keyword should address the probable queries made by users from diverse categories. Moreover, a keyword should have a popular meaning that most of the users and data responsible party agree on (Kalantari *et al.* 2010).

Currently, the keyword metadata element is created by the metadata responsible parties in two ways. In the first approach, they use a library of search words during metadata creation. Using this kind of library has been common within different SDI initiatives. For instance, the metadata authors or responsible parties in Australia need to select the search words from the ANZLIC search words library within the ANZMET Lite and xMet Client tools. These search words are also recommended to the users when discovering the datasets over the Web, as shown by the orange arrow in Figure 6-11. These search words are described in Table 6-3.

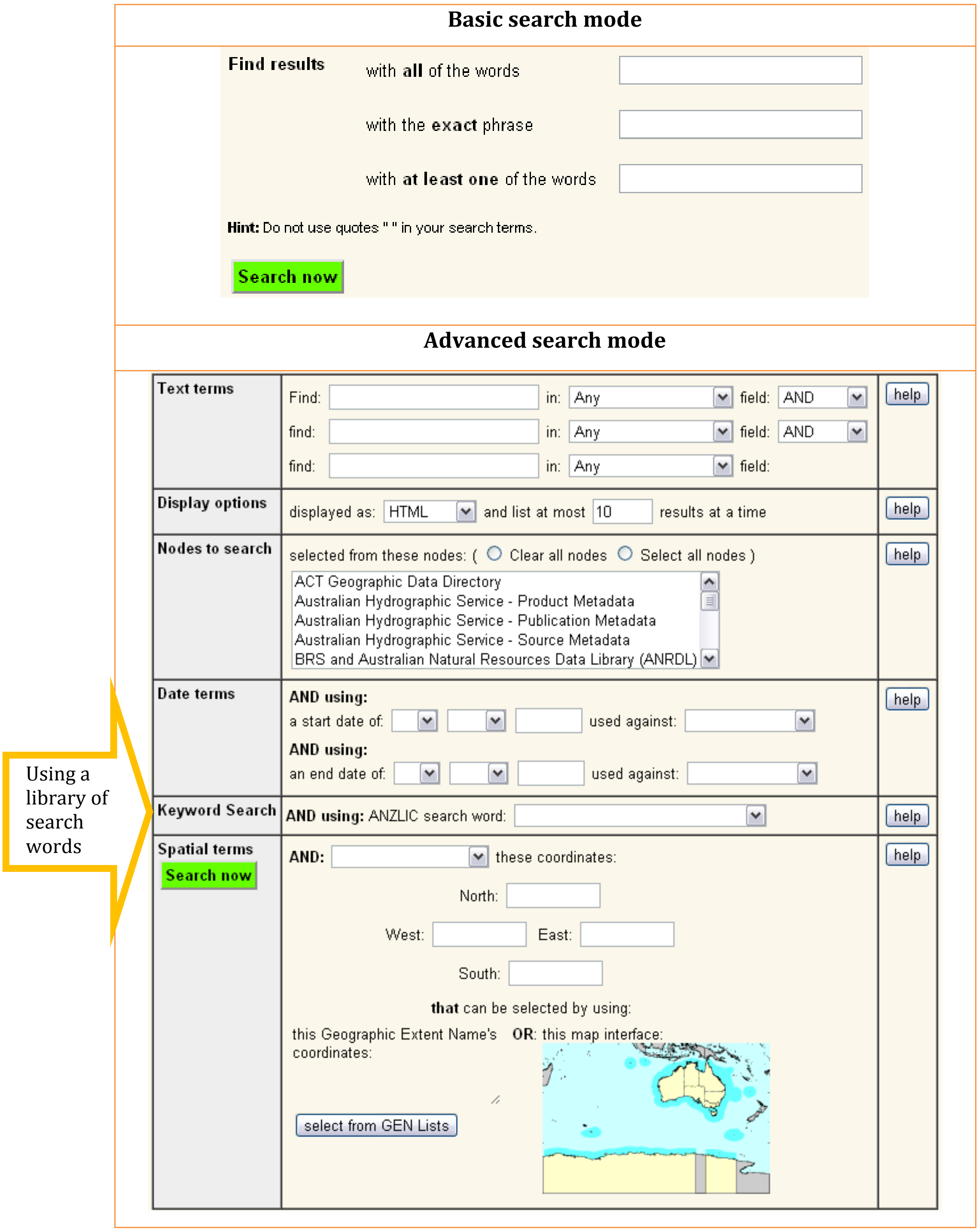


Figure 6-11: Basic and advanced search modes in Australia Spatial Data Directory (ASDD) Portal[26]

[26] http://asdd.ga.gov.au/asdd/search.html

Table 6-3: The ANZLIC library of search words

Search words used in the ANZLIC library of search words			
AGRICULTURE	AGRICULTURE Crops	AGRICULTURE Horticulture	AGRICULTURE Irrigation
AGRICULTURE Livestock	ATMOSPHERE	ATMOSPHERE Air Quality	ATMOSPHERE Greenhouse
ATMOSPHERE Ozone	ATMOSPHERE Pressure	BOUNDARIES	BOUNDARIES Administrative
BOUNDARIES Biophysical	BOUNDARIES Cultural	CLIMATE AND WEATHER	CLIMATE AND WEATHER Climate change
CLIMATE AND WEATHER Drought	CLIMATE AND WEATHER El Nino	CLIMATE AND WEATHER Extreme weather events	CLIMATE AND WEATHER Meteorology
CLIMATE AND WEATHER Radiation	CLIMATE AND WEATHER Rainfall	CLIMATE AND WEATHER Temperature	DEMOGRAPHY
DISEASE	ECOLOGY	ECOLOGY Community	ECOLOGY Ecosystem
ECOLOGY Habitat	ECOLOGY Landscape	ENERGY	ENERGY Coal
ENERGY Electricity	ENERGY Petroleum	ENERGY Renewable	ENERGY Use
FAUNA	FAUNA Exotic	FAUNA Insects	FAUNA Invertebrates
FAUNA Native	FAUNA Vertebrates	FISHERIES	FISHERIES Aquaculture
FISHERIES Freshwater	FISHERIES Marine	FISHERIES Recreational	FLORA non-specific
FLORA Exotic	FLORA Native	FORESTS	FORESTS Agriforestry
FORESTS Natural	FORESTS Plantation	GEOSCIENCES	EOSCIENCES Geochemistry
GEOSCIENCES Geology	GEOSCIENCES Geomorphology	GEOSCIENCES Geophysics	GEOSCIENCES Hydrogeology
HAZARDS	HAZARDS Cyclones	HAZARDS Drought	HAZARDS Earthquake
HAZARDS Fire	HAZARDS Flood	HAZARDS Landslip	HAZARDS Manmade
HAZARDS Pests	HAZARDS Severe local storms	HAZARDS Tsunamis	HEALTH
HERITAGE	HERITAGE Aboriginal	HERITAGE Architectural	HERITAGE Natural
HERITAGE World	HUMAN ENVIRONMENT	HUMAN ENVIRONMENT Economics	HUMAN ENVIRONMENT Housing
HUMAN ENVIRONMENT Livability	HUMAN ENVIRONMENT Planning	HUMAN ENVIRONMENT Structures and Facilities	HUMAN ENVIRONMENT Urban Design
INDUSTRY	INDUSTRY Manufacturing	INDUSTRY Mining	INDUSTRY Other
INDUSTRY Primary	INDUSTRY Service	LAND	LAND Cadastre
LAND Cover	LAND Geodesy	LAND Geography	LAND Ownership
LAND Topography	LAND Use	LAND Valuation	MARINE
MARINE Biology	MARINE Coasts	MARINE Estuaries	MARINE Geology and Geophysics
MARINE Human Impacts	MARINE Meteorology	MARINE Reefs	MINERALS
MOLECULAR BIOLOGY	MOLECULAR BIOLOGY Genetics	OCEANOGRAPHY	OCEANOGRAPHY Chemical
OCEANOGRAPHY Physical	PHOTOGRAPHY AND IMAGERY	PHOTOGRAPHY AND IMAGERY Aerial	PHOTOGRAPHY AND IMAGERY Remote Sensing
PHOTOGRAPHY AND IMAGERY Satellite	POLLUTION	POLLUTION Air	POLLUTION Noise
POLLUTION Soil	POLLUTION Water	SOIL	SOIL Biology
SOIL Chemistry	SOIL Erosion	SOIL Physics	TRANSPORTATION
TRANSPORTATION Air	TRANSPORTATION Land	TRANSPORTATION Marine	UTILITIES
VEGETATION	VEGETATION Floristic	VEGETATION Structural	WASTE
WASTE Greenhouse gas	WASTE Heat	WASTE Liquid	WASTE Sewage
WASTE Solid	WASTE Toxic	WATER	WATER Groundwater
WATER Hydrochemistry	WATER Hydrology	WATER Lakes	WATER Quality
WATER Rivers	WATER Salinity	WATER Supply	WATER Surface
WATER Wetlands			

Although, using this kind of a search words library will standardise the process of keyword allocation among the metadata responsible parties, it can restrict the end users to select a keyword from a defined range in which they may or may not be familiar with. Also, this approach to keyword creation requires the end users to have the same knowledge and insight as the responsible parties have about the whole library keywords and the areas they cover.

In contrast to the first approach, in the second approach the keywords are created based on the responsible parties' own opinions about the datasets. For instance, Figure 6-12 illustrates a metadata record harvested by the Australian Urban Research Infrastructure Network (AURIN) metadata system from Australian Bureau of Statistics (ABS) WFS data source[27] accessible through the Western Australia Shared Land Information Platform (SLIP). As it can be seen, the keywords that have been entered by the data responsible party during publishing the dataset over the Web are stored in 'keyword' element. These keywords are not included in any library.

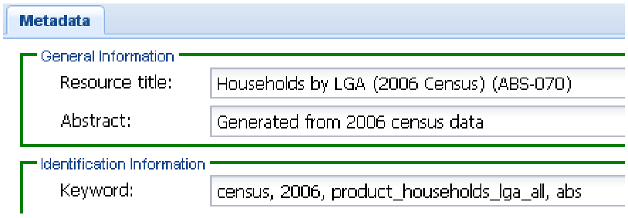


Figure 6-12: A sample harvested metadata record including the keyword- AURIN metadata system

From the current keyword creation approaches it can be understood that the data discovery systems are limited to some defined keywords and there is no approach for improving the content of keywords over the time. Also, according to the results of the tools assessment discussed in Section 5.3.3 in Chapter 5, the data discovery systems do not support the interaction of end users for sharing their knowledge about the datasets as a new helpful type of metadata. Therefore, there is a need for a new approach within the data catalogues to assist the metadata responsible parties, as well as, the end users in finding and sharing the most effective and popular keywords for describing the spatial datasets.

Having already explored in Section 3.3.2 in Chapter 3, it was shown that folksonomy feature of Web 2.0, as a new form of metadata that are created by users, can facilitate the generation of good keywords for any sharable resources. Folksonomic metadata consists of words that users generate and attach to content, which are also known as tags (Alexander 2006). Tagging allows ranking and data organisation to directly utilise inputs from end users (Xu *et al.* 2006).

The folksonomy concept is new in the spatial arena and can be considered as the potential way to improve the spatial metadata content (Kalantari *et al.* 2010). However, according to Passant and Laublet (2008) tagging raises issues from an information retrieval point of view, including the ambiguity and heterogeneity of tags since tags, from a machine point of view, do not carry any semantics about what they represent, while a human can interpret such semantics when

[27] https://www2.landgate.wa.gov.au/ows/wfsabs_4283/wfs?SERVICE=WFS&request=getCapabilities

tagging or reading. In response to this issue, Xu *et al.* (2006) also propose a set of criteria for identifying the most appropriate tags through eliminating noise and spam. These criteria for a good tag combination contains five items including high coverage of multiple facets, high popularity, least effort, uniformity (normalisation), and the exclusion of certain types of tags.

According to the advantages of tagging and folksomony for creating a new type of metadata, discussed previously in Section 3.3.2 in Chapter 3, this section focuses on the design of a new automatic approach rooted in these two features of Web 2.0 to improve the content of the descriptive keyword element of the spatial metadata through to the end users' interaction. The new approach also concentrates on the end users' experience for rating the popular keywords identified automatically and adding new values to the descriptive keyword metadata element. Consequently, according to Figure 6-13 two complementary models namely 'indirect' and 'direct' have been designed for the automatic spatial metadata enrichment approach.

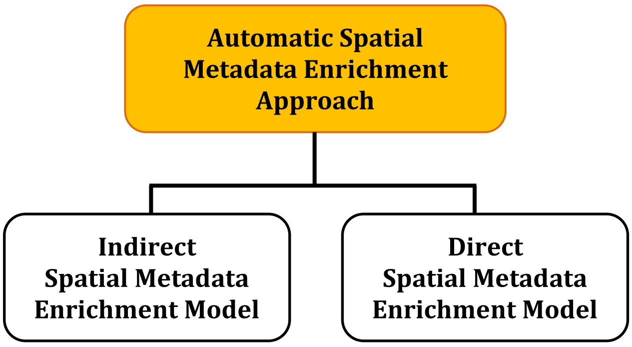


Figure 6-13: Models designed for the automatic spatial metadata enrichment approach

The indirect model is designed to recognise and tag the most popular keywords for describing datasets through monitoring the end users' behaviour during the data discovery process without their understanding. However, the direct model enables the end users to interact with the metadata records and rate (agree and disagree with) the tagged keywords and also directly add value to the keyword metadata element. Through the direct model, the end users are also able to comment on the datasets and create a new additional metadata.

The tagged keywords resulted from both direct and indirect models will be also visualised in a tag cloud, as discussed in Section 3.3.2 in Chapter 3. Tag cloud technique has been selected to visualise the outputs of automatic spatial metadata enrichment approach because of a number of distinct tasks that can be accomplished using the clouds, as introduced by Rivadeneira *et al.* (2007):

- Search: locating (or determining the absence of) a specific target or alternative target. This is often as a means to get more detailed information about the target. The target here is a spatial dataset.

- Browsing: casually exploring the cloud without a specific target or purpose, often drilling down on multiple discovered targets as they pique the end users' interest.
- Impression formation and impression presentation: the cloud can be scanned to get a general idea about a subject (spatial data catalogue). Visually prominent items may carry more weight in this initial impression, but other less prominent items also serve to enrich the impression.
- Recognition or matching: recognising the entire cloud as data that describes a subject.

Using the tag cloud which includes the most popular keywords describing the datasets within a spatial data discovery system, will enhance and facilitate the spatial dataset discovery and retrieval process (Kalantari *et al.* 2010).

The indirect and direct models forming the spatial metadata enrichment approach are discussed in the following sections. As described in the research design and methods chapter, in order to prove the concept of the automatic spatial metadata enrichment approach, a prototype system has been implemented within two different environments: the GeoNetwork as an open source spatial data catalogue application and Model Information Knowledge Environment (MIKE) as an example of a data product – data modelling environment. The prototype system development will be also explored in the next chapter.

6.5.1 Indirect Spatial Metadata Enrichment Model

The indirect model is concentrated on monitoring search words that are utilised by the spatial data users, analysing them and then employing the most popular ones to enrich the content of descriptive keyword metadata element in an automated manner. Therefore, this model is streamlined in three stages:

1. Monitoring Search Word
2. Recording Search Word
3. Assigning Search Word.

The stages are here described below.

- **Stage 1: Monitoring Search Word**

The spatial data catalogue systems consisting of data and metadata repositories (distributed or centralised) such as the GeoNetwork opensource, ASDD, and the INSPIRE Geoportal typically provide the end users with the services to discover and access the data. As mentioned earlier, through these systems one of the common ways of data discovery is to query the metadata records via search words. Then, the discovery service will search and retrieve all the corresponding metadata records with that search word. The end users will be able to view the results by opening the metadata files and deciding on which data is more suitable for their needs. Finally, through the access service they will be able to access the required data or be aware of the access and use policies.

The main aim of Stage 1 (Figure 6-14) is to automatically identify the relevant search words (on a scale of highly relevant to low relevant) against retrieved metadata records.

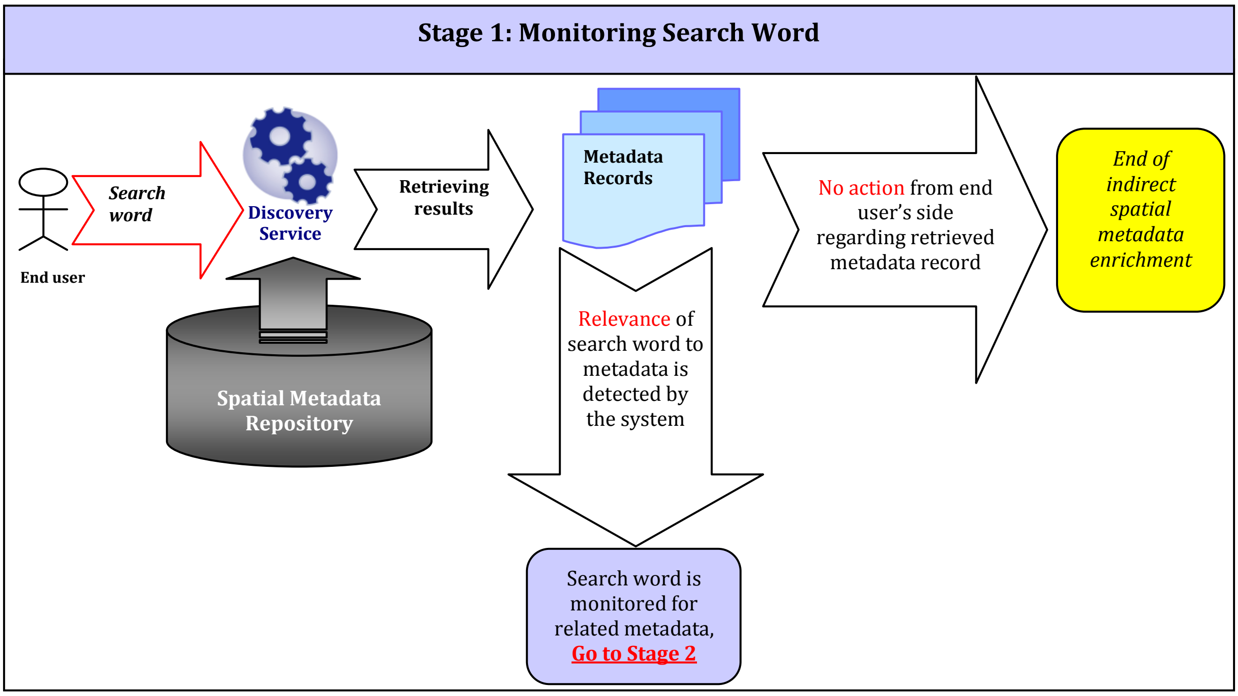


Figure 6-14: Stage 1 of the indirect spatial metadata enrichment model

The identification of the search words' degree of relevance would be identified based on the end users' behaviours during the discovery process. Monitoring the end users' behaviours might be undertaken in different ways based on the functionalities supported by any spatial data discovery system. For instance, when a user searches for a dataset within the GeoNetwork opensource catalogue it provides the user with a list of metadata records. Each retrieved metadata record is provided with the title and abstract of the dataset. The user is also allowed to see complete metadata records by clicking a button, namely 'Metadata' (Figure 6-15). Clicking this button by the end user means that the end user is interested to see more information about the retrieved metadata record and the search word is somehow relevant to this metadata record. Also, there is another button, namely 'Download' within the GeoNetwork metadata retrieval page to access the actual dataset (again, Figure 6-15). Clicking this button by the end user confirms that the search word is most likely relevant to the retrieved metadata record.

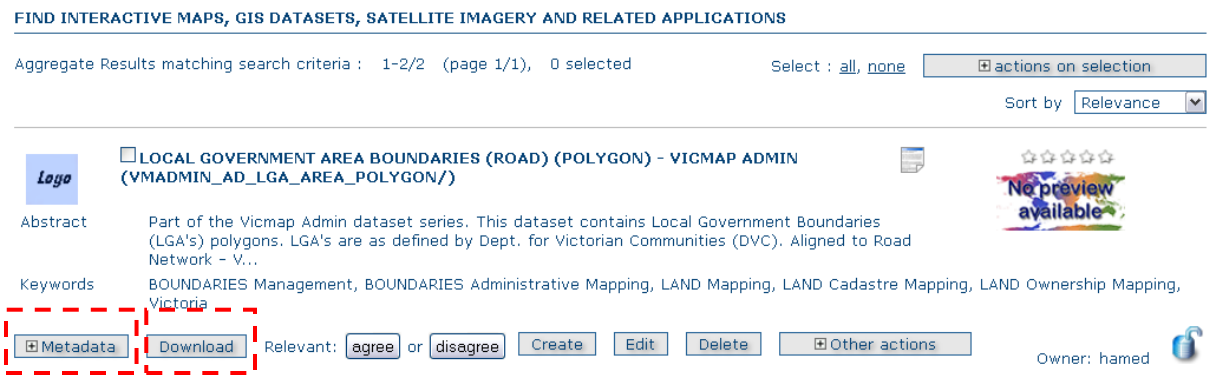


Figure 6-15: Some of the functionalities provided by the GeoNetwork opensource catalogue which can be used for monitoring the end users' behaviour

Also, Table 6-4 outlines the scale suggested for monitoring the relevance of search words to retrieved metadata within the GeoNetwork opensource catalogue. As already mentioned, the end users' behaviours regarding retrieved metadata records might be different in various data discovery systems. However, the function of accessing more information about a metadata record has been implemented within most of these systems.

Table 6-4: The scale suggested for monitoring the relevance of search words to retrieved metadata within the GeoNetwork opensource catalogue

User's behaviour	Relevance Scale
No action from end users' side regarding retrieved metadata record	Not Relevant
Clicking the 'Metadata' button next to retrieved metadata record to view the complete metadata and explore more information about the dataset	Low Relevant
Spending 'Enough' time (in the prototype system 3 minute time has been defined as enough time) with the metadata details open for retrieved metadata record	Somewhat Relevant
Clicking the 'Download' button next to retrieved metadata record to access the dataset	Highly Relevant

As a result, in this stage of the indirect enrichment model any used search word from 'Not Relevant' to 'Highly Relevant' is monitored by the system to be used in the automatic spatial metadata enrichment approach. The search words identified as 'Highly Relevant' to 'Low Relevant' are used and sent to Stage 2 of the indirect enrichment model. Also, the search words identified as 'Not Relevant' together with the 'Relevant' search words will be recorded by the system to create a suggestion list of search words which provides the subsequent users with previously used search words while typing a search word in the spatial data discovery search box.

- **Stage 2: Recording Search Word**

In this stage, any search word that has been automatically identified as 'Relevant' (the whole scale outlined in Table 6-4 excluding 'Not Relevant') to any retrieved metadata record in Stage 1, would be recorded and stored in a temporary table (Figure 6-16).

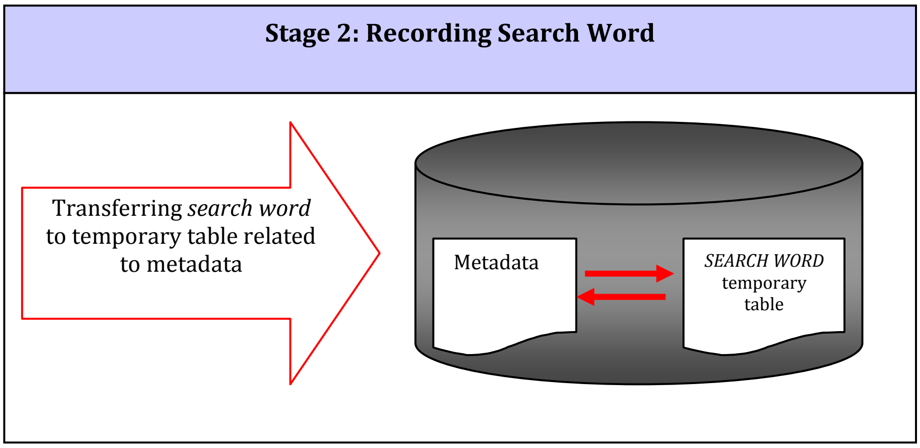


Figure 6-16: Stage 2 of the indirect spatial metadata enrichment model

This temporary table is related to the corresponding metadata records through a foreign key namely 'Metadata UUID[28]', which contains the unique identifier for each metadata record. The temporary table also contains other columns including 'Dataset Title', 'Search Word', 'Number of Repetition', and 'Weight'. Table 6-5 outlines the description of these columns.

Table 6-5: The columns of the temporary table to record relevant search words

Column	Description
UUID	Unique identifier of relevant metadata record
Dataset Title	Title of the relevant dataset
Search word	Search word monitored and recorded by the system during the data discovery process
Number of Repetition	Number of times the search word has been identified as 'relevant' (on a scale of highly relevant to low relevant) by the system
Weight	Sum of weights which have been assigned to search word systematically based on the end users' behaviours

Depending on the end users' behaviours during the dataset discovery process, a specific weight needs to be considered for recording the search word that is relevant to the retrieved metadata. There have been different methods for designing a weighting system for the purpose of decision making such as the Analytic Hierarchy Process (AHP) and its generalisation, the Analytic Network Process (ANP), as discussed by Saaty (2004). However, because the main focus of this research was on testing the use of Web 2.0 features to involve the end users and improve the content of spatial metadata, the investigation into these methods was considered out of the scope of research. Therefore, in this stage a weighting system was suggested for the GeoNetwork opensource catalogue to automatically assign different weights to each relevant search word following how the user interacts with the retrieved metadata records (according to Table 6-4). Table 6-6 presents this weighting system.

[28] Universally Unique Identifier

Table 6-6: Weighting system for recording search word within the GeoNetwork opensource catalogue based on the indirect metadata enrichment model

End user's behaviour	Proposed weight
Clicking the 'Metadata' button to explore more information about the dataset	0.25
Spending 'Enough' time (in the prototype system 3 minute time has been defined as enough time) with the metadata details open	0.50
Clicking the 'Download' button to access the dataset	0.75

Accordingly, the main aim of this stage can be defined as:

> *'Automatically recognising the popularity of the same search words that different users agree on them for finding the specific spatial datasets.'*

Increasing use of the same search word illustrates its popularity.

- ## Stage 3: Assigning Search Word

Among the search words recorded in Stage 2, any that meet a specific threshold of weight will be assigned to their spatial metadata records and stored in the descriptive keywords element (Figure 6-17). This specific threshold can be defined depending on the spatial data responsible parties' decision.

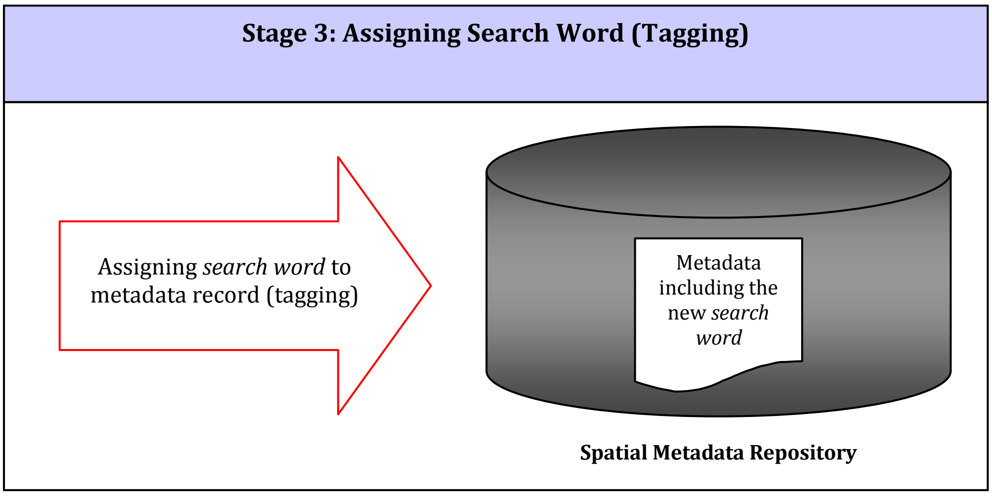


Figure 6-17: Stage 3 of the indirect spatial metadata enrichment model

As already mentioned, the assigned search words will be shown in a tag cloud within the data discovery system. For instance, Figure 6-18 illustrates the search words that might be assigned to a road dataset.

Road 1:2000 transport node freeways **network** streets
transportation highways **road network** address Centreline

Figure 6-18: An example tag cloud for a road dataset

The search words are shown in the tag cloud differently following their weights and indicate their popularity as well. Within the tag cloud, the tagged search words that are used more frequently by the end users will be highlighted and shown in a bold format (such as 'road network' in Figure 6-18). Also, the tagged search words are hyperlinked to their actual metadata records. By clicking each search word, its relevant metadata record will be popped up quickly in the data discovery user interface.

In addition to the search words recognised automatically using the indirect enrichment model, the end users will be provided with some means to rate the existing tagged search words and add values to metadata keyword element directly through the direct enrichment model. This model is explained as following.

6.5.2 Direct Spatial Metadata Enrichment Model

Contrary to the indirect metadata enrichment model, the end users are able to undertake the following tasks using the direct enrichment model:

- Tag a dataset with words they feel best describes what it is about.
- Agree/disagree with the relevance of their used search word of the retrieved metadata – through responding to an embedded question in the data discovery interface next to the retrieved metadata record.
- Agree/disagree with the relevance of formerly tagged search words (by previous users) of the retrieved metadata – through responding to an embedded question in the data discovery interface next to the retrieved metadata record.

Accordingly, the end users will be involved in the metadata enrichment process and the knowledge and experiences of the users regarding the spatial datasets will be shared. Moreover, the tagging process will help the users easily and quickly find their tagged datasets again within the data discovery system.

Similar to the indirect metadata enrichment model, a weighting system is proposed for recording the search words in the direct enrichment model. This weighting system automatically assigns different weights to each tagged search word following how the end user interacts with the retrieved corresponding metadata records. Table 6-7 outlines the weighting system suggested for the GeoNetwork opensource catalogue based on the direct metadata enrichment model.

Table 6-7: Weighting system for recording search word within the GeoNetwork opensource catalogue based on the direct metadata enrichment model

End user's behaviour	Proposed weight
Adding a new tag to the retrieved metadata record	1.0
'Agreeing' with an existing tagged search word related to the retrieved metadata record	1.0
'Disagreeing' with an existing tagged search word related to the retrieved metadata record	-1.0

As a result of this model, the end users will tag the datasets according to their awareness of dataset and also the intention of using that dataset which is usually related to their circumstances. These tags are visualised in the same tag cloud used in the indirect enrichment model. For instance, the tagging process for 1:2000 map of Melbourne city can be imagined here (Figure 6-19). The end users of this map can be Melbourne citizens, tourists, students, or decision makers from different organisations.

Melbourne Melbourne CBD Melbourne parcels Melbourne roads
Melbourne buildings 2000 Melbourne streets Melbourne map
Melbourne city 2012 Melbourne bus stops Melbourne railway
Melbourne freeways Melbourne tollways Melbourne parks Melbourne POI
Melbourne dataset 1:2000 Melbourne libraries Melbourne universities
1/2000 Melbourne Melbourne museums

Figure 6-19: Example tag cloud for Melbourne map 1:2000

Using the direct metadata enrichment model could also facilitate tagging the same new search word to a group of metadata records describing the same type of dataset. For instance, the 'cadastre' search word could be tagged to the metadata of cadastral maps of various areas. However, in order to implement this capability the following preconditions need to be considered:

1. In the spatial data catalogue, there are metadata records that describe the same type(s) of spatial data. For instance, in an Australian spatial data catalogue (e.g. ASDD) there are metadata records for cadastral maps of different states.
2. The spatial metadata records stored in the data catalogue are classified into different groups based on their topic. For instance, the metadata records for cadastral maps are stored in 'Land' group.

Once the above preconditions are provided, a new search word that has been assigned to a dataset (in a specific group) and has met the threshold might be assigned to other datasets in the same group. Assigning the search word to the datasets in the same group needs to be undertaken by the end users through the suggestions provided by the data catalogue system. For instance, when an end user is assigning a new search word (e.g. 'cadastre') to a dataset entitled 'Victorian cadastral map' the data catalogue system suggests him/her to assign the same keyword to other datasets stored in the group containing the Victorian cadastral map (such as 'Queensland cadastral map').

In addition to a generation of a tag cloud for each spatial dataset, a common tag cloud will also be generated within the data discovery system. It will consist of the most popular search words already assigned to spatial datasets using the indirect and direct models. Using this tag cloud as a discovery option, the end users will be able to review the most popular and previously successful and agreed search words and select any that are also suitable for their needs to retrieve the associated datasets promptly and simply. For instance, Figure 6-20 illustrates a sample tag cloud for Australian Spatial Data Discovery (ASDD) portal. As shown in this example, the popular search words are gathered from separate metadata records and stored in a single tag cloud that will help the end users to review the existing popular search words at a glance and then choose them based on their own requirements.

Figure 6-20: Example tag cloud for the ASDD

In this tag cloud, the end users will be able to retrieve the metadata associated to any tag through clicking on the tag that acts like a hyperlink. In addition, the end users can have the possibility to visualise the tags which are assigned to the same dataset by hovering the mouse over a tag in the cloud. Figure 6-21 illustrates this concept.

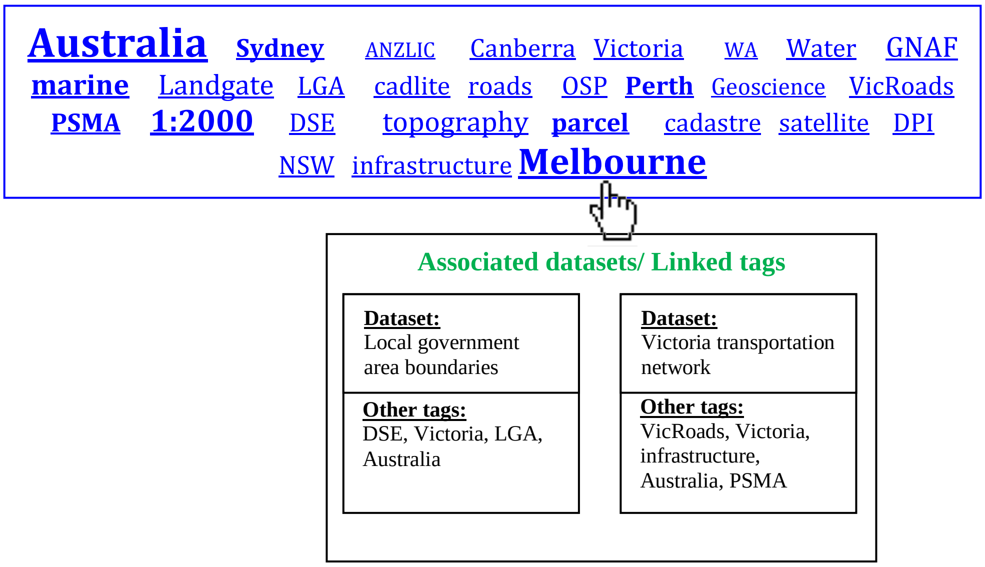


Figure 6-21: Visualising a search word's associated datasets and linked tags

One of the critical considerations for developing the automatic spatial metadata enrichment approach would be the involvement of the spatial data responsible parties to assess the recorded search words and remove the noise and spam before they are assigned to the metadata records as new 'metadata keyword' values. As discussed in Section 5.2.1 in Chapter 5, the case study investigations confirmed that most of the organisations in the case study region – Australia prefer to examine any automatically generated and updated metadata before the metadata is stored and published. The assessment of recorded search words within the automatic spatial metadata enrichment approach will result in the prevention of assigning the ones with irrelevant meanings to the metadata records. In other words, the data responsible parties will filter out the unrelated and meaningless search words and agree with the relevant ones to be visualised in the tag cloud.

6.6 Chapter Summary

This chapter developed a framework to address the identified challenges regarding spatial metadata management and automation arisen from the literature review, case study investigations and selected tools assessment. As part of this framework, the chapter designed a generic spatial data lifecycle based on the recommended information lifecycle by the Australian Government Information Management Office (AGIMO) and then mapped the ISO 19115: 2003 metadata elements against this generic lifecycle. Based on the results of mapping metadata elements, a 'lifecycle-centric spatial metadata creation approach' to create metadata along with the spatial data lifecycle steps was designed. As another part of the framework, the chapter designed an 'automatic spatial metadata updating (synchronisation) approach' rooted in a GML-based integrated data model for dataset and metadata storage. This approach would be adopted within the spatial data lifecycle for updating spatial data and metadata in real time. Moreover, the technical requirements for implementing this approach were discussed in detail. Finally, as the last part of the framework this chapter designed an 'automatic spatial metadata enrichment approach' based on Web 2.0 features. This approach aimed to improve the content of keyword metadata element through the interaction with the end users seeking spatial data over the Web.

According to the approaches designed in this chapter for automatic spatial metadata updating and enrichment, the next chapter will explore the implementation of two prototype systems in order to prove the concept of designed approaches.

CHAPTER 7

AUTOMATIC SPATIAL METADATA UPDATING AND ENRICHMENT PROTOTYPE SYSTEMS IMPLEMENTATION

7 AUTOMATIC SPATIAL METADATA UPDATING AND ENRICHMENT PROTOTYPE SYSTEMS IMPLEMENTATION

7.1 Introduction

This chapter discusses the implementation of the automatic spatial metadata updating and enrichment prototype systems to address the fifth research objective. The chapter first explores the identified use cases, implementation architecture and capabilities of the automatic spatial metadata updating prototype system integrated with the GeoNetwork opensource catalogue. Next, the chapter reviews the identified use cases, implementation architecture and capabilities of the automatic spatial metadata enrichment prototype system within two different environments: the GeoNetwork and Model Information Knowledge Environment (MIKE). The development of both indirect and direct metadata enrichment models within these two environments will be discussed in this chapter.

7.2 Automatic Spatial Metadata Updating (Synchronisation) Prototype System

This section aims to explore the implementation of a prototype system based on the conceptual design and technical requirements of the automatic spatial metadata updating (synchronisation) approach discussed in Section 6.4, Chapter 6. In fact, proving the new GML-based metadata synchronisation concept proposed in this thesis is the main purpose of implementing this prototype system.

Referring to the research design, discussed in Section 4.4 in Chapter 4, the prototype system was decided to be integrated with the GeoNetwork opensource catalogue for a number of reasons. Firstly, the GeoNetwork is widely used within the case study region – Australia according to the results of the case study investigations. Secondly, the GeoNetwork has been used in Australia as the main platform for building up spatial metadata tools such as BlueNet MEST (ANZ-MEST), as discussed in Section 5.3.2 in Chapter 5. In addition, since the GeoNetwork is open source software the contributions to this software can be shared among the GeoNetwork community of users and developers, so that others can use and build upon the work provided by this thesis. Moreover, the GeoNetwork currently lacks functionality for the automatic updating of metadata records in conjunction with dataset modification (Rajabifard *et al.* 2009).

With the purpose of integration of the automatic spatial metadata updating prototype system with the GeoNetwork, apart from the technical requirements mentioned in Section 6.4.2 in Chapter 6 it was required to develop a new Web service (entitled 'SYNC') to synchronise the metadata catalogues stored in both GeoNetwork and spatial databases via CSW update operations. Also, the prototype system and GeoNetwork interfaces integration would be performed through adding a visual modification facility (a new button) to the GeoNetwork metadata update interface. This facility enables the GeoNetwork users to simply access the prototype system interface.

The SYNC service would be called from both prototype system and GeoNetwork interfaces. On the first interface, it would be linked to the 'Save' button (Figures 7-1), so that by pressing this button the following two requests would be sent to the servers respectively:

- A WFS-T request through mapping software application (e.g. OGC compliant map server) to update the dataset and metadata in the spatial database.
- A CSW update request through the GeoNetwork CSW server to synchronise the metadata record in the GeoNetwork database with its corresponding record in the spatial database.

On the GeoNetwork metadata update interface, the SYNC service would be linked to the 'Save' and 'Save and close' buttons (again, Figure 7-1), so that by pressing each of these buttons the following request would be sent to the CSW server:

- CSW update request to synchronise the metadata record in the spatial database with its corresponding record in the GeoNetwork database.

Figure 7-1 illustrates the role of SYNC service within the context of automatic spatial metadata updating prototype system.

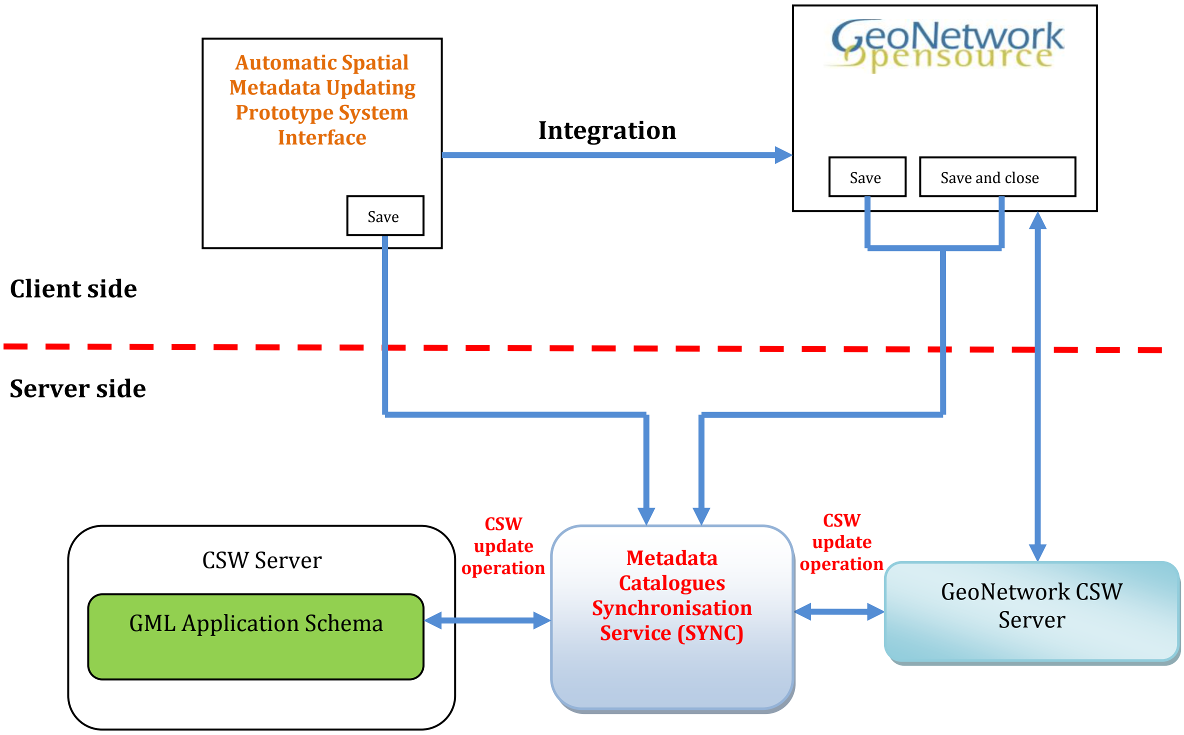


Figure 7-1: The role of metadata catalogues synchronisation service (SYNC) within the automatic spatial metadata updating prototype system

According to the conceptual design of the automatic spatial metadata updating approach and also the need for integrating this approach with the GeoNetwork environment, the following nine use cases were defined for implementing a prototype system:

Use case 1 – user stores the spatial dataset and the ISO 19115: 2003 compliant metadata using an integrated data model and also inserts the corresponding metadata catalogue into the GeoNetwork database. The integrated data model needs to be developed based on the components of an integrated data model already discussed in Section 6.4.2 in Chapter 6; including spatial database, dataset and metadata tables, relationship between these tables and a middleware to accommodate the dataset and metadata (using the GML). Borehole Geology (boresg) dataset provided by the Victorian Department of Sustainability and Environment (DSE) and its related metadata were selected for the purpose of this use case.

Use case 2 – user logs in to the GeoNetwork opensource catalogue and searches for the vector dataset, which is supposed to be modified (boresg dataset).

Use case 3 – once the boresg dataset metadata is found and retrieved the user clicks on the ‘Edit’ button underneath the metadata to retrieve the GeoNetwork metadata update interface. This button is already included in the GeoNetwork data discovery interface.

Use case 4 – once the metadata update interface is opened the user clicks on a button namely 'Map-based Modification' to open the 'automatic spatial metadata updating prototype system' interface. This button needs to be added to the GeoNetwork metadata update interface. The button will be the connector for the prototype system interface with the GeoNetwork environment.

Use case 5 – through the prototype system interface the user views both features and metadata for the same dataset (boresg) and modifies the features using an 'Edit Toolbar' and clicks the 'Save' button when finished.

Use case 6 – once the 'Save' button is clicked, user sees the updated metadata values in real time for the following elements based on the changes to the dataset:

- Metadata date stamp
- Date of last update
- Lineage statement
- North bounding latitude
- South bounding latitude
- West bounding longitude
- East bounding longitude

These metadata elements were chosen to be employed in the prototype system to prove the concept of automatic spatial metadata updating approach. Of these elements, the lineage statement metadata element documents the history of dataset changes based on the following rule.

When (date and time) + Who (GeoNetwork user) + What action (Updated/Added/Deleted) + Which feature (of the dataset)

Through this use case not only the metadata values stored in the spatial database will be updated automatically and simultaneously but also the corresponding metadata values stored in the GeoNetwork database will be synchronised in real time.

Use case 7 – user is notified about the search words that have been recorded for the metadata using the automatic spatial metadata enrichment prototype system and makes a decision for accepting them or rejecting the meaningless search words or spam. The automatic spatial metadata enrichment prototype system will be explored in Section 7.3. Through this use case, both prototype systems developed in this thesis will be integrated and the user responsible for updating dataset and metadata will be able to assess the 'keyword' metadata values enriched by the end users.

Use case 8 – user closes the prototype system interface when dataset modification is finished and returns to the GeoNetwork metadata update interface. By clicking the 'Reset' button, already provided by the GeoNetwork, the synchronised metadata values affected by the dataset modification will be retrieved in this interface.

Use case 9 – optionally, user edits the metadata values directly in the GeoNetwork metadata update interface and clicks the 'Save' or 'Save and close' button and both metadata catalogues in the GeoNetwork and spatial databases will be synchronised.

For addressing the above use cases, the next section designs and develops the prototype system implementation architecture.

7.2.1 Implementation Architecture

The architecture designed for implementation of the automatic spatial metadata updating (synchronisation) prototype system is illustrated in Figure 7-2. As it can be seen, this architecture contains three main layers (Olfat *et al.* 2012a):

- Storage Layer
- Service Layer
- Application Layer.

The storage layer includes two main databases to store the dataset and metadata: GeoNetwork and spatial databases. The service layer contains three Web servers that support the OGC Web services (WFS-T and CSW) and SYNC service. The application layer provides the users with a user-friendly graphical interface integrated with the GeoNetwork metadata update interface to modify datasets to make the real-time metadata updating happen.

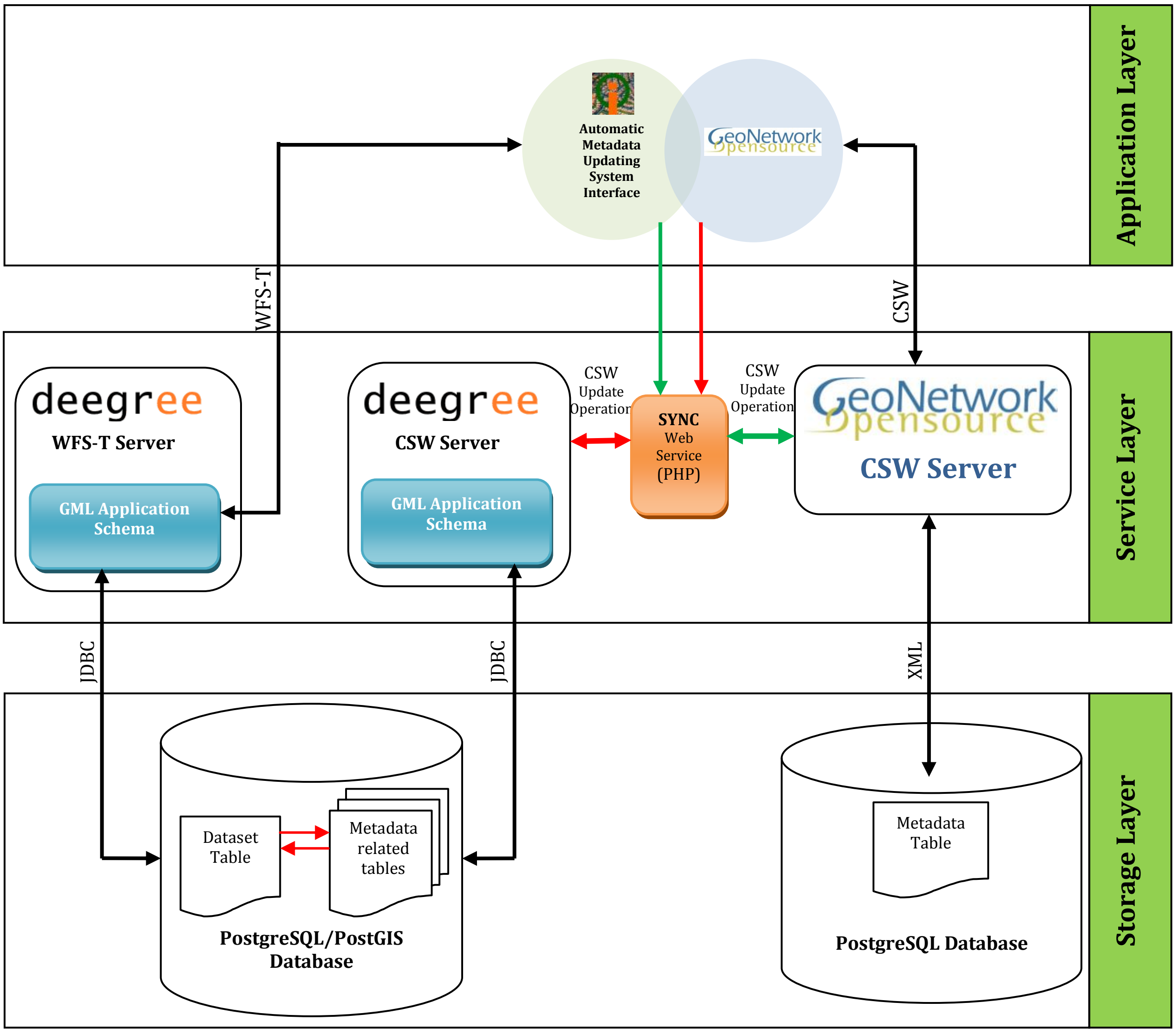


Figure 7-2: Implementation architecture for automatic spatial metadata updating prototype system to address the use cases (Olfat *et al.* 2012a)

The characteristics of these layers are now described in detail:

- ## Storage Layer

This layer provides the databases to store spatial data and metadata for the prototype system. As described earlier, there are two main databases that are created and maintained in the storage layer. As one of the major requirements for developing the prototype system has been to use the open source technologies, the combination of 'PostgreSQL' and 'PostGIS' was chosen to build the storage layer. PostgreSQL is a powerful, open source object-relational database system (PostgreSQL 2011) which is spatially enabled though PostGIS (PostGIS 2011). The following databases are then created:

➢ GeoNetwork Database

The aim of creating this database was providing a database to be connected to the GeoNetwork CSW server and enable the users to store and maintain metadata records through the GeoNetwork interface. Also, the GeoNetwork database would be used by the 'SYNC' service to synchronise the metadata records in both databases. In fact, the GeoNetwork database has been considered to address:

- part of the first use case (see page 199) for inserting a sample metadata catalogue
- part of the sixth use case (see page 200) for storing the updated metadata values affected by the dataset modification
- the eighth use case (see page 201) for retrieving the updated metadata values in the GeoNetwork metadata update interface
- the last use case (see page 201) for editing the metadata values directly from the GeoNetwork metadata update interface.

By default, the GeoNetwork installs with its database configured to McKoi SQL (an open source SQL database system written in Java (Mckoi 2011)) and following the implementation architecture (see Figure 7-2) this database was migrated to PostgreSQL. GeoNetwork database includes different tables such as 'metadata' (for storing metadata records), 'settings' (for storing the GeoNetwork configurations), and 'users' (for storing the users' information). Among these tables, 'metadata' table has been used within the prototype system implementation architecture for synchronising metadata catalogues stored in both GeoNetwork and spatial databases.

Metadata catalogue of the sample dataset (Borehole Geology (boresg) dataset provided by the DSE, Victoria) was inserted into the GeoNetwork database using the CSW insert operation.

➢ Spatial Database

This database was created to store both spatial dataset and metadata tables in the same environment to address the first use case (see page 199) and to build three of the four required components of an integrated data model already discussed in Section 6.4.2 in Chapter 6 (including spatial database, dataset and metadata tables, and relationship between the tables). Also, the spatial database would connect to the deegree Web servers (deegree will be discussed in the next layer) as well as the SYNC service to address the sixth use case (see page 200) relating to updating metadata at the same time with dataset modification.

To create the spatial database the following two major requirements were identified:

- Requirement 1 – according to the ISO/TS 19139: 2007 Standard, the spatial metadata records should be stored in the XML structure. In order to relate these metadata to their relevant datasets for achieving the real-time metadata updating in conjunction with any change to the dataset, it was required to store metadata in a relational database such as PostgreSQL. To do so, there was a need for a data model that could map metadata values from an XML structure to relational tables.

- Requirement 2 – the database that maintains this data model should be supported by a CSW Web server in order to insert, update, and delete the metadata values.

Accordingly, among the existing open source Web servers supporting the OGC services, 'deegree-CSW' was the only choice that could meet the above requirements at the stage of this research. A sample database was provided by deegree to store metadata records that could be accessible once deegree-CSW was installed. This database was supported by a GML Application Schema (csw_postgres.xsd) that maps the metadata elements to the database tables and columns. Among the metadata tables, 'md_metadata' was the root table. In this table, each metadata record included a unique 'id' and was related to one record in the 'fileidentifier' table, which stored the unique value of the file identifier for each metadata record.

Although these tables had no physical relationships with each other in the database the relationships between them were defined in the GML Application Schema.

In this layer, in addition to metadata tables created in the spatial database, a sample dataset table (boresg) was inserted into the database to build the integrated data model. The metadata of this sample dataset, which was already provided by the DSE, was also imported in the database through a CSW insert operation.

To make the relationship between the dataset table and the 'md_metadata' table, it was required to create a new column in the dataset table as the foreign key to the 'md_metadata' table. It was not necessary to add a constraint to the dataset table for this foreign key, this relationship was defined in the 'Dataset + Metadata' GML Application Schema. This schema was entitled 'DataMetadata.xsd' and was different from csw_postgres.xsd. The schema was specifically developed as the complementary component of the integrated data model for addressing the first use case of the automatic metadata updating prototype system (see page 199). It aimed at mapping the dataset and metadata elements to GML and connecting to deegree-WFS-T server for the purposes of dataset modification and real-time metadata updating. This schema will be explained in further detail in the next section.

- **Service Layer**

In order to address the use cases 1, 6 and 9 (see pages 199 – 201), this layer aims to provide the required OGC compliant Web services for the following groups of operations:

1. inserting, deleting, and updating metadata records in the GeoNetwork database through the CSW service
2. inserting, deleting, and updating the spatial dataset and its related metadata in the spatial database through the WFS-T service
3. inserting, deleting, and updating metadata records in the spatial database through the CSW service
4. synchronising the same metadata catalogues stored in the GeoNetwork and spatial databases.

The existing Web servers in the geospatial community that could handle the first three operations were employed for the prototype system development. However, for the last operation a new service namely 'SYNC' was designed and developed in this research. The Web servers and related services used in this layer are detailed as following.

➢ **GeoNetwork CSW Server**

Once the GeoNetwork is installed, it contains the CSW server which addresses the first group of required operations related to the service layer development. This server supports the OGC CSW service version 2.0.2. The CSW operations can be accessed using POST and GET methods, as well as, SOAP[29] encoding and includes GetCapabilities, DescribeRecord, GetRecordById, GetRecords and Transaction (GeoNetwork 2011a).

In order to install the GeoNetwork, its WAR[30] file version 2.6.3 was downloaded from the GeoNetwork Website[31] and put into the Web application folder of the OpenGeo Suite version 2.4.3 to be compiled. The OpenGeo Suite is a Web mapping platform built on powerful, cutting-edge, and open source geospatial components (OpenGeo 2011) and was selected for compiling the GeoNetwork for two main reasons. Firstly, this platform installed with the PostgreSQL and PostGIS required for the prototype system development and secondly, it provided the proxy needed for OpenLayers (used in the application layer). Some of the tasks that OpenLayers performed (including the WFS requests) required the use of a proxy script because of restrictions in JavaScript on the use of XMLHTTPRequest making requests to remote servers.

---

[29] Simple Object Access Protocol
[30] Web Application Archive
[31] http://geonetwork-opensource.org/

➢ WFS-T Web Server

According to the third requirement for implementing the automatic metadata updating approach (see Section 6.4.2 in Chapter 6), there was a need to define an open source server that supports the compliant OGC WFS-T for complex features. As a result, an investigation was undertaken into the following open source servers which supported the OGC compliant Web services for simple and complex features. The aim of this study was choosing the appropriate server to address the second group of required operations related to the service layer development (see page 205).

GeoServer

The GeoServer is a Java-based software server that allows users to view and edit spatial data. This server, through the Application Schema support (app-schema) extension, provides support for complex features in the GeoServer WFS (GeoServer 2011a), as illustrated in Figure 7-3. However, it was realised that there was no support for the WFS-T (update, insert, and delete capabilities) for complex features in the GeoServer at the stage of this research.

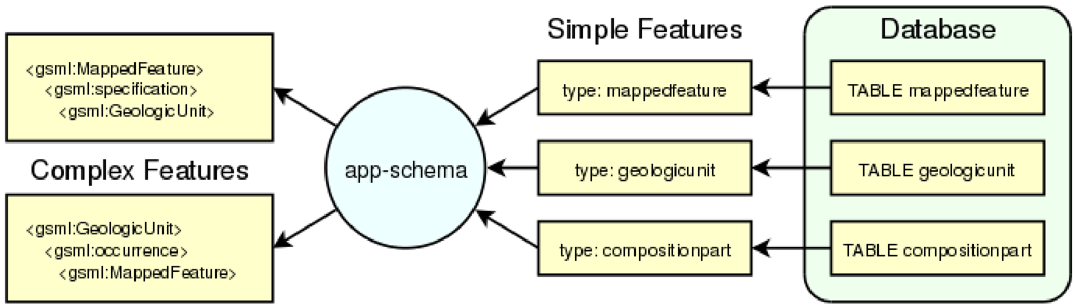


Figure 7-3: Three tables in a database are accessed using GeoServer simple feature support and converted into two complex feature types (GeoServer 2011a)

deegree

deegree is Free Software protected by the GNU Lesser General Public License (GNU LGPL) based on a Java Framework of which its entire architecture is developed using standards of the OGC and ISO/TC 211 (deegree 2010b). WFS requests supported by deegree include GetCapabilities, DescribeFeatureType, GetFeature, Transaction for Simple and Complex features, GetFeatureWithLock and LockFeature. Figure 7-4 shows the overall architecture of the modules involved in deegree WFS.

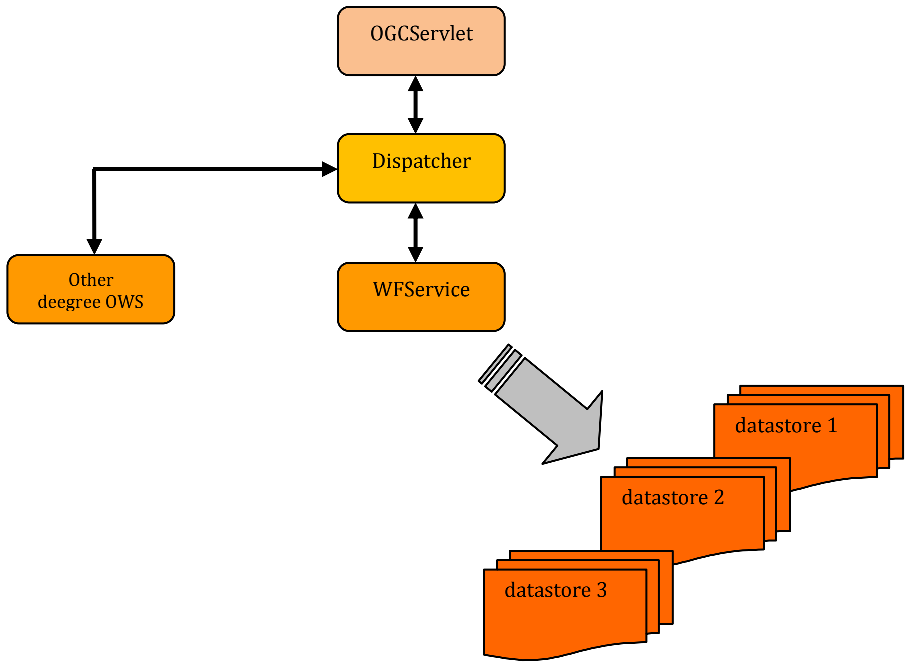


Figure 7-4: deegree WFS architecture overview (deegree 2010b)

Once the capabilities of the above servers were compared against the prototype system's technical requirements, deegree was selected to be utilised as the WFS-T server. According to the deegree's recommendation, Apache Tomcat, which is an open source software implementation of the Java Servlet and JavaServer Pages technologies (Apache 2011), was chosen as the Web server to host deegree WFS-T.

After setting up the Web server, a GML Application Schema ('DataMetadata.xsd' mentioned earlier) was developed within deegree server to address the following purposes:

- including the characteristics of spatial dataset, as well as, its related metadata elements
- mapping the relationships between dataset and metadata (complex features)
- mapping the spatial and non-spatial metadata elements to the related tables/columns in the database
- setting the WFS-T capabilities for each element.

In fact, data sources (datasets and metadata) must be assigned to feature types in order to be servable by the WFS. A data store was configured by specifying a GML Application Schema that defined the structure of the served feature types, as well as, access and mapping information to the physical data source. On start up, the deegree WFS scans certain directories for .xsd files. The feature types defined in the 'annotated GML Application Schema' are backed by the respective data source and served by the WFS.

The most complex and sophisticated part of configuring deegree WFS was to create data store configurations. This was the process of mapping feature types (and their properties) to the entities of a physical data source (sample ‘boresg’ dataset table and related metadata tables). These basic concepts recommended by deegree (2010b) were considered during the GML Application Schema development:

- The configuration of a data store has to be a valid GML 3 application schema document. The mapping information is contained in annotation elements.
- There is a 1:1 relation between feature types and the entities of a data source (table), but it is possible to map properties of a feature type to a related table. This allows, for example, properties that have cardinality > 1.
- Feature types can be complex, i.e. one or more properties can be defined to contain features themselves. This allows complex hierarchies of feature types that may even be recursive (a feature type can contain features of the same type as sub-features).
- In order to map, the table name should be UPPERCASE and the feature type names should be in UpperCamelCase. The properties and property types should be in lowerCamelCase (this is a proposed convention, but it is not mandatory).

For instance, Figure 7-5 gives an overview from a sample part of the GML Application Schema (DataMetadata.xsd) set up for the ‘boresg’ dataset according to deegree WFS instructions. It shows the method to define a dataset element and its WFS-T settings and the way to map the dataset attributes and geometry to the database and most importantly, it illustrates how to map the dataset metadata to its content root element (MD_Metadata).

Defining Dataset

```
element name="Boresg" type="app:BoresgType" substitutionGroup="gml:_Feature">
xsd:annotation>
  <xsd:appinfo>
    <deegreewfs:table>BORESG</deegreewfs:table>
    <deegreewfs:gmlId prefix="BORESG_">
      <deegreewfs:MappingField field="GID" type="INTEGER"/>
      <deegreewfs:IdGenerator type="DB_SEQ">
        <deegreewfs:param name="sequence">boresg_gid_seq</deegreewfs:param>
      </deegreewfs:IdGenerator>
      <deegreewfs:IdentityPart>false</deegreewfs:IdentityPart>
    </deegreewfs:gmlId>
    <deegreewfs:visible>true</deegreewfs:visible>
```

WFS-T Settings

```
    <deegreewfs:transaction update="true" delete="true" insert="true"/>
    <deegreewfs:DefaultSRS>EPSG:4326</deegreewfs:DefaultSRS>
  </xsd:appinfo>
</xsd:annotation>
</xsd:element>
<xsd:complexType name="BoresgType">
  <xsd:complexContent>
    <xsd:extension base="gml:AbstractFeatureType">
      <xsd:sequence>
```

Mapping Attributes

```
        <xsd:element name="gid" type="xsd:integer">
        <xsd:element name="localname" type="xsd:string">
```

Mapping Metadata Element

```
        <xsd:element name="metadata" type="gml:FeaturePropertyType" minOccurs="0" maxOccurs="1">
          <xsd:annotation>
            <xsd:appinfo>
              <deegreewfs:IdentityPart>false</deegreewfs:IdentityPart>
              <deegreewfs:Content type="app:MD_Metadata">
                <deegreewfs:Relation>
                  <deegreewfs:From fk="true">
                    <deegreewfs:MappingField field="FK_MD_METADATA" type="INTEGER"/>
                  </deegreewfs:From>
                  <deegreewfs:To>
                    <deegreewfs:MappingField field="ID" type="INTEGER"/>
                  </deegreewfs:To>
                </deegreewfs:Relation>
              </deegreewfs:Content>
            </xsd:appinfo>
          </xsd:annotation>
        </xsd:element>
```

Mapping Geometry

```
        <xsd:element name="the_geom" type="gml:GeometryPropertyType" minOccurs="0" maxOccurs="1">
          <xsd:annotation>
            <xsd:appinfo>
              <deegreewfs:IdentityPart>false</deegreewfs:IdentityPart>
              <deegreewfs:Content>
                <deegreewfs:SRS>EPSG:4326</deegreewfs:SRS>
                <deegreewfs:MappingField field="THE_GEOM" type="GEOMETRY" srs="4326"/>
              </deegreewfs:Content>
            </xsd:appinfo>
          </xsd:annotation>
        </xsd:element>
      </xsd:sequence>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>
```

Figure 7-5: GML Application Schema (DataMetadata.xsd) designed for the 'boresg' dataset to address the first use case (explained on page 199)

This GML Application Schema is applied by the deegree WFS-T server as a feature type that directs the server to present relevant data for a dataset and its metadata from the database at the client side, as well as, to reflect and store the changes (updated metadata values right after dataset modification) at the server side. As a result, the GML Application Schema is a significant part of metadata integrated data model developed in this research (to address the first use case explained on page 199).

Although deegree WFS-T was beneficial for automatic metadata updating approach implementation, it suffered from a limitation for connecting to several databases. Only one database was allowed to be defined in the GML Application Schema as the back end. The way that PostgreSQL database was defined in the schema is shown below:

```
<xsd:annotation>
        <xsd:appinfo>
                <deegreewfs:Prefix>app</deegreewfs:Prefix>
                <deegreewfs:Back end>POSTGIS</deegreewfs:Back end>
                <deegreewfs:DefaultSRS>EPSG:4326</deegreewfs:DefaultSRS>
                <JDBCConnection xmlns="http://www.deegree.org/jdbc">
                        <Driver>org.postgresql.Driver</Driver>
                        <Url>jdbc:postgresql://localhost:54321/csw</Url>
                        <User>postgres</User>
                        <Password>postgres</Password>
                        <SecurityConstraints/>
                        <Encoding>iso-8859-1</Encoding>
                </JDBCConnection>
        <deegreewfs:SuppressXLinkOutput>false</deegreewfs:SuppressXLinkOutput>
        </xsd:appinfo>
</xsd:annotation>
```

Since one of the initial assumptions was to store datasets and metadata in a common database to achieve the integrated metadata data model the above limitation did not affect the prototype system implementation.

Another important consideration for database mapping in the GML Application Schema was to avoid the duplication of metadata for each dataset feature (record). In this regard, the ‘deegreewfs:SuppressXLinkOutput’ value in the PostgreSQL mapping was set to ‘false’, otherwise features with the same gml:id would be repeated throughout the document.

By setting up the WFS-T server, which supported the complex features transaction, the main requirements of the automatic metadata updating approach such as transferring dataset and metadata in GML format to the client side and transferring the changes in geometries and metadata values back to the server side were addressed. Next, to synchronise the corresponding metadata catalogues in the spatial database and the GeoNetwork database an OGC compliant CSW Web server was required. The next section explores this requirement.

- **CSW Web Server**

As discussed earlier in relation to storage layer, 'deegree-CSW Web server', which addresses the storage of metadata values in a relational database by some means, was selected to be employed in the service layer. In fact, this server addresses the third group of required operations related to the service layer development (see page 205).

deegree-CSW is an implementation of the OpenGIS Catalogue Services Specification 2.0.2 - ISO 19115/ISO 19119 Metadata Application Profile (1.0.0) which supports GetCapabilities, DescribeRecord, GetRecords, GetRecordByID, Transaction and Harvest operations (deegree 2010a). In the prototype system, deegree-CSW was responsible for sending the CSW requests to the spatial database, and the GeoNetwork-CSW dealt with sending the CSW requests to the GeoNetwork database.

- **SYNC Web Service**

As described earlier, the aim of SYNC service was to synchronise the corresponding metadata catalogues stored in both GeoNetwork and spatial databases. This would result in keeping the metadata record stored in the GeoNetwork database current with its relevant spatial dataset. This service was implemented based on the OGC CSW service version 2.0.2 supported by both GeoNetwork and deegree.

To implement the SYNC service, PHP language was chosen because it is a widely used scripting language that is especially suited for Web development and can be embedded into HTML (PHP 2011). Prior to starting the script development there was a need to identify the required steps for the metadata catalogues synchronisation task. This task included the following two main parts:

- posting the CSW update request from deegree to GeoNetwork
- posting the CSW update request from GeoNetwork to deegree.

The recognised steps to undertake the above tasks are described in **Appendix 8**. Based on the steps explained in this Appendix the PHP code was scripted for implementation of SYNC service.

- **Application Layer**

This layer provides the required graphical user interfaces (GUI) for addressing the use cases 2, 3, 4, 5, 6, 7, 8 and 9 (see pages 199 — 201). Following the integration of the prototype system with the GeoNetwork environment for the use cases 2, 3, 4, 8, and 9 the original GeoNetwork interfaces were used and customised. In addition, a separate GUI was designed and developed for the prototype system to address the use cases 5, 6 and 7.

To be user-friendly and easy-to-use were the most significant characteristics considered for designing a GUI for the prototype system. This GUI not only should 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 visualise them in different sections. In addition, the interface should include an ‘Edit Toolbar’ that enables the users to modify the dataset and save the changes by sending the WFS-T request to the deegree Web server. Moreover, the GUI needs to visualise the search words that have been recorded by the automatic spatial metadata enrichment prototype system and met the threshold weight. This is the integrating point for both prototype systems developed in this thesis. The automatic spatial metadata enrichment prototype system will be discussed in Section 7.3. Figure 7-6 illustrates the data flow for the automatic spatial metadata updating process.

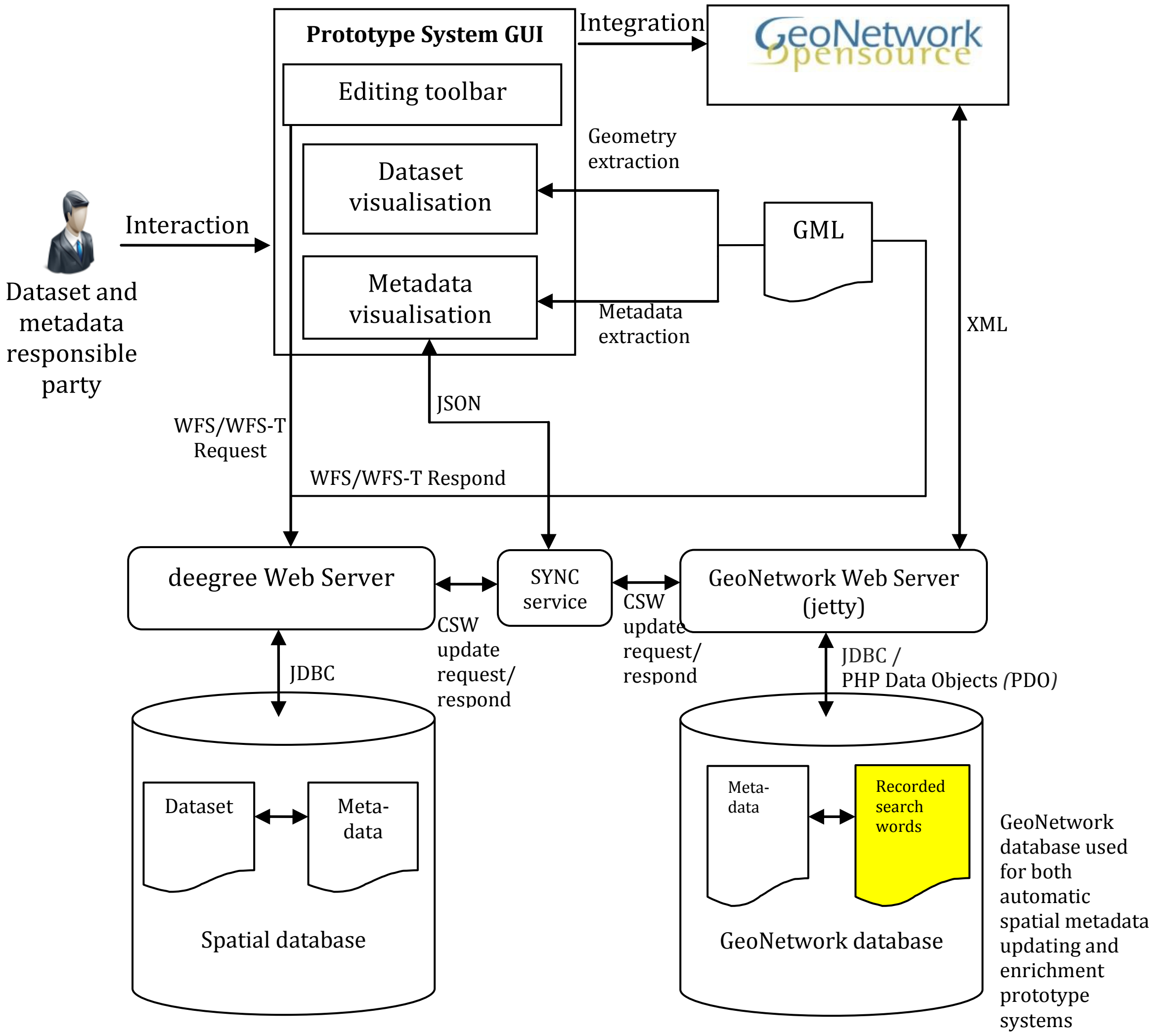


Figure 7-6: Data flow for the automatic spatial metadata updating prototype system (Olfat *et al.* 2012b). Also, the data flow for the integration of automatic spatial metadata enrichment prototype system with the automatic spatial metadata updating system is illustrated.

To develop the application layer, different International Standards and technologies such as XSL (EXtensible Stylesheet Language), XML (EXtensible Markup Language), GML (Geographic Markup Language), JSON (JavaScript Object Notation), HTML (Hypertext Markup Language), JavaScript based on OpenLayers, Ext and GeoExt libraries/frameworks, and PHP (for developing Web services) were employed. Figure 7-7 illustrates the overall view of the automatic metadata updating prototype system GUI.

As it can be seen in the Figure 7-7, following the fifth use case (see page 200), the GUI consists of two main sections as following:

- Dataset modification section
- Metadata updating section.

The first section is the place where dataset features would appear in vector format. The features can be identified and more importantly; they can be edited by the data responsible party.

The second section belongs to those metadata elements that might be updated at the same time with dataset modification. This section also includes some core metadata elements to identify metadata, such as file identifier, title, and point of contact. The metadata elements included in any tab in this section are explored in more detail in the next section.

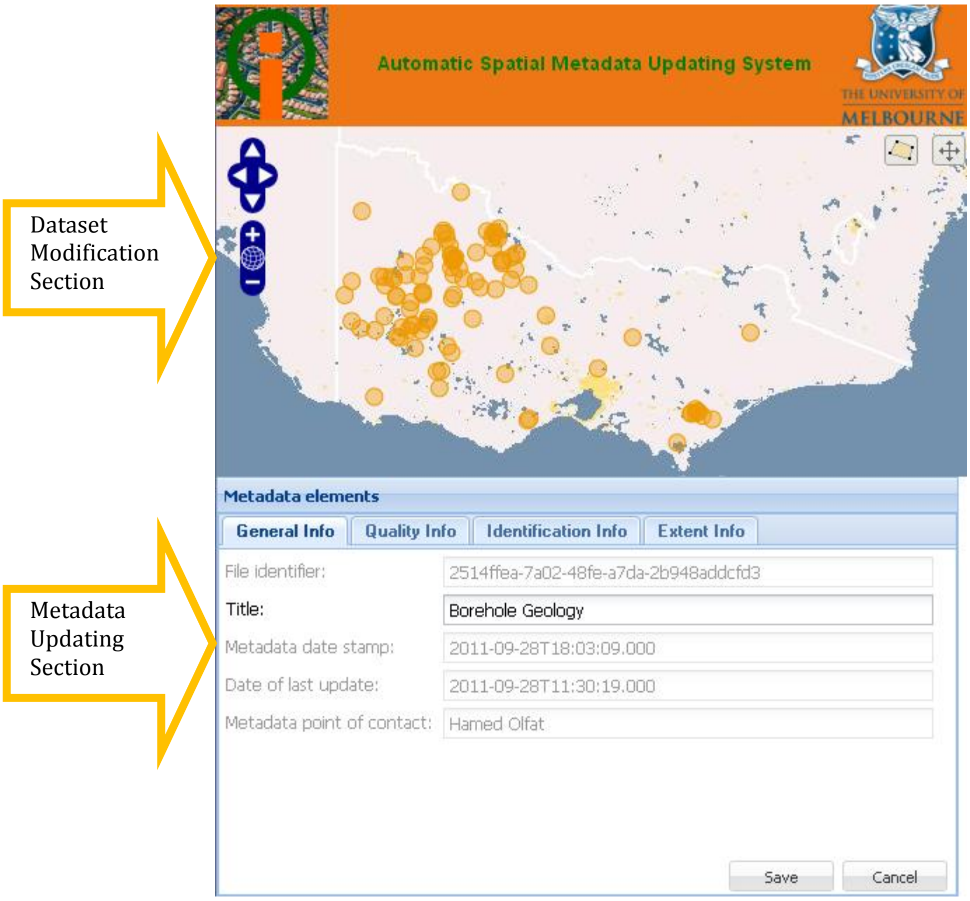


Figure 7-7: Automatic spatial metadata updating prototype system GUI to address the use cases 5, 6 and 7 (explained on page 200)

In order to address the fourth use case (see page 200) and integrate the prototype system with the GeoNetwork metadata update interface (accessible through http://IP:Port/geonetwork/srv/en/metadata.update), a new button labelled 'Map-based Modification' was added to this interface, as illustrated in Figure 7-8. When the user logs in the GeoNetwork and accesses the update interface for any metadata record, by clicking this button the automatic metadata updating GUI including relevant dataset and metadata (see Figure 7-7) would appear.

Figure 7-8: New button added to the GeoNetwork metadata update interface to address the use case 4 (explained on page 200)

In addition, as already discussed in the service layer, the SYNC service was assigned to 'Save' and 'Save and close' buttons in GeoNetwork metadata update interface (again, Figure 7-8). By clicking these buttons after the metadata modification, the metadata record in the spatial database would be updated automatically. Also, to support the two-way synchronisation, the SYNC service was assigned to the 'Save' button on the automatic spatial metadata updating GUI (see Figure 7-7) and by clicking this button the metadata elements affected by the dataset modification would be updated in the GeoNetwork database as well.

Accordingly, the major functions expected from the GUI are listed below:

- GML to JavaScript Extraction
- Form population
- Edit dataset
- WFS-T preparation
- Calling SYNC service and reloading features.

As illustrated earlier in Figure 7-6, the GML document as the output of metadata integrated data model that accommodates dataset geometries, attributes and metadata values would be extracted to JavaScript. Therefore, the first function would be relating to this particular extraction for pulling out geometries and metadata and managing them in two different sections shown in Figure 7-7.

After the GML extraction, the second function would be populating metadata forms. In this regard, the metadata elements' mapping path stored in the GML Application Schema was used to demonstrate the latest metadata value from the back end.

Another important function of the interface would be preparing the edit tools to modify the dataset. Therefore, a tool was developed based on the OpenLayers library to modify features.

Once the dataset is edited, the new geometries and metadata values should be sent to the back end through a WFS-T request. This happens when the user clicks the 'Save' button on the interface (see Figure 7-7). At this time, the ISO metadata elements that are affected by the dataset modification would be updated and stored in the spatial database. The metadata elements employed in this prototype system were chosen from Table 6-1 in Chapter 6 (except the metadata date stamp). These elements along with their automation approaches are outlined in Table 7-1:

Table 7-1: The ISO19115: 2003 metadata elements used in the automatic spatial metadata updating proyotype system

Tab	Metadata element	Type of Synchronisation	Automation approach
General Info	Metadata date stamp Date of last update	Client Side	When the dataset is edited by the user and the 'Save' button is clicked the values for these two elements would be extracted from the user interface through a JavaScript code and assigned to the edited features (of the dataset); so that they would be passed to the WFS Transaction.
Quality Info	Lineage statement	Client Side	The value for this element is generated automatically after pressing the 'Save' button on the Client Side and documents the history of changes.
Extent Info	North bounding latitude South bounding latitude West bounding longitude East bounding longitude	Server Side	To automate the values of the dataset bounding box, a trigger was generated in PostgreSQL/PostGIS.

The next section presents the metadata tabs developed in the prototype system GUI.

➢ Metadata Tabs

To demonstrate the metadata values that are affected by the dataset modification, four tabs were developed in the interface (Figure 7-9).

Figure 7-9: Metadata tabs in the GUI

Each tab consisted of different elements, as follows:

General Info Tab

According to Figure 7-10, this tab contains some general metadata elements as following:

- File identifier
- Title
- Metadata date stamp
- Date of last update

- Metadata point of contact.

Out of these elements, only two of them could be affected by the dataset modification: ‘Metadata date stamp’ and ‘Date of last update’. When the dataset was edited by the user (data responsible party) and ‘Save’ button was clicked, the values for these two elements would be extracted from the user interface (client side) through a JavaScript code and assigned to the edited features (of the dataset) so that they would be passed to the WFS-T. Based on the ISO 19115: 2003 Standard, date would be in "YYYY-MM-DD'T'HH:MM:SS" Format.

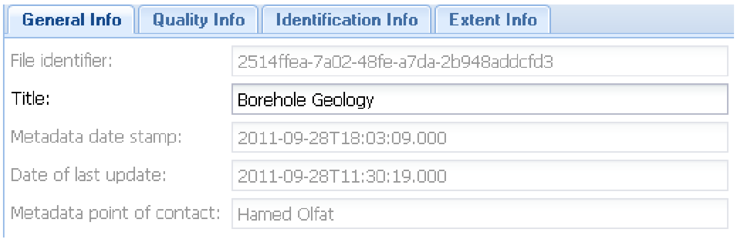


Figure 7-10: General Info tab in the GUI

The other two elements in this tab including the ‘File identifier’ and ‘Metadata point of contact’ were read-only elements. These elements aimed to give some information about the identity of the metadata record. The element ‘Title’ presented dataset title and could be updated manually by the user.

Quality Info Tab

As illustrated in Figure 7-11, this tab includes only one element: ‘Lineage statement’.

This element, as part of data quality information, would document the history of modifications undertaken by privileged users who have logged in the GeoNetwork to access the dataset and metadata update interface.

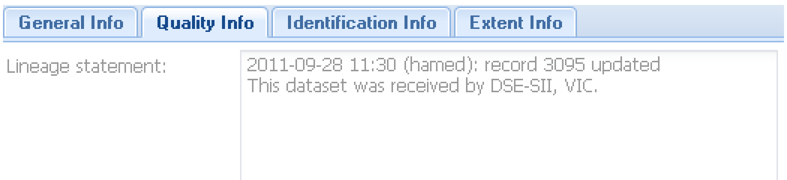


Figure 7-11: Quality Info tab in the GUI

As mentioned earlier, this element documents the history of changes based on the following rule:

When (date and time) + Who (GeoNetwork’s user) + What action (Updated/Added/Deleted) + Which feature (of the dataset)

This function would assist the data responsible parties and users to have a better knowledge of the dataset's latest developments. Also, when the dataset is transferred to other interested parties the new users would be aware of dataset's maintenance history through this automatic documentation.

Identification Info Tab

As shown in Figure 7-12, this tab includes two different metadata elements: 'Reference system' and 'Possible new keywords'.

Reference system value is based on the reference system that has been set for the dataset during its storage process at the back end (spatial database – PostGIS). If the reference system changes at the back end, this value would be updated as well.

The 'possible new keywords' were considered in this tab to address the seventh use case (see page 200). This section of the interface includes the search words that were assigned to the metadata through the indirect and direct models of the automatic spatial metadata enrichment prototype system. This prototype system will be explored in Section 7-3.

Presenting the search words informs the metadata responsible party about the most popular and user-agreed search words relevant to the metadata record. The weight column in this tab illustrates the popularity of the search word among the end users. Depending on the metadata responsible party's decision, the highest weighted search words can be inserted into metadata information as the new 'descriptive keyword' through the GeoNetwork metadata update interface. This would result in enriching the content of metadata and facilitating datasets discovery process. The search words are shown in this tab through a Web service developed by PHP.

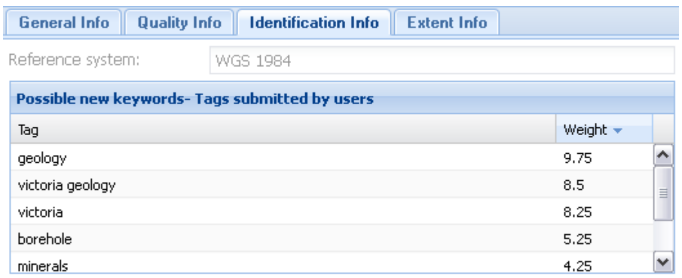


Figure 7-12: Identification Info tab in the GUI

Extent Info Tab

This tab embraces four metadata elements which present the geographical extent (bounding box) of the dataset including 'North bounding latitude', 'South bounding latitude', 'West bounding longitude' and 'East bounding longitude', as illustrated in Figure 7-13.

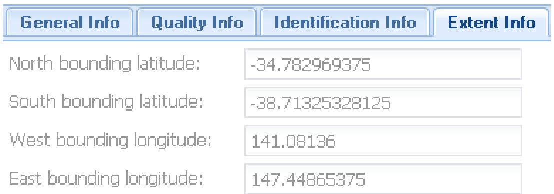


Figure 7-13: Extent Info tab in the GUI

In order to automate the values of the dataset bounding box, a trigger was scripted in PostgreSQL for dataset table (boresg). This trigger, which is illustrated in **Appendix 8**, would generate the most current bounding box of the dataset and would store this value in 'geom' column of 'ex_geogrbbox' table for the related metadata record. 'ex_geogrbbox' table is designed as part of deegree-CSW database for storing the bounding box. Through this trigger, whenever the geographic extent of the dataset is changed due to the dataset modification process the new values of extent would be calculated automatically and stored in the metadata record. These new values would be also transferred in GML and visualised in the prototype system GUI.

The functionalities implemented for the automatic spatial metadata updating approach would most likely provide the geospatial community with a variety of benefits including reduction of the burden of manual metadata updating after dataset modification, facilitation of the interoperability by publishing the datasets in GML and regardless of any specific dataset format, and enablement of data responsible parties to publish and share datasets along with attributes and metadata in a single document. This approach also helps to avoid missing, incomplete, out-of-date and unreliable metadata. Moreover, having the metadata synchronisation approach in place could give a peace of mind to data responsible parties, due to the metadata always being current with dataset changes. The synchronisation could also provide a better discovery service to users seeking spatial datasets over the Web by providing them with the most recent version of metadata.

The next section reviews some of the main challenges that the implementation of prototype system faced.

7.2.2 Summary of Prototype System Implementation Main Challenges

The first important challenge was regarding the support of dataset and metadata integration by the GML technology. The GML supports this integration at the feature-level and not at the dataset-level; which was counted as a limitation for implementing the prototype system. To avoid the redundancy of metadata in the GML document, the GML Application Schema was customised to generate metadata elements for one feature of the dataset as the reference metadata for the remaining features of the same dataset.

In addition, the lack of a standard data model to store the ISO 19115: 2003 metadata elements in a relational database was another obstacle to implement the prototype system. This data model is part of the integrated metadata data model.

Other technical key challenges are also summarised here:

- difficulty in developing and debugging the XML schema for complex features
- lack of maturity for complex features: GeoServer did not support the WFS-T on complex features, deegree supported it within strict constraints (same database, and foreign key between the simple features composing the complex feature)
- implementation of standards: deegree implemented the CSW update operation as a delete followed by an insert, which had an effect on preserving a unique identifier for a given metadata record
- cross-browser compatibility: OpenLayers XML parsing gave different results in different browsers.

The next section reviews the implementation of the automatic spatial metadata enrichment prototype system.

7.3 Automatic Spatial Metadata Enrichment Prototype System

As discussed in Section 4.4.3 in Chapter 4, in order to prove the concept of proposed automatic metadata enrichment approach (see Section 6.5, Chapter 6), a prototype system was designed and developed within two different environments: GeoNetwork as an open source spatial data catalogue application and Model Information Knowledge Environment (MIKE) as an example of data product – data modelling environment. This section explores the implementation of the prototype system within these two environments.

7.3.1 Implementation within the GeoNetwork opensource

In order to develop a prototype system within the GeoNetwork opensource catalogue to provide the indirect and direct metadata enrichment models (discussed earlier in Sections 6.5.1 and 6.5.2 in Chapter 6) the following use cases were defined:

Use case 1 – the user logs in the GeoNetwork opensource catalogue and starts discovering the existing metadata records by typing a 'search word' in the search box already provided by the GeoNetwork. By pressing the 'Search' button the matching metadata records will be found and retrieved.

Use case 2 – While the user starts to type a search word in the search box, the previous used search words that match with the typed letters will be suggested to the user along with their frequency use.

Use case 3 – When metadata record(s) are retrieved for any input search word, if the user clicks on 'Metadata' button underneath any retrieved record to see the details of metadata, the input search word will be recorded by the prototype system and stored in a temporary table linked to the metadata table in the GeoNetwork database. This search word will be recorded with 0.25 unit of weight, according to the weighting system discussed in Section 6.5.1, Chapter 6. Also, if the user spends more than 3 minutes with the metadata details open another 0.50 weight will be recorded for the input search word according to the weighting system.

Use case 4 – When metadata record(s) are retrieved for any input search word, if the user clicks on 'Download' button underneath of any retrieved record to access the actual dataset the input search word will be recorded by the prototype system and stored in a temporary table linked to the metadata table in the GeoNetwork database. This search word will be recorded with 0.75 weight, according to the weighting system discussed in Section 6.5.1, Chapter 6.

Use case 5 – When the metadata record(s) are retrieved for any input search word, the user sees a question ('Relevant?') next to each retrieved metadata that asks about the relevance of metadata with the input search word. This question contains two buttons namely 'Agree' and 'Disagree'. If the user clicks the 'Agree' button, the input search word will be recorded by the prototype system and stored in a temporary table linked to the metadata table in the GeoNetwork database. This search word will be recorded with 1 unit of weight, according to the weighting system discussed in Section 6.5.2, Chapter 6.

However, if the user clicks the 'Disagree' button the input search word will still be recorded by the prototype system and stored in a temporary table and 1 unit of weight will be deducted from its weight. In some cases, if the recorded search word has no weight its weight will be minus zero.

Use case 6 – when the user clicks the 'Agree' button a pop-up will open to visualise the search words already recorded for the metadata record along with their weights. The user will be able to agree with any of them by ticking a box next to the recorded search word. Also, the user will be able to insert a new search word to be tagged to the metadata. When the user's decision is submitted, the weight of the agreed search words will be increased by 1 unit and the newly inserted tag will gain 1 unit of weight, according to the weighting system discussed in Section 6.5.2, Chapter 6.

Use case 7 – when the user clicks the 'Disagree' button a pop-up will open to show the search words already recorded for the metadata record along with their weights; so that the user will be able to disagree with any of them by ticking a box next to the recorded search word. Also, the user will be able to insert a new search word to be tagged to the metadata. When the user's decision is submitted, the weight of the disagreed search words will be reduced by 1 unit and the newly inserted tag will gain 1 unit of weight, according to the weighting system discussed in Section 6.5.2, Chapter 6.

Use case 8 – in the GeoNetwork data discovery interface, the user is able to access a tag cloud that visualises the recorded search words, which met the weight threshold (1 unit of weight for this prototype system). The search words in this cloud are shown along with their weights. There is a direct relationship between the weight and the size of the search word shown in the tag cloud. As the weight increases, the search word will be larger.

Figure 7-14 illustrates the use diagram designed based on the above use cases.

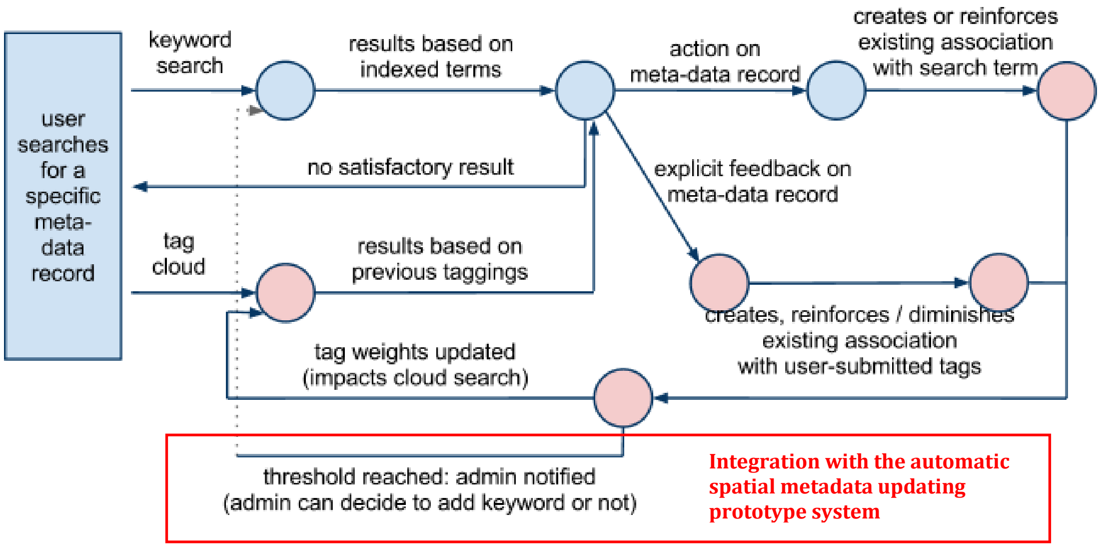


Figure 7-14: Use diagram designed for agree/disagree/tag cloud add-on to address the use cases

To address the above use cases in the automatic spatial metadata enrichment prototype system and thus enhance the end users' experience, the following two add-ons were designed and implemented within the GeoNetwork environment:

- Agree/disagree/tag cloud add-on
- Suggestion list add-on.

The add-ons have been developed in a way that could be simply installed by other users on a working GeoNetwork platform.

The prerequisite to implement the add-ons was to build a GeoNetwork development environment. This environment was built up using the following components:

- OpenGeo Suite, a self-contained, single-click installation, open source GIS stack containing a database (PostgreSQL/PostGIS), a map server (GeoServer) and a Web server (Jetty)
- GeoNetwork (WAR file distribution)
- TortoiseSVN, a tool to checkout source code
- Eclipse, an integrated development environment (IDE) for server side Java debugging
- Maven, a companion tool to build and distribute Java projects
- Firefox and Firebug for client side JavaScript debugging.

Once the GeoNetwork development environment was set up, the architecture for add-ons implementation was designed. The overall view of the architecture is illustrated in Figure 7-15.

As it can be shown, the architecture consists of three main layers: storage, service, and application. The technologies used in each layer are also outlined in Figure 7-15.

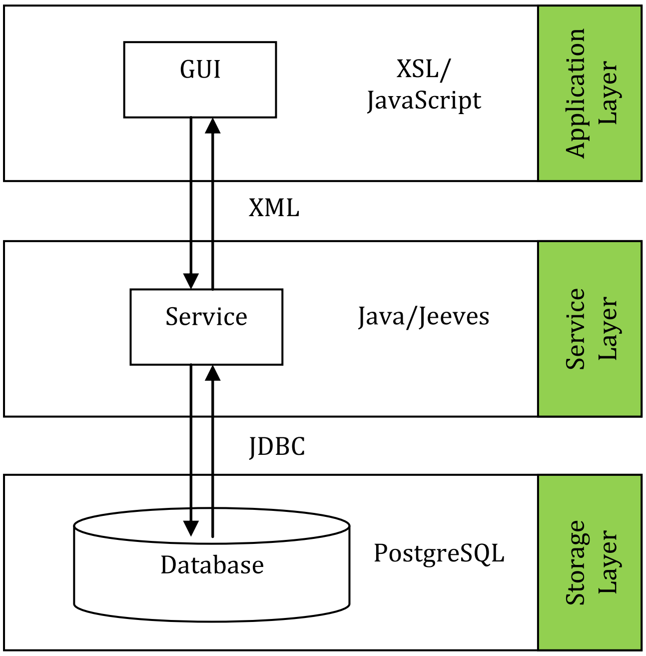


Figure 7-15: Implementation architecture of add-ons within the GeoNetwork to address the use cases (explained on pages 221 and 222)

In the storage layer, the PostgreSQL was selected as the database for storing the required data used by the add-ons. Also, following the GeoNetwork architecture Java and Jeeves services were selected as the main players in the service layer. Moreover, following the GeoNetwork architecture the XSL and JavaScript (based on the ExtJS library) were chosen for implementing the GUI in the application layer.

The characteristics of the add-ons are explained in the following sections:

- **Agree/Disagree/Tag Cloud Add-on**

In order to address the use cases 3 to 8 (see pages 221 and 222), this add-on automatically observes the end users' interaction with the spatial data discovery process within the GeoNetwork and collects the end users' feedback on metadata records. It monitors every interaction of end users including exploration of metadata details, spending time on reading metadata, and downloading the actual spatial data. It also asks the end users to submit new search words, and agree or disagree with the existing tagged search words. Therefore, this add-on provides the end users with a way to enrich metadata records with tagged search words (implicitly or explicitly), so that subsequent users benefit from this enrichment in their own searches. The tagged search words are represented in a cloud, which is a graphically weighted list of search words (tags).

Following the indirect and direct metadata enrichment approaches, the following two functions were considered for this add-on to record a search word as 'keyword' metadata to describe spatial data:

- Implicit recording: the search words are recorded automatically (with different weights) by the data catalogue system based on the end users' behaviours already outlined in Table 6-6, Chapter 6. This results in addressing the use cases 3 and 4 (see page 221).
- Explicit recording: the search words are recorded by the end users of the data catalogue system. This function is accessible via two buttons entitled 'Agree' and 'Disagree' embedded in metadata record details within the GeoNetwork data discovery page, as illustrated in Figure 7-16. 'Agree' allows the users to reinforce an existing tagged search word or assign a new search word to a metadata record, as illustrated in Figure 7-17. 'Disagree' enables the users to diminish the weight of an existing tagged search word or assign a new search word to a metadata record, similar to Figure 7-17. The weighting system for recording search words is designed based on Table 6-7, already described in Chapter 6. The explicit assignment would be visible both at the top of the metadata record details (when closed) and at the bottom of the metadata record details (when open). This results in addressing the use cases 5, 6 and 7 (see pages 221 and 222).

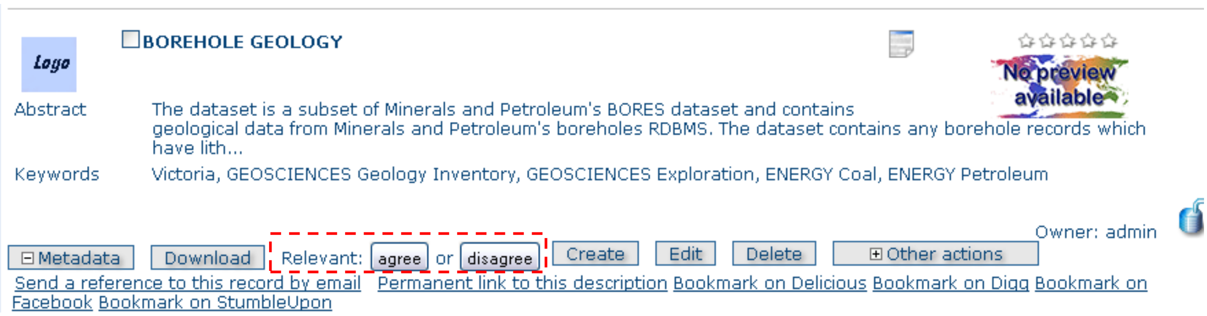


Figure 7-16: Implementing Agree/Disagree buttons with regard to direct metadata enrichment model to address the fifth use case (explained on page 221)

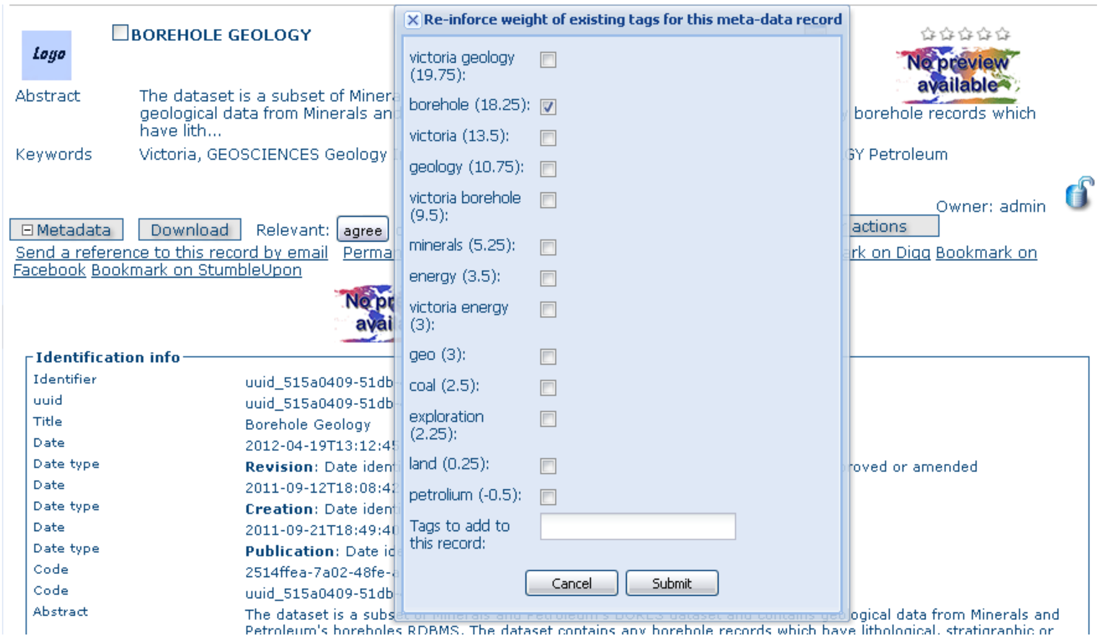


Figure 7-17: Implementing Agree/Disagree functions to agree or disagree with the existing tagged search words and add new search word to address the use cases 6 and 7 (explained on page 222)

Once the relevant search words are recorded, they should be assigned to their related metadata records as a new 'keyword' metadata value. As already described in Section 6.5.1 in Chapter 6, the assignment of recorded search words depends on the specific threshold considered by the data catalogue administrator or metadata/data responsible party. In this prototype system the threshold has been considered as 1 unit of weight.

Finally, the assigned search words (tags) were shown through a tag cloud in this prototype system. Following the conceptual design of metadata enrichment, the search words that were more frequently used and agreed by the users were viewed in a highlighted format within the cloud. Also, the weight of each search word was listed next to it within the cloud, which confirms its popularity in the discovery process. Figure 7-18 illustrates the tag cloud generated within the GeoNetwork for this prototype system. A click on a search word in the tag cloud will search for the metadata record(s) associated with this search word. This results in addressing the last use case (see page 222).

Figure 7-18: Tag cloud within the GeoNetwork to visualise most popular search words to address the last use case (explained on page 222)

To implement this prototype system, according to Figure 7-15, a PostgreSQL database was developed in the storage layer to include a table for storing recorded search words, their weights, and their relationships with the metadata records. Also, the following services were developed within the GeoNetwork in the service layer:

- SELECT: A Java service that returns the tagged search words and weights for the representation of the tag cloud.
- SELECT: A Java service that returns the metadata records associated to a specific tagged search word.
- SELECT: A Java service that returns the tagged search words and weights for a given metadata record.
- UPDATE: A Java service to insert/delete/update the weight of a search word associated to a metadata record.

In the application layer, the related XSL within the GeoNetwork, which contains the search form, was modified to include a button (link) that represents the tag cloud (see Figure 7-18) and two other buttons to 'Agree' or 'Disagree' with the relevance of search word to the retrieved metadata records (see Figure 7-16).

Finally, at this stage a Web service was developed by PHP in order to notify the metadata responsible party about the tagged search words related to each metadata record. Through this service the metadata responsible party would be able to give approval to those search words most suitable for describing the dataset to be inserted in the 'keyword' metadata element. Accordingly, this results in improvement of the content of metadata and facilitating the dataset discovery process. As already discussed in Section 7.2, this service would be called during the automatic spatial metadata updating process and acts as a link between both prototypes implemented in this thesis.

In the enrichment prototype system, in addition to 'agree/disagree/tag cloud' add-on, another add-on was implemented within the GeoNetwork to facilitate and accelerate the spatial data discovery process. The implementation of this add-on is reviewed in the next section.

- **Suggestion List Add-on**

The add-on was designed and implemented within the GeoNetwork to address the second use case (explained on page 221) and provide the end user with a suggestion list based on previously searched terms. This happens while typing the search word in 'WHAT?' box, as illustrated in Figure 7-19. Using this add-on, all subsequent searches benefit from previous searches. This facility most likely provides a user-generated context for metadata improvement and automation.

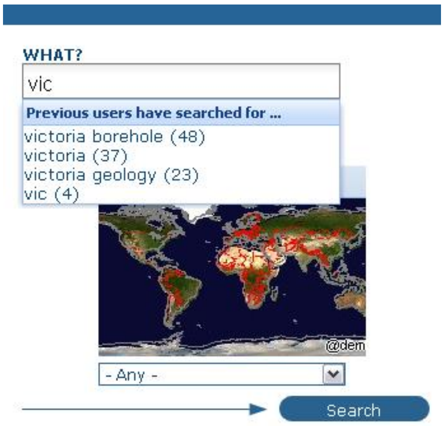


Figure 7-19: Suggestion list add-on within the GeoNetwork to address the second use case (explianed on page 221)

Following Figure 7-19, this add-on uses the PostgreSQL in the storage layer to store the search words. In this regard, a new table was added to this database to capture previously searched terms.

In addition, in the service layer the following two services were developed within the GeoNetwork to address the main aim of this add-on:

- SELECT: A Java service with 1 parameter (string being searched) that returns a weighted/sorted list of previously searched words containing the parameter string.

- UPDATE: A Java service with 1 parameter (string searched) that persists in a search of the database, and keeps a count of the total number of times this search word has been used. The count will be used to weight the result list.

In the application layer, this add-on uses the ComboBox class of ExtJS library. The data store behind the ComboBox is an AJAX[32] call to the retrieval service. In addition, to implement this layer the related XSL within the GeoNetwork that contains the search field was modified.

Through the add-ons developed within the GeoNetwork the end users are able to interact with the data discovery system for creating and improving the content of ‘keyword’ metadata element. They can also share their knowledge about datasets by agreeing or disagreeing with the relevance of the existing tagged search words or adding new search words to datasets. Also, the tag cloud potentially provides the end users with the capabilities to visit the most popular search words, gain ideas regarding the available data, and access the datasets faster and simpler.

The next section reviews the implementation of the automatic spatial metadata enrichment within the MIKE.

7.3.2 Implementation within the Model Information Knowledge Environment (MIKE)

MIKE has been developed by the Victorian Department of Primary Industries (DPI) for management and registry of instances of biophysical and socio-economic modelling work in Victoria, Australia. According to 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. MIKE is also 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 contacts (including full details e.g. address and phone), model limitation, data inputs, abstract, background, history, purpose, programming language, Web page, author, keywords, access and application type and etc (Williams *et al.* 2009).

The automatic metadata enrichment prototype system was implemented within the MIKE as a new business function to enable the end users to interact with the model discovery system to

[32] Asynchronous JavaScript and XML

improve the content of metadata and deliver better search and discovery of models via a tag cloud tool.

For implementing the new business function, there was a need to discover the software environment of the MIKE development. Table 7-2 summarises the MIKE development environment, which was accessed for the prototype system implementation.

Table 7-2: MIKE programming and database environment

Software Requirement	Type of Software
Operating System	Windows XP SP3
Web Server	Windows IIS 5.1 & Tomcat 5.5
Database	Microsoft SQL Server 2008
Spatial Database Engine	ESRI ArcSDE 9.3.1
Spatial Web Server	ESRI ArcIMS 9.3.1
Web Application	Java jdk 1.6.0_20 , Microsoft .NET 3.5 SP1 , Visual Basic for Applications , HTML
Web Development	Microsoft Visual Web Developer 2008 Express Edition

Figure 7-20 shows the overall view of the MIKE system GUI (home page). This figure also illustrates the menu relating to the prototype system in red box (‘Model Discovery’).

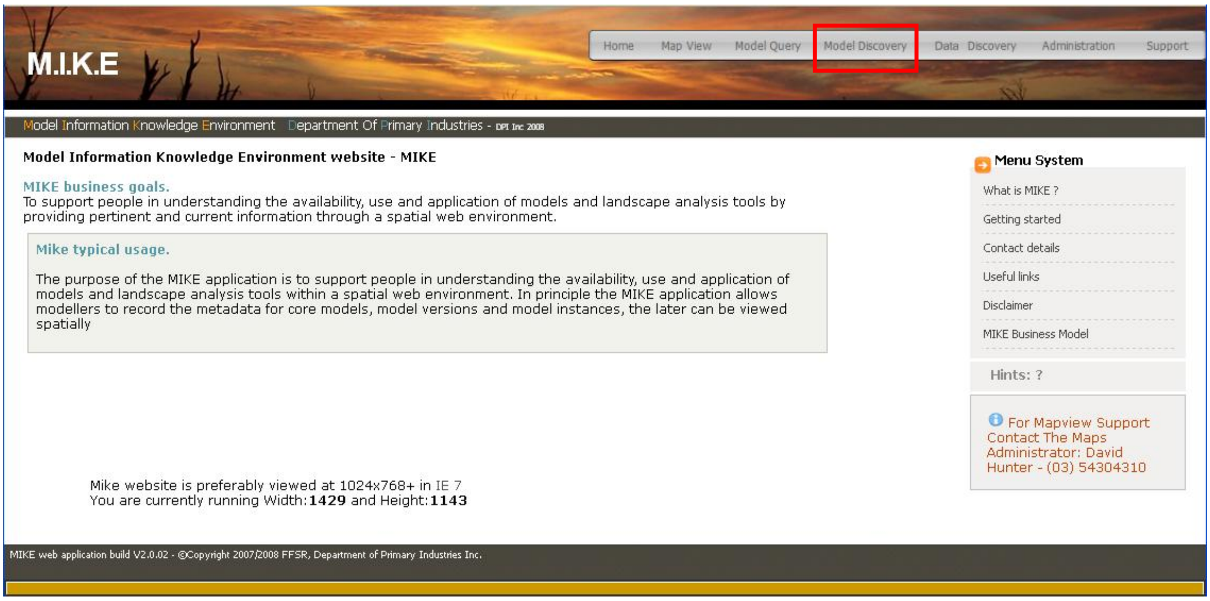


Figure 7-20: The MIKE system GUI (home page)

In order to develop a prototype system within the MIKE to provide the indirect and direct metadata enrichment models, discussed earlier in Sections 6.5.1 and 6.5.2 in Chapter 6, the following use cases were defined:

Use case 1 – user accesses the MIKE model discovery interface and starts searching for the existing models by typing a ‘search word’ in the search box. By pressing a button namely ‘Discover Models’ the matching metadata records will be found and retrieved. There was no interface designed for model discovery within the version of MIKE accessed for the prototype system development; therefore, this interface was needed to be designed and developed for the prototype system.

Use case 2 – when model metadata record(s) are retrieved for any input search word, if the user clicks on each retrieved record a new page will be opened, which includes the details of model metadata.

Use case 3 – once the user finishes the exploration of the model metadata details and wants to close the metadata page, by clicking the 'Exit this Model' button on the page, a pop-up will be shown to ask about the relevance of this metadata record to the user's requirements. This pop-up screen will contain two buttons, 'YES' and 'NO'.

Use case 4 – if the user clicks the 'NO' button, the pop-up will be closed and the input search word will not be recorded for the retrieved model metadata. However, if the 'YES' button is clicked, the input search word will be recorded for the retrieved model metadata and stored in a temporary table with 1 unit of weight (repetition).

Use case 5 – when user clicks the 'YES' button a pop-up screen will be opened to visualise the search words already recorded for the model metadata record, so that the user will be able to agree with any by ticking a box next to the recorded search word. Also, the user will be able to insert a new search word to be tagged to the metadata. When the user's decision is submitted, the weight of the agreed search words will be increased for 1 unit and the newly inserted search word will gain 1 unit of weight.

Use case 6 – in the model metadata details interface, the user is able to see a tag cloud including the recorded search words for the metadata, which met the weight threshold (10 units of weight for this prototype system). Also, in this interface the user will be able to click a button, 'comments' to leave a note on the model metadata or to visit others' comments.

Use case 7 – in the model discovery interface, the user is able to access a tag cloud that visualises the recorded search words that met the weight threshold. There is a direct relationship between the weight and the size of search word in the tag cloud. As the weight increases, the search word will be shown bigger. Also, if the user clicks on any of the search words in the cloud, a pop-up screen will be opened to show the related models of which this search word has been tagged to, along with the frequency of using this search word.

To implement a prototype system to address the above use cases, the system architecture was designed based on the automatic metadata enrichment approach discussed in Sections 6.5.1 and

6.5.2 in Chapter 6, as well as, the MIKE programming and database environment outlined in Table 7-2. The architecture was designed in three layers: storage, service, and application as illustrated in Figure 7-21.

The storage layer consisted of a Microsoft SQL Server 2008 database. This database included a table for storing metadata values for the models that was originally designed by the MIKE's development team. Also, in this database two other tables were designed and added to the MIKE data model to be used for the prototype system. The characteristics of these tables will be discussed later. The service layer included Windows IIS as the application server and also ASP.NET and VB.NET were the programming languages used in this layer. Moreover, the required indirect and direct services for metadata enrichment such as model search and retrieval, search word monitoring, recording, assignment, and tag cloud generation were implemented in this layer. The application layer contained a root Web page entitled 'Model Discovery', which was designed as the main interface of the prototype system, and some other Web pages, to call the required services from the service layer. These pages will be presented later.

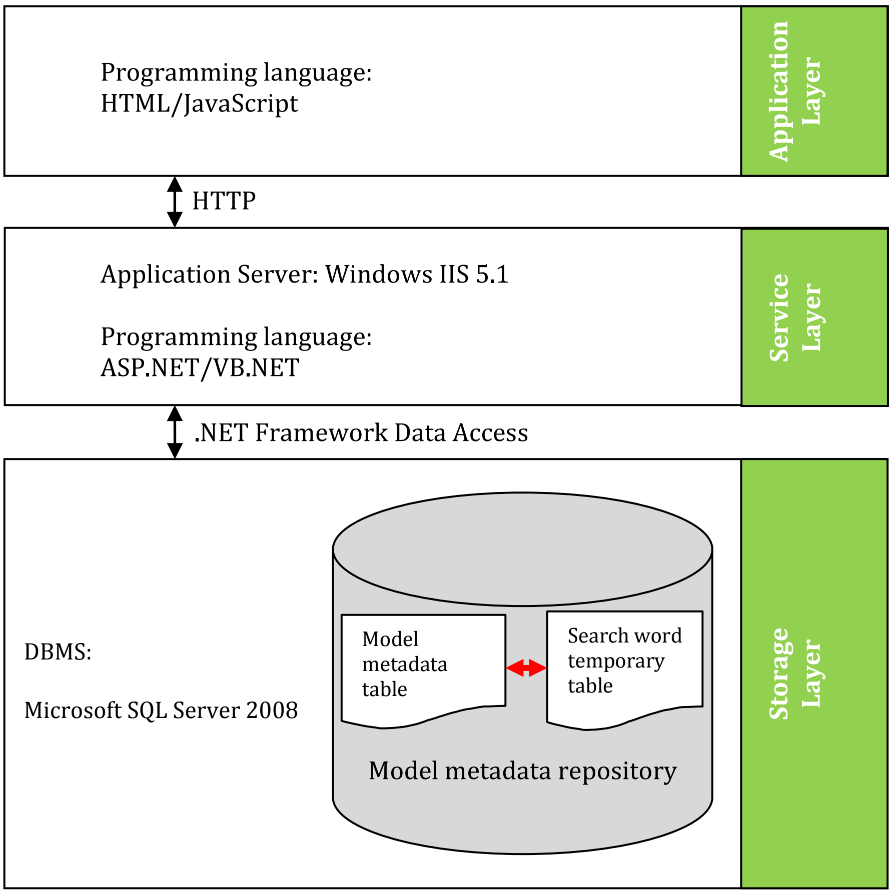


Figure 7-21: Implementation architecture of automatic metadata enrichment prototype system within MIKE to address the use cases (explianed on pages 230 and 231)

Once the implementation architecture was designed, the indirect and direct metadata enrichment models were implemented within the MIKE as discussed below.

- **Indirect Metadata Enrichment Model**

The implementation of three stages involved in the indirect metadata enrichment model within the MIKE is detailed here.

➢ **Stage 1: Monitoring Search Word**

To implement this stage and address the first use case (see page 230), there was a need to design a new Web page within the MIKE dedicated to the prototype system. Therefore, a new Web Form entitled 'Model Discovery' was included within the MIKE development project in the application layer. The link to this page was also embedded within the MIKE main menu, as already highlighted in Figure 7-20.

The first component of the model discovery page was a service to provide a user-friendly and easy-to-use environment for discovering the existing models. Figure 7-22 illustrates the main components of this service including a text box for typing the search word, a button for sending the discovery request to the server and database, and finally a grid containing three columns (Select Item, Model ID, and Model Name) for showing the discovery results.

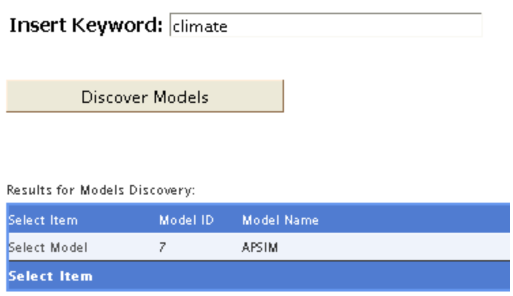


Figure 7-22: Components of the model discovery service to address the first use case (explained on page 230)

This service matches the typed search word (e.g. climate) with the abstract of the models stored in the database and retrieves the corresponding metadata records (e.g. APSIM[33]). If the discovery process provided any results, the search word should be monitored by the system for the retrieved metadata records. As a technical view, the search word would be stored in a 'session' component that enables the storage and retrieving values for a user as the user navigates ASP.NET pages in a Web application.

---

33 Agricultural Production Systems sIMulator

Once the search word was monitored, the system needed to understand the end user's point of view regarding the relevance of search word to the retrieved metadata records. In this regard and in order to address the second use case (see page 231), it was necessary to design another Web page to view the details of retrieved metadata records for the end user's information. Accordingly, a new Web Form namely 'Model Details' (Figure 7-23) was designed and included within the MIKE development project. From a technical perspective, by clicking any metadata record retrieved during the discovery process, its ModelID would be transferred to Model Details page through a 'Query String' to facilitate monitoring and recording the search word for that record.

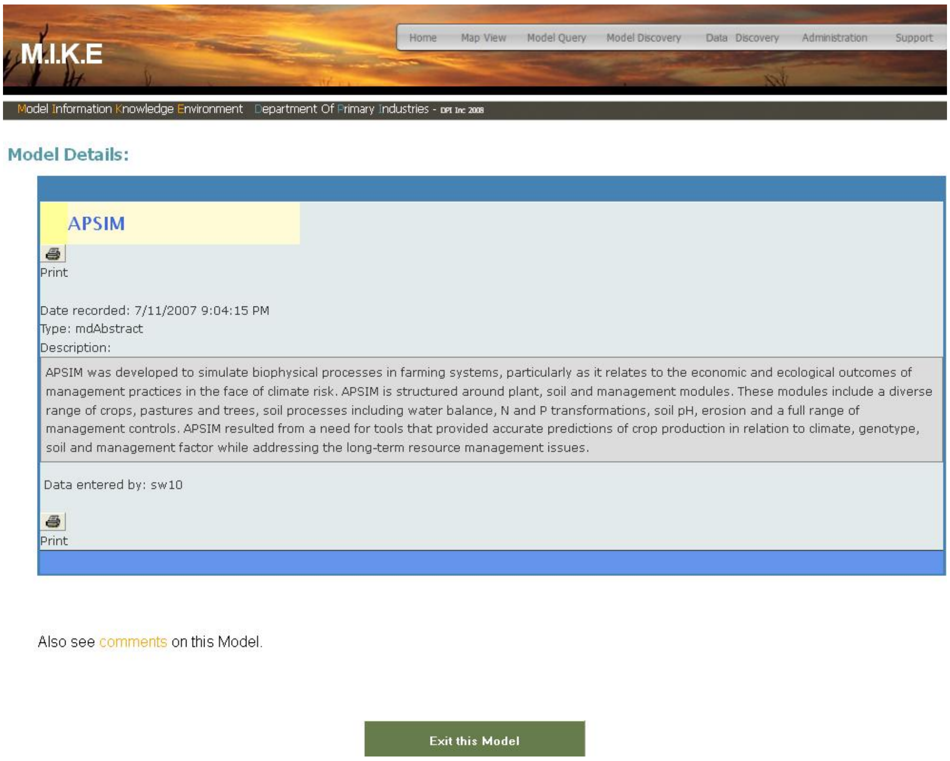


Figure 7-23: Overall view of model metadata retrieval page (e.g. for APSIM model) to address the second use case (explained on page 231)

As indicated in Figure 7-23, a button has been implemented in Model Details page entitled 'Exit this Model', which addresses the following two purposes:

- closing the page
- opening a pop-up screen to ask the end user about the suitability and relevance of the retrieved metadata record during the discovery process against the model that was initially sought. The result of this query would confirm the relevance between the search word and the retrieved metadata (and certainly its related model). Therefore, the system would be able to understand whether to record or not to record the search word as a new value of the retrieved metadata record (stage 2 of indirect enrichment model).

To address the second purpose (and the third use case explained on page 231), a new Web Form entitled ‘Pop-up’ was designed and included within the MIKE development project. Figure 7-24 shows the pop-up screen.

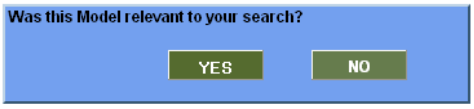


Figure 7-24: Pop-up screen to recognise the relevance of search word to the retrieved model metadata based on the end user's interaction (addressing the third use case)

According to the fourth use case (see page 231), if the end user clicks the ‘NO’ button it means that the search word used for model discovery was not appropriate for describing the retrieved model in the end user’s point of view and thus the pop-up screen would be closed. In this case, the end user would be able to return to the model discovery page to explore the other retrieved metadata records or change the search word. On the other hand, if the end user clicks the ‘YES’ button it means that the search word used for model discovery was appropriate for describing the retrieved metadata record in the end user’s opinion. Therefore, this search word should be recorded by the system following the second stage of indirect enrichment model, and is reviewed in the next section.

➢ Stage 2: Recording Search Word

In this stage, any search word that has been identified as relevant to any model metadata record would be recorded by the system to address the fourth and fifth use cases (see page 231). According to the conceptual design of metadata enrichment (see Figure 6-16, Chapter 6), to implement this stage there was a need to design a temporary table for storing search word and its related metadata characteristics. Therefore, a new table namely ‘ModelKeyword_tbl’ was designed in the MIKE database. This table contained 5 columns, as follows:

- ID: this column is the Primary Key and Identity of the table.
- Search word: this column stores the search word that is recorded automatically by the system, as well as, the search word that is inserted by the end user as a new tag (based on the direct enrichment model).
- Weight (Repetition): this column stores the frequency of assigning the search word to a particular metadata.
- ModelID: this column stores the ID of the model of which the search word has been recorded.
- NavigateURL: this column stores the URL of the model of which the search word has been recorded.

Once the 'ModelKeyword_tbl' was designed it was added to the MIKE data model.

Whenever a pop-up screen opens in stage 1, the ModelID would be also transferred to this screen through a 'Query String'. This would enable the system to record the characteristics of the model if the end user clicks the 'YES' button (see Figure 7-24).

Therefore, after clicking the 'YES', the system stores the relevant search word (which is already stored in a session, according to stage 1) in a 'Search word' column and also stores the 'ModelID' and 'NavigateURL' of the related model in the relevant columns of 'ModelKeyword_tbl' table. At this stage, to address the fourth use case (explained on page 231) the system also stores 1 unit as the weight (repetition) of the end user's agreement on the relevance of the recorded search word to the retrieved metadata.

Once the search words were recorded by the system, it was required to identify the most popular ones based on their weight (repetition) and assigning them to the metadata record as the new metadata values. The next section reviews the third stage of the indirect enrichment model.

- **Stage 3: Assigning Search Word (Tagging)**

According to the indirect metadata enrichment model design, in this stage any recorded search word that met a threshold was assigned to its related model metadata as a new value. Moreover, the assigned search words were visualised in a tag cloud within the Model Discovery (to address the seventh use case explained on page 231) and Model Details (to address the sixth use case explained on page 231) pages.

In this prototype system, any search word that had more than 10 units of weight (repetition), according to the MIKE metadata administrator's point of view, was allowed to be assigned to its related metadata record as a new value (model keyword) and shown in the tag cloud. Figure 7-25 illustrates the tag cloud generated within the MIKE Model Discovery page. As the amount of agreements on the relevance of each search word to the metadata record(s) increases the search word is shown in a bolder and highlighted format (e.g. Cat).

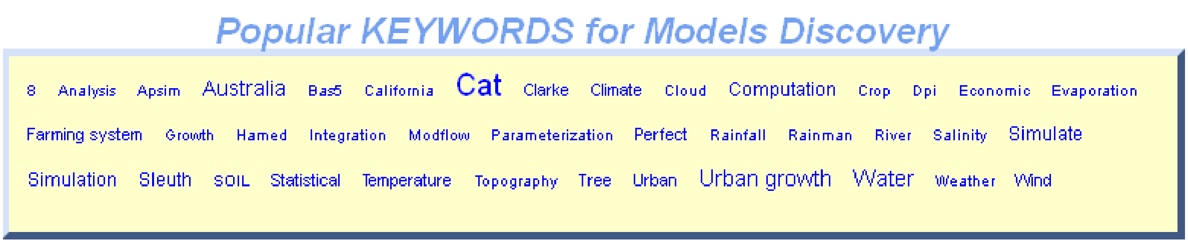


Figure 7-25: Tag cloud implemented within the MIKE Model Discovery page (addressing the seventh use case)

By clicking each search word (tag) within the cloud a pop-up screen would be displayed to present its related models, as well as, frequency of repetition of this search word. The frequency illustrates how many times the relevance of the search word to the model has been agreed by different end users explicitly or implicitly. For instance, Figure 7-26 illustrates the pop-up screen for 'Water' search word, once it is clicked in the tag cloud (see Figure 7-25).

The KEYWORD "Water " has been tagged to the following Model(s):

View Item	Model ID	Model Name	Frequency of Using This KEYWORD for This Model
View	7	APSIM	8
View	1	Modflow-96	26
View	8	CropSyst	16

Figure 7-26: Popup screen to show related models to each search word within the cloud (addressing the seventh use case explained on page 231)

The next section explores the implementation of the direct metadata enrichment model.

• Direct Metadata Enrichment Model

According to the conceptual design of the direct metadata enrichment model (see Section 6.5.2 in Chapter 6), there was a need for the functions to facilitate the end users' interaction to improve the content of 'keyword' metadata element. The end users interact in this process either based on their former knowledge of models or according to what they understand from reading the metadata details.

In this regard, the following two functions were implemented within the prototype system to address the fifth and sixth use cases (explained on page 231):

- agreeing with the previous tagged search words and tagging a new search word
- commenting on the models.

In order for the end users to agree with the previous search words, a function was developed to retrieve and visualise previously tagged search words when the end user clicks the 'YES' button in stage 1 of the indirect model (see Figure 7-24). This would enable the end users to agree with the existing search words and assign a new search word. Figure 7-27 illustrates the interface of this function.

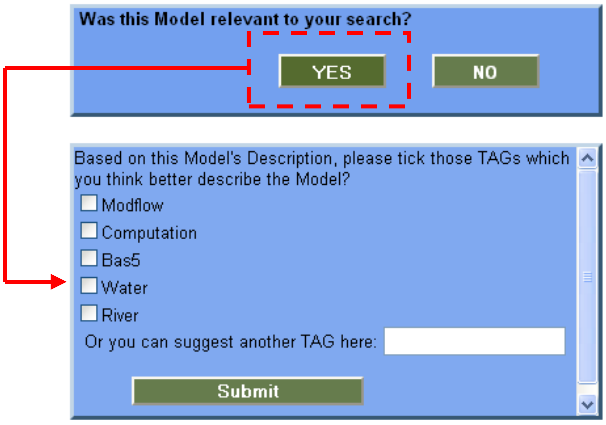


Figure 7-27: Enabling the end users to agree with the existing tagged search words and assign new search word (addressing the fifth use case explained on page 231)

In addition, to improve the content of metadata for each model another function was implemented within the Model Details page that enabled the end users to comment on the models. This function would give the opportunity to the end users to share their information, awareness and ideas on the fitness for use, advantages and disadvantages of the existing models. The comment could be another new type of metadata that enriches the model metadata based on the end users' experience and feedback.

The prerequisite to implement this function was to add a new table into the MIKE database. This table namely 'Comment_tbl2' contained 5 columns including ID, user_name (name of commenter-optional), modelID, and comment. This table was also inserted into the MIKE data model.

In order to comment, the end users could click the 'comments' button in Model Details page (see Figure 7-23) and post their comments through a pop-up screen, which is illustrated in Figure 7-28.

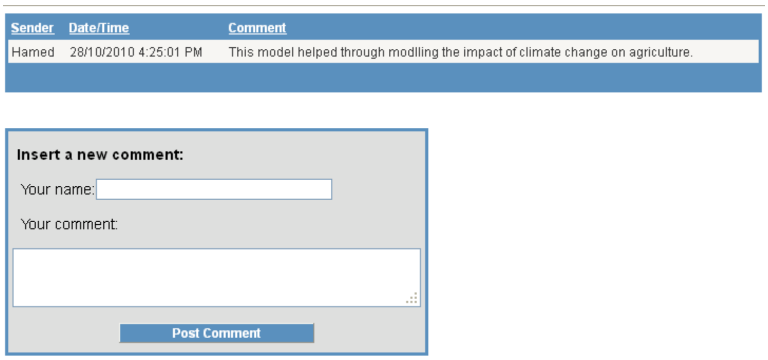


Figure 7-28: Pop-up screen to show and record comments on models to address the sixth use case (explained on page 231)

Based on the implementation of the indirect and direct metadata enrichment models, Figure 7-29 shows the overall view of the Model Discovery interface within the MIKE system.

Figure 7-29: Model Discovery interface within the MIKE system

The new functionalities implemented within the MIKE system have the potential to aid the MIKE interface to be more user-friendly by increasing the end users' interaction in the model discovery process. Also, the models discovery process can be facilitated via a tag cloud including the most popular search words. The end users will be able to explore the tag cloud and get an idea regarding the available and commonly used environmental models within Victoria, Australia. Both modellers and end users are also able to share their knowledge about models by posting comments on models.

According to the popularity of the GeoNetwork opensource among the geospatial community, and as another important research output, the next section explains the lessons learned during the implementation of automatic spatial metadata updating and enrichment prototype systems within the GeoNetwork environment. Sharing these lessons will likely result in facilitating further design and development of similar metadata automation approaches within this environment.

7.4 Lessons Learned from Prototype Systems Implementation within the GeoNetwork opensource

Some technical issues and challenges arose during the implementation of the automatic spatial metadata updating and enrichment prototype systems within the GeoNetwork opensource environment.

The main issue with the automatic metadata updating prototype system implementation was the mixed storage mechanism (XML in the relational database) within the GeoNetwork. This issue prevents the adoption of modern software development approaches for the user interface. However, the GeoServer, deegree and OpenLayers projects are at the forefront of standard implementation and were robust and modular components on top of which this prototype system could be built.

The current GeoNetwork's data model is not normalised according to the relational database paradigm. An entire metadata XML string is stored in a single database column. This means that accessing a single metadata element can only be undertaken using a parsing mechanism (XSL is the current mechanism). The entire technology stack (data model, service layer, user interface) is affected by this storage method. All operations on metadata data (extraction, checks, and updates) are cumbersome and difficult to maintain because they are locked in the XML paradigm. Bringing a relational and normalised data model to metadata management would be a significant step towards modern data management and software development techniques, but would require a large amount of work and a paradigm shift. Many software challenges encountered during the spatial metadata updating prototype system development were directly or partially related to this sole issue.

Moreover, as another challenge affecting the implementation of both prototype systems, the GeoNetwork's user interface is a direct consequence of the storage method (XML in a single column) – it has to perform the rendering of XML to HTML on the fly, using XSL transformations. Due to massive increases in interpreters speed, JavaScript has become the de facto standard for developing Web application interactions. Consequently, users are now expecting a level of interactivity far beyond what was the norm only several years ago. Conversely, developers are now expecting to implement user interactions using modern JavaScript frameworks (jQuery, ExtJS, etc.). Even though XSL and JavaScript do not directly compete as coding technologies, having a layer of XSL transformation is a burden on the development time, maintainability and evolution of the GeoNetwork's user interface components.

Furthermore, XSL, McKoi, and Jeeves are ageing technologies whose relevance is diminishing. This is demonstrated by the occurrences of the term XSL in English texts between 1995 and 2008 (Google 2011). This means that community support, documentation and the number of examples online are slowly decreasing. It can be argued that the use of these technologies is preventing a wider circle of contributors – whose skill-set is within more modern technologies – from advancing the GeoNetwork project.

In conclusion, it can be emphasised that although the GeoNetwork is a modular base with strong features such as harvesting, indexing, multi-criteria search and CSW cataloguing, it suffers from a number of issues. The main issues include the lack of a modern user interface and robust data model. Also, deviations from GeoNetwork's standard use (new functionalities) come with a steep learning curve due to an abundance of technologies. These issues might be the obstacles for more adoption of this open source project, from both a user and developer community perspectives. This means that a major overhaul of the project (basically re-developing a normalised data model and a modern user interface) is needed to ensure that the systemic issues identified above are dealt with, so that the project can be opened to a larger audience of contributors for renewed project sustainability.

7.5 Chapter Summary

This chapter explored the implementation of the automatic spatial metadata updating and enrichment prototype systems. These systems were developed to prove the concept of proposed approaches discussed in Sections 6.4 and 6.5 in Chapter 6. In this regard, the chapter first defined the use cases and designed the architecture for implementing the automatic spatial metadata updating. The chapter then explored the technologies and considerations used to implement this architecture in three main layers: storage, service and application. Next, the chapter reviewed the implementation of the automatic spatial metadata enrichment prototype system within two different environments: GeoNetwork opensource and MIKE. Different use cases were defined for each environment and accordingly the implementation architecture for each environment was designed and developed to address the indirect and direct metadata enrichment models already discussed in Sections 6.5.1 and 6.5.2 in Chapter 6. Finally, this chapter shared some of the main challenges encountered during the implementation of the automatic spatial metadata updating and enrichment prototype systems within the GeoNetwork environment. Sharing these challenges could facilitate further design and development of similar systems within the GeoNetwork.

The next chapter reviews the results that arose from the evaluation of automatic spatial metadata updating and enrichment prototype systems.

CHAPTER 8

PROTOTYPE SYSTEMS EVALUATION

8 PROTOTYPE SYSTEMS EVALUATION

8.1 Introduction

This chapter discusses the results of evaluating the automatic spatial metadata updating and enrichment prototype systems against the set of criteria suggested in research design, as mentioned in Section 4.4.4, Chapter 4. The information provided in this chapter leads to addressing the last research objective regarding gathering end users' and experts' feedback to recommend improvement to the proposed approaches for automating metadata and prototype systems designed and implemented to achieve this.

8.2 Evaluation of Automatic Spatial Metadata Updating Prototype System

The evaluation of automatic spatial metadata updating (synchronisation) prototype system implemented within the GeoNetwork opensource catalogue was undertaken based on the set of criteria suggested for this purpose, as presented in Table 4-2 in Chapter 4.

The set of criteria was proposed based on the requirements of an automatic spatial metadata updating approach identified during the research and included three main categories: system functionality, usability, and efficiency. The system functionality was relating to: the users' willingness to apply the integrated data model into the existing datasets and metadata in their organisations; overall effectiveness of integrated data model in facilitating datasets publication and related interoperability issues; usefulness of system in updating the dataset and metadata in real time; overall effectiveness of the system in improving the dataset discoverability; confidence of users to maintain metadata records using the system; and satisfaction of users with metadata elements supported by the system. The system usability criteria was developed based on the System Usability Scale (SUS) introduced by Brooke (1996) to explore the users' willingness to utilise the system; system ease of use (user-friendliness); users' confidence to use the system; need for learning a lot about the system before using it; need for technical support to work with the system; quickness of learning the system; the quality of functionalities integration within the system; and sufficiency of system functionalities. Finally, the system efficiency was related to understand the users' opinions whether time and cost might be saved for spatial metadata updating using the system.

According to the research design, a Web-based questionnaire was developed based on the set of criteria and distributed among the participating organisations of the initial survey of the research (outlined in Table 4-1, Chapter 4). A copy of the questionnaire is provided in **Appendix 4**. The URL of the questionnaires was also shared in professional social networks. In total, 13 organisations participated in the evaluation of automatic spatial metadata updating prototype system, as already outlined in Table 4-6 in Chapter 4.

Figure 8-1 illustrates the distribution of participating organisations in the evaluation of automatic spatial metadata updating prototype system internationally. As it can be seen, 77% of participants were from Australia. Then, the United States, Tanzania, and Pakistan had around 8% amount of contribution.

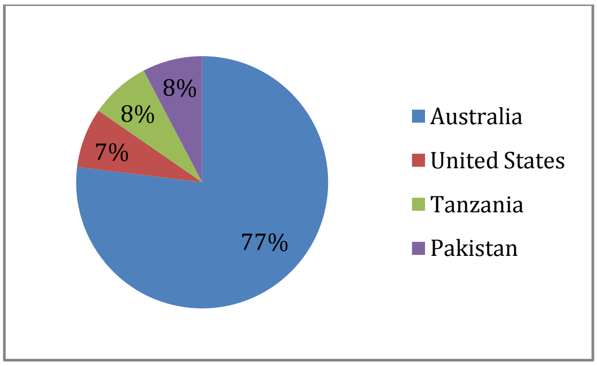


Figure 8-1: Countries participated in the evaluation of automatic spatial metadata updating prototype system

The following sections review the results of prototype system evaluation.

8.2.1 System Functionality

Information was first collected on the effectiveness of integrated data model designed and developed in this thesis based on the GML for storing both spatial data and metadata. The design of this data model was explored in detail in Section 6.4.2, Chapter 6 and the implementation of the integrated data model was also reviewed in Section 7.2.1, Chapter 7.

The participating organisations were asked to explain their willingness to apply the integrated data model into the existing datasets and metadata in their organisations. As illustrated in Figure 8-2, 69% of respondents agreed and other 8% strongly agreed that they desire to apply the integrated metadata data model into their existing systems. In addition, 23% of respondents did not state any specific willingness or opposition regarding this matter. None of them disagreed with this idea. Therefore, it was confirmed that most of the participants (77%) would like their organisations to adopt the integrated data model.

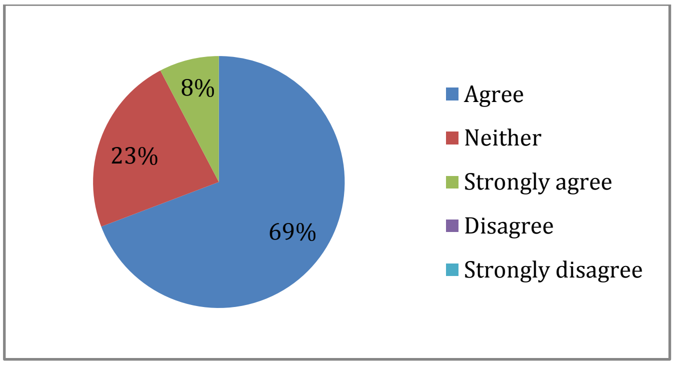


Figure 8-2: Participating organisations' extent of agreement on adopting the integrated metadata data model

The survey instrument also gathered information regarding the participating organisations' extent of agreement on the effective role of integrated data model to address the dataset's interoperability issues. As shown in Figure 8-3, 46% of participating organisations agreed and 8% strongly agreed with this expected role of new data model. Two independent groups of respondents (each around 8%) disagreed and strongly disagreed with this role. Moreover, the remaining 31% neither agreed nor disagreed. As a result, it was understood that publishing the dataset along with its metadata in GML format among the end users and spatial systems could be a potential alert to tackle dataset's interoperability issue. The role of GML in addressing the spatial data interoperability issue has been already acknowledged by the geospatial community (Zhang *et al.* 2003, Peng and Zhang 2004).

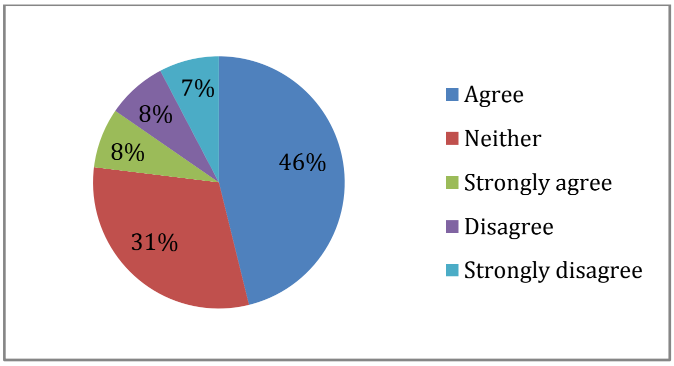


Figure 8-3: Participating organisations' extent of agreement on the effective role of integrated metadata data model in addressing datasets interoperability issues

Next, the participating organisations were asked to compare the GML-based integrated data model with the existing data models used in their organisations and share their views. They expressed that the new data model provided easier access to dataset and metadata. Also, using this data model, the dataset and metadata could be updated simultaneously. Moreover, the data model developed a foundation for metadata automation, harvesting and dealing faster with queries. On the other hand, some explained that the need for changing the organisation's policy regarding spatial data and metadata storage and software compatibility could be the main potential challenge affecting the new data model's adoption.

In addition, in this part of the survey, information was collected on the effectiveness of automatic spatial metadata updating (synchronisation) approach.

The participating organisations were asked to illustrate their extent of agreement on the usefulness of adopting synchronisation functionality within their organisations' existing metadata management system. As shown in Figure 8-4, 54% of them agreed and other 23% strongly agreed that adoption of synchronisation functionality would be beneficial to their organisations. However, 8% of them disagreed with the expected value of synchronisation approach acceptance. Also, the remaining 15% neither agreed nor disagreed with the expected value. Based on majority of the participating organisations' point of view (77% in total), it was understood that adopting synchronisation functionality, as proposed in this thesis, could be worthwhile for enhancing the organisations' existing metadata management system.

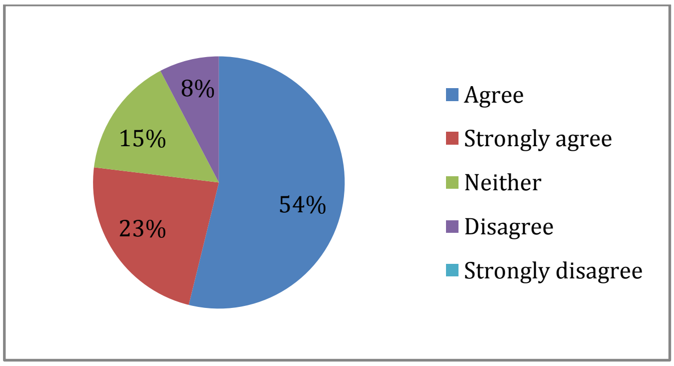


Figure 8-4: Participating organisations' extent of agreement on the usefulness of adopting synchronisation approach

Then, the participating organisations' views regarding the role of synchronisation approach developed to improve the dataset discoverability were gathered. As illustrated in Figure 8-5, 77% of participating organisations agreed that developed synchronisation approach facilitated the dataset discovery process through sharing the most up-to-date version of metadata. 15% of respondents did not have any particular opinion about the expected role of synchronisation approach. But 8% of participating organisations disagreed that this approach would play an important role in dataset search enhancement. According to the results, it was recognised that the synchronisation approach could play a key role in advancing the dataset discoverability. This result was already acknowledged by Westbrooks (2004b) who argued that the lack of tools for metadata synchronisation hinders the spatial data discovery and access.

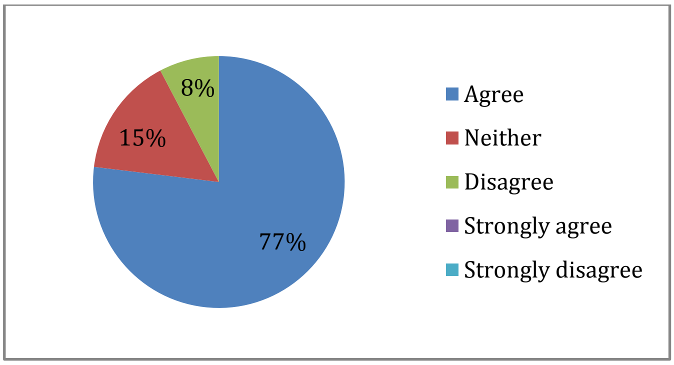


Figure 8-5: Participating organisations' extent of agreement on the role of synchronisation approach in improving the dataset discoverability

Also, the survey instrument gathered information on the participating organisations' confidence to use the system based on synchronisation approach to update metadata for the datasets they maintain. As a result, 69% of respondents acknowledged that they would be sufficiently confident to use the system to update metadata records. On the other hand, the remaining 31% were not completely certain that they would use the system due to some concerns. The uncertainty regarding possibility of integrating synchronisation functionality with participants' existing systems, lack of support for some modification functionalities (e.g. attributes editing tool), and lack of support for updating feature-level metadata were the main concerns.

Finally, in this part of the survey the participating organisations were asked to explain their views regarding the metadata elements supported by the automatic spatial metadata updating prototype system (including metadata date stamp, date of last update, lineage statement, north bounding latitude, south bounding latitude, west bounding longitude, and east bounding longitude). It was found that 77% of participating organisations confirmed that supported elements would satisfy their needs whenever a change happened to the dataset. However, the remaining 23% expected additional metadata elements to be supported by the system, such as dataset projection, storage format, type of modifications applied to dataset, and how the modification are applied (in terms of used software, parameters, and process/workflow).

Once the prototype system level of effectiveness was evaluated by the participating organisations, the survey focused on assessing the usability of the system. The next section explored the results of responses analysis regarding the system usability.

8.2.2 System Usability

As discussed in Section 4.4.4 in Chapter 4, a customised version of the System Usability Score (SUS) including 8 statements was used for evaluating the prototype system usability. These statements were mapped to the SUS original statements and outlined in Table 4-3, Chapter 4.

The way to calculate the score of each statement was also discussed in Section 4.3.2, Chapter 4. However, since Brooke (1996) cautioned that 'SUS scores for individual items are not meaningful on their own', the overall SUS score needed to be calculated to measure the prototype system usability. This overall score will be later outlined in this section.

In this part of the survey, the participating organisations were asked to state their willingness to use the demonstrated prototype system for metadata updating in their organisations. According to Figure 8-6, 54% of participating organisations agreed and 8% of them strongly agreed that they would like to use the automatic spatial metadata updating (synchronisation) system more frequently than the existing metadata management systems in their organisations; 23% neither agreed nor disagreed with using this system over their existing systems. However, 8% disagreed and remaining 7% strongly disagreed with using the presented system. As a result, the SUS score calculated for the relevant statement was equal to 2.461 (out of 4).

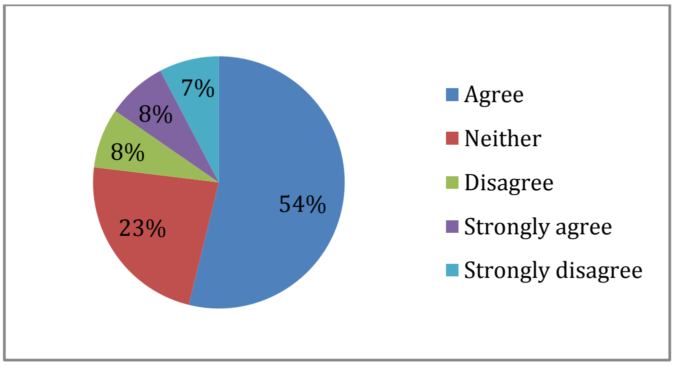


Figure 8-6: Participating organisations' extent of agreement on willingness to use the system

Next, the participating organisations were asked to share their views regarding the user-friendliness of the automatic spatial metadata updating system. According to Figure 8-7, 77% of participating organisations agreed and 15% of them strongly agreed that the system is user-friendly. The remaining 8% neither agreed nor disagreed with the system user-friendliness. Accordingly, the SUS score calculated for the related statement was equal to 3.077.

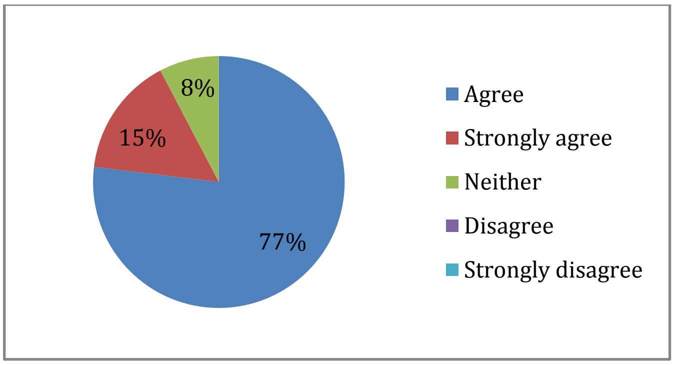


Figure 8-7: Participating organisations' extent of agreement on system ease of use (user-friendliness)

The participating organisations were also asked to state their level of confidence to use the automatic spatial metadata updating system. According to Figure 8-8, 54% of participating organisations agreed and other 8% of them strongly disagreed that they were confident to use the system. The remaining 38% shared no specific opinion regarding this subject. Hence, the SUS score calculated for the relevant statement was equal to 2.38.

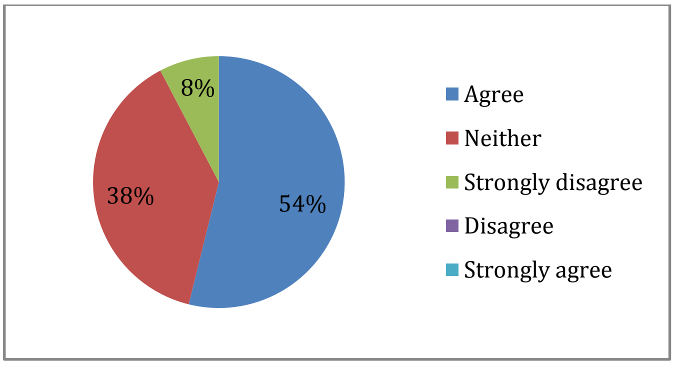


Figure 8-8: Participating organisations' extent of agreement on the confidence to use the system

In addition, the survey instrument gathered data regarding the participating organisations' extent of agreement on the need for learning a lot of things before using the automatic spatial metadata updating system. According to Figure 8-9, 31% of participating organisations agreed that they needed to learn a lot about the system before they could get going with it. However, 46% of them disagreed and another 8% strongly disagreed with this prerequisite for working with the system. Also, the remaining 15% did not have any specific point of view. Therefore, the SUS score calculated for the related statement was equal to 2.31.

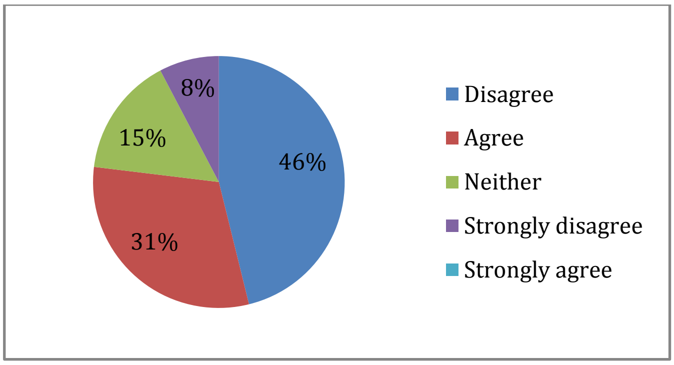


Figure 8-9: Participating organisations' extent of agreement on the need for learning a lot of things before using the system

The participating organisations were also asked whether they needed the support of a technical person to be able to work with the automatic spatial metadata updating system. As shown in Figure 8-10, 8% of participating organisations agreed that they would need a technical person's support. However, 62% of respondents disagreed and other 15% strongly disagreed with the need for this level of support from a technical person to work with the system. The remaining 15% neither agreed nor disagreed. Thus, the SUS score calculated for the related statement was equal to 2.846.

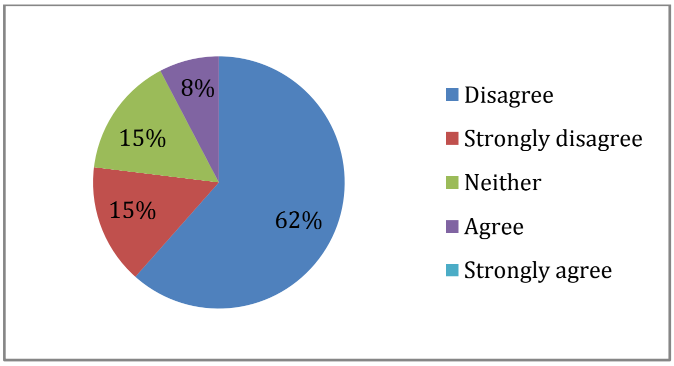


Figure 8-10: Participating organisations' extent of agreement on the need for technical support to work with the system

The participating organisations also shared their point of view about the quickness of learning about the automatic spatial metadata updating system. According to Figure 8-11, 69% of participating organisations agreed and other 23% strongly agreed that people responsible for metadata could learn to use this system very quickly. The remaining 8% shared no specific idea regarding this matter. As a result, the SUS score calculated for the related statement was equal to 3.154.

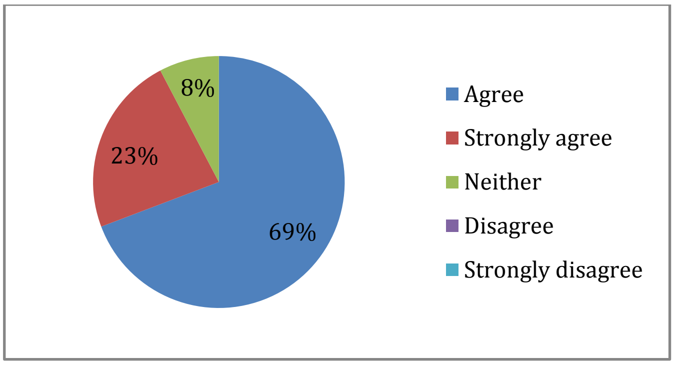


Figure 8-11: Participating organisations' extent of agreement on the quickness of learning the system

The participating organisations were asked to share their extent of agreement on the quality of functionalities integration within the automatic spatial metadata updating system. These functionalities included the services for dataset and metadata updating within the prototype system interface and the GeoNetwork interface. According to Figure 8-12, 77% of respondents agreed that the functionalities are well integrated. Also, the remaining 23% had no particular opinion about the quality of functionalities integration. Therefore, the SUS score calculated for the related statement was equal to 2.769.

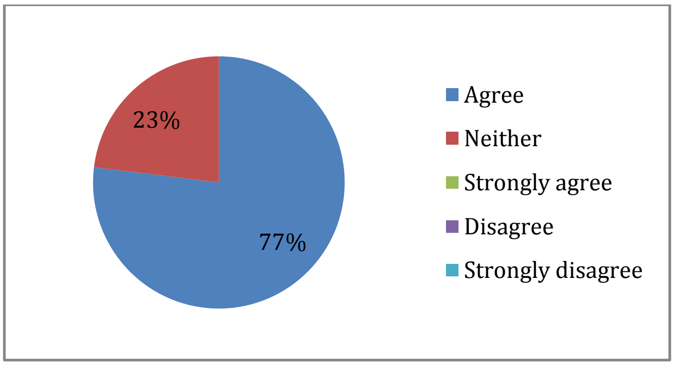


Figure 8-12: Participating organisations' extent of agreement on the quality of functionalities integration within the system

The survey instrument collected data regarding the participating organisations' extent of agreement on the sufficiency of functionality to make the automatic spatial metadata updating system a useful tool. As illustrated in Figure 8-13, 61% of respondents agreed and other 8% strongly agreed that there was enough functionality in the system. On the other hand, 8% of them disagreed with the sufficiency of functionality in the system. From their point of view, the system needed to support more editing tools for spatial data and attributes modification. The remaining 23% neither agreed nor disagreed with the level of supported functionality. Hence, the SUS score calculated for the related statement was equal to 2.692.

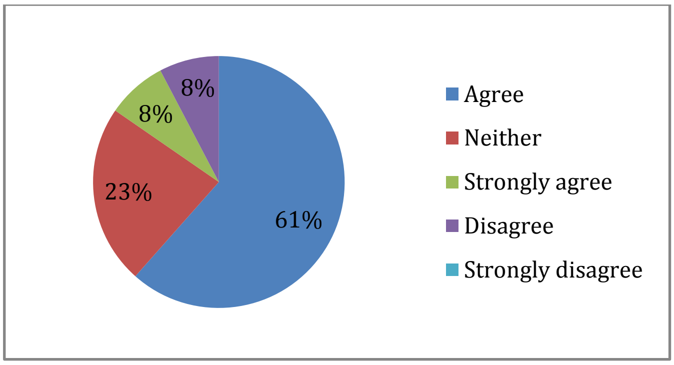


Figure 8-13: Participating organisations' extent of agreement on sufficiency of system functionalities

Following the SUS scores achieved by the eight statements customised for evaluating the usability of this prototype system, the overall SUS score was equal to 67.8 out of 100 by multiplying the sum of scores for these statements by 3.125 (see Section 4.4.4 in Chapter 4). According to the rule-of-thumb standard, the so-called 'university grade analog', discussed by Bangor *et al.* (2008), the SUS score of this prototype system sits between 'OK' and 'GOOD' ratings, among 'worst imaginable', 'poor', 'ok', 'good', 'excellent', and 'best imaginable' ratings. Based on this standard, in terms of the acceptability, the prototype system is located in 'high marginal' range among 'not acceptable', 'low marginal', 'high marginal', and 'acceptable' ranges. This means that the automatic spatial metadata updating prototype system should be considered a candidate for increased scrutiny and continued improvement in terms of usability to pass at least the SUS score of 70, if it is supposed to be converted to an 'acceptable' fully-functional product.

In this part of the survey, the participating organisations also shared their positive and negative views regarding the automatic spatial metadata updating approach and associated prototype system. In their opinion, facilitating and speeding up the spatial data and metadata updating process (especially in a Web-based environment) is one of the positive aspects of the approach. Also, they stated that this approach would reduce time and effort required for metadata updating and decrease human error, which commonly happens in manual metadata entry. In their view the proposed approach provided an appropriate integration of dataset and metadata. Moreover, based on their opinion the approach gave peace of mind to data responsible parties due to the metadata being always current with the dataset changes. Furthermore, they found the prototype system very easy-to-use, easy-to-follow and well-integrated.

On the other hand, lack of support for maintaining spatial topology and attributes, limited dataset editing capabilities, lack of tests with large volumes of data, and uncertainty about the software compatibility and integration of this system with the other existing systems in the organisations were the main negative aspects of the approach outlined by the participating organisations.

8.2.3 System Efficiency (Time and Cost)

In this part of the survey information was gathered on the potential efficiency of automatic spatial metadata updating approach against the existing approaches in the organisations.

In this regard, the participating organisations were asked to describe the current status of dataset and metadata updating in their organisations. They stated the time usually spent on updating the values of the following subset of ISO 19115: 2003 metadata elements using their existing metadata systems for each dataset:

- Metadata date stamp
- Date of last update
- Lineage statement
- North bounding latitude
- South bounding latitude
- West bounding longitude
- East bounding longitude.

These elements have been selected because the automatic spatial metadata updating prototype system has employed them for the metadata synchronisation purpose.

As illustrated in Figure 8-14, 31% of participating organisations acknowledged that updating the above metadata values after any modification to the dataset took around 5 minutes per dataset. Also, 15% of them agreed that this task takes about half an hour per dataset using their existing systems. Another 15% indicated that it takes 10 minutes for updating the above metadata values for each dataset in their organisations. Also two independent groups of participating organisations (each around 8%) indicated that they usually spend between 1 minute and 15 minutes to update the values of the asked metadata elements using the existing systems. The remaining (23%) did not answer the related question. Accordingly, the average time for updating the above metadata values after any modification to the dataset was equal to 11.6 minutes per dataset.

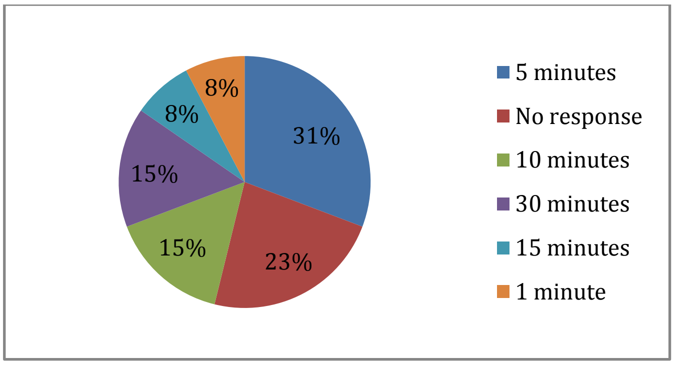


Figure 8-14: Required time for updating subset of ISO 19115: 2003 metadata values in different organisations

However, as discussed earlier in Section 7.2.1 in Chapter 7, these metadata values could be automatically and simultaneously updated whenever the dataset is modified using the automatic spatial metadata updating approach. Therefore, it was understood that applying this approach to the existing metadata management systems within the organisations could result in improving the efficiency of metadata updating in terms of time saving by 11.6 minutes per dataset (as the average time). In order to confirm this, the participating organisations were asked to state their idea regarding the efficiency of automatic spatial metadata updating approach in terms of time. According to Figure 8-15, 77% of the participating organisations agreed and another 8% strongly agreed that using the automatic metadata updating system would likely result in saving time for maintaining metadata records in their organisations. The remaining 15% did not mention any specific opinion regarding the role of system in saving time. Accordingly, it was considered affirmative that in comparison with the existing metadata management systems in the organisations the automatic spatial metadata updating approach could improve the efficiency of metadata updating in terms of time.

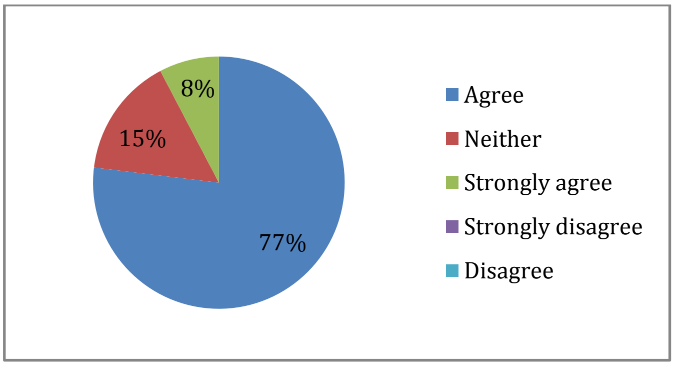


Figure 8-15: Participating organisations' extent of agreement on efficiency of automatic spatial metadata updating approach in terms of time

Because only a subset of ISO 19115: 2003 metadata elements were employed in the automatic spatial metadata prototype system to prove the associated concept, measuring the actual cost benefit resulted from the system in the participating organisations was complicated and the respondents could not answer the relevant questions in the questionnaire (in Part 5, **Appendix 4**). However, the respondents were asked to share their general opinion regarding the role that the prototype system could play in improving the efficiency of metadata updating in terms of cost. According to Figure 8-16, 69% of them agreed that utilising the automatic metadata updating system would likely result in cost savings for their organisation. Also, the remaining 31% had no specific opinion regarding this matter.

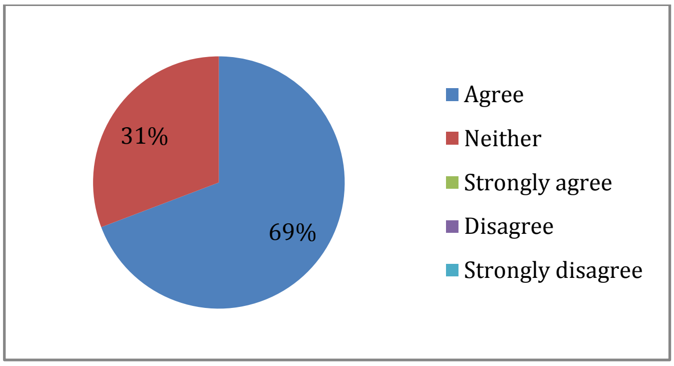


Figure 8-16: Participating organisations' extent of agreement on efficiency of automatic spatial metadata updating approach in terms of cost

As a result, it was considered that the proposed automatic spatial metadata updating approach in this thesis could improve the efficiency of metadata updating not only in terms of time but also in terms of cost.

The next section explores the results of evaluating the automatic spatial metadata enrichment prototype system implemented within the GeoNetwork opensource catalogue.

8.3 Evaluation of Automatic Spatial Metadata Enrichment Prototype System

Similar to the evaluation of automatic spatial metadata updating prototype system, the evaluation of the automatic spatial metadata enrichment prototype system implemented within the GeoNetwork opensource catalogue (including 'agree/disagree/tag cloud' and 'suggestion list' add-ons) was undertaken based on the set of criteria proposed for this purpose, as already outlined in Table 4-4, Chapter 4.

The set of criteria was suggested based on the requirements of an automatic spatial metadata enrichment approach identified during the research and included two main categories: system functionality, and usability. The system functionality related to: the preferences regarding spatial data discovery methods; overall effectiveness of system in improving the content of keyword metadata and data discovery; overall effectiveness of designed search word weighting system; overall effectiveness of user-generated search words in data discovery against the original search words embedded in metadata; need for an approval process for user-generated search words; and overall effectiveness of a tag cloud in facilitating dataset search and discovery. The system usability criteria was developed based on the System Usability Scale (SUS) introduced by Brooke (1996) to explore the users' willingness to utilise the system, system ease of use (user-friendliness), need for technical support to work with the system, and the quality of add-ons integration within the system.

According to the research design discussed in Section 4.4.4 in Chapter 4, a Web-based questionnaire was designed based on the set of criteria and distributed among the participating organisations of the initial survey of the research (outlined in Table 4-1, Chapter 4). A copy of the questionnaire is provided in **Appendix 5**. The URL of the questionnaires was also shared in professional social networks. In total, 26 organisations from different countries participated in the evaluation of prototype system, as already outlined in Table 4-7, Chapter 4.

Figure 8-17 illustrates the distribution of participating organisations in the evaluation of enrichment prototype system internationally. As it can be seen, 50% of participating organisations were from Australia. The United States was the next most important player with 30% and Italy, Indonesia, Pakistan, Spain, and Tanzania had the same 4% amount of contribution.

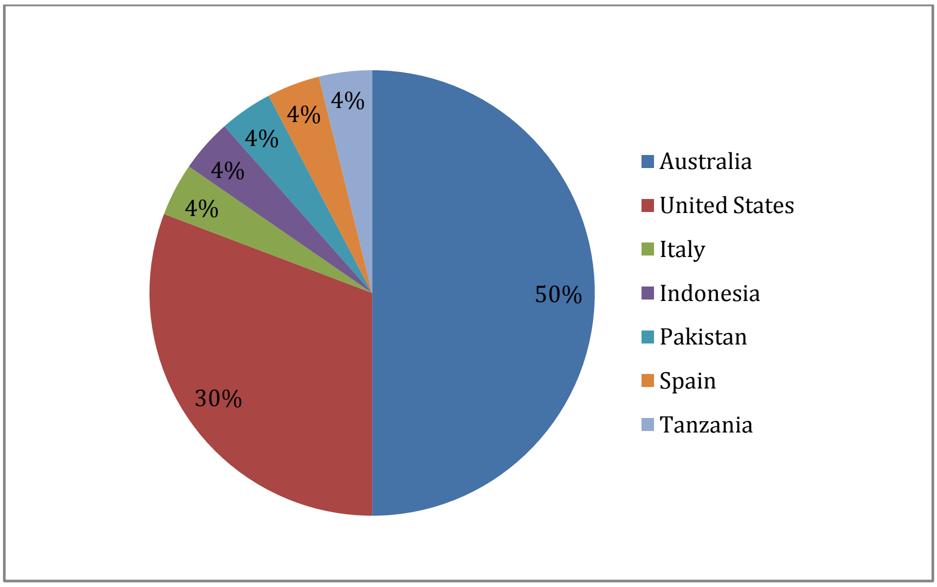


Figure 8-17: The countries participated in the evaluation of automatic spatial metadata enrichment prototype system

The following sections review the results of prototype system evaluation.

8.3.1 System Functionality

This part was aiming to evaluate the effectiveness of prototype system in terms of its functionality.

First, the survey instrument gathered information on the participating organisations' preferences for dataset discovery within a data catalogue system. The main purpose for collecting this information was to realise the expectations of geospatial community regarding data discovery process. The related question was asked prior to the introduction and presentation of automatic spatial metadata enrichment prototype system within the GeoNetwork opensource catalogue to the participating organisations. According to Figure 8-18, the results showed that 39% of the respondents liked the data catalogue that suggested the related previously searched words and their frequency whilst typing a search word in the search box. Also, 22% of respondents preferred to only type a search word in search box and then push the search button. In addition, 9% of respondents confirmed that accessing a cloud of most popular search words (tag cloud) would be desirable for them. 17% of respondents agreed that using the basic search equipped with suggestion list functionality would be appropriate. Moreover, other 9% liked to access all proposed discovery functionalities including basic search, suggestion list and tag cloud. The remaining 4% also chose the basic search and tag cloud as their desirable search functionalities.

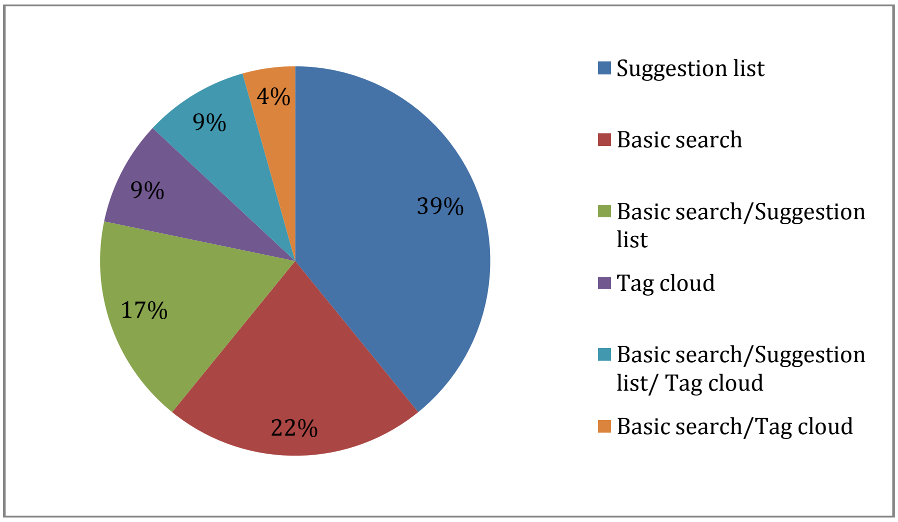


Figure 8-18: Participating organisations' preferences regarding spatial data discovery methods

Some of the participating organisations shared their suggestions regarding other desirable discovery methods. These suggestions have been classified in the following 5 categories:

1. monitoring the search words used by any end user in data discovery and suggesting those to him/her in the next search attempts

2. searching based on a combination of search words, spatial extent, and temporal extent
3. using ontology-based engines for searching
4. typing the search word while a search is started of the datasets (based on resource title metadata element) and a drop down list is populated
5. typing a description of the required information and the catalogue system suggests the datasets that could meet their needs.

Once the above information was collected, the participating organisations were informed comprehensively about the add-ons designed and implemented for the automatic spatial metadata enrichment approach (‘agree/disagree/tag cloud’ and ‘suggestion list’). Then, they were asked mainly about the effectiveness of the add-ons during the survey.

In this regard, the information was gathered on the extent of participating organisations’ agreement on the effective and helpful role of add-ons in interacting with the end users to improve the content of metadata and datasets search and discovery process. According to Figure 8-19, the results showed that 61% of respondents agreed and other 27% strongly agreed that the expected role of add-ons to improve the content of metadata keyword and dataset discovery had been achieved. In addition, 8% had a neutral opinion about the add-ons’ role and 4% disagreed with this expected role. Overall, a high extent of agreement (88% in total) with the effective role of add-ons in interacting with the end users to advance metadata keyword content and facilitate data discovery process confirmed the successful proof of automatic spatial metadata enrichment concept.

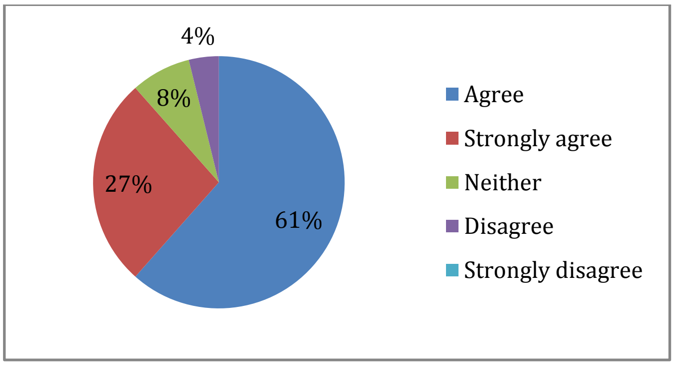


Figure 8-19: Participating organisations’ extent of agreement on the effectiveness of add-ons in interacting with end users to improve metadata keyword and data discovery

The participating organisations were also asked whether they think it would be useful to apply the automatic spatial metadata enrichment into their existing spatial data discovery systems. As illustrated in Figure 8-20, 77% of them agreed and other 8% strongly agreed that adoption of automatic metadata enrichment approach would be valuable for their organisations; 8% of them shared no specific opinion; and the remaining 7% disagreed with the expected role of automatic metadata enrichment within their existing discovery systems. As a result, it could be understood that most of the participating organisations (85% in total) found the designed and developed indirect and direct metadata enrichment models derived from end users' interaction useful to improve the current data discovery systems.

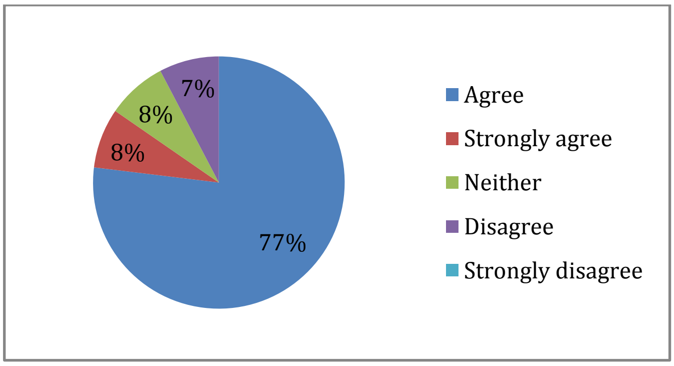


Figure 8-20: Participating organisations' extent of agreement on the usefulness of applying the automatic spatial metadata enrichment into their existing data discovery systems

In addition, the participating organisations shared their views regarding the role of 'suggestion list add-on', discussed in Section 7.3.1 in Chapter 7, to accelerate the data discovery task and share the previous users' interaction. As shown in Figure 8-21, 61% of respondents agreed and other 27% strongly agreed that suggesting the previous searched terms while typing the keyword during the discovery process would result in facilitating the data discovery. But 8% did not have any specific idea and the remaining 4% disagreed with the positive role of this add-on in improving the data search. Therefore, it could be understood that most of the participating organisations (88% in total) agreed on the positive role of suggestion list add-on in facilitating the data discovery process.

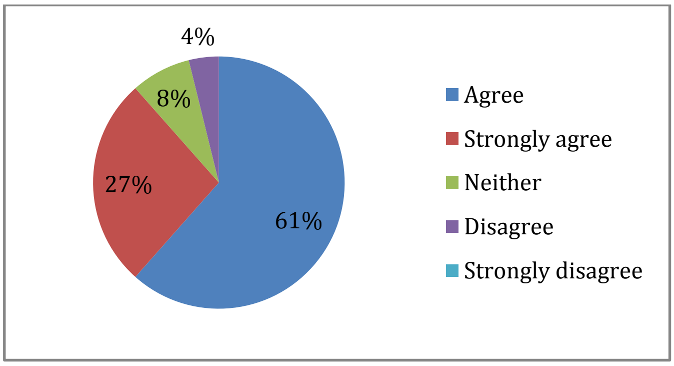


Figure 8-21: Participating organisations' extent of agreement on the positive role of suggestion list add-on in accelerating data discovery task and sharing the previous users' interaction

Next, the search word weighting system proposed for the prototype system (see Tables 6-6 and 6-7 in Chapter 6) was introduced to the participating organisations and they were asked to explain their extent of agreement on the logic of weighting system. The results illustrated that 73% of participating organisations agreed and other 4% strongly agreed with the weighting system logic. Another 15% had no specific idea and 8% disagreed with the proposed system logic. Figure 8-22 shows the extent of respondents' agreement on the logic of developed search word weighting system.

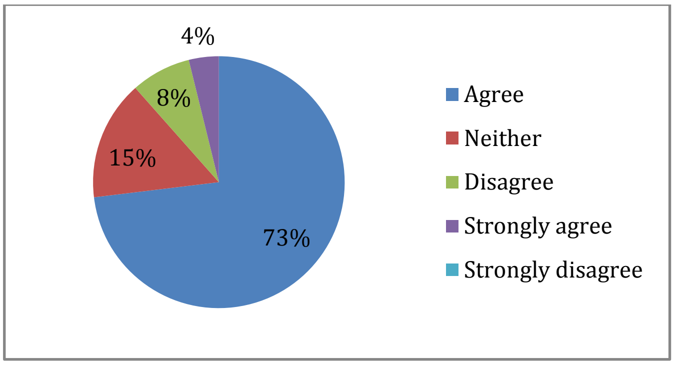


Figure 8-22: Participating organisations' extent of agreement on developed search word weighting system

Although, the overall extent of agreement (77%) proved that the designed weighting system was valid in the participating organisations' point of view; they were encouraged to propose any other factors that should have been considered in this system. The participating organisations contributed extensively and shared their ideas and suggestions, as grouped and summarised in Table 8-1.

Table 8-1: Summary of participating organisations' suggestions for improving the search word weighting system

Suggestion topic	Suggestion description
Categorising and weighting search words based on the end user's purpose of search	Any end user's query is specific to his/her interest/purpose. For instance, if a user searches 'waterbodies' for pollution study, he/she is not interested in being led to the same data that might have been used by a real estate agent to site a housing development. This user would be looking for data that included sampling, flow and other hydrological attributes while the real estate agent is more interested in viewshed and proximity to services. Therefore, there is a need for recognition of user's interest and purpose while searching and categorising and weighting the search words accordingly.
Increasing weight of dataset 'download'	Downloading dataset by the user is solid evidence that the search word was useful for finding dataset. Only assigning a weight of 0.75 to the related search word devalues the human decision on the relevance of search word to the downloaded dataset.
Controlling same user's behaviour for agreeing/ disagreeing	Agreeing/disagreeing more than once per search word by the same user could decrease the accuracy of weighting system. Thus, there should be a process for controlling same user's behaviour for agreeing/disagreeing.
Removing the positive weight for exploring metadata details	The action of pressing the metadata button to expand the information of a given dataset does not necessarily mean that the user was looking for this exact dataset. It could be realised that the user was looking for this dataset, but there is a small or even a higher percentage of people just exploring the catalogue to see what is available. Therefore, recording weight for exploring metadata details might add more noise to the results of the weighting process than making it more accurate.
Replacing considered weight for 'enough time spent on reading metadata' with 'how much time spent on reading metadata'	Considering a range of weights based on the time user spends on reading metadata details would be a more accurate method for weighting search word. More time spent on metadata details indicates more relevance between search word and related dataset. The users usually exit the irrelevant pages quickly while searching.
Engaging user credibility in weighting system	A different weighting system should be applied for trusted and mistrusted users of any catalogue system. In this case, there is a need for a user authentication system within the data catalogue.

Once the participating organisations' feedback on the weighting system was gathered, they were asked to indicate if they agree that search words provided by the users' interaction in the discovery process are more practical for finding the datasets than the search words already provided in the metadata record (different ways of creating search words were explored in Section 6.5, Chapter 6). According to Figure 8-23, the results showed that 34% neither agreed nor disagreed that the user-generated search words are more practical than the search words included in the metadata; 27% agreed and other 8% strongly agreed that the search words created by the users' interaction are more effective for data discovery purpose. On the other hand, the remaining 31% disagreed with the effective role of user-generated search words in comparison with the search words originally included in the metadata. After removing the neutral responses from the analysis, it was shown that there is only a slightly positively skewness toward agreement with the higher effectiveness of user-generated search words in data discovery against the original search words embedded in metadata.

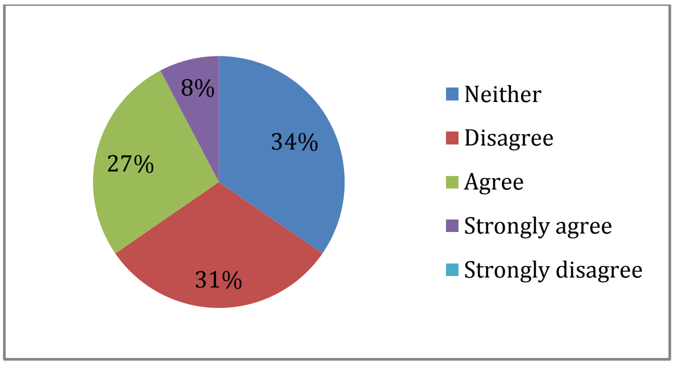


Figure 8-23: Participating organisations' extent of agreement on higher effectiveness of user-generated search words in data discovery against the original search words embedded in metadata

In the next step, information was gathered on the extent of participating organisations' agreement on the need for an approval process before considering the user-generated search word as a new metadata value. According to Figure 8-24, 27% of participating organisations strongly supported this idea. Also, 46% acknowledged that such a process should exist. However, other 19% disagreed with the idea and confirmed that the data responsible parties should trust user-generated search words. Other 8% neither agreed nor disagreed with the proposed idea. The majority of positive responses acknowledged the importance of the service implemented within the automatic spatial metadata updating prototype system (see Figure 7-12 in Chapter 7) to allow the responsible parties to review and approve the user-generated search words as new metadata values.

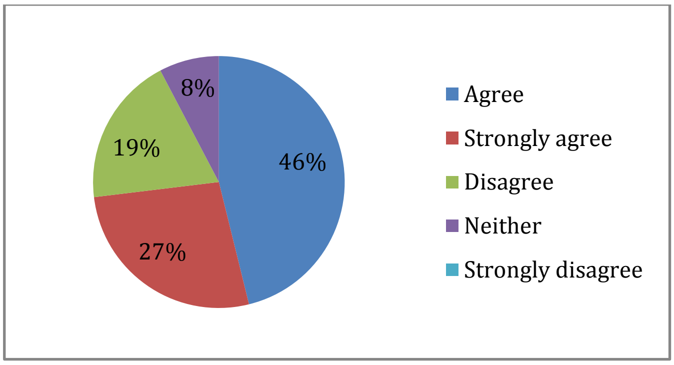


Figure 8-24: Participating organisations' extent of agreement on an approval process for user-generated search words

Next, the participating organisations were asked to share their level of agreement on the role of 'spatial tag cloud' as a valuable addition to support the search and discovery functionality of a data catalogue system. As illustrated in Figure 8-25, 46% of respondents agreed and other 12% strongly agreed with the effective role of tag cloud in data discovery. In contrast, 11% of them disagreed with the expected role of tag cloud. The remaining 31% had neutral opinion regarding this matter. Accordingly, it was understood that using the tag cloud, which presents the most popular search words and their weights, could play a significant role in improving the current spatial data search and discovery approach.

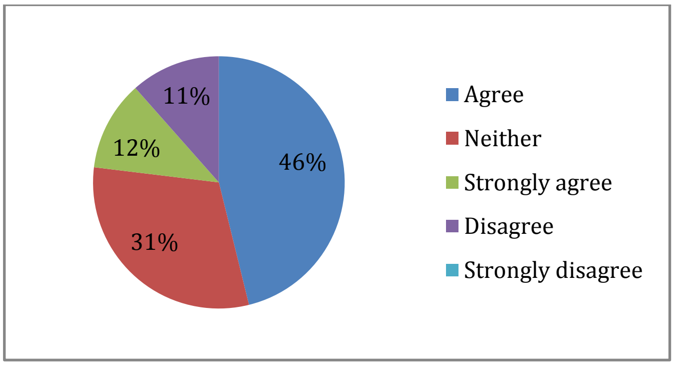


Figure 8-25: Participating organisations' extent of agreement on the role of tag cloud in facilitating dataset search and discovery

Once the add-ons' level of effectiveness was evaluated by the participating organisations, the survey concentrated on assessing the usability of the add-ons. The next section explores the results of responses analysis regarding the system usability.

8.3.2 System Usability

As discussed in Section 4.4.4 in Chapter 4, a customised version of System Usability Score including 4 statements was used for evaluating the prototype system usability. These statements were already mapped to the SUS original statements and outlined in Table 4-5, Chapter 4. The way to calculate the score of each statement was also discussed in Section 4.3.2, Chapter 4. As discussed earlier, to assess the usability of the prototype system the overall SUS score needs to be calculated based on the individual scores. This score will be later outlined in this section.

In this part, the participating organisations were first asked to state their willingness to use the demonstrated system for data discovery in their organisations. According to Figure 8-26, 46% of participating organisations agreed and the other 4% strongly agreed that they would like to use the developed system for data discovery purpose more than their existing data discovery systems. In addition, 31% of respondents neither agreed nor disagreed with this matter. However, the remaining 19% disagreed that they would like to use the system more than their existing data discovery systems. As a result, the SUS score calculated for the related statement was equal to 2.65 (out of 4).

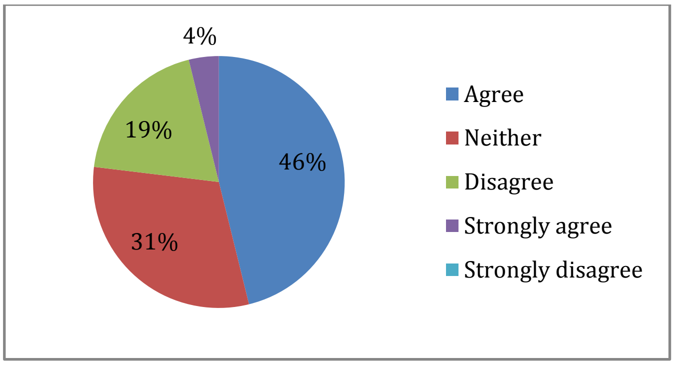


Figure 8-26: Participating organisations' extent of agreement on willingness to use the system

Next, the participating organisations were asked to indicate their level of agreement on the user-friendliness of the prototype system. As illustrated in Figure 8-27, 77% of participating organisations agreed and other 4% also strongly agreed that the prototype system is user friendly. Other 19% of respondents shared no specific opinion. Therefore, the SUS score calculated for the related statement was equal to 2.843.

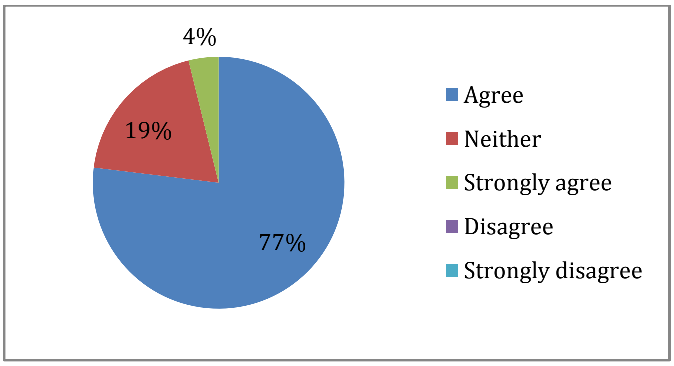


Figure 8-27: Participating organisations' extent of agreement on the system user-friendliness

The participating organisations were also asked whether they need the support of a technical person to be able to work with the automatic spatial metadata enrichment system. According to Figure 8-28, 65% of respondents disagreed and other 19% strongly disagreed with the need for a technical support. On the other hand, 12% agreed that they would need technical support. The remaining 4% indicated no specific idea regarding this matter. Therefore, the SUS score calculated for the relevant statement was equal to 2.921.

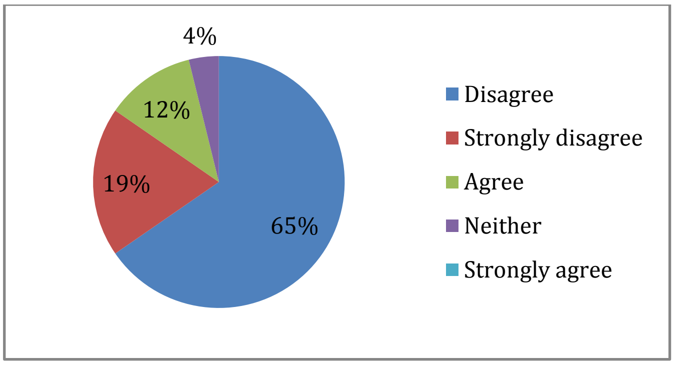


Figure 8-28: Participating organisations' extent of agreement on the need for technical support to work with the system

Then, the participating organisations were asked to indicate their point of view regarding the quality of add-ons ('agree/disagree/tag cloud' and 'suggestion list' add-ons) integration in the prototype system. As illustrated in Figure 8-29, 62% of participating organisations agreed that the add-ons and related functionalities were well-integrated within the system. Other 38% of them had no specific opinion regarding this matter. Thus, the SUS score calculated for the relevant statement was equal to 2.615.

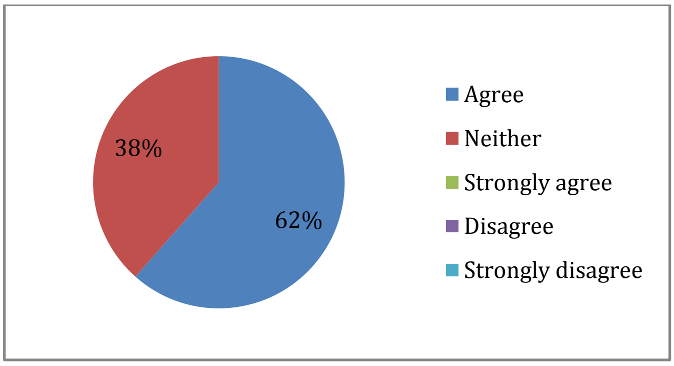


Figure 8-29: Participating organisations' extent of agreement on the quality of add-ons integration within the system

According to the SUS scores achieved in the above 4 statements, the overall SUS score of the automatic spatial metadata enrichment prototype system was calculated and was equal to 68.9 out of 100 by multiplying the sum of scores for these statements by 6.25 (as discussed in Section 4.4.4 in Chapter 4). According to the rule-of-thumb standard, discussed by Bangor *et al.* (2008), the SUS score of this prototype system sits between the 'ok' and 'good' ratings among 'worst imaginable', 'poor', 'ok', 'good', 'excellent', and 'best imaginable' ratings. Based on this standard, in terms of the acceptability the prototype system is located in 'high marginal' range among 'not acceptable', 'low marginal', 'high marginal', and 'acceptable' ranges. This means that the automatic spatial metadata enrichment prototype system should be improved in terms of usability to pass at least the SUS score of 70, if it is supposed to be converted to an 'acceptable' fully-functional product. This result was very similar to the result of evaluating the usability of the automatic spatial metadata updating prototype system.

In this section of the survey, the participating organisations also shared their opinions regarding the positive aspects of the automatic spatial metadata enrichment approach. From their point of view, this automatic approach improved the effectiveness and efficiency of data discovery by reducing time and effort used to add metadata using conventional approaches. In addition, the proposed approach boosted the relevance of search words to datasets and therefore more accuracy in search results would be achieved. Building a standardised search word thesaurus was reported as another potential benefit of the new approach. Moreover, the proposed approach gets one step closer to bridging the gap between spatial data users and providers through engaging the users in metadata enrichment process. Furthermore, the search functionality provided by the new approach is seen consistent with current Web experience.

However, some negative aspects of the automatic enrichment approach were reported by the participating organisations in this part of survey. The approach mainly assumed that users are sufficiently well-informed to make suitable choices which may not be the case in reality. Also, the approach generated additional overhead activity for data responsible parties to verify and approve the user-generated search words. There is also a risk of associating irrelevant, incorrect, and misspelled search words to datasets by end users. In addition, this approach required a certain amount of use to be valuable, which may be an obstacle for a rarely-used catalogue to adopt it. Moreover, this approach may interrupt flow from search to accessing the data because of more clicks required. Another negative aspect of the approach could be making the newer datasets (with a low ranking) less frequently discovered and used. In terms of presentation, as number of search words increased, the tag cloud may get more complicated to use.

There were also different views regarding the use of GeoNetwork for prototyping the automatic metadata enrichment approach. Most of the participating organisations considered this as a positive aspect of prototype system. On the other hand, a few of respondents stated that GeoNetwork would not be a good platform for implementing the prototype system because it is technologically obsolete, and in some point of the time a full re-engineering would be needed. This disadvantage of the GeoNetwork raised by the participating organisations was already pointed out in Section 7.4 in Chapter 7 during the exploration of lessons learned from the implementation of automatic spatial metadata updating and enrichment prototype system within the GeoNetwork.

In the last stage of the survey, information was gathered on expected improvements to the functionality and user interface of the prototype system. The proposed suggestions are grouped and outlined in Table 8-2.

Table 8-2: Expected improvements in automatic spatial metadata enrichment prototype system in participating organisations' point of view

Improvement suggestion
In order for the geospatial community to access, use and assess the prototype system, the add-ons in current status should be released along with their installation instructions and shared within the GeoNetwork community.
From a search perspective, users have a very clear goal in mind when they search using a specific word. Therefore, the proposed enrichment approach requires more emphasis on semantics and interpreting a user's interest from a combination of search words used. This will result in facilitating the dataset discovery process, as well as, recording much more relevant search words to enrich metadata content.
The prototype system should support feedback on data quality and data use.
In the current status of prototype system, most frequently used search words are likely to refer to most important and common datasets. Visualising a list of most accessed datasets would be also a valuable addition to the system.
In addition to agreeing and disagreeing with the existing search word, another service can be developed to enable rating the search words. In this case, the users can indicate how much they rate the relevance of search word (e.g. by using the star icon) to a particular dataset.
Personal preferences regarding the tagged search words can be retrieved during the discovery process. This requires the data catalogue to support user authentication system. When the user logs in to the catalogue, the search words recommended by him/her would be retrieved during the search process even if these words have not been approved by the data responsible party.
Implementing the prototype system within GeoNode open source project and sharing the source code with its community could be an improvement for the prototype system.

Having explored the results of prototype systems evaluation, the next section discusses how these results have addressed the fourth and fifth research objectives.

8.4 Chapter Summary and Conclusions

This chapter explored the results of prototype systems evaluation to make sure that the research objectives regarding the design and development of automatic spatial metadata updating and enrichment have been successfully addressed.

The automatic spatial metadata updating (synchronisation) prototype system evaluation survey participated by thirteen organisations internationally confirmed that the GML-based integrated data model, as the core component of this prototype system to support dataset format agnostic metadata automation approach, has been successfully received by the participating organisations. The ability of this data model to address the challenges regarding spatial data interoperability, dataset format dependency, detached metadata and dataset storage, and real-time metadata and dataset updating has been acknowledged by the participating organisations. However, moving toward adoption of the integrated data model for spatial data and metadata storage requires a paradigm shift in the organisations. The metadata records which have been previously stored as flat files (e.g. in XML) need to be stored in the relational database and linked to the spatial dataset stored in the same or a different database. A GML Application Schema also needs to be designed and developed to couple the dataset and metadata. It should be emphasised that no data redundancy would be caused by the integrated data model and the dataset and metadata would only be transferred over the Web using the GML and WFS technologies.

The analysis of the responses to the automatic spatial metadata updating (synchronisation) questionnaire also confirmed that the metadata synchronisation approach designed and developed in this thesis could be worthwhile for improving the organisations' existing metadata management system, as well as, the dataset discovery process through sharing the most up-to-date version of metadata. Nevertheless, the participating organisations raised concern about the technological consistency between the metadata synchronisation functionality and their existing Web-based GIS systems. Because the proposed automatic metadata synchronisation approach has been designed and developed module-based and is founded on the open source technologies (e.g. GeoNetwork, deegree, PostgreSQL/PostGIS, OpenLayers, etc.) as well as, the OGC Standards (GML, WFS, and CSW), it has the potential to be integrated with the Web-based GIS systems. Also, the proposed open source technologies could be replaced with other proprietary or open source technologies that are compliant with the standards used in the prototype system. Toward this integration, it is assumed that the GIS platform supports the OGC compliant standards and Web Services including the GML, WFS-T and CSW.

Lack of support for some spatial data modification tools was also raised as one of the limitations of the automatic spatial metadata updating prototype system. According to the main aim of prototyping technique, discussed in Section 4.4.3 in Chapter 4, which is proving the concept of the proposed automatic spatial metadata updating approach, only some of the dataset edit tools supported by the OpenLayers JavaScript library were used. However, according to the technologies used in the prototype system implementation architecture the system has the potential to support more modification tools.

Some of the participating organisations also mentioned that lack of support for synchronising the metadata at the feature-level could be seen as another limitation of the prototype system. However, following the implementation of the prototype system discussed in Section 7.2 in Chapter 7, whenever a dataset is modified in the prototype system the 'lineage statement' metadata element would be updated to show the history of changes to the features of the dataset. This means that although this research has focused on metadata automation at the dataset-level, the capability to monitor and document changes to features, has also been considered by the prototype system.

The results of the automatic spatial metadata updating prototype system evaluation also showed that metadata elements employed in the prototype system to be updated along with the dataset changes could satisfy the needs of most of the participating organisations. Some of the respondents expected additional metadata elements to be supported by the system such as dataset projection, storage format, type of modifications applied to dataset, and how the modification are applied. Examining the possibility to support the requested elements has been considered as part of the future plan to improve the system.

The analysis of evaluation survey results showed that adoption of automatic spatial metadata updating approach by the participating organisations could result in improving the efficiency of metadata updating not only in terms of time but also in terms of cost.

The automatic spatial metadata enrichment prototype system evaluation survey participated by 26 organisations internationally confirmed that the indirect and direct metadata enrichment models could potentially improve the content of spatial metadata through the users' interaction. The survey responses analysis also showed that the data discovery process could be facilitated through adoption of metadata enrichment models that generate a cloud of most popular search words for finding datasets.

The weighting system suggested for the prototype system implementation, discussed in Sections 6.5.1 and 6.5.2 in Chapter 6, was well received by the participating organisations. However, some suggestions were raised to improve the weighting system, such as increasing the weight of dataset download, and removing the positive weight for metadata exploration.

Although, most of the participating organisations (85%) approved the effective role of prototype system in enriching the content of metadata keyword and showed their willingness to adopt the developed add-ons in their organisations' data discovery systems, the results showed only slightly positively skewness toward agreement, with the higher effectiveness of user-generated search words in data discovery against the original search words embedded in metadata. This was an unexpected finding that might be linked to the fact that a tagging concept is new in the spatial arena and user-generated tags suffer from the ambiguity and heterogeneity issues. The spatial data users need more time to experience the discovery systems equipped with the tag cloud to understand the potential benefit or drawback. As a result, the majority of participating organisations in the automatic spatial metadata enrichment prototype system evaluation survey agreed that user-generated tags need to be verified and approved by the data responsible parties. This is similar to the result of case study investigations, discussed in Section 5.2.1 in Chapter 5, in which most of participating organisations preferred the semi-automatic metadata generation process rather than a fully-automatic process in order to verify and approve the generated metadata.

The findings of the prototype system evaluation also showed that the automatic spatial metadata enrichment approach could be significantly improved through integrating with an ontology-based search engine that interprets a user's interest from a combination of search words used. This could also result in solving existing problems of semantic heterogeneity in keyword-based search engines, facilitating the dataset discovery process and recording more relevant search words to enrich metadata content.

Following the results achieved from the evaluation of automatic spatial metadata updating and enrichment prototype systems, it could be realised that the approaches designed and developed in this thesis to address the fourth and fifth research objectives have been successfully proved. However, comparing the System Usability Score (SUS) achieved from the evaluation of automatic spatial metadata updating and enrichment prototype systems (respectively 67.8 and 68.9 out of 100) with the information provided by rule-of-thumb standard, showed that the prototype systems needs improvement in terms of usability to be an acceptable product. Addressing the usability improvements along with the other enhancements and expectations suggested by the participating organisations, discussed throughout this chapter, will help facilitate developing the prototype systems into fully-functional products.

The next chapter will present the final conclusions of the research by firstly examining the overall achievements in response to the initial research questions and stated objectives. It also discusses possible future research and key recommendations.

CHAPTER 9

CONCLUSIONS AND RECOMMENDATIONS

9 CONCLUSIONS AND RECOMMENDATIONS

9.1 Introduction

Current approaches cannot effectively and efficiently manage metadata creation, updating and improvement for the ever-growing amount of spatial data created and exchanged between people or organisations within the Spatial Data Infrastructures (SDIs) and data sharing platforms. Metadata is commonly collected and created in a separate process from the spatial data lifecycle. This requires the metadata author or responsible party to put in extra effort to gather necessary data for metadata creation. Also, metadata and related spatial data are often stored and maintained separately using a detached data model. This issue results in avoiding the automatic and simultaneous metadata updating when a dataset is modified. In addition, the end users are disconnected from metadata creation and improvement process.

This research investigated the design and development of a framework and associated approaches and tools to facilitate and automate spatial metadata creation, updating (in real time with dataset modification), and enrichment (through the end users' interactions). The framework and technical solutions were developed and tested via prototype systems using GML and Web 2.0 technologies.

This concluding chapter examines the outcomes achieved during this research, reflects on the original research problem and suggests directions for future research efforts.

9.2 Research Aim and Objectives

As detailed in Chapter 1, the central aim of this thesis was to

'design and develop a framework to facilitate and automate the spatial metadata management process, which:

> *a) integrates the metadata creation process with the spatial data lifecycle and reduces the burden of metadata author's intervention while metadata creation*
>
> *b) more effectively and efficiently supports the updating of metadata whenever the datasets are modified*

c) enriches the content of metadata through end users' interaction in data discovery process.'

In order to respond to this aim and address the main challenges regarding the spatial metadata management and automation identified through the case study investigations in this research, a framework was designed in Section 6.2, Chapter 6. The framework consisted of three complementary approaches, including:

- 'lifecycle-centric spatial metadata creation'
- 'automatic spatial metadata updating (synchronisation)'
- 'automatic spatial metadata enrichment'.

The framework also contained a newly integrated data model for storing dataset and metadata.

The lifecycle-centric spatial metadata creation approach was designed to create metadata in conjunction with the spatial data lifecycle steps (planning and policy making, data collection, dataset creation, storage, publication, discovery and access, utilisation, and maintenance). This approach maps the ISO 19115: 2003 metadata elements against the spatial data lifecycle steps, and hence the responsible parties involved in each step could collaborate in creating and maintaining the metadata values associated to that step. Accordingly, the metadata created in this way is more likely to be accurate and up-to-date.

The automatic spatial metadata updating (synchronisation) approach was designed to update metadata concurrent with any modification of the dataset during the spatial data lifecycle. In fact, the metadata synchronisation approach is an automatic process by which properties of a spatial dataset are read from both back end (where dataset is stored) and front end (where the modification environment is up and running) and written into its spatial metadata at the same time as any change to the dataset. As discussed in Section 7.2 in Chapter 7, this approach was successfully implemented and tested via a prototype system using the open source environments (including the GeoNetwork, PostGIS, PostgreSQL, deegree, OpenLayers, and etc.). The findings of the prototype system evaluation, discussed in Section 8.2 in Chapter 8, also confirmed that the automatic spatial metadata updating approach could support the updating of metadata whenever the dataset is modified more effective and efficient than the current approaches.

The automatic spatial metadata enrichment approach including two complementary indirect and direct models was design-rooted in Web 2.0 features (folksonomy and tagging) to improve the content of descriptive keyword metadata element through the end users' interaction. As discussed in Section 7.3 in Chapter 7, this approach was successfully developed and tested via a prototype system within two different environments: GeoNetwork as an open source spatial data catalogue application and MIKE, as an example of a data product – data modelling environment. The results of the prototype system evaluation, discussed in Section 8.3 in Chapter 8, also confirmed that the automatic spatial metadata enrichment approach not only could enrich the content of metadata through the end users' interaction but also could facilitate the data discovery process through monitoring, recording and visualising the most popular search words.

The integrated data model was designed to store spatial dataset and metadata and support the real-time spatial dataset and metadata updating, in particular. As discussed and presented in Section 6.4.2 in Chapter 6, this data model was rooted in the GML technology to accommodate both spatial data and metadata values. The data model including a GML Application Schema was successfully developed and tested with a sample dataset and metadata from the Victorian Department of Sustainability and Environment (DSE), which was reviewed in Section 7.2.1 in Chapter 7. The automatic spatial metadata updating prototype system evaluation also confirmed that the integrated data model was well received by the participating organisations.

In the following sections, the objectives of the research and their related outcomes will be reviewed and discussed.

9.2.1 Objective 1: To Investigate the Underlying Principles and Components of Spatial Metadata Automation

Investigating the spatial metadata underlying principles, as discussed in Chapter 2, and metadata automation related works in two main areas: digital library and information science community and the geospatial community, as discussed in Chapter 3, assisted the research to recognise a number of existing challenges regarding the spatial metadata management and automation. The identification of the issues arisen by using a detached data model for dataset and metadata storage, the challenges linked to the existing manual and semi-automatic metadata creation and updating approaches, the problem of disconnection of end users from the spatial metadata creation and improvement, and the issue of semi-automatic metadata solutions dependency on the dataset format, were some of the findings achieved when addressing the first Objective.

The outcome of Objective 1 also highlighted the metadata standard considered in this research (ISO 19115: 2003), metadata exchange format (XML), a number of key methods in spatial metadata automation (including metadata extraction from dataset, harvesting, inference, and computation), and state-of-the-art technologies (such as Web 2.0 features, open source software (e.g. GeoNetwork, deegree, and PostGIS), and open standards (e.g. GML, WFS, and CSW) which could be utilised for developing the technical solutions in this thesis. This information had a significant influence on designing a framework and associated approaches and tools to address the aim of this research.

9.2.2 Objective 2: To Study the Current Status of Spatial Metadata Management in the Geospatial Community and Identify the Requirements of Metadata Automation and its Related Challenges

The qualitative case study investigations provided a detailed understanding of the current status of spatial metadata management and also the metadata automation requirements in the context of Australia – the selected case study region.

The research identified that there are still multiple parties involved in creating, updating, storing and sharing metadata for the participating organisations and a manual metadata approach was used by the majority of them. It also discovered a number of challenges involved in spatial metadata management including heavily involvement of metadata author in gathering information about metadata; lack of user friendly tools for cataloguing metadata; complication of current metadata standards and data models; ineffectiveness and inefficiency of existing semi-automatic metadata creation approaches in addressing users' needs; and lack of rigorous approaches for updating spatial data and metadata simultaneously.

Moreover, the research recognised the requirements for an automatic spatial metadata creation and updating approach including metadata creation at the time of spatial data collection; dynamic generation of metadata elements from a dataset (e.g. dataset extent); storing and delivering metadata with the dataset; simultaneous updating spatial data and its related metadata; advancing the data discovery process through the automatic improvement of the metadata content; ability to integrate with spatial data tools (e.g. Web-based GIS and data catalogue systems); and compliance with relevant metadata schemas in particular ANZLIC and ISO 19115: 2003. The case study confirmed that the participating organisations preferred the metadata to be generated automatically and then verified and approved by a person.

Along with the case study investigations, examining a number of selected spatial metadata management tools (e.g. GeoNetwork, CatMDEdit, and ANZMet Lite) against the set of criteria including the support for integration of metadata creation with spatial data lifecycle; integrated metadata data model; automatic metadata creation and updating; and interaction with end users to improve the metadata content; highlighted the main challenges relating to the spatial metadata management and automation. These challenges included consideration of metadata management outside the spatial data lifecycle; using detached metadata data model; lack of support for real-time spatial data and metadata updating; dependency of metadata automation methods on dataset format; and limited interaction with end users for metadata creation and improvement.

The outcome of Objectives 1 and 2 in particular assisted with the design and development of a framework to address the identified challenges and fulfilment of Objectives 3, 4, and 5.

9.2.3 Objective 3: To Design an Approach for Integrating Metadata Creation Process with the Spatial Data Lifecycle

Designing a new approach for the integration of metadata creation process with the spatial data lifecycle was considered as part of the framework developed to address the main challenges identified in this research. This approach namely 'lifecycle-centric spatial metadata creation' identified the ISO 19115: 2003 metadata elements that need to be created within any step of the spatial data lifecycle. In this regard, a generic spatial data lifecycle was designed and employed by the new approach informed by the 'Information Lifecycle' recommended by the Australian Government Information Interoperability Framework. The generic spatial data lifecycle consisted of eight steps including planning and policy making, data collection, spatial dataset creation, storage, publication, discovery and access, utilisation, and maintenance.

The research also realised that highest number of ISO 19115: 2003 metadata elements should be created within the spatial dataset creation step. Planning and policy making, dataset maintenance, publication, data collection, dataset storage, utilisation, and discovery are respectively the next steps with the highest number of elements. The research showed that using the lifecycle-centric spatial metadata creation approach the metadata could be completed over time in conjunction with the spatial data lifecycle and therefore, it would be more likely to be accurate and up-to-date. Moreover, this approach could reduce the burden of metadata creation for metadata authors by involving the spatial data responsible parties and interacting with the end users in creating and updating metadata values.

The outcome of this Objective recognised the spatial metadata elements that need to be involved in fulfilment of the Objectives 4 and 5.

9.2.4 Objective 4: To Design Approaches for Automatic Spatial Metadata Updating and Enrichment

Designing a new approach for updating spatial metadata automatically was considered as another part of the framework developed to address the main challenges identified in the research. This approach namely 'automatic spatial metadata updating (synchronisation)' focused on automating the process of updating a subset of ISO 19115: 2003 metadata values whenever the vector dataset is modified, regardless of the format of dataset. The research identified a number of technical requirements to develop such a new approach. The main requirement was an integrated data model built upon GML technology for storing and delivering the spatial metadata and dataset. A mapping software application to generate an integrated data model and support dataset and metadata updating, a user-friendly interface to view dataset and metadata from integrated data model and modify the dataset, and synchronisation scripts to update metadata based on dataset changes were also the other technical requirements considered for implementing this approach.

Similarly, designing a new approach for improving the content of metadata in an automated manner was considered as the last part of the framework developed to address the main challenges identified in the research. This approach namely 'automatic spatial metadata enrichment' concentrated on Web 2.0 features (folksonomy and tagging) to involve end users seeking spatial data to improve the content of 'keyword' metadata element. The research designed two complementary models. The first model namely 'indirect' monitored the end users' behaviour against the retrieved metadata during the data discovery process, recorded the search words that were relevant to datasets (based on a weighting system) and finally assigned the popular search words to metadata 'keyword' element. The second model namely 'direct' allowed the end users to tag a dataset with words they feel best describe what it is about, and agree/disagree with the relevance of their used search words or formerly tagged search words (by previous users) to the retrieved metadata.

The outcome of Objective 4 (including the automatic spatial metadata updating and enrichment conceptual designs together with technical requirements) has also contributed to the fulfilment of Objective 5.

9.2.5 Objective 5: To Implement Prototype Systems for Automatic Spatial Metadata Updating and Enrichment

Implementation of prototype systems for automatic spatial metadata updating and enrichment approaches contributed to prove the new concepts proposed in Objective 4.

The automatic spatial metadata updating (synchronisation) prototype system was implemented using three-layer architecture including: storage, service, and application layers. In the storage layer two databases were built; one for storing the spatial dataset together with metadata and another one for storing the corresponding metadata in the GeoNetwork. In the service layer, the Web servers for supporting the required services (including WFS-Transactional, and CSW) were employed. Also, a new Web service namely 'SYNC' was developed in the service layer to synchronise the metadata catalogues stored in both databases in the storage layer. In the application layer, an interface was developed and integrated with the GeoNetwork interface to display the spatial data and metadata coming in GML format (as the output of integrated data model developed for bundling dataset and metadata). This interface also provided the end users with the facility to modify the vector dataset and see the reflection on a subset of ISO 19115: 2003 metadata elements (e.g. date of revision, lineage statement, and bounding box) automatically and simultaneously. Through the 'SYNC' Web service the corresponding metadata catalogue stored in the GeoNetwork database was also updated with any change to the dataset. In order to implement the system the open source technologies (e.g. GeoNetwork opensource catalogue, OpenLayers, deegree WFS-T and CSW server, PostgreSQL, PostGIS, OpenLayers, and GeoExt) were used.

The automatic spatial metadata enrichment prototype system was implemented within two environments: GeoNetwork as an open source spatial data catalogue application and Model Information Knowledge Environment (MIKE) by DPI-VIC as an example of data product – data modelling environment. For developing the prototype system, three-layer architecture including storage, service and application layers was designed. For developing the system within the GeoNetwork the open source technologies (e.g. PostgreSQL, and ExtJS) were used; however, for development of the system with the MIKE the commercial technologies (e.g. Microsoft SQL Server 2008, and ASP.NET) were utilised. Also, the system within the GeoNetwork was formed as two new add-ons namely 'agree/disagree/tag cloud' and 'suggestion list' and the system with the MIKE was implemented through a new Web page namely 'Model Discovery'. As a result, both indirect and direct metadata enrichment models were developed within the GeoNetwork and MIKE. The most popular search words could be identified automatically by the system and agreed or disagreed by the end users.

Since the GeoNetwork was open source software, the automatic spatial metadata updating and enrichment prototype systems were developed modular to be shared within the GeoNetwork community and used by end users and developers.

The outcome of Objective 5 assisted with the fulfilment of Objective 6.

9.2.6 Objective 6: To Evaluate the Automatic Spatial Metadata Updating and Enrichment Prototype Systems and Identify the Areas which Need Improvement

The evaluation of prototype systems gathered end users' and experts' feedback to recommend improvements to the proposed approaches for automating metadata in Objective 4 and prototype systems designed and implemented to achieve this in Objective 5.

The outcome of Objective 6 showed a successful implementation of the new automatic spatial metadata updating and enrichment approaches. It was acknowledged that the new GML-based integrated data model developed within the automatic spatial metadata updating (synchronisation) system for storing spatial dataset and metadata has the ability to address the challenges regarding spatial data interoperability, dataset format dependency, detached metadata and dataset storage, and real-time dataset and metadata updating. Also, it was realised that the automatic spatial metadata updating approach could be worthwhile for facilitating and improving the organisations' existing metadata updating system, as well as, the dataset discovery process through maintaining and sharing the most up-to-date version of metadata. However, support for more metadata elements to be updated concurrent with dataset modification, implementing more dataset edit tools, and support for updating feature-level metadata in real time with dataset modification were identified as the main areas for improving the automatic spatial metadata updating approach and system.

In addition, it was shown that the indirect and direct metadata enrichment models proposed by the automatic spatial metadata enrichment approach could connect the end users to spatial metadata creation and improvement process, and hence enrich the content of metadata through the end users' interaction. The data discovery process could also be facilitated through the adoption of metadata enrichment models that generate a cloud of the most popular search words for finding datasets. Some areas that needed improvement were also identified including the capabilities of spatial data discovery engine, the current weighting system for recording relative search words, and the quality and validity of user-generated search words. It was also understood that the automatic spatial metadata enrichment approach could be significantly enhanced through integration with an ontology-based data discovery system, which interprets a user's interest from a combination of search words used.

The outcome of Objective 6 together with the outcomes of Objectives 3, 4, and 5 contributed to address the research problem.

9.3 Conclusion on Research Problem

The research problem was that

'Current spatial metadata management approaches are:

> *a) not integrated with the spatial data lifecycle*
>
> *b) not effective and efficient for automatic and continuous updating the metadata content whenever the datasets are modified*
>
> *c) not sufficiently interactive with the end users to involve them in creating and improving the content of metadata.'*

This research problem was addressed through designing, developing, and testing a new framework capable of integrating metadata creation with the spatial data lifecycle via a 'lifecycle-centric spatial metadata creation' approach, updating metadata in real time with dataset modification via an 'automatic spatial metadata updating (synchronisation)' approach, and involving the end users in creating and improving the content of keyword metadata via an 'automatic spatial metadata enrichment' approach.

The new framework took advantages of a new GML-based integrated data model for storing and bundling spatial data and metadata. This data model replaced the detached data model for storing and delivering spatial data and metadata separately and had the potential to address the challenges regarding spatial data interoperability, dataset format dependency, and real-time dataset and metadata updating.

In addition, the new framework was facilitated through Web 2.0 features (folksonomy and tagging) to interact with end users for creating and improving the content of keyword metadata element. Based on this, the end users could share their knowledge about datasets, visit the most popular search words, gather ideas regarding the available data, and access the datasets faster and simpler.

9.4 Recommendations for Future Research

According to the investigations undertaken throughout this research it is recommended that future research efforts could be directed in the following areas:

- Firstly, as was discussed, the spatial metadata creation process needs to be integrated with the steps involved in the spatial data lifecycle through the 'lifecycle-centric spatial metadata creation approach'. The scope of this research did not enable further investigation into designing and developing a system to support the proposed lifecycle-centric spatial metadata creation approach. Therefore, it is suggested that a comprehensive system to create and maintain the ISO 19115: 2003 metadata in parallel with the spatial data lifecycle and rooted in the existing most applicable and agreed metadata automation methods will provide a useful contribution for addressing the problem of missing or incomplete metadata. This happens through recognising the specific step to generate and update metadata within the data lifecycle, and reducing the burden of metadata creation for metadata authors through involving the spatial data responsible parties and interacting with the end users in creating and updating metadata values.
- Secondly, although the evaluation of automatic spatial metadata updating (synchronisation) prototype system indicated the successfulness of relevant approach in

updating vector spatial data and metadata in real time, the number of considered ISO 19115: 2003 metadata elements need to be increased. As a result, an investigation into appropriate automation methods for updating more metadata elements concurrent with dataset modification is suggested to be undertaken.

- Thirdly, as was raised by some of the participating organisations during the evaluation phase, there is a need to maintain metadata values at the feature-level, in addition to the dataset-level, at the same time with any change to the dataset. Although, this research focused on documenting the history of changes of a dataset at the feature-level, a further investigation would be required to define the metadata elements that should be considered at the feature-level and to design the approaches that update the content of those elements concurrent with dataset modification.
- Fourthly, as the research identified, to implement a robust GML-based integrated data model for spatial data and metadata storage there is a need for storing the (XML-based) ISO 19115: 2003 compliant metadata records in a relational database. Accordingly, an investigation into designing and developing a data model that supports the storage of XML-based metadata records in a relational database will considerably facilitate the implementation of the integrated data model proposed in this research for spatial data and metadata storage and real-time maintenance.
- Fifthly, while a weighting system was proposed for the purpose of prototyping the automatic spatial metadata enrichment approach within this research and also agreed by most of the participating organisations in the evaluation process, there is a need for further investigation into designing a comprehensive weighting system that can be applied to any spatial data discovery system for recording most popular search words.
- Sixthly, since metadata improves the discoverability of both spatial data and services and this research has mainly focused on automating the metadata for data products, an investigation into metadata automation for spatial services shared in networked environments is suggested for future research.
- Seventhly, an investigation into improving the quality and efficiency of spatial data and services discovery and retrieval with respect to the volume of metadata sets maintained by different authorities that provide catalogue services, would be suggested as another area for future research.
- Finally, as was highlighted during the evaluation phase, the automatic spatial metadata enrichment approach proposed in this research could be improved by integration with an ontology-based data discovery system. It is suggested that the integration will contribute to address the issues of ambiguity and heterogeneity of user-generated search

words and facilitate the spatial data discovery through recording, assigning and visualising the most relevant search words for describing datasets.

REFERENCES

REFERENCES

Adelman, C., Kemmis, S. and Jenkins, D., 1980. Rethinking case study: notes from the second Cambridge conference. *In:* Simon, H. ed. *Towards a Science of the Singular.* Norwich: Centre for Applied Research in Education, University of East Anglia.

AGIMO, 2006. *Australian Government Information Interoperability Framework.* Australian Government Information Management Office (AGIMO), Department of Finance and Administration, Australian Government.

Ahonen-Rainio, P., 2005. *Visualization of Geospatial Metadata for Selecting Geographic Datasets.* (PhD). Helsinki University of Technology Publications in Cartography and Geoinformatics.

Al-Khalifa, H. S. and Davis, H. C. 2007a. Exploring the Value of Folksonomies for Creating Semantic Metadata. *International Journal on Semantic Web & Information Systems (IJSWIS)*, 3(1), 13-39.

Al-Khalifa, H. S. and Davis, H. C., 2007b. FAsTA: A Folksonomy-Based Automatic Metadata Generator. *2nd European Conference on Technology Enhanced Learning.* Crete, Greece.

Albertoni, R., Bertone, A. and Martino, M. D., 2004. Visual Analysis of Geographic Metadata in a Spatial Data Infrastructure. *15th International Workshop on Database and Expert Systems Applications (DEXA'04).* IEEE Computer Society, 861–865.

Alexander, B. 2006. Web 2.0: A New Wave of Innovation for Teaching and Learning? *Educause Review*, 41(2).

Anderson, G. and Moreno-Sanchez, R. 2003. Building web-based spatial information solutions around open specifications and open source software. *Transaction in GIS*, 7(4), 447-466.

Ankolekar, A., Krotzsch, M., Tran, D. T. and Vrandecic, D. 2008. The two cultures: Mashing up Web 2.0 and the Semantic Web. *Web Semantics*, 6(1).

ANZ-MEST, 2012. *TRAC site for the 'ANZ-MEST software' (previously called 'BlueNetMEST software'), a branch of GeoNetwork* [online]. Available from: http://anzlicmet.bluenet.utas.edu.au/ [Accessed 5 July 2012].

ANZLIC, 1996. *ANZLIC Guidelines: Core Metadata Elements Version 1.*

ANZLIC, 2009. *Spatial Metadata* [online]. Australia and New Zealand Land Information Council (ANZLIC). Available from: http://www.anzlic.org.au/infrastructure_metadata.html [Accessed 3 July 2009].

Apache, 2011. *Apache Tomcat Homepage* [online]. Available from: http://tomcat.apache.org/ [Accessed 12 July 2011].

ASDD, 2010. *Australian Spatial Data Directory (ASDD)* [online]. Available from: http://asdd.ga.gov.au/asdd/ [Accessed 14 March 2010].

AURIN, 2012. *AURIN Portal* [online]. Available from: http://aurin.org.au/ [Accessed 1 June 2012].

Baird, K. and Jorum-Team, 2006. Automated Metadata. [online]. Available from: www.jorum.ac.uk/docs/pdf/automated_metadata_report.pdf [Accessed 14/07/09].

Balfanz, D., 2002. Automated geodata analysis and metadata generation. *In:* Erbacher, R. F., *et al.* eds. *Visualization and Data Analysis.* San Jose, CA, USA, 285-295.

Ballari, D., Maganto, A., Nogueras, J., Pascual, A. and Bernabé, M., 2006. Experiences in the Use of an ISO19115 Profile within the Framework of the Spanish SDI. *GSDI-9 World Conference.* Santiago, Chile.

Bangor, A., Kortum, P. T. and Miller, J. T. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, 24(6), 574-594.

Batcheller, J. K. 2008. Automating Geospatial Metadata Generation - An Integrated Data Management and Documentation Approach. *Computers and Geosciences*, 34, 387-398.

Batcheller, J. K., Dunfey, R. I., Reitsma, F. and Gittings, B. M., 2007. Automating Geospatial Metadata Using ESRI's ArcGIS and Microsoft's .NET. *10th AGILE International Conference on Geographic Information Science.* Aalborg University, Denmark.

Batcheller, J. K., Gittings, B. M. and Dunfey, R. I. 2009. A Method for Automating Geospatial Dataset Metadata. *Future Internet*, 1, 28-46.

Bateman, S., Gutwin, C. and Nacenta, M., 2008. Seeing things in the clouds: the effect of visual features on tag cloud selections. *The nineteenth ACM conference on Hypertext and hypermedia.* New York, USA, 193-202.

Beard, K., 1996. A Structure for Organizing Metadata Collection. *Third International Conference/Workshop on Integrating GIS and Environmental Modelling.*

Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The Semantic Web. *Scientific American*.

Bernstein, A. and Klein, M., 2002. Towards high-precision service retrieval. *In:* Horrocks, I. and Hendler, J. eds. *The Semantic Web – First International Semantic Web Conference (ISWC 2002).* Sardinia, Italy, 84–101.

Biernacki, P. and Waldorf, D. 1981. Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods and Research*, 10(2), 141-163.

BlueNet, 2008. *Instructions on the use of the AODN's MEST (Metadata Entry and Search Tool)* [online]. Available from: http://www.bluenet.org.au/MEST_help.html [Accessed 25 August 2011].

BlueNet, 2010. *The BlueNet Metadata Entry and Search Tool (MEST)* [online]. Available from: http://www.bluenet.org.au/MEST_intro.html [Accessed 20 March 2010].

Bonaccorsi, A. and Rossi, C. 2003. Why Open Source software can succeed. *Research Policy*, 32(7), 1243-1258.

Borjas, B., Florczyk, A. J., Lopez-Pellicer, F. J., Nogueras-Iso, J. and Zarazaga-Soria, F. J., 2011. Automatic Metadata Generation for the Web Geo-resources. *INSPIRE Conference 2011: INSPIREd by 2020 - Contributing to smart, sustainable and inclusive growth.* Edinburgh, Scotland.

Boulanger, A. 2005. Open-source versus proprietary software: is one more reliable and secure than the other? *IBM Systems Journal*, 44(2), 239-248.

Bradford, D. L., 2007. *A Geospatial Data Catalog and Metadata Management Tools for the U.S. Environmental Protection Agency's Western Ecology Division.* (Master of Science Geography Program).

Bray, A. and Ramage, S., 2012. *OGC Market Report Open Standards and INSPIRE.* OGC.

Brooke, J., 1996. SUS: a "quick and dirty" usability scale. *In:* Jordan, P. W., *et al.* eds. *Usability Evaluation in Industry.* London: Taylor & Francis, 189-194.

Cardinaels, K., Meire, M. and Duval, E., 2005. Automating Metadata Generation: The Simple Indexing Interface. *14th International World Wide Web Conference.* Chiba, Japan, 1140-1141.

Carey, J. W. 1993. Linking qualitative and quantitative methods: Integrating cultural factors into public health. *Qualitative Health Research*, 3, 298–318.

Carney, D., Smith, J. and Place, P., 2005. *Topics in Interoperability: Infrastructure Replacement in a System of Systems.* Software Engineering Institute, Carnegie Mellon University.

CatMDEdit, 2011a. *About CatMDEdit* [online]. Available from: http://catmdedit.sourceforge.net/ [Accessed 25 August 2011].

CatMDEdit, 2011b. *User's Manual- Management of Repositories* [online]. Available from: http://iaaa.cps.unizar.es/software/index.php/Management_of_repositories#Extracting_automatic_metadata_from_data_files [Accessed 27 August 2011].

Chi, Y.-L. 2009. A consumer-centric design approach to develop comprehensive knowledge-based systems for keyword discovery. *Expert Systems with Applications*, 36(2), 2481–2493.

Coetzee, S. 2012. Reference model for a data grid approach to address data in a dynamic SDI. *GEOINFORMATICA*, 16(1), 111-129.

Coleman, D. and Nebert, D. 1998. Building a North American Spatial Data Infrastructure. *Cartography and Geographic Information Systems*, 25(3), 151-160.

Cooper, A. K., Coetzee, S., Kaczmarek, I., Kourie, D., Iwaniak, A. and Kubik, T., 2011a. Challenges for quality in volunteered geographic information. *AfricaGEO 2011*. Cape Town, South Africa.

Cooper, A. K., Rapant, P., Hjelmager, J., Laurent, D., Iwaniak, A., Coetzee, S. and Moellering, H., 2011b. Extending the formal model of a spatial data infrastructure to include volunteered geographical information. *25th International Cartographic Conference (ICC).* Paris, France.

Craglia, M. and Annoni, A., 2007. INSPIRE: An Innovative Approach to the Development of Spatial Data Infrastructures in Europe. *In:* Onsrud, H. ed. *Research and Theory in Advancing Spatial Data Infrastructure Concepts.* ESRI Press, 93-105.

Creswell, J. W., Clark, V. L. P., Gutmann, M. L. and Hanson, W. E., 2003. Advanced mixed methods research designs. *In:* Tashakkori, A. and Teddlie, C. eds. *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, California: Sage Publications, 209-240.

Crompvoets, J., Rajabifard, A., Bregt, A. and Williamson, I. 2004. Assessing the worldwide development of national spatial data clearinghouses. *International Journal of Geographic Information Science (IJGIS),* 18(7), 1-25.

Daratech, 2011. *GIS/Geospatial sales up 10.3% to US$4.4 billion Growth forecast to top 8.3% in 2011.* Cambridge, MA, USA: Daratech, Inc.

deegree, 2010a. *deegree Web Catalogue Service v.2.5.* Bonn: Dept. of Geography, Bonn University.

deegree, 2010b. *deegree Web Feature Service v2.5* [online]. Available from: http://download.deegree.org/deegree2.5/docs/wfs/html/deegree_wfs_documentation_en.html [Accessed 10 June 2011].

Dempsey, L. and Heery, R., 1997. *Specification for resource description methods Part 1: A review of metadata: a survey of current resource description formats.*

Denning, P. J., Comer, D., Gries, D., Mulder, M., Tucker, A., Turner, A. J. and Young, P. 1989. Computing as a discipline. *Communications of the ACM,* 32(1), 9-23.

Devillers, R., Gervais, M., Bédard, Y. and Jeansoulin, R., 2002. Spatial Data Quality: From Metadata to Quality Indicators and Contextual End-User Manual. *OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management.* Istanbul.

DublinCore, 2010. *Dublin Core Metadata Element Set, Version 1.1* [online]. Available from: http://dublincore.org/documents/dces/ [Accessed 12 February 2012].

Duval, E., Hodgins, W., Sutton, S. and Weibel, S. L., 2002. Metadata Principles and Practicalities. *D-Lib Magazine.*

Eagleson, S., 2002. *GIS applied to administrative boundary design.* (Doctor of Philosophy). The University of Melbourne.

Egenhofer, M., 2002. Toward the Semantic Geospatial Web. *In:* Voisard, A. and Chen, S.-C. eds. *10th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS).* McLean, VI, USA, 1-4.

ESRI, 2002a. Creating a Custom Metadata Synchronizer. *An ESRI ® White Paper* [online]. Available from: http://www.esri.com/library/whitepapers/pdfs/metadata.pdf [Accessed 10/09/09].

ESRI, 2002b. Metadata and GIS. *An ESRI ® White Paper* [online]. Available from: http://www.esri.com/library/whitepapers/pdfs/metadata-and-gis.pdf [Accessed 11/07/09].

ESRI, 2009. *ArcGIS Server Geoportal Extension 9.3.1* [online]. Available from: http://webhelp.esri.com/geoportal_extension/9.3.1/ [Accessed 27 October 2011].

ESRI, 2010. *Metadata Synchronization* [online]. Available from: http://webhelp.esri.com/arcgiSDEsktop/9.3/index.cfm?TopicName=Metadata_synchronization [Accessed 27 October 2011].

ESRI, 2012. *The ArcGIS metadata format* [online]. Available from: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//003t0000002w000000 [Accessed 2 July 2012].

Ezigbalike, C. and Rajabifard, A., 2009. Indicators for Assessing Spatially Enabled Government Services. *GSDI 11 World Conference.* Rotterdam, The Netherlands.

Fegeas, R. G., Cascio, J. L. and Lazar, R. A. 1992. An Overview of FIPS 173, The Spatial Data Transfer Standard. *Cartography and Geographic Information Systems,* 19(5), 278-229.

FGDC, 2000. *Content Standard for Digital Geospatial Metadata Workbook*. Federal Geographic Data Committee (USA).

FGDC, 2009. *North American Profile of International Standard on Geographic Information Metadata* [online]. Available from: http://www.fgdc.gov/nap/metadata [Accessed 10 August 2009].

Florczyk, A. J., Lopez-Pellicer, F. J., Nogueras-Iso, J. and Zarazaga-Soria, F. J. 2012. Automatic Generation of Geospatial Metadata for Web Resources. *International Journal of Spatial Data Infrastructures Research*, 7, 151-172.

Friesen, N., Roberts, A. and Fisher, S. 2002. CanCore: Learning Object Metadata. *Canadian Journal of Learning and Technology*, 28(3), 43-53.

Fuggetta, A. 2003. Opensourcesoftware- an evaluation. *Journal of Systems and Software*, 66(1), 77-90.

GeoNetwork, 2011a. *CSW service* [online]. Available from: http://geonetwork-opensource.org/stable/developers/xml_services/csw_services.html [Accessed 08 September 2011].

GeoNetwork, 2011b. *GeoNetwork User Manual v2.6.4* [online]. Available from: http://geonetwork-opensource.org/manuals/2.6.4/users/ [Accessed 26 October 2011].

GeoNetwork, 2011c. *Homepage* [online]. Available from: http://geonetwork-opensource.org/ [Accessed 14 October 2011].

GeoNetwork, 2012a. *Metadata Exchange Format* [online]. Available from: http://geonetwork-opensource.org/manuals/trunk/eng/developer/mef/index.html [Accessed 14 February 2012].

GeoNetwork, 2012b. *What is GeoNetwork opensource?* [online]. Available from: http://geonetwork-opensource.org/manuals/2.6.4/users/quickstartguide/geographic_management/index.html [Accessed 29 March 2012].

GeoNode, 2012a. *GeoNode's Architecture Overview* [online]. Available from: http://docs.geonode.org/en/latest/developers/architecture.html [Accessed 2 February 2012].

GeoNode, 2012b. *GNIP 21 – Create new vector layers and add/edit/delete layer features* [online]. Available from: https://github.com/GeoNode/geonode/wiki/GNIP-21-%E2%80%93-Create-New-Vector-Layers-and-Add-Edit-Delete-Layer-Features [Accessed 3 July 2012].

GeoServer, 2011a. *Application Schema Support* [online]. Available from: http://docs.geoserver.org/stable/en/user/data/app-schema/index.html [Accessed 5 March 2011].

GeoServer, 2011b. *Complex Features* [online]. Available from: http://docs.geoserver.org/stable/en/user/data/app-schema/complex-features.html [Accessed 15 February 2011].

Giger, C. and Najar, C., 2003. Ontology-Based Integration of data and Metadata. *6th AGILE Conference on Geographic Information Science*. Lyon, France.

Gill, T., 2000. Metadata and the World Wide Web. *Introduction to Metadata, Pathways to digital Information, Online Edition, Version 2.1*.

Gillham, B., 2000. *Case study research methods*. London: Continuum.

Göbel, S. and Lutze, K., 1998. Development of Meta Databases for Geospatial Data In The WWW. *6th ACM International Symposium on Advances in Geographic Information Systems 1998*. Washington DC, USA, 94-99.

Gold, E. B., Sternfeld, B., Kelsey, J. L., Brown, C., Mouton, C., Reame, N., Salamone, L. and Stellato, R. 2000. Relation of demographic and lifestyle factors to symptoms in a multi-racial/ethnic population of women 40-55 years of age. *American Journal of Epidemiology, Oxford University Press*, 152, 463-473.

Goodchild, M. F. 2007. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, 2, 24-32.

Google, 2011. *Google books- Ngram Viewer* [online]. Available from: http://books.google.com/ngrams/graph?content=xsl&year_start=1995&year_end=2011&corpus=0&smoothing=3 [Accessed 18 October 2011].

Grasso, M. and Craglia, M., 2010. *D 2.2.3 European Open Source Metadata Editor (revised 2010-12-20).*

Green, D. and Bossomaier, T., 2002. *Online GIS and Spatial Metadata.* London: Taylor & Francis.

Greenberg, J., 2003a. Metadata and the World Wide Web. *Encyclopaedia of Library and Information Science.* 2 ed. New York: Marcel Dekker, 1876-1888.

Greenberg, J., 2003b. Metadata Generation: Processes, People and Tools. *Bulletin of the American Society for Information Science and Technology.* 18-21.

Greenberg, J. 2004. Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications. *Internet Cataloging- Taylor & Francis,* 6(4), 59-82.

Greenberg, J., Pattuelli, M. C., Parsia, B. and Robertsin, W. D. 2001. Author-generated Dublin Core metadata for web resources: a baseline study in an organization. *Journal of Digital Information,* 2(2).

Greenberg, J., Spurgin, K. and Crystal, A., 2005. *Final Report for AMeGA (Automatic Metadata Generation Applications) Project.*

Greenberg, J., Spurgin, K. and Crystal, A. 2006. Functionalities for Automatic Metadata Generation Applications: A Survey of Metadata Experts' Opinions. *International Journal of Metadata Semantics and Ontology,* 1(1), 3-20.

Greenfield, T. K., Midanik, L. T. and Rogers, J. D. 2000. Effects of Telephone Versus Face-to-Face Interview Modes on Reports of Alcohol Consumption. *Addiction 95,* 277-284.

Guptill, S. C., 1999. Metadata and Data Catalogues In Geographical Information Systems. *In:* Longley, P. A., *et al.* eds. *Geographical Information Systems: Principles, Techniques, Management and Applications.* second ed.: John Wiley & Sons, Inc, 677–692.

Guy, M., Powell, A. and Day, M., 2004. *Improving the Quality of Metadata in Eprint Archives* [online]. Available from: http://www.ariadne.ac.uk/issue38/guy/ [Accessed 9 February 2012].

Hall, G. B. and Leahy, M. G., 2008. *Open Source Approaches in Spatial Data Handling.* Berlin: Springer.

Han, H., Giles, C. L., Manavoglu, E., Zha, H., Zhang, Z. and Fox, E. A., 2003. Automatic Document Metadata Extraction using Support Vector Machines. *3rd ACM/IEEE-CS Joint Conference on Digital Libraries.* 37-48.

Hartley, J., 2004. Case Study Research. *In:* Cassell, C. and Symon, G. eds. *Essential Guide To Qualitative Methods In Organizational Research.* London: SAGE Publications Ltd.

Hatala, M. and Forth, S., 2003. A Comprehensive System for Computer-Aided Metadata Generation. *12th International Conference of The World Wide Web Consortium (WWW2003).* Budapest.

Hill, L., Janee, G., Dolin, R., Frew, J. and Larsgaard, M. 1999. Collection metadata solutions for digital library applications. *Journal of the American Society for Information Science,* 50(13), 1169 -1181.

Howe, D., 1996. *Free on-line Dictionary of Computing (FOLDOC)* [online]. Available from: http://wombat.doc.ic.ac.uk/ [Accessed 6 February 2012].

INSPIRE, 2009. *INSPIRE Metadata Implementing Rules: Technical Guidelines Based on EN ISO 19115 and EN ISO 19119.* INSPIRE.

INSPIRE, 2012. *INSPIRE Geo Portal* [online]. Available from: http://inspire-geoportal.ec.europa.eu/ [Accessed 20 June 2012].

ISO9241-11, 1998. *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability.* Geneva, Switzerland.

ISO19115, 2003. *Geographic Information - Metadata.* Geneva, Switzerland: International Standardization Organization (ISO).

Iso, J. N., Barrera, J. u., Crespo, F. G., Laiglesia, S. and Medrano, P. R. M., 2008. Integrating catalog and GIS tools: access to resources from CatMDEdit thanks to gvSIG. *4as Jornadas Internacionales gvSIG.* Valencia.

ISO/TC211/WG1, 1998. *Geographic Information – Part 1: Reference Model.* Geneva, Switzerland.

ISO/TS19139, 2007. *Geographic information — Metadata — XML schema implementation.* Geneva, Switzerland: International Standardization Organization (ISO).

Iwaniak, A., Kaczmarek, I., Kubik, T., Lukowicz, J., Paluszynski, W., Kourie, D. G., Cooper, A. K. and Coetzee, S., 2011. AN INTELLIGENT GEOPORTAL FOR SPATIAL PLANNING. *25th International Cartographic Conference (ICC).* Paris, France.

Jaccard, J. and Wan, C. K., 1996. *LISREL approaches to interaction effects in multiple regression.* Thousand Oaks, CA: Sage Publications.

Jokela, S., 2001. *Metadata enhanced content management in media companies.* (Doctor of Technology). Helsinki University of Technology.

Jokela, T., Koivumaa, J., Pirkola, J., Salminen, P. and Kantola, N. 2006. Methods for quantitative usability requirements: a case study on the development of the user interface of a mobile phone. *Personal and Ubiquitous Computing*, 10(6), 345-355.

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, 119-129.

Kalantari, M., Rajabifard, A. and Olfat, H., 2009. Spatial Metadata Automation: A New Approach. *In:* Ostendorf, B., *et al.* eds. *Spatial Science Conference 2009 (SSC2009).* Adelaide, Australia, 629-635.

Kaplan, B. and Maxwell, J. A., 1994. Qualitative Research Methods for Evaluating Computer Information Systems. *In:* Anderson, J. G., Aydin, C. E. and Jay, S. J. eds. *Evaluating Health Care Information Systems: Methods and Applications.* Sage, Thousand Oaks, CA, 45-68.

Klien, E., Einspanier, U., Lutz, M. and Hübner, S., 2004. An Architecture for Ontology-Based Discovery and Retrieval of Geographic Information. *7th Conference on Geographic Information Science (AGILE 2004).* .

Krechmer, K. 2006. Open standards requirements. *International Journal of IT Standards and Standardization Research*, 4(1), 43-61.

Kresse, W. and Fadaie, K., 2004. *ISO Standards for Geographic Information.* Springer.

Kubik, T. and Iwaniak, A., 2010. Building and maintaining metadata repositories with the aid of ontology tools and technologies. *GSDI-12 World Conference.* Garden City, Singapore.

Lacasta, J., Nogueras-Iso, J., Béjar, R., Muro-Medrano, P. R. and Zarazaga-Soria, F. J. 2007. A web ontology service to facilitate interoperability within a spatial data infrastructure: applicability to discovery. *Data and Knowledge Engineering*, 63 (3), 947–971.

Land Information Council of Jamaica, M. o. A., 2008. WHAT IS METADATA. [online]. [Accessed 20/08/09].

Landgate, 2012. *About SLIP Enabler* [online]. Available from: https://www2.landgate.wa.gov.au/web/guest/about-slip [Accessed 15 June 2012].

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

Linkedin, 2012. *Linkedin Learning Centre* [online]. Available from: http://learn.linkedin.com/what-is-linkedin/ [Accessed 9 August 2012].

Litwin, L. and Rossa, M., 2011. *Geoinformation Metadata in INSPIRE and SDI: Understanding. Editing. Publishing.* Springer.

Lutz, M., 2005. *Ontology-based Discovery and Composition of Geographic Information Services.* (PhD). Munster University.

Mack, N., Woodsong, C., MacQueen, K. M., Guest, G. and Namey, E., 2005. *Qualitative Research Methods: A Data Collector's Field Guide.* North Carolina: Family Health International.

Maedche, A. and Staab, S. 2001. Ontology Learning for the Semantic Web. *IEEE Intelligent Systems*, 16(2), 72-79.

Manso-Callejo, M., Wachowicz, M. and Bernabé-Poveda, M., 2009. Automatic Metadata Creation for Supporting Interoperability Levels of Spatial Data Infrastructures. *GSDI - 11 Conference.* Rotterdom, The Netherlands.

Manso-Callejo, M., Wachowicz, M. and Bernabé-Poveda, M., 2010. The design of an automated workflow for metadata generation. *In:* Sánchez-Alonso, S. and Athanasiadis, I. N. eds. *COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE.* Spain: Springer, 275 - 287.

Manso, M. A., Nogueras-Iso, J., Bernabé, M. A. and Zarazaga-Soria, F. J., 2004. Automatic Metadata Extraction from Geographic Information. *7th Conference on Geographic Information Science (AGILE 2004).* Heraklion, Greece.

Mathes, A., 2004. *Folksonomies - Cooperative Classification and Communication Through Shared Metadata* [online]. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.135.1000&rep=rep1&type=pdf.

Mathys, T., 2004. The Go-Geo! Portal metadata initiatives. *Geographical Information Science Research UK 12th Annual Conference.* University of East Anglia, Norwich, UK, 148-154.

Maxwell, J. A., 1996. *Qualitative Research Design: An Interactive Approach.* Thousand Oaks, CA: Sage Publications.

McCarty, C., House, M., Harman, J. and Richards, S. 2006. Effort in phone survey response rates: The effects of vendor and client-controlled factors. *Field Methods,* 18, 172-188.

McClelland, M. 2003. Metadata Standards for Educational Resources. *IEEE Computer,* 36(11), 107-109.

Mckoi, 2011. *Homepage* [online]. Available from: http://www.mckoi.com/Mckoi%20SQL%20Database.html [Accessed 12 June 2011].

Moellering, H., J.G.L.Aalders, H. and Crane, A., 2005. *World Spatial Metadata Standards.* Elsevier on behalf of The International Cartographic Association.

Morris, S., Nagy, Z. and Tuttle, J., 2007. *North Carolina Geospatial Data Archiving Project.* NCSU Libraries and North Carolina Centre for Geographic Information & Analysis.

Murugesan, S. 2007. Understanding web 2.0. *IT Professional,* 9(4), 34–41.

Myers, M. D. 1997. Qualitative Research in Information Systems. *MIS Quarterly,* 21(2), 241-242.

Najar, C., 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.

Najar, C., Rajabifard, A., Williamson, I. and Giger, C., 2007. A Framework for Comparing Spatial Data Infrastructures: An Australian-Swiss Case Study. *In:* Onsrud, H. J. ed. *Research and Theory in Advancing Spatial Data Infrastructure Concepts.* Redlands, California, USA: ESRI Press, 201-213.

Nebert, D. D., 2000. *Developing Spatial Data Infrastructures: The SDI Cookbook.* Version 1.0 ed.: GSDI-Technical Working Group.

Nebert, D. D., 2004. *Developing Spatial Data Infrastructures: The SDI Cookbook ver 2.0.* Technical Working Group Chair, GSDI.

Neteler, M. and Mitasova, H., 2002. *Open Source Gis: A Grass Gis Approach.* Boston: Kluwer Academic Publishers.

Newman, W. L., 1994. *Social Research Methods.* Boston: Allyn and Bacon.

Nichol, G. E., *et al.*, 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.

Niemann, B., 2011. Data.gov Geospatial Platform Needs 'Data Life Cycle' Management.

NISO, 2004. Understanding Metadata. *NISO Press Booklets* [online]. Available from: http://www.niso.org/publications/press/UnderstandingMetadata.pdf.

Nogueras-Iso, J., Zarazaga-Soria, F. J. and Muro-Medrano, P. R., 2005. *Geographic Information Metadata for Spatial Data Infrastructures - Resources, Interoperability and Information Retrieval.* Springer.

Nunamaker, J. F., Chen, M. and Purdin, T. D. M. 1990-91. Systems development in information systems research. *Journal of Management Information Systems,* 7(3), 89-106.

O'Leary, D., 1988. Expert System Prototyping as a Research Tool. *In:* Turban, E. and Watkins, P. eds. *Applied Expert Systems.* North Holland, 17-32.

OBM, 2010. *A-16 Supplemental Guidance - Geospatial Data Lifecycle* [online]. Available from: http://www.fgdc.gov/policyandplanning/A16Draft/A16_SG_GeospatialDataLifecycle [Accessed 31 July 2011].

OGC, 2003. *OGC Reference Model.* OGC 03-040.

OGC, 2007. *OpenGIS® Geography Markup Language (GML) Encoding Standard* [online]. Open Geospatial Consortium. Available from: http://portal.opengeospatial.org/files/?artifact_id=20509 [Accessed 6 August 2009].

OGC, 2009. *OpenGIS® Standards and Related OGC documents* [online]. Available from: http://www.opengeospatial.org/standards [Accessed 12 July 2009].

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H. and Williamson, I. 2012a. A GML-based Approach to Automate Spatial Metadata Updating. *International Journal of Geographical Information Science,* DOI:10.1080/13658816.2012.678853.

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H. and Williamson, I. P. 2012b. Spatial Metadata Automation: A Key to Spatially Enabling Platform. *International Journal of Spatial Data Infrastructures Research,* 7, 173-195.

Olfat, H., Kalantari, M., Rajabifard, A. and Williamson, I. 2012c. Towards a Foundation for Spatial Metadata Automation. *Journal of Spatial Science,* 57(1), 65-81.

Olfat, H., Kalantari, M., Rajabifard, A., Williamson, I. P., Pettit, C. and Williams, S., 2010a. Exploring the Key Areas of Spatial Metadata Automation Research in Australia. *GSDI-12 World Conference.* Garden city, Singapore.

Olfat, H., Rajabifard, A. and Kalantari, M., 2010b. Automatic Spatial Metadata Update: a New Approach. *XXIV FIG International Congress 2010.* Sydney, Australia.

OpenGeo, 2011. *Homepage* [online]. Available from: http://opengeo.org/ [Accessed 20 February 2011].

OSDM, 2009a. *ANZLIC Metadata Profile Short User Guide for the ANZMet Lite* [online]. Available from: http://www.osdm.gov.au/Metadata/08_ANZLIC_MetaProfile_Short_%20ANZMetLite.pdf/?id=1002 [Accessed 24 August 2011].

OSDM, 2009b. *ANZMet Lite Metadata Entry Tool User Guide* [online]. Available from: http://spatial.gov.au/sites/default/files/legacy/osdm.gov.au/Metadata/ANZLIC%2Bmetadata%2Bresources/ANZMet%2BToolkit%2B(final%2Bdraft%2B-%2B07.2009)/10_ANZMet_Lite_UserGuide.pdf/index.pdf [Accessed 20 September 2011].

OSDM, 2009c. *ANZMet Toolkit (final draft - 07.2009)* [online]. Available from: http://www.osdm.gov.au/Metadata/ANZLIC+metadata+resources/ANZMet+Toolkit+(final+draft+-+07.2009)/default.aspx [Accessed 20 October 2011].

Pasca, M., Petriglia, L., Mattioni, F., Torchio, M. and Mariotti, C., 2009. Experiences in the Creation and Updating of INSPIRE Compliant Metadata Catalogue. *GSDI - 11 Conference.* Rotterdom, The Netherlands.

Passant, A. and Laublet, P., 2008. Meaning of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data. *Workshop on Linked Data on the Web (LDOW), International Word Wide Web Conference.* Beijing, China.

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.

Peng, Z.-R. 2005. A proposed framework for feature-level geospatial data sharing: a case study for transportation network data. *Geographic Information Science,* 19, 459-481.

Peng, Z.-R. and Zhang, C. 2004. The Roles of Geography Markup Language (GML), Scalable Vector Graphics (SVG), and Web Feature Service (WFS) Specifications in The Development of Internet Geographic Information Systems (GIS) *Geographical Systems,* 6(2), 95-116.

Phillips, A., Williamson, I. and Ezigbalike, C., 1998. The Importance of Metadata Engines in Spatial Data Infrastructures *26th Annual Conference of AURISA* Perth, Western Australia.

PHP, 2011. *Homepage* [online]. Available from: http://www.php.net/ [Accessed 19 September 2011].

PostGIS, 2011. *Homepage* [online]. Available from: http://postgis.refractions.net/ [Accessed 18 August 2011].

PostgreSQL, 2011. *Homepage* [online]. Available from: http://www.postgresql.org/about/ [Accessed 07 September 2011].

Pountain, D., 2003. The Penguin Dictionary of Computing. New York: Penguin Putnam.

PSMA, 2012. *Enhancing PSMA Australia's Datasets Fact Sheet* [online]. Available from: http://www.psma.com.au/aboutpsma/documents/PSMA_Data_Enhancement_Fact_Sheet.pdf [Accessed 21 June 2012].

Rajabifard, A., ed., 2007. *Towards a Spatially Enabled Society.* Melbourne: The University of Melbourne.

Rajabifard, A., Feeney, M.-E. F. and Williamson, I. P. 2002. Directions for the Future of SDI Development. *International Journal of Applied Earth Observation and Geoinformation,* 4(1), 11-22.

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, 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.

Rautenbach, V., Coetzee, S. and Iwaniak, A. 2012. Orchestrating OGC web services to produce thematic maps in a spatial information infrastructure. *Computers, Environment and Urban Systems*.

Richardson, R. and Smeaton, A. F., 1995. *Using WordNet in a knowledge-based approach to information retrieval.* Dublin City University.

Rivadeneira, A. W., Gruen, D. M., Muller, M. J. and Millen, D. R., 2007. Getting Our Head in the Clouds : Toward Evaluation Studies of Tagclouds. *CHI 2007.*

Saaty, T. L. 2004. Decision making - the Analytic Hierarchy and Network Processes (AHP/ANP). *JOURNAL OF SYSTEMS SCIENCE AND SYSTEMS ENGINEERING,* 13(1), 1-35.

Sadeghi-Niaraki, A., Rajabifard, A., Kim, K. and Seo, J., 2010. Ontology Based SDI to Facilitate Spatially Enabled Society. *GSDI 12 World Conference.* Singapore.

Saraf, G., 2008. Semantic Web Unfolding the Underlying Technology. [online]. Available from: http://www.scribd.com/doc/4300645/Semantic-Web [Accessed 01/05/09].

Schindler, U. and Diepenbroek, M. 2008. Generic XML-based framework for metadata portals. *Computers & Geosciences,* 34(12), 1947–1955.

Shanzhen, Y., Qi, L. and Jicheng, C., 1999. An interoperability GIS model based on the spatial information infrastructure. *Geoinformatics'99 Conference.* Ann Arbor, 1-5.

Shirky, C., 2005. *Ontology is overrated: categories, links, and tags* [online]. Available from: www.shirky.com/writings/ontology_overrated.html [Accessed 24 February 2010].

Sinclair, J. and Cardew-Hall, M. 2008. The folksonomy tag cloud: when is it useful? *Journal of Information Science,* 6(1), 15-23.

Steinacker, A., Ghavam, A. and Steinmetz, R. 2001. Metadata Standards for Web-Based Resources. *IEEE Multimedia Journal,* 8(1).

Tashakkori, A. and Teddlie, C., 2003. *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, California: Sage Publications.

Taussi, M., 2007. *Automatic Production of Metadata Out of Geographic Datasets.* (MSc). Helsinki University of Technology.

Taylor, C., 2003. An introduction to Metadata. [online]. Available from: http://www.library.uq.edu.au/iad/ctmeta4.html [Accessed 10/05/09].

Teddlie, C. and Tashakkori, A., 2003. Major issues and controversies in the use of mixed methods in the social and behavioral sciences. *In:* Tashakkori, A. and Teddlie, C. eds.

Handbook of mixed methods in social and behavioral research. Thousand Oaks, California: Sage Publications, 3-50.

Thomas, M., Caudle, D. M. and Schmitz, C. M. 2009. To tag or not to tag? *Library Hi Tech,* 27(3), 411-434.

Thomas, R. M., 2003. *Blending qualitative and quantitative research methods in theses and dissertations.* Thousand Oaks, California: Corwin Press.

Timpf, S., Raubal, M. and Kuhn, W., 1996. Experiences with Metadata. *7th International Symposium on Spatial Data Handling, SDH'96.* Delft, The Netherlands, 12B.31-12B.43.

Tsou, M. H., 2002. An Operational Metadata Framework for Searching, Indexing, and Retrieving Distributed Geographic Information Services on The Internet. *Lecture Notes in Computer Science.* Springer, Berlin, 313-332.

Turnitsa, C. and Tolk, A., 2006. Battle Management Language: A Triangle with Five Sides. *Simulation Interoperability Standards Organization (SISO) Spring Simulation Interoperability Workshop (SIW).* Huntsville, AL.

UN, 2009. *REPORT ON THE ACTIONS TAKEN ON RESOLUTIONS OF THE SEVENTEENTH UNITED NATIONS REGIONAL CARTOGRAPHIC CONFERENCE FOR ASIA AND THE PACIFIC.* Bangkok: United Nations Statistics Division and the Permanent Committee on Geographical Information System Infrastructure for Asia and the Pacific (PCGIAP).

Velikova, G., *et al.* 1999. Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. *Journal of Clinical Oncology*, 17, 998-1007.

Voss, J., 2007. Tagging, Folksonomy & Co – Renaissance of Manual Indexing. *10th international Symposium for Information Science.*

West, L. A. and Hess, T. J. 2002. Metadata as a Knowledge Management Tool: Supporting Intelligent Agent and End User Access to Spatial Data. *Decision Support Systems,* 32, 247- 264.

Westbrooks, E. L., 2004a. Distributing and Synchronizing Heterogeneous Metadata for the Management of Geospatial Information Repositories for Access. *Metadata In Practice.*

Westbrooks, E. L., 2004b. Distributing and Synchronizing Heterogeneous Metadata in Geospatial Information Repositories for Access. *In:* Hillman, D. I. and Westbrooks, E. L. eds. *Metadata in Practice.* Chicago.

Williams, S., Pettit, C., Hunter, D. and 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. *Spatial Data Infrastructure Convergence: Research, Emerging Trends, and Critical Assessment.* Delft, 137-150.

Williamson, I. P., Rajabifard, A. and Feeney, M.-E. F., 2003. *Developing Spatial Data Infrastructures: from Concept to Reality.* London: Taylor & Francis.

Wilson, E., 1998. Manuals go click. *The Age,* 30/6/1998, p.

Wyatt, J. C. 2000. When to use web-based surveys. *Journal of the American Medical Informatics Association,* 7(4), 426-430.

xMET, 2011. *xMET extensible Metadata Editing Tool* [online]. Available from: http://www.mymaps.gov.au/xmet/ [Accessed 25 August 2011].

Xu, Z., Fu, Y., Mao, J. and Su, D., 2006. Towards the Semantic Web: Collaborative Tag Suggestions. *Collaborative Web Tagging Workshop at the WWW 2006.* Edinburgh, Scotland.

Yeung, A. K. W. and Hall, G. B., 2007. *Spatial database systems : design, implementation and project management.* Dortrecht, Netherlands: Springer.

Yin, R. K., 1994. *Case study research: design and methods.* 2nd ed. Thousand Oaks, CA: Sage Publications.

Zarazaga-Soria, F. J., Lacasta, J., Nogueras-Iso, J., Torres, M. P. and Muro-Medrano, P. R. 2003. A Java Tool for Creating ISO/FGDC Geographic Metadata. *Geodaten- und Geodienste-Infrastrukturen - von der Forschung zur praktischen Anwendung,* 18, 17-30.

Zhang, C., Peng, Z.-R., Li, W. and Day, M. J. 2003. GML-Based Interoperable Geographical Database. *Cartography*, 32(2), 1-16.

APPENDIX 1

PUBLICATIONS RELATING TO RESEARCH

APPENDIX 1 - PUBLICATIONS RELATING TO RESEARCH

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H., and Williamson, I.P. 2012. **A GML-based Approach to Automate Spatial Metadata Automation**. *International Journal of Geographical Information Science (IJGIS*), DOI:10.1080/13658816.2012.678853.

Olfat, H., Kalantari, M., Rajabifard, A., and Williamson, I.P. 2012. **Towards a Foundation for Spatial Metadata Automation**. *Journal of Spatial Science*, 57(1), 65-81.

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H., and Williamson, I.P. 2012. **Spatial Metadata Automation: A Key to Spatially Enabling Platform**. *International Journal of Spatial Data Infrastructures Research (IJSDIR)*, 7, 173-195.

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, 119-129.

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***, GSDI-12 World Conference, 18-22 October 2010, Garden City, Singapore.

Olfat, H., Rajabifard, A., Kalantari, M., 2010. ***Automatic Spatial Metadata Update: a new approach***, FIG International Congress 2010, 11-16 April 2010, Sydney, Australia.

Kalantari, M., Rajabifard, A. and Olfat, H., 2009. **Spatial Metadata Automation: A New Approach**. *In:* Ostendorf, B., *et al.* eds. *Spatial Science Conference 2009 (SSC2009).* Adelaide, Australia, 629-635.

Olfat, H., Rajabifard, A., Kalantari, M., 2009. **A Synchronisation Approach to Automate Spatial Metadata Updating Process**, *Coordinates Magazine*, VI (3), 27-32.

Olfat, H. 2012. **Design and Development of a 'Process-based Spatial Metadata Creation Approach'**, *Spatial Data Infrastructure Asia & the Pacific Newsletter*, 9(8), 4-5.

Olfat, H. 2011. **A critical need for any SDI initiative: Automatic Spatial Updating and Enrichment**, *Spatial Data Infrastructure Asia & the Pacific Newsletter*, 8(10), 4-5.

Olfat, H. 2009. **Spatial Metadata Automation**, *Spatial Data Infrastructure Asia & the Pacific Newsletter*, 6(12), p.5.

APPENDIX 2

SUPPLEMENTARY INFORMATION RELATING TO SPATIAL METADATA AUTOMATION BACKGROUND (CHAPTER 3)

APPENDIX 2 - SUPPLEMENTARY INFORMATION RELATING TO SPATIAL METADATA AUTOMATION BACKGROUND (CHAPTER 3)

Table 3-I: Summery of the information that may be obtained from raster files (Manso *et al.* 2004)

Format	W / H	Bound Box	Pixel resolution	Bits /pixel	Bands /dimensions	Max, min statistical	Other statistics	Horizontal Units	Projection	Datum	Ellipsoid	Other
Gif	*	*(1)	*(1)	*	*							
Png	*	*(1)	*(1)	*	*							
Jpg	*	*(1)	*(1)	*	*							
Iff	*	*(1)	*(1)	*	*							
Tiff	*	*(1)	*(1)	*	*							
GeoTiff	*	*	*	*	*			*	*	*	*	
Bmp	*	*(1)	*(1)	*	*							
Pcx	*	*(1)	*(1)	*	*							
Psd	*	*(1)	*(1)	*	*							
Ras	*	*(1)	*(1)	*	*							
Sid	*	*(1) (2)	*(1) (2)	*	*			*(2)	*(2)			
Ecw	*	*(1) (2)	*(1) (2)	*	*			*(2)	*(2)	*(2)		
JP2000	*	*(1)	*(1)	*	*							
GeoJP2 (JP2k)	*	*	*	*	*			*	*	*	*	
DOQ2 (USGS)	*	*	*	*	*			*	*	*		
DTED	*	*	*	*	*			*	*	*	*	
DTE-Socet set	*	*	*	*	*			*				
GRD ESRI	*	*	*			*(3)		*	*	*	*	
GRD surfer	*	*	*			*						
DEM - USGS	*	*	*	*		*		*	*	*		
MicroDEM	*	*	*	*	*	*		*	*	*		
E00 grd	*	*	*(4)		*	*(4)	*(4)	*(4)	*(4)	*(4)	*(4)	
NTIF	*	*	*	*	*			*	*			*
ADF grd	*	*	*(4)		*	*(4)						
Lan erdas	*	*	*	*	*							
IMG erdas	*	*	*	*	*	*						
PIX	*	*	*		*							
ERS	*	*	*	*	*	*						
DOC (Idrisi)	*	*	*	*		*		*	*	*	*	

[1] In case of the existence of a world file.
[2] If version file contains this information.
[3] If all the data file has been read and its values may be computed.
[4] If all the file sections are presented

Table 3-II: Summary of information that may be obtained from vector files (Manso *et al.* 2004)

Format	BndBox	Nº of layers	Name of layers	Nº of different features	Name & number of features	Horizontal Units	Projection	Date	Ellipsoid
E00 arc	*					*	*	*	*
ADF arc	*			*	*	*	*	*	*
DGN	*	*		*	*				
DXF	*	*	*	*	*				
SHP	*	*	*	*	*				
MIF	*			*	*	*(1)	*(1)		
TAB	*			*	*	*	*	*	*
DWG	*	*	*	*	*				
VEC (idrisi)	*			*	*	*	*		
BIN	*	*	*	*	*				

[1]Equivalent semantics to the names of the layers

Table 3-III: Qualified Dublin Core element set used in the study (Batcheller 2008, Batcheller *et al.* 2007)

Core element name	Element refinement	Description
Title	–	Title
	Alternative	Alternative title
Description	Abstract	A brief narrative summary
Language		Language
Subject	Keywords	Main theme(s)
Date	Created	Date of creation
	Modified	Last date of update
	Period.name	Name of a specific interval. Used here to define frequency of update
Creator		Originating person/organisation
Publisher		Distributing person/organisation
Contributor		Contributing person/organisation
Format		Digital manifestation of resource
Type	Dataset	Nature of content
Rights	Access Rights	Access restrictions
Coverage	Spatial.Box.name	Name of geographic extent of dataset
	Spatial.Box.projection	Spatial reference system of dataset
	Spatial.Box.northlimit	Limits of dataset extent in coordinates
	Spatial.Box.eastlimit	
	Spatial.Box.southlimit	
	Spatial.Box.westlimit	
Identifier		Online linkage to dataset
Relation		A reference to a related resource
Source		A reference to a resource from which the present resource is derived

Table 3-IV: The UK GEMINI 32 metadata element set- optional elements in italics (Batcheller *et al.* 2009)

Element	Description
Title	Dataset name
Alternative title	*Alternative name*
Dataset language	Language used
Abstract	Narrative summary describing the dataset
Topic category	Main themes of dataset (high-level categories)
Subject	Topic of the dataset content (low-level categories)
Date	Data capture period
Dataset reference date	Date of dataset publication
Originator	*Originating person or organisation*
Lineage	*Dataset pedigree*
West bounding coordinate	Western limit of dataset
East bounding coordinate	Eastern limit of dataset
North bounding coordinate	Northern limit of dataset
South bounding coordinate	Southern limit of dataset
Extent	Geographic identifier of dataset
Vertical extent information	*Vertical domain of dataset*
Spatial reference system	Name of spatial reference system
Spatial resolution	*Capture precision of data*
Spatial representation type	*Method of spatial representation*
Presentation type	*Method of data manifestation*
Data format	Digital format of data
Supply media	*Method of data supply*
Distributor	Distributing organisation
Frequency of update	Prescribed frequency of data update
Access constraint	*Rights of data access*
Use constraints	*Rights of data use*
Additional information source	*Source of further details about dataset*
Online resource	*Online sources of dataset*
Browse graphic	*Illustrative sample of dataset*
Date of update of metadata	Last date of metadata update
Metadata standard name	*Name of metadata standard and profile used*
Metadata standard version	*Metadata version used*

Table 3-V: Automatic metadata element production enabling dynamic interoperability (Manso-Callejo *et al.* 2009)

MD_Metadata: Packed	Metadata element	(P) Produced[34]; (M) Multiple values)	(R:Raster, V:Vector, D:DTM)	Explanation
	dateStamp	P		Date, time of metadata generation
distributionInfo:distributionFormat	version	P		Some stores have versions
distributionInfo:distributionFormat	fileDecompressionTechnique	P		Some stores (mainly imagery) use
contentInfo:MD_CoverageDescription	contentType	P	R	Raster and grid data must be distinguished
contentInfo:MD_CoverageDescription: dimension:MD_RangeDimension	sequenceIdentifier	M	R	Band number
contentInfo:MD_CoverageDescription: dimension: MD_Band	maxValue	M	R	Pixel or cell maximum value
contentInfo:MD_CoverageDescription: dimension: MD_Band	minValue	M	R	Pixel or cell minimum value
contentInfo:MD_CoverageDescription: dimension: MD_Band	units	M	R	When grid data have associated units as DTM
contentInfo:MD_CoverageDescription: dimension: MD_Band	bitsPerValue	M	R	Number of bits used to encode band
contenInfo:MD_ContentInformation:MD_FeatureCatalogueDescription	includedWithDataset	P		When grid data have associated categories like classified image
contenInfo:MD_ContentInformation	featuresTypes	P		Name for every feature type
spatialRepresentationInfo: MD_GridSpatialRepresentation	numberOfDimensions	M	R	Usually 2
spatialRepresentationInfo: MD_GridSpatialRepresentation	cellGeometry	M	R	Depend on whether grid or image is
spatialRepresentationInfo:MD_GridSpatialRepresentation	transformationParameterAvailability	M	R	If GI store information is about control points
spatialRepresentationInfo:MD_GridSpatialRepresentation:axisDimensionProperties:MD_Dimension	dimensionName	M	R	Grid and image, row or column
axisDimensionProperties:MD_Dimension	dimesionSize	M	R	Count of rows or columns
spatialRepresentationInfo: MD_GridSpatialRepresentation: axisDimensionProperties:MD_Dimension: resolution	value	M	R	Pixel size on grid and raster rectified
axisDimensionProperties:MD_Dimension: resolution	units	M	R	Units of measure of resolution if known
spatialRepresentationInfo:MD_VectorSpatialRepresentation	topologyLevel	M	V	For vector data, topology information about data.
spatialRepresentationInfo: MD_VectorSpatialRepresentation: geometricObjects: MD_GeometricObjects	geometricObjectType	M	V	Type of geometry elements. One per type informing about type.
spatialRepresentationInfo: MD_VectorSpatialRepresentation: geometricObjects: MD_GeometricObjects	geometricObjectCount	M	V	Count of object elements per type

[34] extracted, calculated or derived

Table 3-V: Automatic metadata element production enabling dynamic interoperability (Manso-Callejo *et al.* 2009)

MD_Metadata: Packed	Metadata element	(P) Produced[34]; (M) Multiple values)	(R:Raster, V:Vector, D:DTM)	Explanation
identification:MD_Identificacion:MD_DataIdentification:citation:CI_Citation	title	P		Title inferred for dataset, based on BBOX, time information and other information that can be extracted, computed or inferred.
identification:MD_Identificacion:MD_DataIdentification:citation:CI_Citation:date:CI_Date	date	P		Date, time of metadata generation
identification:MD_Identificacion:MD_DataIdentification:citation:CI_Citation:date:CI_Date	dateType	P		Creation
identification:MD_Identificacion:MD_DataIdentification:resourceFormat:MD_Format	name	P		Same as distribution Format
identification:MD_Identificacion:MD_DataIdentification:resourceFormat:MD_Format	version	P		Same as distribution Format
identification:MD_Identificacion:MD_DataIdentification:resourceFormat:MD_Format	fileDecompressionTechnique	P		Same as distribution Format
identification:MD_Identificacion: MD_DataIdentification: resourceSpecificUsage: MD_Usage	userDeterminedLimitations	P		Some datasets stores use limitations. In this case, extract and include
identification:MD_Identificacion: MD_DataIdentification: resourceConstraints: MD_LegalConstraints	useConstraints	P		Some datasets stores use constraints. In this case, extract and include
identification:MD_Identificacion:descriptiveKeywords:MD_Keywords	keyword	P		Based on raster datasets analysis, content type can be inferred. Then some keywords included in thesaurus can be inferred.
identification:MD_Identificacion:descriptiveKeywords:MD_Keywords	type	P		Keyword type
identification:MD_Identificacion:descriptiveKeywords:MD_Keywords:thesaurusName:CI_Citation	title	P		Thesaurus Name
identification:MD_Identificacion:descriptiveKeywords:MD_Keywords:thesaurusName:CI_Citation	date	P		Thesaurus keywords date
identification:MD_Identificacion:descriptiveKeywords:MD_Keywords:thesaurusName:CI_Citation	dateType	P		Thesaurus date type
identification:MD_Identificacion	spatialRepresentationType	P		From dataset representation type can be extracted.
identification:MD_Identificacion:spatialResolution:MD_Resolution	distance	M	R	For raster and rectified grid, pixel size can be used to define resolution distance
identification:MD_Identificacion:spatialResolution:MD_Resolution:equivalentScale:MD_RepresentativeFraction	denominator	M	R	From resolution distance denominator can be computed, but only one is needed.
identification:MD_Identification:extent:EX_Extent: geographicElement: EX_GeographicExtent:EX_GeographicBoundingBox	extentTypeCode	P		1 (true: inclusion)

Table 3-V: Automatic metadata element production enabling dynamic interoperability (Manso-Callejo *et al.* 2009)

MD_Metadata: Packed	Metadata element	(P) Produced[34]; (M) Multiple values)	(R:Raster, V:Vector, D:DTM)	Explanation
identification:MD_Identification:extent:EX_Extent: geographicElement: EX_GeographicExtent:EX_GeographicBoundingBox	westBoundLongitude	P		West Longitude computed from dataset identifying source CRS.
identification:MD_Identification:extent:EX_Extent: geographicElement:EX_GeographicExtent:EX_GeographicBoundingBox	eastBoundLongitude	P		East Longitude computed from dataset identifying source CRS.
identification:MD_Identification:extent:EX_Extent: geographicElement: EX_GeographicExtent:EX_GeographicBoundingBox	southBoundLatitude	P		South Latitude computed from dataset identifying source CRS.
identification:MD_Identification:extent:EX_Extent: geographicElement:EX_GeographicExtent:EX_GeographicBoundingBox	northBoundLatitude	P		North Latitude computed from dataset identifying source CRS.
identification:MD_Identification:extent:EX_Extent: geographicElement: EX_GeographicExtent:EX_GeographicDescription: geographicIdentifier:RS_Identifier:authority:CI_Citation	title	P		Geographic identifier (toponym) computed by gazetteer reverse query
identification:MD_Identification:extent:EX_Extent: geographicElement: EX_GeographicExtent:EX_GeographicDescription: geographicIdentifier:RS_Identifier:authority:CI_Citation:date:CI_Date	date	P		Date time extracted from gazetteer metadata or database used to compute toponym
identification:MD_Identification:extent:EX_Extent: geographicElement:EX_GeographicExtent:EX_GeographicDescriptio: geographicIdentifier:RS_Identifier:authority:CI_Citation:date:CI_Date	dateType	P		publication
identification:MD_Identification:extent:EX_Extent: verticalElement:Ex_VerticalExtent	minimumValue	M	D	For grid dataset containing DTM, lower value.
identification:MD_Identification:extent:EX_Extent: verticalElement:Ex_VerticalExtent	maximumValue	M	D	For grid dataset containing DTM, upper value.
referenceSystemInfo:MD_ReferenceSystem:referenceSystemIdentifier:identifier:RS_Identifier	codeSpace	P		EPSG
referenceSystemInfo:MD_ReferenceSystem:referenceSystemIdentifier:identifier:RS_Identifier	version	P		EPSG database version
referenceSystemInfo:MD_ReferenceSystem:referenceSystemIdentifier:identifier:RS_Identifier	code	P		EPSG code computed by CRS extracted from dataset and translated to this codeSpace
referenceSystemInfo:MD_ReferenceSystem:referenceSystemIdentifier:identifier:RS_Identifier:authority:CI_Citation	title	P		EPSG Coordinate Reference Systems database
referenceSystemInfo:MD_ReferenceSystem:referenceSystemIdentifier:identifier:RS_Identifier:authority:CI_Citation:date:CI_Date	date	P		EPSG Database date
referenceSystemInfo:MD_ReferenceSystem:referenceSystemIdentifier:identifier:RS_Identifier:authority:CI_Citation:date:CI_Date	dateType	P		revision

Table 3-VI: Overview of the automatically generated metadata elements (Manso-Callejo *et al.* 2010)

MD_Metadata: Packed	Metadata element	(G) Extract, compute, infer; (C) dataset dependent; (I) imposed	(R:Raster, V:Vector, D:DTM)	Function: (D:Discover, E:evaluate, R:Retrieve, U:use)
	language	G		
	characterSet	I		
	hierarchyLevel	G		
	metadataStandardName	I		
	metadataStandardVersion	I		
	dataSetURI	G		
	dateStamp	G		
distributionInfo:transferOptions	unitsOfDistribution	G		R
distributionInfo:transferOptions	transferSize	G		R
distributionInfo:distributionFormat	name	G		R
distributionInfo:distributionFormat	version	G		R
distributionInfo:distributionFormat	fileDecompressionTechnique	G		R
contentInfo:MD_CoverageDescription	contentType	G	R	U
contentInfo:MD_CoverageDescription: dimension:MD_RangeDimension	sequenceIdentifier	C	R	U
contentInfo:MD_CoverageDescription: dimension:MD_RangeDimension	descriptor	C	R	U
contentInfo:MD_CoverageDescription: dimension: MD_Band	maxValue	C	R	U
contentInfo:MD_CoverageDescription: dimension: MD_Band	minValue	C	R	U
contentInfo:MD_CoverageDescription: dimension: MD_Band	units	C	R	U
contentInfo:MD_CoverageDescription: dimension: MD_Band	bitsPerValue	C	R	U
contenInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	complianceCode	G		U
contenInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	language	G		U
contenInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	includedWithDataset	G		U
contenInfo:MD_ContentInformation	featuresTypes	G		U
spatialRepresentationInfo: MD_GridSpatialRepresentation	numberOfDimensions	C	R	U
spatialRepresentationInfo: MD_GridSpatialRepresentation	cellGeometry	C	R	U
spatialRepresentationInfo: MD_GridSpatialRepresentation	transformationParameterAvailability	C	R	U
spatialRepresentationInfo: MD_GridSpatialRepresentation: axisDimensionProperties:MD_Dimension	dimensionName	C	R	U
axisDimensionProperties:MD_Dimension	dimensionSize	C	R	U
spatialRepresentationInfo: MD_GridSpatialRepresentation: axisDimensionProperties:MD_Dimension: resolution	value	C	R	U
axisDimensionProperties:MD_Dimension: resolution	units	C	R	U
spatialRepresentationInfo: MD_VectorSpatialRepresentation	topologyLevel	C	V	U

Table 3-VI: Overview of the automatically generated metadata elements (Manso-Callejo *et al.* 2010)

MD_Metadata: Packed	Metadata element	(G) Extract, compute, infer; (C) dataset dependent; (I) imposed	(R:Raster, V:Vector, D:DTM)	Function: (D:Discover, E:evaluate, R:Retrieve, U:use)
spatialRepresentationInfo: MD_VectorSpatialRepresentation:	geometricObjectType	C	V	U
geometricObjects: MD_GeometricObjects				
spatialRepresentationInfo: MD_VectorSpatialRepresentation: geometricObjects: MD_GeometricObjects	geometricObjectCount	C	V	U
identification:MD_Identificacion: MD_DataIdentification: citation: CI_Citation	title	G		D
identification:MD_Identificacion: MD_DataIdentification:citation: CI_Citation: date: CI_Date	date	G		D
identification:MD_Identificacion: MD_DataIdentification:citation: CI_Citation: date: CI_Date	dateType	G		D
identification:MD_Identificacion: MD_DataIdentification	credit	G		D
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	filename	G		D
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	fileDescription	G		D
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	fileType	G		D
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	name	G		D
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	version	G		D
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	fileDecompressionTechn ique	G		D
identification:MD_Identificacion: MD_DataIdentification: resourceSpecificUsage: MD_Usage	userDeterminedLimitati ons	G		E
identification:MD_Identificacion: MD_DataIdentification: resourceConstraints: MD_LegalConstraints	useConstraints	G		E

APPENDIX 3

RESEARCH QUESTIONNAIRE FOR CASE STUDY INVESTIGATIONS

APPENDIX 3 - RESEARCH QUESTIONNAIRE FOR CASE STUDY INVESTIGATIONS

spatial metadata automation

Centre for SDIs and Land Administration
Department of Geomatics
The University of Melbourne
August 2009

Part 1- Introduction

The Spatial Metadata Automation Research Project in the Centre for Spatial Data Infrastructure (SDI) and Land Administration (CSDILA) investigates automation of metadata sets so that when spatial data is updated, metadata related to that data is also automatically updated.

This project is an Australian Research Council (ARC) Linkage Project involving researchers from the CSDILA at the University of Melbourne and industry partners.

The research team has prepared a questionnaire to investigate the key areas of research in the metadata domain for organisations dealing with spatial data. This survey aims to identify current processes, issues and problems when creating and updating metadata records.

This questionnaire consists of six sections including questions regarding participant's information, organisational information, applications of metadata, creation and updating metadata, metadata management tools, sharing metadata and research questions on metadata. The questions are designed for short answers or selection of multi-choice answers, however, in some instances a possibility for further explanation is also provided.

In advance we would like to thank you for participation and input to this survey.

Part 2 - Participant's information

1. Your name:

2. Your email address:

3. Name of your organisation:

Part 3 - Organisational Spatial Data and Metadata Characteristics

4. How many spatial datasets does your organisation maintain?

5. What are your organisation activities regarding metadata for spatial datasets?

□Management □Coordination □Ownership □Security □Other (please specify)

6. What are your organisation activities regarding spatial metadata?

□Create □Update (edit, manipulate, delete) □Share with other organisations □Storing in data clearinghouse □Acquire metadata from other activities □Other (please specify)

Part 4 - Spatial Metadata Applications

7. To better conceptualise 'metadata automation', some of the most common applications of spatial metadata have been defined in this section. If your organisation has other types of applications, please name those with a short explanation.

NO.	Spatial Metadata Application	Definition
1	Data Discovery	Which datasets hold the data we are interested in? These enable organisations to know and publicise what data holdings they have. Discovery metadata is the minimum amount of information to be provided about the nature and content of the data resource. This falls into broad categories to answer the 'what, why, when, who, where and how' questions about geospatial data.
2	Data Exploration	Does the identified dataset contain sufficient information to enable appropriate analysis to be made for our purposes? This is documentation to be provided with the data to ensure that others use the data correctly and wisely. It should provide information to enable an inquirer to ascertain that data fit to given purpose does exist, to evaluate its properties, and to reference some point of contact for more information. The exploration metadata is also important when evaluating the quality of the dataset. Data quality measures have been created in order to help this evaluation process.

3	Data Exploitation	How to obtain and use this dataset? This helps end users and provider organisations to effectively store, reuse, maintain and archive their data holdings. This documentation should include properties required to access, transfer, interpret, and apply the data in the end application where it is exploited. Each of these purposes requires different levels of information and organisations should look at their overall needs and requirements before developing their metadata systems.
4	…………………….	…………………………………………

Part 5 - Spatial Metadata Creation and Updating

8. What are your organisation metadata input methods?

□Fully manual □Fully automatic □Mixed manual/automatic Method (semi automatic method) □Other (please specify)

9. For what percentage of your organisation datasets do you receive metadata from third parties and generate your own metadata by conversion? And in which format(s) do you receive them?

Format	Percentage
TXT	
XML	
HTML	
PDF	
Other formats (please specify)	

10. What standard do you use for creating and recording metadata

□ FGDC □ANZLIC version 2 □ISO 19115 ANZLIC profile 1.1 □Other standard (please specify)

11. What format do you store your metadata records

□TXT □XML □HTML □PDF □Other formats (please specify)

12. What is your mechanism to update spatial data?

Updating Team
□An internal team in your organisation □Third party (External team) □Other (please specify)

13. Do you update spatial data and its metadata at the same time?

□YES □NO

If not, please explain the spatial metadata updating time period.

Metadata Updating Time Period
□Fortnightly □Monthly □Yearly □ Other (please specify)

14. Is the spatial metadata updated by the team that updates spatial data?

□YES □NO □ Special arrangement (please explain)

Part 6 - Spatial Metadata Management Tools

15. What metadata tools do you use in your organisation?

Metadata tool	Description

Part 7 - Spatial Metadata Sharing

16. Do you share your metadata of resources with other parties?

□YES □NO

17. What percentage of your metadata currently can be accessed by other parties?

18. Please state through what mechanism(s) you publish your metadata of a digital data resource. (Note: more than one option may be selected.)

□Through web based data catalogue □Via online internal search engine □Via external search engines (e.g. google and yahoo) □Other (please specify)

Part 8 - Spatial Metadata Management Issues and Challenges/ Desirable Automation Workflow

19. What are the issues associated with generating and updating spatial metadata?

□Generation and updating spatial metadata is time consuming, monotonous and a labour-intensive process.
□Generation and updating spatial metadata is expensive.
□Current metadata standards and data models are complicated.
□The automatic approaches for generating spatial metadata do not cover all users' needs.
□Spatial data and its related metadata cannot be updated at the same time and it would spend more time and energy through more data redundancy.
□Lack of user friendly tools
□Multiple parties involved in the metadata creation and update process
□Other (please specify)

20. What spatial metadata creation workflow would you prefer?

□Fully automatic
□Initially metadata representation automatically generated, then edited by a person.
□A person enters some metadata, which is then automatically processed/encoded and/or augmented.
□A person creates all metadata
□Other (please specify)

21. Are there any features you think would be desirable in an automatic spatial metadata creating and updating application?

The research team appreciates for your cooperation.

APPENDIX 4

RESEARCH QUESTIONNAIRE FOR EVALUATING THE AUTOMATIC SPATIAL METADATA UPDATING PROTOTYPE SYSTEM

APPENDIX 4 - RESEARCH QUESTIONNAIRE FOR EVALUATING THE AUTOMATIC SPATIAL METADATA UPDATING PROTOTYPE SYSTEM

spatial metadata automation

Questionnaire for the evaluation of automatic spatial metadata updating prototype system

Centre for SDIs and Land Administration
Department of Infrastructure Engineering
The University of Melbourne

February 2012

Part 1- Introduction

The Spatial Metadata Automation Research Project in the Centre for Spatial Data Infrastructures (SDIs) and Land Administration (CSDILA) explores different approaches for automating the spatial metadata and the impacts 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.

This project is an Australian Research Council (ARC) Linkage Project which involves researchers from the CSDILA at the University of Melbourne and industry partners.

As part of the objectives of this project, the research team have designed and developed the following prototype system:

- Automatic Spatial Metadata Updating

This questionnaire consists of 4 sections including questions regarding participant's information, evaluation of the functionality, usability, and efficiency of the automatic metadata updating prototype system. The questions are designed for short answers or selection of multi-choice answers, however, in some instances a possibility for further explanation is also provided.

In advance we would like to thank you for participation and input to this survey.

Part 2- Participant's Information

Your name: Your email address:

Name of your organisation:

Division/unit and position title:

Part 3- Evaluation of System Functionality

This section aims at evaluating the effectiveness of the '**Automatic Spatial Metadata Updating Prototype System**' which is implemented within the GeoNetwork opensource. This prototype system is designed and developed to integrate the spatial dataset and metadata in a common medium environment (GML document) and exchange them through a Web service among the spatial systems and end users. With any change the dataset's responsible party makes to the dataset, its related metadata record would be updated automatically and at the same time.

The users of this system are the metadata and dataset responsible parties and custodians in the organisations.

Before answering the questions in this section, please watch this short video as an introductory to the whole system. **Please enlarge the video and set youtube to play this video in 'HD' quality.**

An overview of the Automatic Metadata Updating prototype system

http://www.youtube.com/watch?v=lyNXsf1x7UQ

Data Model

The **automatic metadata updating approach** is based on an **integrated metadata data model**. Through this data model, dataset and metadata from different sources are coupled and mapped to a single medium environment and exchanged among the spatial systems and end users. Accordingly, with any change to the dataset the relevant metadata values (e.g. extent, date of revision) would be updated at the same time and the new values for the dataset and metadata would be sent and stored at the backend concurrently. Therefore, this data model would reduce the burden of human interactions for updating the metadata records after modifying the datasets.

In fact, through this data model a single GML file consisting of the latest version of the **geometry**, **attributes** and **metadata** for each dataset is published to the end users through a Web service on request.

Figure 1 illustrates the simplified **integrated metadata data model**, which underlies the automatic metadata updating process.

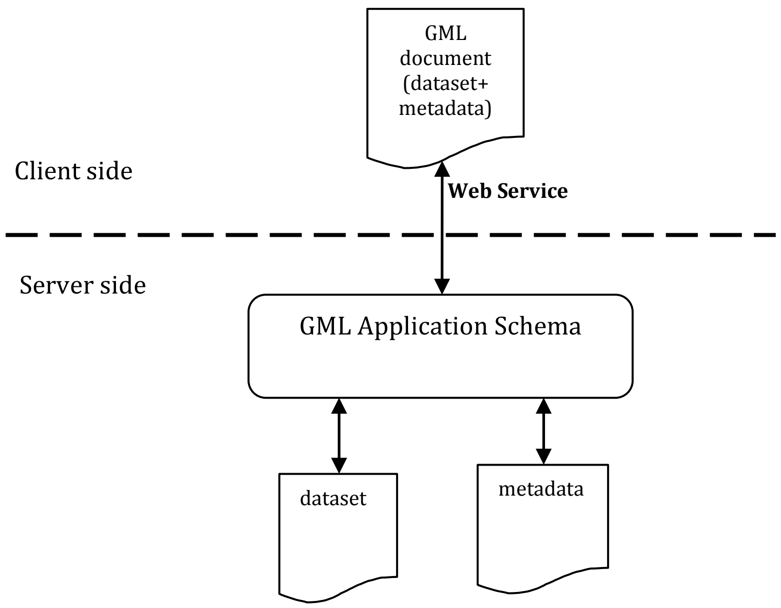


Figure 1. Simplified Integrated Metadata Data Model

1. To what extent do you agree with the following statement?
'It is useful to apply the **integrated metadata data model** (Figure 1) introduced by the automatic metadata updating system into the existing datasets and metadata in your organisation.'

Strongly disagree	1	2	3	4	5	Strongly agree

2. To what extent do you agree with the following statement?
'The **integrated metadata data model** (Figure 1) introduced by the automatic metadata updating system addresses the datasets interoperability issues and facilitates publishing of the datasets along with their metadata.'

Strongly disagree	1	2	3	4	5	Strongly agree

3. What are the advantages of the **integrated metadata data model** against the existing data models you are using for storing metadata in your organisation? Please list them here.

4. Do you perceive any disadvantages regrading the **integrated metadata data model** (Figure 1)?
□Yes □No
If so, please describe them:

Dataset Modification and the Automatic Metadata Updating

According to the video you watched earlier, when you access the metadata update page for any metadata record, by pressing the 'Map-based Modification' button you are able to edit the relevant dataset as well. Also, through the dataset modification interface you are able to edit features, and save the changes and at the same time the relevant metadata values are updated. If you go back to the metadata update page in the GeoNetwork, and push the 'Reset' button the latest details of the metadata would be shown.

5. To what extent do you agree with the following statement?
'It is useful to implement this functionality within your organisation's existing systems to manage and update the dataset and metadata automatically in real time.'

Strongly disagree						Strongly agree
	1	2	3	4	5	

6. To what extent do you agree with the following statement?
'Utilising the automatic metadata updating system facilitates sharing the most up-to-date version of metadata and therefore it improves the existing dataset discovery process.'

Strongly disagree						Strongly agree
	1	2	3	4	5	

7. According to the overall functionalities of the automatic metadata updating approach, are you sufficiently confident to use the system to update metadata for datasets you maintain?
□Yes □No
If not, what are your other expectations from this system?

8. If you assume that you already created the metadata record for your dataset, does updating the following elements satisfy your needs whenever a change happens to the dataset?

- Metadata date stamp
- Date of last revision
- Lineage statement (history of the dataset)
- Bounding box

□ Yes □ No

9. If your answer to the previous question was 'NO', what other metadata elements do you or your unit think would change at the same time with dataset modification and you recommend to be considered in the metadata updating system?

Part 4- Evaluation of System Usability

Please answer the following questions regarding the automatic metadata updating system usability.

10. I would like to use this system more frequently than the existing metadata management system in my organisation.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree
1 2 3 4 5

11. I perceive that the system is user friendly.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree
1 2 3 4 5

12. I feel confident using the system.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree
1 2 3 4 5

13. I need to learn a lot of things before I can get going with this system.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree
1 2 3 4 5

14. I need the support of a technical person to be able to use this system.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree
1 2 3 4 5

15. I perceive that people responsible for metadata could learn to use this system very quickly.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree
1 2 3 4 5

16. Various functions in this system are well integrated.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree

1 2 3 4 5

17. There is enough functionality in the system to make it a useful tool.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree

1 2 3 4 5

18. Please list the perceived negative aspects of the system:

19. Please list the perceived positive aspects of the system:

20. Do you have any suggestions to improve the user interface and/or functionality?

21. What are your other expectations from an automated metadata updating system?

Part 5- Evaluation of System Efficiency

22. For how many datasets in your organisation do you maintain metadata?
………….. (Please give an approximate number.)

23. What is the dataset updating time period in your organisation?
□Fortnightly
□Monthly
□Yearly
□ Other (please specify)

24. Imagine that you already created a metadata record in the GeoNetwork or another entry tool. In this case, how much time do you approximately spend on updating the following elements after any modification to each dataset?

- Metadata date stamp
- Date of last update
- Lineage statement
- Bounding box

……………. Minutes (in total)

25. To what extent do you agree with the following statement?
'Using the automatic metadata updating system will result in saving time for maintaining metadata records in your organisation.'

Strongly disagree						Strongly agree
	1	2	3	4	5	

26. Are the metadata records updated by the same team who updates the datasets?
□ Yes □ No

27. How much budget do you approximately consider for dataset updating every year?
……….. $ (US or Australian)
28. How much budget do you approximately consider for metadata updating every year?
……….. $ (US or Australian)

29. If the dataset and metadata are updated through an automated real-time approach by the team who only maintains the datasets, how much money do you think will be saved for your organisation every year?
……….. $ (US or Australian)

30. To what extent do you agree with the following statement?
'Utilising the automatic metadata updating system could likely result in cost savings for your organisation.'

Strongly disagree						Strongly agree
	1	2	3	4	5	

Thank you for taking our survey. Your response is very important to us.

APPENDIX 5

RESEARCH QUESTIONNAIRE FOR EVALUATING THE AUTOMATIC SPATIAL METADATA ENRICHMENT PROTOTYPE SYSTEM

APPENDIX 5 - RESEARCH QUESTIONNAIRE FOR EVALUATING THE AUTOMATIC SPATIAL METADATA ENRICHMENT PROTOTYPE SYSTEM

spatial metadata automation

Questionnaire for the evaluation of
automatic spatial metadata enrichment prototype system

Centre for SDIs and Land Administration
Department of Infrastructure Engineering
The University of Melbourne

February 2012

Part 1- Introduction

The Spatial Metadata Automation Research Project in the Centre for Spatial Data Infrastructures (SDIs) and Land Administration (CSDILA) explores different approaches for automating the spatial metadata and the impacts 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.

This project is an Australian Research Council (ARC) Linkage Project which involves researchers from the CSDILA at the University of Melbourne and industry partners.

As part of the objectives of this project, the research team have designed and developed the following prototype system:

- Automatic Spatial Metadata Enrichment

Therefore, this survey aims to evaluate the effectiveness, efficiency and satisfaction of the above prototype system based on the opinion of the spatial data and metadata custodians and users.

This questionnaire consists of 3 sections including questions regarding participant information, and the evaluation of the functionality and usability of the automatic metadata enrichment prototype system. The questions are designed for short answers or selection of multi-choice answers, however, in some instances a possibility for further explanation is also provided.

In advance we would like to thank you for participation and input to this survey.

Part 2- Participant's Information

1. Your name:

2. Your email address:

3. Name of your organisation:

4. Division/unit and position title:

Part 3- Evaluation of System Functionality

Please answer the following questions before reading the introduction of this section.

5. Which of the following ways do you prefer for discovering the datasets in a data catalogue system? A combination of choices can also be selected. □ Only typing a keyword in a box and then pushing the 'Search' button □ Whilst typing a keyword in the search box, the related previously searched keywords and their frequency can be suggested by the system. □ Accessing a cloud of the most popular keywords which are generated based on the previous users' interaction in the discovery system. □ Other (Please specify)

This section aims at evaluating the perceived effectiveness of the 'Automatic Spatial Metadata Enrichment Prototype System' which has been implemented within the GeoNetwork opensource. This prototype system is designed and developed to interact with end users to facilitate the spatial data discovery through improving the datasets search words (descriptive keyword metadata element). In fact, automatic metadata enrichment involves improving content of the metadata through monitoring tags and search words that are used by end users for finding datasets. Creating metadata by monitoring the users' interaction is based on the Folksonomy concept.

In this system, two models have been designed and developed: indirect and direct models. The indirect model relies on the system to monitor and record the most popular search words and tag them to datasets as new metadata values. However, the direct model is based on users' interaction in the system to tag the datasets with the search words they think best describe those.

In order to implement the above two models, two add-ons have been designed and developed within the GeoNetwork environment, as following:

- **Suggestion list add-on**: provides of 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.

- **Agree/disagree/Tag cloud add-on**: automatically observes users' interaction with the GeoNetwork and collects users' 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. Tag cloud also provides a visual representation of user-generated search words and tags' relative importance in the GeoNetwork's user interface. The Cloud is enriched by the user interactions on metadata records. Through this means a user-generated taxonomy can be established.

The users of this system are the users seeking datasets over the Web and through data catalogue systems as well as the data custodians who maintain the metadata records.

Before going through the evaluation process, please watch the following short video. This video represents the role of the automatic metadata enrichment prototype system (three above add-ons) in the datasets discovery process within the GeoNetwork. **Please enlarge the video and set youtube to play this video in 'HD' quality.**

An overview of the Automatic Metadata Enrichment prototype system
http://www.youtube.com/watch?v=KFmlPy-Iv9I

Now, please answer the following questions.

6. To what extent do you agree with the following statement?
'Utilising the **automatic spatial metadata enrichment system** (indirect and direct models) will likely result in interacting with end users to improve the content of metadata keyword and datasets search and discovery process.'

Strongly disagree						Strongly agree
	1	2	3	4	5	

7. To what extent do you agree with the following statement?
'It is useful to apply the automatic metadata enrichment system into your organisation's existing data discovery system.'

Strongly disagree						Strongly agree
	1	2	3	4	5	

8. To what extent do you agree with the following statement?
'Suggesting the previous searched terms while typing the keyword during the discovery process would result in accelerating the data discovery task and sharing the previous users' interaction.'

Strongly disagree						Strongly agree
	1	2	3	4	5	

9. Based on the video you watched earlier, the current keyword weighting logic to record the search word for the dataset's metadata (stage 2 of the tagging process) is as following:
Indirect weighting process:
- Pressing the 'metadata' button to explore more information about the dataset: adding 0.25 of weighting unit to the search word which was used for the discovery purpose.
- Spending 'enough' time with the metadata details open: adding 0.5 of weighting unit to the search word which was used for the discovery purpose.
- Pressing the 'download' button to access the dataset: adding 0.75 of weighting unit to the search word which was used for the discovery purpose.

Direct weighting process:
- 'Agreeing' with an existing search word: adding 1 weighting unit to that search word.
- 'Adding' a new search word (tag): adding 1 weighting unit to the new search word.
- 'Disagreeing' with an existing search word: reducing 1 weighting unit from that search word (in some cases, the search word weight might become minus zero).

To what extend do you agree with the above search word weighting logic?

Strongly disagree						Strongly agree
	1	2	3	4	5	

10. Following the question 9, do you have any suggestions to improve the current search word weighting logic? In your opinion, what other factors should be considered in the search word weighting process?

11. Following the question 9, which generic factors should be considered in any discovery system for identifying the **most appropriate weight** that allows the search word to be tagged to the dataset's metadata as a new metadata value?

12. To what extent do you agree with the following statement?
'The search words provided by the users' interaction in the discovery process are more practical for finding the datasets than the search words already provided in the metadata record.'

Strongly disagree	1	2	3	4	5	Strongly agree

13. To what extent do you agree with the following statement?
'The search words provided by the users' interaction in the discovery process should be verified and approved by the data custodian before being considered as a new metadata value (descriptive keyword).'

Strongly disagree	1	2	3	4	5	Strongly agree

14. To what extent do you agree with the following statement?
'Using the Tag Cloud to present the most popular search words and their weights is a valuable addition to supporting the search and discovery functionality of a data catalogue system.'

Strongly disagree	1	2	3	4	5	Strongly agree

Part 4- Evaluation of System Usability

Based on the video you watched earlier, please answer the following questions regarding the usability of the automatic metadata enrichment system.

15. I think that I would like to use this system for searching and discovering datasets more frequently than the current system in my organisation.

Strongly disagree	1	2	3	4	5	Strongly agree

16. I perceive that the system is user friendly.

Strongly disagree	1	2	3	4	5	Strongly agree

17. I think that I need the support of a technical person to be able to use this system.

Strongly disagree	1	2	3	4	5	Strongly agree

18. I think that the various functions (add-ons) in this system are well integrated.

Strongly disagree	1	2	3	4	5	Strongly agree
	☐	☐	☐	☐	☐	

19. List the perceived negative aspects of the system:

20. List the perceived positive aspects of the system:

21. Do you have any suggestions to improve the user interface and/or functionality?

22. What are your other expectations from an automated metadata enrichment system?

Thank you for taking our survey. Your response is very important to us.

APPENDIX 6

SUPPLEMENTARY INFORMATION RELATING TO SELECTED SPATIAL METADATA MANAGEMENT TOOLS (CHAPTER 5)

APPENDIX 6 - SUPPLEMENTARY INFORMATION RELATING TO SELECTED SPATIAL METADATA MANAGEMENT TOOLS (CHAPTER 5)

Table 5-I: Arguments Passed In to IMetadataSynchroniser::Update (ESRI 2002a)

itemDesc	Value	Description
Boilerplate	Null	A special case that is called only the first time synchronisation occurs. This should be used to add boilerplate text, such as documentation hints or fixed contact information for an organisation that should be changed again by synchronisation.
CoverageEntity[i]	IArcInfoTable	Provides access to INFO tables and a coverage feature class' feature attribute table.
CoverageFeatureClass[i]	ICoverageFeatureClass	Provides information about a coverage feature class including the type of feature class and whether it has an attribute table or topology.
DatasetLocation	String	Location of the dataset on disk, or connection information for ArcSDE data or ArcIMS services.
DatasetName	String	Name of the dataset derived from the file name or the table name.
DatasetSize	String	Size of the dataset on disk. (For some file-based data sources only.)
DDExtent	IEnvelope	Envelope containing the dataset's extents in decimal degrees.
Entity[i]	IClass	Provides access to an object class, such as a table or feature class (when synchronizing the class). For example, this is used to obtain detailed descriptions of fields and subtypes.
EntityBrief[i]	IClass	Provides access to an object class, such as a table or feature class (when synchronizing a dataset that classes participate in, like topologies and networks). For example, a feature dataset has a list of the datasets it contains, but doesn't include detailed descriptions of them.
Environment	String	Name and version of the operating system and software being used.
FeatureClass[i]	IFeatureClass	Provides access to a feature class. Used to record feature information.
GeoForm	String	Mode in which the spatial data is represented. The following are four possible values: • Raster digital data • Remote sensing image • Tabular digital data • Vector digital data These values are only appropriate for use in metadata following FGDC metadata standards. The ISO synchroniser ignores this value and determines the type of data based on the NativeForm value; it is recommended that custom synchronisers not using FGDC standards do the same.
GeometryType	String	Type of data stored in the dataset. The following are two possible values: • Raster • Vector
Language	String	Two-letter language code derived from the default input locale of the operating system. For example, "en" is the code for English locales. This is presumed to be the language of the data and metadata.
MetadataDate	String	The current date in YYYYMMDD format.

Table 5-I: Arguments Passed In to IMetadataSynchroniser::Update (ESRI 2002a)

itemDesc	Value	Description
MetadataStandard	String	The name of the metadata standard supported by ArcCatalog, for example, FGDC-STD-001-1998. The default value provided is only appropriate for use in metadata following FGDC metadata standards.
NativeExtent	IEnvelope	Envelope containing the dataset's extents in projected coordinates. If the dataset is not projected, the envelope contains its extent in decimal degrees.
NativeForm	String	Type of dataset, for example shapefile, personal geodatabase table, or raster dataset. This string is the same as that displayed in the Type column in ArcCatalog's contents view, and is derived from the NativeType of the dataset.
NetworkRule[i]	IRule	Provides information about a connectivity rule in a geometric network.
NetworkSchema	INetSchema	Provides information about the schema of a geometric network, such as element classes, ancillary roles, and weights.
OperatingSystem	String	The name of the operating system on the computer used to create or update the metadata.
RasterBand	IRasterBand	Provides access to information about a raster band including its attribute table and colour map.
RasterDataset	IRasterDataset	Provides information about a raster dataset, such as its format and compression type.
Relationship[i]	IRelationshipClass	Provides access to a relationship class. For example, this is used to record detailed information about a relationship class when synchronizing the relationship class' metadata.
RelationshipBrief[i]	IRelationshipClass	Provides access to a relationship class. This is used to obtain basic information about a relationship class when synchronizing a dataset that participates in the relationship class.
Server	String	The ArcIMS uniform resource locator (URL), for ArcIMS Image, ArcMap Image, and Feature services.
Service	String	The ArcIMS service name.
ServiceType	String	The ArcIMS service type; Image for Image and ArcMap Image services, Feature for Feature services.
ServiceFCType	String	The geometry type of the feature class in an ArcIMS Feature service.
ServiceFCName	String	The name of the feature class in an ArcIMS Feature service.
Software	String	The name and version of the software used to create or update the metadata.
SpatialReference	ISpatialReference	Provides access to the dataset's spatial reference.
Tin	ITin	Provides access to a TIN dataset.

Table 5-II: Metadata elements generated automatically using CatMDEdit (CatMDEdit 2011b)

ISO 19115: 2003 metadata elements	SHP	DGN	ECW	FICC	Geo TIFF	GIF/ GFW	JPG /JGW	PNG /PGW
MD_Metadata.identificationInfo> MD_Data|dentification.exte nt> EX_Extent.geographicElement> EX_GeographicBoundingBox.northBoundLatitude, EX_GeographicBoundingBox.southEIoundLatitude, EX_GeographiclaeundingBox.eastEIoundLengitude, EX_GeographicBoundingBox.westBoundLongitude	*	*	*	*	*	*	*	*
MD_Metadata.identificationInfo> MD_Dataldentification.spatialRepresentationType	*	*	*	*	*	*	*	*
MD_Metadata.spatialRepresentationInfo>MD_VectorSpatialRepresentation.geometric0bjects>MD_Geometric0bjectsgeometricObjectType	*	*		*				
MD_Metadata.spatialRepresentationlnfo>MD_VectorSpatilRepresentation.geometric0bjects>MD_GeometricObjectsgeometricObjectCount	*	*		*				
MD_Metadata.spatialRepresentationlnfo>MD_GridSpatialRepresentation.number0fDimensions			*		*	*	*	*
MD_Metadata.spatialRepresentationInfo>MD_Gridspatia1Representation.axisDimensionProperties>MD_Dimension.dimensionName			*		*	*	*	*
MD_Metadata.spatialRepresentationInfo>MD_GridSpatialRepresentation.axisDimensionProperties>MD_DimensiondimensionSize			*		*	*	*	*
MD_Metadata.distributionInfo>MD_Distribution.transfer0ptions>MD_DigitalTransfer0ptions.onLine>CI_0nIineResource.Iinkage			*		*	*	*	*
MD_Metadata.distributionlnfo> MD_Distribution.transfer0ptions> MD_DigitalTransfer0ptions.transferSize			*		*	*	*	*
MD_MetaclatadistributionInfo> MD_Distribution.distributionFormat> MD_Format.name			*		*	*	*	*
Feature catalogue directory	*							

Table 5-III: Metadata elements which can be automated through ANZMetLite (OSDM 2009a)

Name	ISO Reference	Collection method
Metadata File Identifier	fileIdentifier(2)	Automated – automatically generated
Metadata Point of Contact	contact (8)	No input required. Derived from contact elements that follow
Metadata Contact Role	role (379)	Automated – pre-set value (default = 'point of contact') otherwise select from a pick list
Metadata Contact Individual Name	individualName (375)	Automated – pre-set value
Metadata Contact Organisation	organisationName (376)	Automated – pre-set value
Metadata Contact Position	positionName (377)	Automated – pre-set value A default setting can be established for an organisation or the details can be entered once and saved to a file for reuse via a pull down menu.
Metadata Hierarchy Level	hierarchyLevel (6)	Automated – pre-set value (otherwise Manual – select from list)
Metadata Hierarchy Level Name	hierarchyLevelName (7)	Automated – pre-set value (otherwise Manual – select from list)
Metadata Standard Name	metadataStandardName (10)	Automated – pre-set value
Metadata Standard Version	metadataStandardVersion (11)	Automated – pre-set value
Metadata Date Stamp	dateStamp (9)	Automated – pre-set value
Metadata Language	language (3)	Automated – pre-set value or select from a list
Resource Title	title (360)	This element would normally be entered manually. ANZMet Lite can derive a title from a database if one is selected on start-up of the tool.
Resource Language	language (39)	Automated – pre-set value (if not English select from list)
West Bounding Longitude	westBoundLongitude (344)	This can be derived if a database is selected with the ANZMet Lite, otherwise Manual – keyboard entry
East Bounding Longitude	eastBoundLongitude (345)	This can be derived if a database is selected with the ANZMet Lite, otherwise Manual – keyboard entry
South Bounding Latitude	southBoundLatitude (346)	This can be derived if a database is selected with the ANZMet Lite, otherwise Manual – keyboard entry
North Bounding Latitude	northBoundLatitude (347)	This can be derived if a database is selected with the ANZMet Lite, otherwise Manual – keyboard entry
Resource Point of Contact	contact (8)	No input required. Derived from the contact elements that follow.
Role	role (379)	Automated – pre-set value (default = 'author'), otherwise select from a pick list
Responsible Party Individual Name	individualName (375)	Automated – pre-set value
Responsible Party Organisation Name	organisationName (376)	Automated – pre-set value
Responsible Party Position Name	positionName (377)	Automated – pre-set value
Resolution	MD_Resolution (59)	This can be derived if a database is selected with ANZMet Lite, otherwise Manually from the elements that follow.
Scale Distance	distance (61)	This can be derived if a database is selected with ANZMet Lite, otherwise Manual (keyboard entry).
Spatial Resolution – Scale	equivalentScale (60)	This can be derived if a database is selected with ANZMet Lite, otherwise Manual.
Data Character Set	characterSet (40)	Automated – pre-set value

APPENDIX 7

RESULT OF MAPPING ISO 19115: 2003 METADATA ELEMENTS AGAINST THE SPATIAL DATA LIFECYCLE STEPS (CHAPTER 6)

APPENDIX 7 - RESULT OF MAPPING ISO 19115: 2003 METADATA ELEMENTS AGAINST THE SPATIAL DATA LIFECYCLE STEPS (CHAPTER 6)

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
1	Metadata entity set information: MD_Metadata	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 2-22).
2	Metadata entity set information: MD_Metadata: fileIdentifier	This is the first element that should be provided for any metadata record. Therefore, 'Planning and policy making' as the first step of data lifecycle is the related step.	Automatic- computation based on a predefined rule (e.g. GeoNetwork)
3	Metadata entity set information: MD_Metadata: language	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite); or extraction **
			Manual- select from an existing list
4	Metadata entity set information: MD_Metadata: characterSet	Planning and policy making	Automatic- predefined value** (e.g. ANZMet Lite)
			Manual- select from an existing list
5	Metadata entity set information: MD_Metadata: parentIdentifier	Planning and policy making	Manual- keyboard entry, or select from an existing list of previous metadata UUIDs (e.g. GeoNetwork)
6	Metadata entity set information: MD_Metadata: hierarchyLevel	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite); or extraction/inference**
			Manual- select from an existing list
7	Metadata entity set information: MD_Metadata: hierarchyLevelName	Planning and policy making	Automatic- system generated following the value for line 6
8	Metadata entity set information: MD_Metadata: contact	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
9	Metadata entity set information: MD_Metadata: dateStamp	This element should be created within data 'Collection' step and updated whenever metadata is 'Modified' within other steps.	Automatic- system generated** (e.g. Arccatalog)
10	Metadata entity set information: MD_Metadata: metadataStandardName	Planning and policy making	Automatic- predefined value** (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
11	Metadata entity set information: MD_Metadata: metadataStandardVersion	Planning and policy making	Automatic- predefined value** (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
11.1	Metadata entity set information: MD_Metadata: dataSetURI	Publication / Maintenance	Manual – keyboard entry Automatic **- harvesting (e.g. GeoNetwork)
12	Metadata entity set information: *Role name*: spatialRepresentationInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 156-185).
13	Metadata entity set information: *Role name*: referenceSystemInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 186-231).
14	Metadata entity set information: *Role name*: metadataExtensionInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 303-319).
15	Metadata entity set information: *Role name*: identificationInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 22-66.5).
16	Metadata entity set information: *Role name*: contentInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 232-265).
17	Metadata entity set information: *Role name*: distributionInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 270-302).
18	Metadata entity set information: *Role name*: dataQualityInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 78-141).
19	Metadata entity set information: *Role name*: portrayalCatalogueInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 268-269).
20	Metadata entity set information: *Role name*: metadataConstraints	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 67-77).
21	Metadata entity set information: *Role name*: applicationSchemaInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 320-333).
22	Metadata entity set information: *Role name*: metadataMaintenance	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 142-155).
23	Identification information: MD_Identification	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 24-35.1).
24	Identification information: *MD_Identification:* citation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 359-409).
25	Identification information: *MD_Identification:* abstract	Creation	Manual – keyboard entry
26	Identification information: *MD_Identification:* purpose	Creation	Manual – keyboard entry
27	Identification information: *MD_Identification:* credit	Creation	Manual- keyboard entry, or select from an existing list of contributors
28	Identification information: *MD_Identification:* status	Creation / Maintenance	Manual- select from an existing list
29	Identification information: *MD_Identification:* pointOfContact	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite) Manual- keyboard entry, or select from an existing list
30	Identification information: *Role name:* resourceMaintenance	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 142-155).

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
31	Identification information: *Role name:* graphicOverview	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 48-51).
32	Identification information: *Role name:* resourceFormat	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 284-290).
33	Identification information: *Role name:* descriptiveKeywords	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 52-55).
34	Identification information: *Role name:* resourceSpecificUsage	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 62-66).
35	Identification information: *Role name:* resourceConstraints	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 67-77).
35.1	Identification information: *Role name:* aggregationInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 66.1-66.5).
36	Identification information: MD_DataIdentification	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 37-46 and 24-35.1).
37	Identification information: MD_DataIdentification: spatialRepresentationType	Creation	Manual- select from an existing list
			Automatic- extraction*** (e.g. ArcCatalog)
38	Identification information: MD_DataIdentification: spatialResolution	Creation	The inputs are provided by its sub elements (lines 59-61).
39	Identification information: MD_DataIdentification: language	Planning and policy making	Automatic –predefined value (e.g. ANZMet Lite)
			Manual-select from an existing list
40	Identification information: MD_DataIdentification: characterSet	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- select from an existing list
41	Identification information: MD_DataIdentification: topicCategory	Creation	Manual – select from an existing list
42	intentionally left blank		
43	intentionally left blank		
44	Identification information: MD_DataIdentification: environmentDescription	Creation	Manual- keyboard entry
			Automatic- extraction from the system (e.g. ArcCatalog)
45	Identification information: MD_DataIdentification: extent	Creation	The inputs are provided by its sub elements (lines 334-358).
46	Identification information: MD_DataIdentification: supplementalInformation	Creation	Manual- keyboard entry
47	Identification information: MD_ServiceIdentification	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 24-35.1).
48	Identification information: MD_BrowseGraphic	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 49-51).

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
49	Identification information: MD_BrowseGraphic: fileName	Publication	Manual- keyboard entry
			Automatic- extraction** (e.g. ArcCatalog)
50	Identification information: MD_BrowseGraphic: fileDescription	Publication	Manual- keyboard entry
			Automatic- extraction/inference** (e.g. ArcCatalog)
51	Identification information: MD_BrowseGraphic: fileType	Publication	Manual- keyboard entry, or select from an existing list
			Automatic- extraction**
52	Identification information: MD_keywords	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 53-55).
53	Identification information: MD_Keywords: keyword	Publication / Maintenance	Manual- keyboard entry, or select from an existing thesaurus
			Automatic for raster-inference***
		Discovery and access	Automatic- tagging keywords (discussed in Section 6.5 in Chapter 6)
54	Identification information: MD_Keywords: Type	Publication	Manual- keyboard entry, or select from an existing thesaurus
55	Identification information: MD_Keywords: ThesaurusName	Publication	Manual- keyboard entry, or select from an existing list
56	Identification information: MD_RepresentativeFraction	It follows its sub elements related step(s).	The inputs are provided by its sub element (line 57).
57	Identification information: MD_RepresentativeFraction: denominator	Creation	Manual- keyboard entry, or select from an existing list
58	intentionally left blank		
59	Identification information: MD_Resolution	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 60-61).
60	Identification information: MD_Resolution: equivalentScale	Creation	Manual- keyboard entry
		Storage	Automatic- computation***
61	Identification information: MD_Resolution: distance	Creation	Manual- keyboard entry
		Storage	Automatic- computation***
62	Identification information: MD_Usage	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 63-66).
63	Identification information: MD_Usage: specificUsage	Planning and policy making / Creation / Utilisation	Manual- keyboard entry
64	Identification information: MD_Usage: usageDateTime	Utilisation	Automatic- system generated
65	Identification information: MD_Usage: userDeterminedLimitations	Utilisation	Manual- keyboard entry
			Automatic- extraction/inference**
66	Identification information: MD_Usage: userContactInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 359-409).

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
66.1	MD_AggregateInformation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 66.2-66.5).
66.2	Identification information: MD_AggregateInformation: aggregateDataSetName	Creation	Manual- keyboard entry
66.3	Identification information: MD_AggregateInformation: aggregateDataSetIdentifier	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 205-208.2).
66.4	Identification information: MD_AggregateInformation: associationType	Creation	Manual- keyboard entry, or select from an existing list
66.5	Identification information: MD_AggregateInformation: initiativeType	Creation	Manual- keyboard entry, or select from an existing list
67	Constraint information: MD_Constraints	It follows its sub elements related step(s)	The inputs are provided by its sub element (line 68).
68	Constraint information: MD_Constraints: useLimitation	Planning and policy making	Manual- keyboard entry
69	Constraint information: MD_LegalConstraints	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 70-72 and 68).
70	Constraint information: MD_LegalConstraints: accessConstraints	Planning and policy making	Manual- select from an existing list
71	Constraint information: MD_LegalConstraintstraints	Planning and policy making	Manual- select from an existing list Automatic- extraction/inference**
72	Constraint information: MD_LegalConstraints: otherConstraints	Planning and policy making	Manual- keyboard entry
73	Constraint information: MD_SecurityConstraints	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 74-77 and 68).
74	Constraint information: MD_SecurityConstraints: classification	Planning and policy making	Manual- select from an existing list
75	Constraint information: MD_SecurityConstraints: userNote	Planning and policy making	Manual- keyboard entry
76	Constraint information: MD_SecurityConstraints: classificationSystem	Planning and policy making	Manual- keyboard entry
77	Constraint information: MD_SecurityConstraints: handlingDescription	Planning and policy making	Manual- keyboard entry
78	Data quality information: DQ_DataQuality	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 79-81).
79	Data quality information: DQ_DataQuality: scope	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 138-141).
80	Data quality information: DQ_DataQuality: *Role name*: report	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 99-127).
81	Data quality information: DQ_DataQuality: *Role name*: lineage	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 82-98).
82	Data quality information: LI_Lineage	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 83-85).

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
83	Data quality information: LI_Lineage: statement	Collection / Creation	Manual- keyboard entry
		Maintenance	Automatic- generated by the system along with the dataset modification (discussed in Section 6-4 in Chapter 6)
84	Data quality information: DQ_DataQuality: *Role name*: processStep	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 86-90).
85	Data quality information: DQ_DataQuality: *Role name*: source	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 92-98).
86	Data quality information: LI_ProcessStep	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 87-91).
87	Data quality information: LI_ProcessStep: description	Creation	Manual- keyboard entry
88	Data quality information: LI_ProcessStep: rationale	Creation	Manual- keyboard entry
89	Data quality information: LI_ProcessStep: dateTime	Creation	Manual- keyboard entry, or select from an existing calendar
90	Data quality information: LI_ProcessStep: processor	Creation	Manual- keyboard entry, or select from an existing list
91	Data quality information: LI_ProcessStep: *Role name*: source	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 92-98).
92	Data quality information: LI_Source	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 93-98)
93	Data quality information: LI_Source: description	Collection	Manual- keyboard entry
94	Data quality information: LI_Source: scaleDenominator	Collection	Manual- keyboard entry, or select from an existing list
95	Data quality information: LI_Source: sourceReferenceSystem	Collection	Manual- keyboard entry, or select from an existing list
96	Data quality information: LI_Source: sourceCitation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 359-409)
97	Data quality information: LI_Source: sourceExtent	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 334-358).
98	Data quality information: LI_Source: *Role name*: sourceStep	It follows its sub elements related step(s).	The inputs are provided by its sub elements(lines 86-90)
99	Data quality information: *DQ_Element*	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 100-107).
100	Data quality information: *DQ_Element:* nameOfMeasure	Creation	Manual- keyboard entry
101	Data quality information: *DQ_Element:* measureIdentification	Creation	Manual- select from an existing list
102	Data quality information: *DQ_Element:* measureDescription	Creation	Manual- keyboard entry
103	Data quality information: *DQ_Element:* evaluationMethodType	Creation	Manual- select from an existing list
104	Data quality information: *DQ_Element:* evaluationMethodDescription	Creation	Manual- keyboard entry

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
105	Data quality information: *DQ_Element:* evaluationProcedure	Creation	Manual- keyboard entry
106	Data quality information: *DQ_Element:* dateTime	Creation	Manual- keyboard entry, or select from an existing calendar
107	Data quality information: *DQ_Element:* result	Creation	Manual- keyboard entry, or select from an existing list
108	Data quality information: DQ_Completeness	Creation	Automatic- using 'Brute force & comparing against Reference' methods*
109	Data quality information: DQ_Completeness Commission	Creation	Automatic- using 'Brute force & comparing against Reference' methods*
110	Data quality information: DQ_CompletenessOmission	Creation	Automatic- using 'Brute force & comparing against Reference' methods*
111	Data quality information: DQ_LogicalConsistency	Creation	Automatic- using 'Brute force' method*
112	Data quality information: DQ_ConceptualConsistency	Creation	Automatic- using 'Brute force' method*
113	Data quality information: DQ_DomainConsistency	Creation	Automatic- using 'Brute force' method*
114	Data quality information: DQ_FormatConsistency	Creation	Automatic- using 'Brute force' method*
115	Data quality information: DQ_TopologicalConsistency	Creation	Automatic- using 'Brute force & comparing against Reference' methods*
116	Data quality information: DQ_PositionalAccuracy	Creation	Manual- keyboard entry, or select from an existing list
117	Data quality information: DQ_AbsoluteExternal PositionalAccuracy	Creation	Manual- keyboard entry, or select from an existing list
118	Data quality information: DQ_GriddedDataPositional Accuracy	Creation	Manual- keyboard entry, or select from an existing list
119	Data quality information: DQ_RelativeInternalPositional Accuracy	Creation	Manual- keyboard entry, or select from an existing list
120	Data quality information: DQ_TemporalAccuracy	Creation	Manual- keyboard entry, or select from an existing list
121	Data quality information: DQ_AccuracyOfATime Measurement	Creation	Manual- keyboard entry, or select from an existing list
122	Data quality information: DQ_TemporalConsistency	Creation	Manual- keyboard entry, or select from an existing list
123	Data quality information: DQ_TemporalValidity	Creation	Manual- keyboard entry, or select from an existing list
124	Data quality information: DQ_ThematicAccuracy	Creation	Manual- keyboard entry, or select from an existing list
125	Data quality information: DQ_ThematicClassification Correctness	Creation	Automatic- using 'comparing against Reference' method*
126	Data quality information: DQ_NonQuantitativeAttribute Accuracy	Creation	Automatic- using 'comparing against Reference' method*
127	Data quality information: DQ_QuantitativeAttribute Accuracy	Creation	Automatic- using 'Stochastic' method*
128	Data quality information: DQ_Result	Creation	Manual- keyboard entry, or select from an existing list

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
129	Data quality information: DQ_ComformanceResult	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 130-132).
130	Data quality information: DQ_ConformanceResult: specification	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 359-409).
131	Data quality information: DQ_ConformanceResult: explanation	Creation	Manual- keyboard entry
132	Data quality information: DQ_ConformanceResult: pass	Creation	Manual- select from an existing list
133	Data quality information: DQ_QuantitativeResult	It follows its sub elements related step(s)	The inputs are provided by its sub elements (lines 134-137).
134	Data quality information: DQ_QuantitativeResult: valueType	Creation	Manual- keyboard entry
135	Data quality information: DQ_QuantitativeResult: valueUnit	Creation	Manual- select from an existing list
136	Data quality information: DQ_QuantitativeResult: errorStatistic	Creation	Manual- keyboard entry
137	Data quality information: DQ_QuantitativeResult: value	Creation	Manual- keyboard entry
138	Data quality information: DQ_Scope	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 139-141).
139	Data quality information: DQ_Scope: level	Creation	Manual- select from an existing list
140	Data quality information: DQ_Scope: extent	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 334-358).
141	Data quality information: DQ_Scope: levelDescription	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 142-155).
142	Maintenance information: MD_MaintenanceInformation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 143-148.1).
143	Maintenance information: MD_MaintenanceInformation: maintenanceAndUpdateFrequency	Creation / Maintenance	Manual- select from an existing list
144	Maintenance information: MD_MaintenanceInformation: dateOfNextUpdate	Creation / Maintenance	Manual- keyboard entry, or select from an existing calendar Automatic- computation based on a predefined time difference
145	Maintenance information: MD_MaintenanceInformation: userDefinedMaintenanceFrequency	Maintenance / Utilisation	Manual- keyboard entry, or select from an existing calendar
146	Maintenance information: MD_MaintenanceInformation: updateScope	Maintenance	Manual- select from an existing list
147	Maintenance information: MD_MaintenanceInformation: updateScopeDescription	Maintenance	Manual- keyboard entry
148	Maintenance information: MD_MaintenanceInformation: maintenanceNote	Maintenance	Manual- keyboard entry
148.1	Maintenance information: MD_MaintenanceInformation: contact	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 359-409).

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
149	Maintenance information: MD_ScopeDescription	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 150-155).
150	Maintenance information: MD_ScopeDescription: attributes	Maintenance	Manual- keyboard entry
			Automatic- harvesting (e.g. using WFS DescribeFeature operation)
151	Maintenance information: MD_ScopeDescription: features	Maintenance	Manual- keyboard entry
			Automatic- harvesting (e.g. using WFS GetFeature operation)
152	Maintenance information: MD_ScopeDescription: featureInstances	Maintenance	Manual- keyboard entry
			Automatic- harvesting (e.g. using WFS GetFeature operation)
153	Maintenance information: MD_ScopeDescription: attributeInstances	Maintenance	Manual- keyboard entry
			Automatic- harvesting (e.g. using WFS DescribeFeature operation)
154	Maintenance information: MD_ScopeDescription: dataset	Maintenance	Manual- keyboard entry
155	Maintenance information: MD_ScopeDescription: other	Maintenance	Manual- keyboard entry
156	Maintenance information: *MD_SpatialRepresentation*	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 156-185).
157	Spatial representation information: MD_GridSpatial Representation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 158-161).
158	Spatial representation information: MD_GridSpatial Representation: numberOfDimensions	Creation	Manual- keyboard entry
		Storage	Automatic- by simple automatic analysis*; Automatic for raster**
159	Spatial representation information: MD_GridSpatial Representation: axisDimensionsProperties	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 179-182).
160	Spatial representation information: MD_GridSpatial Representation: cellGeometry	Creation	Manual- select from an existing list
		Storage	Automatic- by simple automatic analysis*; Automatic for raster**
161	Spatial representation information: MD_GridSpatial Representation: transformationParameter Availability	Creation	Manual- select from an existing list
			Automatic for raster**
162	Spatial representation information: MD_Georectified	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 163-169 and 158-161).
163	Spatial representation information: MD_Georectified: checkPointAvailability	Creation	Manual- select from an existing list

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
164	Spatial representation information: MD_Georectified: checkPointDescription	Creation	Manual- keyboard entry
165	Spatial representation information: MD_Georectified: cornerPoints	Creation	Automatic- by simple automatic analysis*
166	Spatial representation information: MD_Georectified: centerPoint	Creation	Automatic- by simple automatic analysis*
167	Spatial representation information: MD_Georectified: pointInPixel	Creation	Automatic- by simple automatic analysis*
168	Spatial representation information: MD_Georectified: transformationDimension Description	Creation	Manual- keyboard entry
169	Spatial representation information: MD_Georectified: transformationDimension Mapping	Creation	Manual- keyboard entry
170	Spatial representation information: MD_Georeferenceable	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 171-175 and 158-161).
171	Spatial representation information: MD_Georeferenceable: controlPointAvailability	Creation	Manual- select from an existing list
172	Spatial representation information: MD_Georeferenceable: orientationParameterAvailability	Creation	Manual- select from an existing list
173	Spatial representation information: MD_Georeferenceable: orientationParameterDescription	Creation	Manual- keyboard entry
174	Spatial representation information: MD_Georeferenceable: georeferencedParameters	Creation	Manual- keyboard entry
175	Spatial representation information: MD_Georeferenceable: parameterCitation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 359-409).
176	Spatial representation information: MD_VectorSpatialRepresentation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 177-178).
177	Spatial representation information: MD_VectorSpatialRepresentation: topologyLevel	Creation	Manual- select from an existing list Automatic for vector**
178	Spatial representation information: MD_VectorSpatial Representation: geometricObjects	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 183-185).
179	Spatial representation information: MD_Dimension	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 180-182).
180	Spatial representation information: MD_Dimension: dimensionName	Creation	Manual- select from an existing list Automatic for raster**-extraction from dataset
181	Spatial representation information: MD_Dimension: dimensionSize	Creation	Manual- keyboard entry Automatic for raster**; Automatic- by simple automatic analysis*
182	Spatial representation information: MD_Dimension: resolution	Creation	Automatic for raster**; Automatic by simple automatic analysis*

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
183	Spatial representation information: MD_GeometricObjects	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 184-185).
184	Spatial representation information: MD_GeometricObjects: geometricObjectType	Creation	Manual- select from an existing list
			Automatic- by simple automatic analysis*; Automatic for vector**
185	Spatial representation information: MD_GeometricObjects: geometricObjectCount	Creation	Automatic- by simple automatic analysis*; Automatic for vector**
186	Reference system information: MD_ReferenceSystem	It follows its sub elements related step(s)	The inputs are provided by its sub element (line 187)
187	Reference system information: MD_ReferenceSystem: referenceSystemIdentifier	Collection / Creation	Manual- keyboard entry, or select from an existing list
		Storage	Automatic - extraction from dataset in the spatial database for vector; by simple automatic analysis for raster*
		Maintenance	
188	intentionally left blank		
189	Reference system information: MD_CRS	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 190-194 and 187).
190	Reference system information: MD_CRS: projection	Creation / Maintenance	Manual- keyboard entry, or select from an existing list
			Automatic for raster - by simple automatic analysis*, Automatic – extraction from dataset**** (e.g. ArcCatalog)
191	Reference system information: MD_CRS: ellipsoid	Creation	Manual- keyboard entry, or select from an existing list
			Automatic for raster - by simple automatic analysis*; Automatic for vector****
192	Reference system information: MD_CRS: datum	Creation	Manual- keyboard entry, or select from an existing list
			Automatic for raster ****- by simple automatic analysis*
193	Reference system information: MD_CRS: *role name*: ellipsoidParameters	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 201-204).
194	Reference system information: MD_CRS: *role name*: projectionParameters	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 215-231).
195	Reference system information: *RS_ReferenceSystem*	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 196-197).
196	Reference system information: *RS_ReferenceSystem:* name	Creation	Manual- keyboard entry
			Automatic - extraction from dataset in the spatial database for vector; by simple automatic analysis for raster*
197	Reference system information: *RS_ReferenceSystem:* domainOfValidity	Creation	Manual- select from an existing list
198	intentionally left blank		

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
199	intentionally left blank		
200	intentionally left blank		
201	Reference system information: MD_EllipsoidParameters	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 202-204).
202	Reference system information: MD_EllipsoidParameters: semiMajorAxis	Creation	Manual- keyboard entry
203	Reference system information: MD_EllipsoidParameters: axisUnits	Creation	Manual- select from an existing list
204	Reference system information: MD_EllipsoidParameters: denominatorOfFlatteningRatio	Creation	Manual- keyboard entry
205	Reference system information: MD_Identifier	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 206-207).
206	Reference system information: MD_Identifier: authority	Planning and policy making	Manual- keyboard entry, or select from an existing list
207	Reference system information: MD_Identifier: code	Planning and policy making	Automatic - extraction from dataset in the spatial database for vector; by simple automatic analysis for raster*
208	Reference system information: RS_Identifier	It follows its sub elements related step(s)	The inputs are provided by its sub elements
208.1	Reference system information: RS_Identifier: codeSpace	Planning and policy making	Manual- keyboard entry, or select from an existing list Automatic***
208.2	Reference system information: RS_Identifier: version	Planning and policy making	Manual- keyboard entry, or select from an existing list Automatic***
209	Reference system information: MD_ObliqueLineAzimuth	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 210-211).
210	Reference system information: MD_ObliqueLineAzimuth: azimuthAngle	Creation	Manual- keyboard entry, or select from an existing list
211	Reference system information: MD_ObliqueLineAzimuth: azimuthMeasurePointLongitude	Creation	Manual- keyboard entry, or select from an existing list
212	Reference system information: MD_ObliqueLinePoint	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 213-214).
213	Reference system information: MD_ObliqueLinePoint: obliqueLineLatitude	Creation	Manual- keyboard entry, or select from an existing list
214	Reference system information: MD_ObliqueLinePoint: obliqueLineLongitude	Creation	Manual- keyboard entry, or select from an existing list
215	Reference system information: MD_ProjectionParameters	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 216-231).
216	Reference system information: MD_ProjectionParameters: zone	Creation	Manual- keyboard entry, or select from an existing list
217	Reference system information: MD_ProjectionParameters: standardParallel	Creation	Manual- keyboard entry, or select from an existing list

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
218	Reference system information: MD_ProjectionParameters: longitudeOfCentralMeridian	Creation	Manual- keyboard entry, or select from an existing list
219	Reference system information: MD_ProjectionParameters: latitudeOfProjectionOrigin	Creation	Manual- keyboard entry, or select from an existing list
220	Reference system information: MD_ProjectionParameters: falseEasting	Creation	Manual- keyboard entry, or select from an existing list
221	Reference system information: MD_ProjectionParameters: falseNorthing	Creation	Manual- keyboard entry, or select from an existing list
222	Reference system information: MD_ProjectionParameters: falseEastingNorthingUnits	Creation	Manual- keyboard entry, or select from an existing list
223	Reference system information: MD_ProjectionParameters: scaleFactorAtEquator	Creation	Manual- keyboard entry
224	Reference system information: MD_ProjectionParameters: heightOfProspectivePoint AboveSurface	Creation	Manual- keyboard entry
225	Reference system information: MD_ProjectionParameters: longitudeOfProjectionCenter	Creation	Manual- keyboard entry
226	Reference system information: MD_ProjectionParameters: latitudeOfProjectionCenter	Creation	Manual- keyboard entry
227	Reference system information: MD_ProjectionParameters: scaleFactorAtCenterLine	Creation	Manual- keyboard entry
228	Reference system information: MD_ProjectionParameters: straightVerticalLongitudeFrom Pole	Creation	Manual- keyboard entry
229	Reference system information: MD_ProjectionParameters: scaleFactorAtProjectionOrigin	Creation	Manual- keyboard entry
230	Reference system information: MD_ProjectionParameters: *role name*: ObliqueLineAzimuthParameter	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 209-211).
231	Reference system information: MD_ProjectionParameters: *role name*: ObliqueLinePointParameter	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 212-214).
232	Content information: *MD_ContentInformation*	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 232-267).
233	Content information: MD_FeatureCatalogue Description	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 234-238).
234	Content information: MD_FeatureCatalogue Description: complianceCode	Creation	Manual- select from an existing list
			Automatic-extraction/computation/inference **
235	Content information: MD_FeatureCatalogue Description: language	Planning and policy making	Manual- select from an existing list
			Automatic-extraction/inference **

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
236	Content information: MD_FeatureCatalogue Description: includedWithDataset	Creation	Manual- select from an existing list
			Automatic- extraction**
237	Content information: MD_FeatureCatalogue Description: featureTypes	Creation	Automatic- extraction** (e.g. using WFS getFeature operation)
238	Content information: MD_FeatureCatalogue Description: featureCatalogueCitation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 359-409).
239	Content information: MD_CoverageDescription	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 240-242).
240	Content information: MD_CoverageDescription: attributeDescription	Creation	Manual- keyboard entry
241	Content information: MD_CoverageDescription: contentType	Creation	Manual- select from an existing list
			Automatic- extraction/inference**
242	Content information: MD_CoverageDescription: *Role name*: dimension	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 256-267).
243	Content information: MD_ImageDescription	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 244-255 and 240-242).
244	Content information: MD_ImageDescription: illuminationElevationAngle	Collection	Manual- keyboard entry
245	Content information: MD_ImageDescription: illuminationAzimuthAngle	Collection	Manual- keyboard entry
246	Content information: MD_ImageDescription: imagingCondition	Collection	Manual- select from an existing list
247	Content information: MD_ImageDescription: imageQualityCode	Collection	Manual- select from an existing list
248	Content information: MD_ImageDescription: cloudCoverPercentage	Collection	Manual- keyboard entry
249	Content information: MD_ImageDescription: processingLevelCode	Collection	Manual- select from an existing list
250	Content information: MD_ImageDescription: compressionGeneration Quantity	Collection	Manual- keyboard entry
251	Content information: MD_ImageDescription: triangulationIndicator	Collection	Manual- select from an existing list
252	Content information: MD_ImageDescription: radiometricCalibrationDataAvailability	Collection	Manual- select from an existing list
253	Content information: MD_ImageDescription: cameraCalibrationInformationAvailability	Collection	Manual- select from an existing list

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
254	Content information: MD_ImageDescription: filmDistortionInformation Availability	Collection	Manual- select from an existing list
255	Content information: MD_ImageDescription: lensDistortionInformation Availability	Collection	Manual- select from an existing list
256	Content information: MD_RangeDimension	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 257-258).
257	Content information: MD_RangeDimension: sequenceIdentifier	Collection	Automatic for raster**- extracting band number***
258	Content information: MD_RangeDimension: descriptor	Collection	Manual- keyboard entry Automatic for raster**
259	Content information: MD_Band	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 260-267 and 257-258).
260	Content information: MD_Band: maxValue	Creation	Automatic for raster**- extracting pixel or cell maximum value***
261	Content information: MD_Band: minValue	Creation	Automatic for raster**- extracting pixel or cell minimum value***
262	Content information: MD_Band: units	Creation	Manual- select from an existing list Automatic for raster**
263	Content information: MD_Band: peakResponse	Creation	Manual- keyboard entry
264	Content information: MD_Band: bitsPerValue	Creation	Automatic for raster**- extracting number of bits used to encode band***
265	Content information: MD_Band: toneGradation	Creation	Manual- keyboard entry
266	Content information: MD_Band: scaleFactor	Creation	Manual- keyboard entry
267	Content information: MD_Band: offset	Creation	Manual- keyboard entry
268	Content information: MD_PortrayalCatalogue Reference:	It follows its sub elements related step(s).	The inputs are provided by its sub element (line 269).
269	Content information: MD_PortrayalCatalogue Reference: portrayalCatalogueCitation	Creation	Manual- keyboard entry
270	Distribution information: MD_Distribution	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 271-273).
271	Distribution information: MD_Distribution: *Role name*: distributionFormat	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 284-290).
272	Distribution information: MD_Distribution: *Role name*: distributor	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 279-283).
273	Distribution information: MD_Distribution: *Role name*: transferOptions	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 274-278).
274	Distribution information: MD_DigitalTransferOptions	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 275-278).

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
275	Distribution information: MD_DigitalTransferOptions: unitsOfDistribution	Publication	Manual- keyboard entry Automatic- extraction/computation/inference **
276	Distribution information: MD_DigitalTransferOptions: transferSize	Publication	Automatic- extraction** Manual- keyboard entry
277	Distribution information: MD_DigitalTransferOptions: onLine	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 396-402).
278	Distribution information: MD_DigitalTransferOptions: offLine	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 291-297).
279	Distribution information: MD_Distributor	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 280-283).
280	Distribution information: MD_Distributor: distributorContact	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 359-409).
281	Distribution information: MD_Distributor: *Role name*: dstributionOrderProcess	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 298-302).
282	Distribution information: MD_Distributor: *Role name*: dstributorFormat	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 284-290).
283	Distribution information: MD_Distributor: *Role name*: dstributorTransferOptions	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 274-278).
284	Distribution information: MD_Format	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 285-290).
285	Distribution information: MD_Format: name	Storage/ Publication	Manual- keyboard entry Automatic- extraction**
286	Distribution information: MD_Format: version	Storage/ Publication	Manual- keyboard entry Automatic- extraction**
287	Distribution information: MD_Format: amendmentNumber	Storage/ Publication	Manual- keyboard entry
288	Distribution information: MD_Format: specification	Storage/ Publication	Manual- keyboard entry
289	Distribution information: MD_Format: fileDecompressionTechnique	Storage/ Publication	Automatic- extraction/inference** Manual- select from an existing list
290	Distribution information: MD_Format: *Role name*: formatDistributor	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 279-283).
291	Distribution information: MD_Medium	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 292-297).
292	Distribution information: MD_Medium: name	Publication	Manual- select from an existing list
293	Distribution information: MD_Medium: density	Publication	Manual- keyboard entry
294	Distribution information: MD_Medium: densityUnits	Publication	Manual- keyboard entry
295	Distribution information: MD_Medium: volumes	Publication	Manual- keyboard entry

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
296	Distribution information: MD_Medium: mediumFormat	Publication	Manual- select from an existing list
297	Distribution information: MD_Medium: mediumNote	Publication	Manual- keyboard entry
298	Distribution information: MD_StandardOrderProcess	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 299-302).
299	Distribution information: MD_StandardOrderProcess: fees	Planning and policy making	Manual- keyboard entry
300	Distribution information: MD_StandardOrderProcess: plannedAvailableDateTime	Planning and policy making	Manual- keyboard entry, or select from an existing calendar
301	Distribution information: MD_StandardOrderProcess: orderingInstructions	Planning and policy making	Manual- keyboard entry
302	Distribution information: MD_StandardOrderProcess: turnaround	Planning and policy making	Manual- keyboard entry
303	Metadata extension information: MD_MetadataExtension Information	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 304-305).
304	Metadata extension information: MD_MetadataExtension Information: extensionOnLineResource	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 396-402).
305	Metadata extension information: MD_MetadataExtension Information: *Role name:* extendedElementInformation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 306-319).
306	Metadata extension information: MD_ExtendedElemen tInformation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 307-319).
307	Metadata extension information: MD_ExtendedElemen tInformation: name	Planning and policy making	Manual- keyboard entry
308	Metadata extension information: MD_ExtendedElemen tInformation: shortName	Planning and policy making	Manual- keyboard entry
309	Metadata extension information: MD_ExtendedElemen tInformation: domainCode	Planning and policy making	Manual- keyboard entry
310	Metadata extension information: MD_ExtendedElemen tInformation: definition	Planning and policy making	Manual- keyboard entry
311	Metadata extension information: MD_ExtendedElemen tInformation: obligation	Planning and policy making	Manual- select from an existing list
312	Metadata extension information: MD_ExtendedElemen tInformation: condition	Planning and policy making	Manual- keyboard entry
313	Metadata extension information: MD_ExtendedElemen tInformation: dataType	Planning and policy making	Manual- keyboard entry
314	Metadata extension information: MD_ExtendedElemen tInformation: maximumOccurrence	Planning and policy making	Manual- keyboard entry
315	Metadata extension information: MD_ExtendedElemen tInformation: domainValue	Planning and policy making	Manual- keyboard entry
316	Metadata extension information: MD_ExtendedElemen tInformation: parentEntity	Planning and policy making	Manual- keyboard entry

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
317	Metadata extension information: MD_ExtendedElemen tInformation: rule	Planning and policy making	Manual- keyboard entry
318	Metadata extension information: MD_ExtendedElemen tInformation: rationale	Planning and policy making	Manual- keyboard entry
319	Metadata extension information: MD_ExtendedElemen tInformation: source	Planning and policy making	Manual- keyboard entry
320	Application schema information: MD_ApplicationSchema Information	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 321-327).
321	Application schema information: MD_ApplicationSchema Information: name	Creation	Manual- keyboard entry
322	Application schema information: MD_ApplicationSchema Information: schemaLanguage	Creation	Automatic- by simple automatic analysis*
323	Application schema information: MD_ApplicationSchema Information: constraintLanguage	Creation	Automatic- by simple automatic analysis*
324	Application schema information: MD_ApplicationSchemaInformation: schemaAscii	Creation	Automatic- by simple automatic analysis*
325	Application schema information: MD_ApplicationSchema Information: graphicsFile	Creation	Automatic- by simple automatic analysis*
326	Application schema information: MD_ApplicationSchema Information: softwareDevelopmentFile	Creation	Automatic- by simple automatic analysis*
327	Application schema information: MD_ApplicationSchemaInformation: softwareDevelopmentFileFormat	Creation	Automatic- by simple automatic analysis*
328	intentionally left blank		
329	intentionally left blank		
330	intentionally left blank		
331	intentionally left blank		
332	intentionally left blank		
333	intentionally left blank		
334	Data type information: EX_Extent	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 335-338).
335	Data type information: EX_Extent: description	Creation	Manual- keyboard entry
336	Data type information: EX_Extent: *Role name*: geographicElement	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 339-349).
337	Data type information: EX_Extent: *Role name*: temporalElement	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 350-353).
338	Data type information: EX_Extent: *Role name*: verticalElement	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 354-358).
339	Data type information: *EX_GeographicExtent*	It follows its sub elements related step(s).	The inputs are provided by its sub element (line 340)
340	Data type information: *EX_GeographicExtent:* extentTypeCode	Creation	Manual- select from an existing list

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
341	Data type information: EX_BoundingPolygon	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 342 and 340).
342	Data type information: EX_BoundingPolygon: polygon	Creation	Manual- keyboard entry, or draw a rectangle
		Storage	Computation based on dataset geometry column
343	Data type information: EX_GeographicBoundingBox	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 344-347 and 340).
344	Data type information: EX_GeographicBoundingBox: westBoundLongitude	Creation	Manual- keyboard entry, or draw a rectangle
		Storage / Maintenance	Automatic- computation*** (e.g. ArcCatalog)
345	Data type information: EX_GeographicBoundingBox: eastBoundLongitude	Creation	Manual- keyboard entry, or draw a rectangle
		Storage / Maintenance	Automatic- computation*** (e.g. ArcCatalog)
346	Data type information: EX_GeographicBoundingBox: southBoundLatitude	Creation	Manual- keyboard entry, or draw a rectangle
		Storage / Maintenance	Automatic- computation*** (e.g. ArcCatalog)
347	Data type information: EX_GeographicBoundingBox: northBoundLatitude	Creation	Manual- keyboard entry, or draw a rectangle
		Storage / Maintenance	Automatic- computation*** (e.g. ArcCatalog)
348	Data type information: EX_GeographicDescription	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 349 and 340).
349	Data type information: EX_GeographicDescription: geographicIdentifier	Creation / Maintenance	Manual – select from an existing list
350	Data type information: EX_TemporalExtent	It follows its sub elements related step(s).	The inputs are provided by its sub element (line 351).
351	Data type information: EX_TemporalExtent: extent	Creation	Manual- keyboard entry, or select from an existing calendar
		Maintenance	Automatic- generated by the system whenever the dataset is modified (e.g. ArcCatalog).
352	Data type information: EX_SpatialTemporalExtent	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 353 and 351).
353	Data type information: EX_SpatialTemporalExtent: *role name:* SpatialExtent	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 339-349)
354	Data type information: EX_VerticalExtent	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 355-358).
355	Data type information: EX_VerticalExtent: minimumValue	Creation	Manual- keyboard entry
		Storage / Maintenance	Automatic- computation***
356	Data type information: EX_VerticalExtent: maximumValue	Creation	Manual- keyboard entry
		Storage / Maintenance	Automatic- computation***

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
357	Data type information: EX_VerticalExtent: unitOfMeasure	Creation	Automatic- predefined value
			Manual- select from an existing list
358	Data type information: EX_VerticalExtent: *role name*: verticalDatum	Creation	Manual- select from an existing list
359	Citation and responsible party information: CI_Citation	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 360-373).
360	Citation and responsible party information: CI_Citation: title	Creation	Manual – keyboard entry
		Storage	Automatic- extraction**
361	Citation and responsible party information: CI_Citation: alternateTitle	Creation / Storage	Manual – keyboard entry
362	Citation and responsible party information: CI_Citation: date	Creation	Manual – select from existing calendar
363	Citation and responsible party information: CI_Citation: edition	Maintenance	Manual – keyboard entry
364	Citation and responsible party information: CI_Citation: editionDate	Maintenance	Manual – select from existing calendar
			Automatic- computation by the system whenever the dataset is modified.
365	Citation and responsible party information: CI_Citation: identifier	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 205-208.2).
366	intentionally left blank		
367	Citation and responsible party information: CI_Citation: citedResponsibleParty	Planning and policy making	Manual- keyboard entry, or select from an existing list
368	Citation and responsible party information: CI_Citation: presentationForm	Creation	Manual- select from an existing list
369	Citation and responsible party information: CI_Citation: series	Creation	Manual- keyboard entry
370	Citation and responsible party information: CI_Citation: otherCitationDetails	Creation	Manual- keyboard entry
371	Citation and responsible party information: CI_Citation: collectiveTitle	Creation	Manual- keyboard entry
372	Citation and responsible party information: CI_Citation: ISBN	Creation	Manual- keyboard entry
373	Citation and responsible party information: CI_Citation: ISSN	Creation	Manual- keyboard entry
374	Citation and responsible party information: CI_ResponsibleParty	It follows its sub elements related step(s).	The inputs are provided by its sub elements (375-379).
375	Citation and responsible party information: CI_ResponsibleParty: individualName	Planning and policy making	Manual- select from an existing list
			Automatic- predefined value (e.g. ANZMet Lite)
376	Citation and responsible party information: CI_ResponsibleParty: organisationName	Planning and policy making	Manual- select from an existing list
			Automatic- predefined value (e.g. ANZMet Lite)

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
377	Citation and responsible party information: CI_ResponsibleParty: positionName	Planning and policy making	Manual- select from an existing list
			Automatic- predefined value (e.g. ANZMet Lite)
378	Citation and responsible party information: CI_ResponsibleParty: contactInfo	It follows its sub elements related step(s).	The inputs are provided by its sub elements (387-392).
379	Citation and responsible party information: CI_ResponsibleParty: role	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- select from an existing list
380	Citation and responsible party information: CI_Address	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 388-392).
381	Citation and responsible party information: CI_Address: deliveryPoint	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
382	Citation and responsible party information: CI_Address: city	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
383	Citation and responsible party information: CI_Address: administrativeArea	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
384	Citation and responsible party information: CI_Address: postalCode	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
385	Citation and responsible party information: CI_Address: country	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
386	Citation and responsible party information: CI_Address: electronicMailAddress	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
387	Citation and responsible party information: CI_Contact	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 388-402).
388	Citation and responsible party information: CI_Contact: phone	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
389	Citation and responsible party information: CI_Contact: address	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list
390	Citation and responsible party information: CI_Contact: onlineResource	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite)
			Manual- keyboard entry, or select from an existing list

Table 6-I: Mapping ISO 19115: 2003 metadata element against the spatial data lifecycle steps

ISO element No.	Element	Proposed Spatial Data Lifecycle Related Step(s)	Proposed Creation Method
391	Citation and responsible party information: CI_Contact: hoursOfService	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite) Manual- keyboard entry, or select from an existing list
392	Citation and responsible party information: CI_Contact: contactInstructions	Planning and policy making	Automatic- predefined value (e.g. ANZMet Lite) Manual- keyboard entry, or select from an existing list
393	Citation and responsible party information: CI_Date	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 394-395).
394	Citation and responsible party information: CI_Date: date	Creation / Maintenance / Publication	Manual – keyboard entry, or select from an existing calendar Automatic- Extraction**
395	Citation and responsible party information: CI_Date: dateType	Creation / Maintenance / Publication	Manual – select from an existing list Automatic- Extraction/inference **
396	Citation and responsible party information: CI_OnlineResource	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 397-402).
397	Citation and responsible party information: CI_OnlineResource: linkage	Publication / Maintenance	Manual – keyboard entry
398	Citation and responsible party information: CI_OnlineResource: protocol	Publication / Maintenance	Manual – keyboard entry
399	Citation and responsible party information: CI_OnlineResource: applicationProfile	Publication	Manual – keyboard entry
400	Citation and responsible party information: CI_OnlineResource: name	Publication	Manual – keyboard entry
401	Citation and responsible party information: CI_OnlineResource: description	Publication	Manual – keyboard entry
402	Citation and responsible party information: CI_OnlineResource: function	Publication	Manual – keyboard entry
403	Citation and responsible party information: CI_Series	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 404-405).
404	Citation and responsible party information: CI_Series: name	Creation	Manual – keyboard entry
405	Citation and responsible party information: CI_Series: issueIdentification	Creation	Manual – keyboard entry
406	Citation and responsible party information: CI_Series: page	Creation	Manual – keyboard entry
407	Citation and responsible party information: CI_Telephone	It follows its sub elements related step(s).	The inputs are provided by its sub elements (lines 408-409).
408	Citation and responsible party information: CI_Telephone: voice	Planning and policy making	Manual – keyboard entry Automatic- predefined value (e.g. ANZMet Lite)
409	Citation and responsible party information: CI_Telephone: facsimile	Planning and policy making	Manual – keyboard entry Automatic- predefined value (e.g. ANZMet Lite)

*According to Taussi (2007)

** According to Manso-Callejo *et al.* (2010)
*** According to Manso-Callejo *et al.* (2009)
**** According to Manso *et al.* (2004)

APPENDIX 8

SUPPLEMENTARY INFORMATION RELATING TO PROTOTYPE SYSTEMS IMPLEMENTATION (CHAPTER 7)

APPENDIX 8 - SUPPLEMENTARY INFORMATION RELATING TO PROTOTYPE SYSTEMS IMPLEMENTATION (CHAPTER 7)

Steps for posting CSW update request from deegree to GeoNetwork:

1. Posting deegree CSW GetRecordById request, as following:

```
URL: http://IP:Port/deegree-csw/services
Post data:
<?xml version="1.0" encoding="UTF-8"?>
<csw:GetRecordById service="CSW" version="2.0.2" outputFormat="application/xml"
outputSchema="http://www.opengis.net/cat/csw/2.0.2"
xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/cat/csw/2.0.2
http://schemas.opengis.net/csw/2.0.2/CSW-discovery.xsd">
<csw:Id>bf7ab961-a69f-48e9-a6da-61c207de86d6</csw:Id>
<csw:ElementSetName>full</csw:ElementSetName>
</csw:GetRecordById>
```

2. Getting the response from deegree
3. Copying the whole lot between <MD_Metadata> and </MD_Metadata> from the above response and inserting them inside the following GeoNetwork CSW update request:

```
URL: http://IP:Port/geonetwork/srv/en/csw
Mime-type: application/xml
Post data:
<?xml version="1.0" encoding="UTF-8"?>
<csw:Transaction service="CSW" version="2.0.2"
xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
xmlns:ogc="http://www.opengis.net/ogc" xmlns:gmd="http://www.opengis.net/gmd"
xmlns:apiso="http://www.opengis.net/cat/csw/apiso/1.0">
  <csw:Update>
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:gml="http://www.opengis.net/gml" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:geonet="http://www.fao.org/geonetwork">
----------------------------------------------------- INSERT HERE------------------------------------------
------
  </gmd:MD_Metadata>
<csw:Constraint version="1.0.0">
        <Filter xmlns="http://www.opengis.net/ogc"
xmlns:gml="http://www.opengis.net/gml">
          <PropertyIsLike wildCard="%" singleChar="_" escapeChar="/">
            <PropertyName>apiso:identifier</PropertyName>
            <Literal>bf7ab961-a69f-48e9-a6da-61c207de86d6</Literal>
          </PropertyIsLike>
        </Filter>
      </csw:Constraint>
  </csw:Update>
</csw:Transaction>
```

4. Posting the above CSW update request to the GeoNetwork.

Steps for posting CSW update request from GeoNetwork to deegree:

1. Posting GeoNetwork CSW GetRecordById request, as following:

```
Url:http://IP:Port/geonetwork/srv/en/csw
Mime-type: application/xml
Post data:
<?xml version="1.0"?>
<csw:GetRecordById xmlns:csw="http://www.opengis.net/cat/csw/2.0.2" service="CSW"
version="2.0.2"
   outputSchema="csw:IsoRecord">
   <csw:Id>bf7ab961-a69f-48e9-a6da-61c207de86d6</csw:Id>
<csw:ElementSetName>full</csw:ElementSetName>
</csw:GetRecordById>
<?xml version="1.0"?>
<csw:GetRecordById xmlns:csw="http://www.opengis.net/cat/csw/2.0.2" service="CSW"
version="2.0.2"
   outputSchema="csw:IsoRecord">
   <csw:Id>bf7ab961-a69f-48e9-a6da-61c207de86d6</csw:Id>
<csw:ElementSetName>full</csw:ElementSetName>
</csw:GetRecordById>
```

2. Getting the response from GeoNetwork

3. Copying the whole lot between <MD_Metadata> and </MD_Metadata> from the above response and inserting them inside the following deegree CSW Update Request:

```
Url: http://IP:Port/deegree-csw/services
<?xml version="1.0" encoding="UTF-8" ?>
 <csw:Transaction service="CSW" version="2.0.2"
xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
xmlns:ogc="http://www.opengis.net/ogc"
xmlns:apiso="http://www.opengis.net/cat/csw/apiso/1.0">
 <csw:Update>
 <MD_Metadata xmlns="http://www.isotc211.org/2005/gmd"
xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:gml="http://www.opengis.net/gml"
xmlns:xlink="http://www.w3.org/1999/xlink">
-----------------------------------------------------INSERT HERE-----------------------------------------
-------
 </MD_Metadata>
 <csw:Constraint version="1.1.0">
 <ogc:Filter>
 <ogc:PropertyIsLike wildCard="%" singleChar="_" escapeChar="/">
  <ogc:PropertyName>apiso:identifier</ogc:PropertyName>
  <ogc:Literal>bf7ab961-a69f-48e9-a6da-61c207de86d6</ogc:Literal>
  </ogc:PropertyIsLike>
  </ogc:Filter>
  </csw:Constraint>
  </csw:Update>
  </csw:Transaction>
```

4. Posting the above CSW Update request to the deegree-CSW.

During the test of the above task for updating metadata record through posting CSW update request to deegree, a critical issue was recognised which might affect the whole process of prototype system implementation. Based on the version of deegree-CSW at this stage of research, the CSW update request was a ‘delete’ operation followed by an ‘insert’. Obviously, this issue resulted in changing the ‘metadata_id’ value in ‘md_metadata’ table and loosing the connection between dataset and metadata whenever CSW update request was posted to deegree.

In order to address this issue, a trigger was scripted in PostGIS for table ‘md_metadata’ to keep the value of ‘fk_md_metadata’ column (foreign key to dataset) in dataset table current with the value of ‘id’ column in ‘md_metadata’ table. This trigger shown below was developed based on ‘metadata file identifier’ value which is a permanent and unique value for any metadata record stored in PostgreSQL.

```
-- Function: fk_md_metadata_update()
-- DROP FUNCTION fk_md_metadata_update();
CREATE OR REPLACE FUNCTION fk_md_metadata_update()
RETURNS trigger AS
$BODY$
DECLARE
id2 integer;
BEGIN
IF TG_RELNAME='md_metadata' THEN
SELECT md_metadata.id INTO id2 FROM md_metadata, fileidentifier WHERE
md_metadata.fk_fileidentifier = fileidentifier.id AND fileidentifier.fileidentifier LIKE
'3e83bbd2-67f1-49d5-aec5-8eb6b4ab312d';
UPDATE boresg SET fk_md_metadata=id2;
END IF;
RETURN NULL;
END;
$BODY$
LANGUAGE plpgsql VOLATILE
 COST 100;
ALTER FUNCTION fk_md_metadata_update() OWNER TO postgres;
```

The trigger developed in PostgreSQL to generate the most current bounding box of the 'boresg' dataset

```
-- Function: boresg_md_propagate_bbox()
-- DROP FUNCTION boresg_md_propagate_bbox();
CREATE OR REPLACE FUNCTION boresg_md_propagate_bbox()
  RETURNS trigger AS
$BODY$DECLARE
tg geometry;
id1 integer;
BEGIN
IF TG_RELNAME='boresg' THEN
SELECT md_metadata.id INTO id1 FROM md_metadata, fileidentifier WHERE
md_metadata.fk_fileidentifier = fileidentifier.id AND fileidentifier.fileidentifier LIKE
'metadata file identifier';
SELECT SetSRID(ST_Extent(the_geom),4326) INTO tg FROM boresg;
UPDATE ex_geogrbbox set geom=tg WHERE id= id1;
END IF;
RETURN NULL;
END;
$BODY$
  LANGUAGE plpgsql VOLATILE
  COST 100;
ALTER FUNCTION boresg_md_propagate_bbox() OWNER TO postgres;
```