The Integration of Multi-source Spatial Datasets in the Context of SDI Initiatives

by

Hossein Mohammadi

December 2008

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

Centre for Spatial Data Infrastructures and Land Administration
Department of Geomatics, Faculty of Engineering
The University of Melbourne,
Melbourne, Victoria 3010, Australia

end of season
last peach in the bowl -
its scent

Angela Smith

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 refereed conference proceedings and journals as listed in Appendix 1.

Hossein Mohammadi

to Mehrnoosh

with love

Abstract

The integration of multi-source spatial data is critical to the delivery of the objectives of spatial services. The demand is also growing dramatically among decision-makers to utilise more effective solutions that are highly dependent on the information. Therefore, spatial information plays a key role in informative spatial decision making. However, it is often difficult and sometimes impossible for users to sensibly integrate datasets from different sources. This is because of the diversity of data standards, specifications and arrangements that are utilised by organisations. Data providers adopt spatial data standards and specifications and establish a data-sharing arrangement based on their requirements, which may differ from those of other organisations. Therefore, multi-source spatial datasets are associated with technical and non-technical inconsistency and heterogeneity.

Spatial Data Infrastructures (SDIs) provide holistic frameworks with a number of technological components that address and overcome the issues and challenges of data integration. SDIs comprise spatial data, people, access network, policies and standards that aim to facilitate the integration of multi-source spatial datasets. As a result, spatial data integration can be investigated and necessary enablers and tools can be developed in the context of SDIs.

The research strategy is designed in such a way as to meet the objectives of the research, namely 'investigation of the potential technical and non-technical barriers to integrate multi-source datasets within SDI initiatives, identification of enablers for effective data integration and design and development of the key tool components'.

To establish the theoretical background of the research, an extensive literature review has been undertaken. The literature review covers the specific areas of spatial data integration, SDIs as sharing platforms, multi-source spatial datasets, data integration and potential barriers and enablers to data integration. The concepts, missions and components of SDI, together with investigating the significance and applications of integration, have provided an understanding of the integration of multi-source spatial datasets and their requirements. The literature review has also informed the research strategy and has been utilised in case study investigation. In order to achieve a better understanding of the issues of and impediments to spatial data integration a number of international and Australian case studies have been conducted. The international case studies have been selected on a voluntarily basis through the support and with the permission of Permanent Committee for GIS Infrastructure in Asia and the Pacific Region-Working Group 3 (PCGIAP-WG3). Seven case study countries in the Asia-Pacific region provided country reports. The Australian case studies included four local councils, two from each state of Victoria and NSW in Australia. Australian federal and state agencies have also provided different datasets for the case study areas. A series of technical visits have also been made in respective of mapping agencies. The analysis of the case studies has identified different technical and non-technical issues that are associated with data integration of five major groups of technical, institutional, policy, legal and social issues.

This analysis of the challenges and issues of data integration has also identified a number of key enablers and tools to overcome the issues. Based on the findings of

the case studies, the research has proposed a data integration toolbox with a number of components including data integration and validation tools, spatial data integration guidelines, integration data model, integration metadata and data specification documents. These components can not only assist practitioners to investigate and identify the issues and problems that are associated with multi-source data integration, but also help in the integration of multi-source datasets.

The design approach follows along with discussion on the development of the toolbox components. The integration of data integration and validation tools has been the major focus of the research. The integration and validation tools are a suite of computer applications that assess and validate the readiness of different datasets against a number of measures and rules. The measures and rules can be defined by SDI administrators or data providers. They include measures, rules and restrictions on spatial and aspatial content of data such as attribute values, geographical extent, quality of data and any policies or restrictions bound to datasets. These tools assess the data and metadata content of data for any possible incompliancy with the measures. The tools then provide a report on the assessment process and a set of technical tools to manipulate the datasets. The tools have been tested through the use test approach with the involvement of real datasets. The use test has shown that the tools provide a customisable and consistent approach which saves time and cost for the assessment and integration of datasets.

The tool can even be employed by users to investigate their own datasets. The users can set rules and define measures in forms of queries or comparing criteria and assess data against these measures and rules.

The data integration guidelines propose a comprehensive document that details the major integration activities. The guideline also outlines a methodology for data evaluation and integration, potential technical and non-technical barriers to spatial data integration, jurisdiction-specific considerations for spatial data integration and possible and available solutions for data integration barriers. The integration data model presents a data model based on the meanings of the geographical phenomena utilising ontologies. It then presents an ontology-based reclassification approach that can be used in the design of an integration data model. Structured and machine-readable metadata and spatial data specification documents have also been proposed by the research. Metadata and data specifications, which contain a conceptual description of features, logical connection between different features and also the constraints that exist between spatial features, will allow automatic information extraction which will be a significant step towards the effective integration of multi-source heterogeneous spatial datasets.

Acknowledgements

It is a pleasure to thank the many people who made this thesis possible.

It is difficult to overstate my gratitude to my supervisor, Assoc. Prof. Abbas Rajabifard. With his enthusiasm and his inspiration, he helped to make the research journey fun for me. Throughout my research period, he provided encouragement, sound advice, good teaching, good company, and lots of good ideas. I would have been lost without him.

I am deeply grateful to my supervisor, Professor Ian Williamson for his detailed and constructive comments, and for his important support throughout this work. His understanding, encouraging and personal guidance have provided a good basis for the present thesis.

During this work, I have collaborated with many members of the Centre for SDIs and Land Administration, for whom I have great regard, and I wish to extend my warmest thanks to all those who have helped me with my work and for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to Andrew Binns, Dr. Saeid Mohsen Kalantari, Shilan Vaez, Dr. Kevin McDougall, Dr. Kate Darlymple, Dr. Payam Ghadirian, Dr. Christine Najar, Dr. Rohan Bennet, Peter Holland, Michael Black, Pual Box, and to Dr. Christian Claussen, Lukasz Grus, Mathew Warnest and Faisal El-Qureshi.

I warmly thank Mrs. Jude Wallace, for her valuable advice and friendly help. Her extensive discussions around my work have been very helpful for this study.

Thanks are also due to the visiting staff to the Department who provided their invaluable advice and research perspectives, especially Professor Stig Enemark, Professor Ian Masser, Professor Harlan Onsrud, Dr. Daniel Stuedler, Assoc. Prof. Milan Konecny, Professor Michael Goodchild; and Mr. Mathew Kistler.

My research would of course not been possible without the support of ARC-Linkage Program and also the valued contributions from the many people in Spatial Information Infrastructure and LandVictoria (Victoria), Departments of Lands and CANRI (NSW), Geoscience Australia and PSMA (ACT) who made this thesis possible.

Gratitude is also due to all the PCGIAP-WG3 members, who have shared their experience with me and made this thesis possible.

I am indebted to my parents, Ata and Azra. They bore me, raised me, supported me, taught me with love, and loved me. No one ever had better parents. My special gratitude is due to my brother, Ehsan, my sister, Sarah and their partners for their loving support.

I would also like to thank my father-in-law and mother-in-law, Ali and Ezat, and my brother-in-laws, Hossein and Mehrdad for their enthusiastic support in my research.

Lastly, and most importantly, I owe my loving thanks to my lovely and beautiful wife, Mehrnoosh. She has lost a lot due to my research. Without her encouragement and understanding, it would have been impossible for me to finish this work. She gave me not only her moral support, but also helped me a lot in doing research.

Hossein Mohammadi

December 2008

Melbourne, Australia

Table of Contents

List of Figures

List of Tables

List of Acronyms

ABS	Australian Bureau of Statistics
AGSDI	Australian Government's Spatial Data Integration
ANZLIC	Australia and New Zealand Land Information Council
ASDD	Australian Spatial Data Directory
ASDI	Australian Spatial Data Infrastructure
BOSSI	Board of Surveying and Spatial Information
BSWM	Bureau of Soils and Water Management
CANRI	Community Access to Natural Resource Information
CGDI	Canadian Geospatial Data Infrastructure
CGDL	Cooperative Geographic Data Library
CI	Catalog Interoperability
CRC.SI	Cooperative Research Centre.Spatial Information
CSDGM	Content Standard for Digital Geospatial Metadata
DAFF	Department of Agriculture, Fishery and Forestry
DAO	Data Access Object
DBMS	DataBase Management System
DCDB	Digital Cadastre DataBase
DCMI	Dublin Metadata Core Initiative
DIF	Directory Interchange Format
DIPNR	Department of Infrastructure, Planning and Natural Resources
DOST	Department of Science and Technology
DPI	Department of Primary Industries
DPWH	Department of Public Works and Highways
DSE	Department of Sustainability and Environment
DTD	Document Type Definition
EICU	Emergency Information Coordination Unit
ETL	Extract, Transform and Load
EMA	Emergency Management of Australia
ESO	Emergency Service Organisation
ESRI	Environmental Systems Research Institute
EUROGI	European Umbrella Organisation for Geographic Information
FGDC	Federal Geographic Data Committee
FIG	Fédération Internationale de Gymnastique (International Federation of Surveyors)
FME	Feature Manipulation Engine
GA	Geoscience Australia
GCMD	Global Change Master Directory
GDA94	Geographic Datum of Australia 94
GEMD	Geospatial and Earth Monitoring Division
GIS	Geospatial Information System
G-NAF	Geocoded National Address File
GSDI	Global Spatial Data Infrastructure
GURAS	Geo-coded Urban and Rural Addressing System
HLURB	Housing and Land Use Regulatory Board
IATFGI	Inter-agency Task Force on Geographic Information
ICSM	Inter-governmental Committee on Surveying and Mapping
ICT	Information and Communication Technology
IDM	Integrated Data Model
INSPIRE	Infrastructure for Spatial Information in Europe initiative
ISO	International Standards Organization
ISO199115/TC211	International Standards Organization/Technical Committee211
JMP	Japanese Metadata Profile

LAN	Local Area Network
LIS	Land Information System
MaCGDI	Malaysian Centre for Geospatial Data Infrastructure
MyGDI	Malaysian Geospatial Data Infrastructure
NaLIS	National Infrastructure for Land Information System
NBII	National Biological Information Infrastructure
NCC	National Computer Center
NEDA	National Economic and Development Authority
NII	National Information Infrastructure
NSDI	National Spatial Data Infrastructure
NSO	National Statistics Office
NSW	New South Wales
OGC	Open GIS Consortium
OSDM	Office of Spatial Data Management
PCGIAP	Permanent Committee for GIS Infrastructure for Asia and the Pacific
PHIVOLCS	Philippine Institute of Vulcanology and Seismology
PIP	Property Information Project
POI	Points of Interest
PSMA	Public Sector Mapping Agency
SDI	Spatial Data Infrastructure
SDMG	Spatial Data Management Group
SDPE	Spatial Data Policy Executive
SDSC	San Diego Supercomputer Center
SEG	Spatially Enabled Government
SERF	Service Entry Resource Format
SII	Spatial Information Infrastructure
SLC	Single Land Cadastre
SPOT	Single Point of Truth
SRS	Spatial Reference System
VBA	Visual Basic for Applications
VGSC	Victorian Government Spatial Committee
VSC	Victorian Spatial Council
VSIS	Victorian Spatial Information Strategy
UML	Unified Modelling Language
UNRCC	United Nations Regional Cartographic Conferences
USGS	US Geological Survey
W3C	Wide Web Consortium
WALIS	Western Australian Land Information System
WAN	Wide Area Network
WOG	Whole of Government
XML	eXtensible Markup Language

Chapter 1

Introduction

1.1. Background to Research

The access and integration of multi-source spatial data are critical to the accomplishement of the objectives of spatial applications. However, access and integration of multi-source spatial data face many technical and non-technical problems that inhibit their use. Considering the complexity of the problems in modern societies including environmental protection and sustainable development, the demand is growing dramatically among decision-makers to utilise more effective solutions that are highly dependent on this information. It has been estimated that 80 per cent of the information that is used for decision making possesses a spatial component (Klinkenberg, 2003). Therefore, spatial information plays significant and key role for informative decision making.

With the advancement of technology in spatial data creation, spatial data is created and owned by many different agencies that utilise spatial data to satisfy their own needs. The fragmentation of spatial data owners causes diversity in policies related to spatial data, and standards and tools to manage and coordinate spatial data. The diversity of approaches in data coordination leads to inconsistency and heterogeneity among multi-source spatial datasets.

From a technical perspective, different data models, technical standards, metadata structures, attribution and tools are utilised in order to obtain, coordinate and share spatial datasets. However, non-technical issues make it more complex and difficult to effectively integrate multi-source spatial datasets. Diversity in institutional arrangements of different organisations and partnership models together with diverse pricing and sharing policies, jurisdictional priorities, legal and social issues hinder effective spatial data integration.

The complexity and diversity of potential technical and non-technical barriers to spatial data integration require a holistic platform to manage the barriers. The holistic framework should also provide respective solutions and enablers to overcome the barriers. Spatial Data Infrastructure (SDI) as a sharing platform aims to facilitate the access and integration of multi-source spatial data within a holistic framework with a number of technological components including policies, standards, access and the interaction between spatial data stakeholders and spatial data.

In this regard, individual countries are also looking to highlight the importance of multi-source spatial data integration within the context of SDI initiatives. For example, Emergency Management of Australia (EMA) identified a lack of integrated data, including spatial data required for emergency management, as a major barrier slowing down the response of rescue teams (Conybeare, 2003; Sharwood, 2005). The Australian spatial information council (ANZLIC) has identified the integration of multi-source spatial datasets as a priority area for implementation of the Australian SDI (ANZLIC, 2003a). The Canadian Geospatial Data Infrastructure (CGDI, 2001) indicated the integration of disparate geospatial information as a major component of CGDI architecture. In Europe, the Infrastructure for Spatial Information in Europe initiative (INSPIRE) recognised spatial data integration (combination) as one of its principles (INSPIRE, 2006). As a result the INSPIRE implementation will gradually harmonise data and information services, eventually allowing the seamless integration of systems and datasets at different levels into a coherent European SDI (INSPIRE, 2002). Countries in the Asia and Pacific region – through the activities of the Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP) –

have identified and seek to address integration issues (PCGIAP, 2006) within a holistic SDI framework.

SDIs aim to assist practitioners to understand and identify the hindrances and barriers to data sharing and integration. Within the holistic framework of SDI, the potential barriers to effective data integration can be identified and also necessary technical tools and non-technical mechanisms can be developed to facilitate the integration of multi-source spatial data.

Many jurisdictions worldwide have recognised the importance of spatial data integration and its benefits to the society. They have also recognised the urgency to develop new standards, technical tools, strategies and policies within SDI to facilitate the effective integration of multi-source spatial data. Many international publications and declarations have recognised spatial data integration as a major issue and priority in the development of SDIs including the 1980 United States National Research Council's study (National Research Council, 1980) which highlights the need for a set of consistent standards, funding programs and jurisdictional coordination for the success of multi-source data integration. The Bogor declaration (FIG, 1996) also urges the establishment of national SDIs to ensure uniform data integration. Parker and Enemark (2005) believe that effective integration of built and natural spatial data is problematic, despite the design and development of SDIs. Many jurisdictions face difficulties in integrating built and natural spatial data (UNRCC-AP, 2003). The future progress of SDI establishment now depends on more thorough investigation of impediments to integration of all spatial datasets and particularly cadastral and topographic datasets (UNRCC-AP, 1997).

1.2. Research Formulation

1.2.1. Statement of Research Problem

Multi-source spatial datasets comply with diverse standards, specifications and arrangements; therefore the integration of multi-source spatial datasets is associated with technical and non-technical issues that hinder the use of spatial datasets in achieving their maximum potential and usability.

The diversity of data integration technical and non-technical issue requires the holistic framework of SDI. Fully functioning SDI to facilitate the maximum use of multi-source spatial datasets, cannot be achieved unless heterogeneities and inconsistencies in the integration of multi-source spatial datasets from both technical and non-technical perspectives are overcome.

Many jurisdictions have fragmented spatial data coordination arrangements and also data custodianship. The fragmentation of data custodians has caused a diversity of approaches in data modelling, data acquisition, maintenance and sharing. Consequently, the lack of a holistic approach to coordinate these activities within a common framework has hampered many of the applications to access, integrate and use spatial data efficiently and easily.

1.2.2. Aims, Research Questions and Objectives

This research project aims to address the gaps in spatial data integration (in terms of the identification of potential technical and non-technical barriers and the provision of available solutions and enablers) which need to be overcome, especially within the context of SDI initiatives. The research will discuss the foundation elements in order to better understand and describe the technical, jurisdictional, institutional, legal and policy perspectives surrounding the integration of multi-source spatial datasets.

This research also aims to develop a data integration toolbox to facilitate the integration of multi-source datasets in the context of SDI initiatives. This will be supported by case studies at Australian and international levels.

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

1. What are the incentives for different jurisdictions to pursue spatial data integration?
2. What are the potential technical and non-technical barriers to effective integration of multi-source spatial datasets?
3. What technical and non-technical developments have been developed in order to address the barriers to data integration?
4. How does SDI assist the development of technological components for effective data integration?
5. What are the key enablers and tools for effective spatial data integration within the context of SDIs?

In answering these questions, the research project has three main objectives:

1. Investigate and identify the potential technical and non-technical barriers to integrate multi-source datasets within SDI initiatives
2. Identify available enablers for effective spatial data integration
3. Design and develop the key components of a spatial data integration toolbox including:
    3.1. A spatial data validation and integration prototype
    3.2. Integration guidelines to address the data integration process together with potential barriers and available enablers and solutions
    3.3. An integration data model design methodology
    3.4. Appropriate metadata content
    3.5. An effective data specification document for data integration
    3.6. Testing strategy

1.3. Justification for Research

Diversity of data providers and consequently diversity of data coordination and management arrangements cause inconsistencies in effective integration of spatial datasets. The importance of the research on integration of multi-source spatial datasets has been highlighted in numerous international publications, declarations and resolutions and in particular UN resolutions. Rajabifard and Williamson (2004a) have promulgated the integration of built and natural datasets within national SDI initiatives as a major concern in the success of national SDI. Resolution 15 of the 14th UN Regional Cartographic Conference (UNRCC) for Asia-Pacific, calls for issues in the integration of cadastral and topographic datasets to be investigated (UNRCC-AP, 1997). The UN Bogor Declaration (FIG, 1999) urges the creation of national SDI to ensure integration and highlights the need for the homogeneity of topographical and cadastral datasets (as two core spatial datasets) to achieve the integration to their maximum potential. The Aguascalientes Statement (Parker & Enemark, 2005) also recommends that there is a need to integrate land administration, cadastre and land registration functions with topographic mapping programs within the context of a wider national strategy for SDIs. UNRCC for Asia-Pacific highlights the need to better understand and appreciate the integration of cadastral and topographic data in SDIs, and highlights the benefits and difficulties of integrating cadastral and land tenure information with topographic information in providing an appropriate basis for supporting sustainable development and environmental management (UNRCC-AP, 2003).

Integration has also been called for in an Australian context with Emergency Management of Australia addressing the lack of comprehensive maps that include all information connected to emergency management as a major barrier which slows down the response of rescue teams and could cost lives (Conybeare, 2003; Sharwood, 2005). A report to the Council of Australian Governments by a high level officials' group recommends the establishment of a nationally consistent system of data collection to ensure a sound knowledge base on natural disasters and disaster mitigation (Australian Government, 2002). ANZLIC – The Spatial Data Infrastructure Standing Committee has identified the integration of multi-source spatial datasets as a priority area for implementation of the Australian SDI (ANZLIC, 2003a).

In order to address both Australian and international concerns in effective spatial data integration, this research will build on international and regional collaboration within the Asia-Pacific region through a partnership with Working Group 3 (spatially-enabled governments – formerly Cadastre) of the Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP). This committee brings together 56 developed and developing countries, forming the basis for the development of case studies. The outcomes such as guidelines and data models include a technical foundation for data sharing and a strategic policy position to support sustainable development initiatives.

1.4. Research Approach

The proposed research design and stages for the study are incorporated into five major steps as illustrated in Figure 1.1. This research can be broadly grouped into five major areas of literature review and strategy development; case studies investigation; analysis; design and development; and outcomes documentation. To establish the theoretical background of the research, an extensive literature review has been undertaken. The research also uses both national and international case studies. The literature review provides the basis for the development of the research strategy and highlights the significant issues that must be taken into consideration through case studies. The developments of the required tools and associated guidelines will be effected based on the outcomes of the case study analysis.

The literature review has been undertaken in the specific areas of spatial data integration, SDIs as sharing platforms, multi-source spatial datasets, data integration and interoperability and potential barriers and enablers. The concepts, missions and components of SDI, together with investigating the significance and applications of integration and interoperability, will provide an understanding of the integration of multi-source spatial datasets and their requirements. The literature review will inform the research strategy and will be utilised in case study investigation.

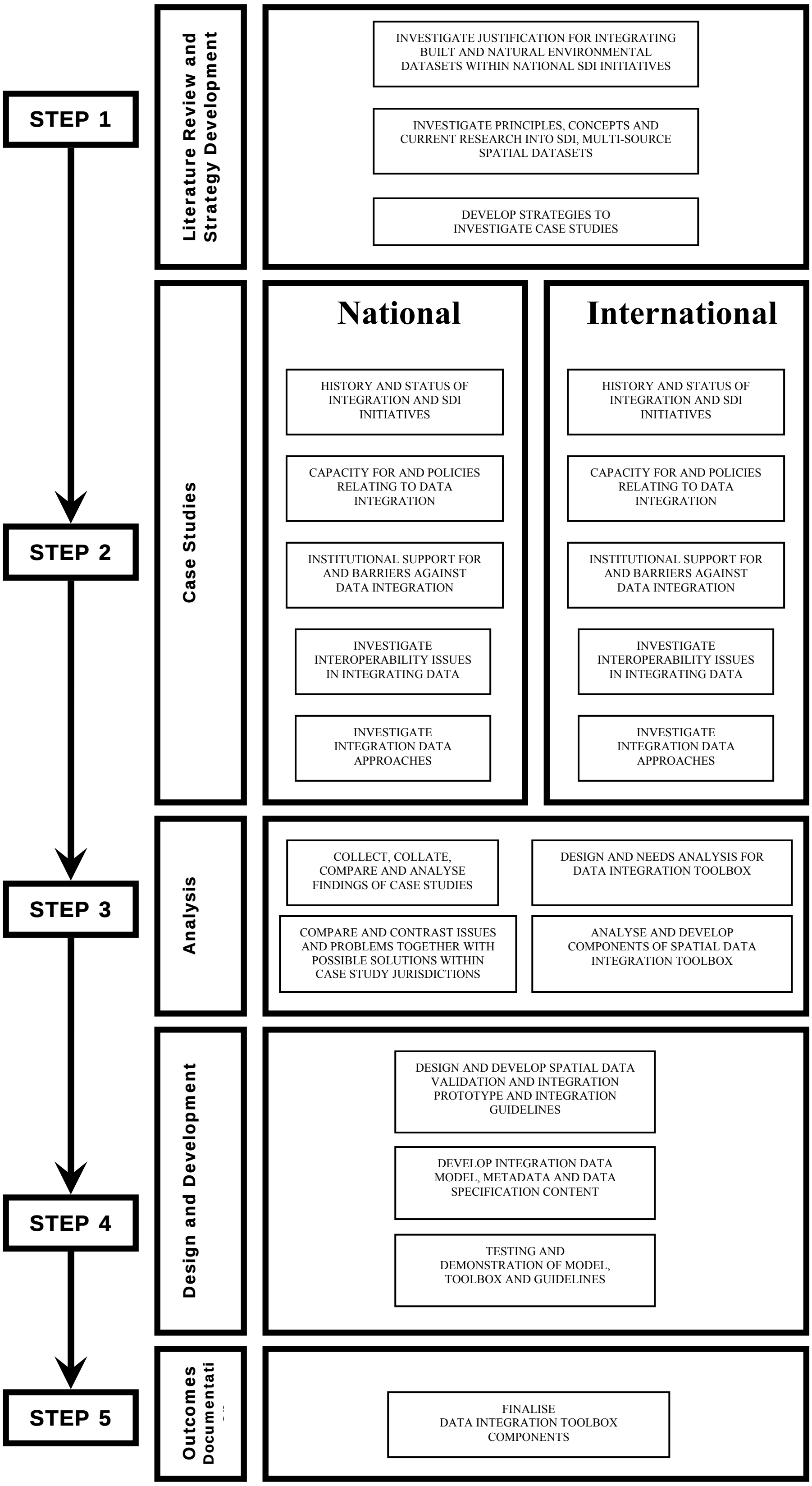


Figure 1.1. Research design and stages

The case studies include a number of countries in the Asia and the Pacific region and also a number of states in Australia. International case studies include seven countries (Japan, Singapore, Australia, Brunei Darussalam, Indonesia, Malaysia and the Philippines) and aim to investigate the challenges of data integration within different countries. Also in order to achieve a better understanding of the issues and impediments of spatial data integration and as part of the case studies, a series of technical visits have been conducted at different Australian jurisdictional levels. The aim of these technical visits is to understand the role of the key players of the spatial data scene and the way in which they coordinate spatial data, and the inconsistencies they or their users encounter in using spatial data. Australian case studies have been selected among research project partners who expressed strong interest in the results and outcomes of the research project. They include the Victoria and NSW's spatial data organisations as well as two national-level agencies (GA and PSMA). Victoria is a state located in the south-eastern corner of Australia. New South Wales (NSW) is Australia's oldest and most populous state, located in the south-east of the country, north of Victoria and south of Queensland (Figure 1.2).

Figure 1.2. Australian states and territories

In this regard, Victoria's major spatial data stakeholder (Department of Sustainability and Environment – DSE), NSW's Department of Lands, Geoscience Australia (GA) and Public Sector Mapping Agency (PSMA) are also visited. Victoria's spatial information is primarily managed by a branch of DSE called Spatial Information Infrastructure (SII). Land Victoria is also a branch of the DSE, which is responsible for balancing development and protecting the state's natural and cultural resources. GA provides a range of national fundamental datasets through its National Mapping Division and is the custodian of Commonwealth data licensed to PSMA Australia. PSMA's main task is to facilitate access to seamless national datasets developed from the data jointly held by all these government agencies.

The analysis (step 3) stage has been done to provide the basis for identification and development of key components for the spatial data integration toolbox. The investigation of data sharing, validation and integration tools is another step that identifies the functionalities of an effective data validation and integration tool as an important component of the toolbox. There are different data models used for integration of spatial datasets. A comparison of existing data models in the case study jurisdictions also highlights the inconsistency among available data models and the approach to be adopted for the development of an effective integration data model.

A key outcome of the research comprises the design and development of a data validation and integration prototype to integrate data from different sources, which is the step 4 of the research. This also addresses data, network, standards and other technical issues related to the technical implementation of the tool.

The data integration guidelines detail the potential technical and non-technical barriers to data integration and propose a methodology for data integration with consideration of the barriers and possible solutions. Metadata and data specification content are also discussed as key significant sources of information on different characteristics of spatial datasets. In addition, a number of recommendations are made to turn metadata and data specifications into useful documents for effective spatial data integration.

The last step of the research is to document supporting literature and the outcomes of case study investigations together with analysis, design and development steps.

1.5. Structure of the Thesis

The thesis has been organised into five parts and eight chapters as shown in Figure 1.3.

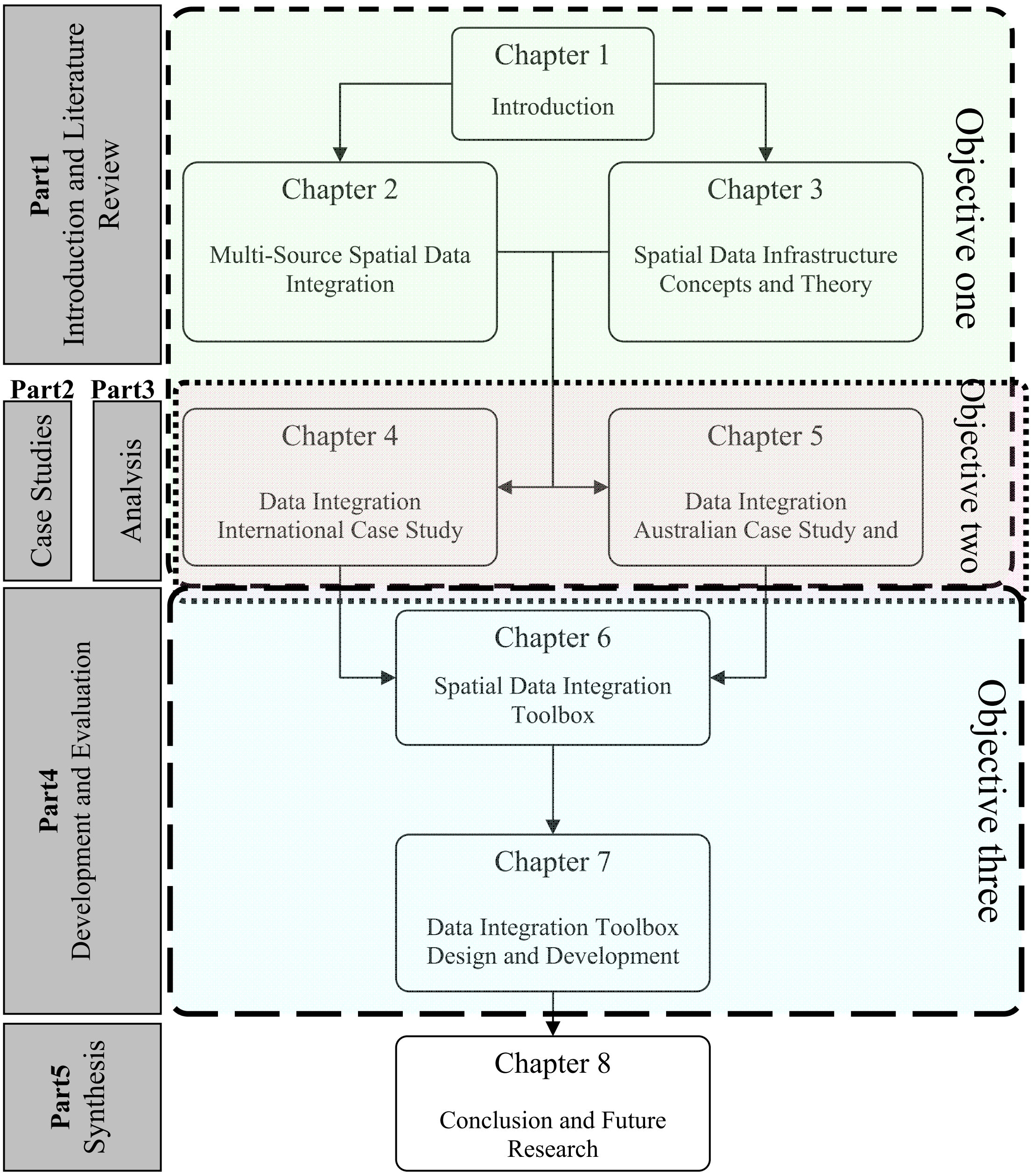


Figure 1.3. Thesis structure and relationship to objectives

Part One is the introduction and includes the statement of the research problem, the aim and objectives. Part One consists of three preliminary chapters of the thesis. In Chapter One a background to the research has been presented which is succeeded by the problem statement, aims and objectives of the research. Part One also justifies the importance of the research and presents the thesis structure. Chapter Two describes the data integration with focus on the requirements of spatial services to integrate multi-source spatial datasets. This chapter will also discuss the issues of and barriers to effective spatial data integration. Chapter Three introduces the spatial information, and the concept and theory of spatial data infrastructures with focus on data sharing, data access and data integration.

Part Two and Three consists of two chapters on investigation and analysis of the case studies. Chapter Four elaborates on the international case studies. This chapter investigates and analyses the case study countries from the Asia-Pacific region with focus on the integration issues and barriers in the context of SDIs. Chapter Five investigates Australian case studies including a number of federal and state organisations particularly PSMA, GA, and the states of Victoria and NSW.

Part Four comprises Chapters Six and Seven. Chapter Six elaborates on the need for a spatial data integration toolbox and Chapter Seven discusses the design and development phase of key components of the toolbox. Chapter Eight concludes the thesis and presents recommendations for future research.

The first three chapters respond to the first objective of the research project, which aims to investigate, identify and understand the potential technical and non-technical barriers to integrate multi-source datasets within SDI initiatives using Australia and a number of international experiences. Chapters Four and Five also help to achieve the first aim which is the identification of spatial data integration challenges and issues. These two chapters also aim to identify available enablers for effective spatial data integration by using a spatial data integration toolbox, which is the second research objective. The third objective of the research project is achieved in Chapters Six and Seven that aim to design and develop the key components of a spatial data integration toolbox.

1.6. Delimitation of Scope and Key Assumptions

The focus of this research is on existing spatial data sharing concepts within SDI initiatives within Australia and case study countries within the Asia-Pacific region. Australia consists of six states and two territories. The federal government of Australia and the two states of Victoria and NSW have been selected as case study jurisdictions. The Australian case studies have been selected among research project partners. The international case studies have participated in the research project through the PCGIAP’s Working Group 3 channel.

The findings and observations in the case studies provided the basis for the main research streams, which mainly focus on technical and non-technical barriers that are valid in the context of case study jurisdictions. The guideline development capitalises on findings and outcomes of the cases studies and discusses a methodology for spatial data integration with consideration of potential technical and non-technical barriers and possible solutions and enablers. The components for the spatial data integration toolbox have also been identified and developed based on the findings and outcomes of the case studies.

1.7. Chapter Summary

This chapter introduces the key problem, research aim and objectives. The research problem was justified and the research methodology was briefly described. A justification of the research methodology has also been provided. The thesis structure has been outlined and some delimitations of the work scope together with assumptions have been discussed.

Chapter 2

Multi-source Spatial Data Integration

2.1. Chapter Aims and Objectives

This chapter aims to elaborate on the concepts and fundamentals of multi-source spatial data integration. This includes the incentives that drive the integration of multi-source spatial data. The chapter also presents the complexity of technical and non-technical issues that hinder effective data integration. Previous attempts to overcome the barriers and issues of spatial data integration are also discussed in this chapter. The need for a holistic platform that overarches, identifies and addresses the data integration barriers is also discussed in this chapter.

2.2. Introduction

Spatial data plays a significant role in much decision making. With the advancement of technology, the creation of spatial data has become much more straightforward and easier (Kasturirangan & Ramamurthy, 2001); hence, many organisations and agencies create and maintain spatial data and provide it to spatial data consumers especially spatial data decision-makers. With new emerging demands from these services, spatial data applications cannot afford to create all data required to meet their users' needs. This is especially so when more value-added and integrated spatial data is required for more complex analysis. Therefore, many of applications collect and integrate spatial data from different sources. The diversity of data providers, data consumers and their needs and characteristics lead to diversity in many aspects including different standards, policies and institutional arrangements (Mohammadi et al., 2008).

The diversity of approaches taken by different stakeholders and providers makes multi-source spatial data inconsistent. The inconsistency emerges in many ways. Some spatial data sets do not comply with common standards and technical specifications. Different institutional arrangements of data providers can be a big obstacle as well. In many cases this leads to weak collaboration and liaison among different stakeholders which hinders effective spatial data integration. This is obviously crucial when some spatial data applications such as emergency management require fast and on-time access to and integration of datasets.

Many studies and investigations have been done to explore different aspects of multi-source spatial data integration (Fonseca et al., 2002; Baker & Young, 2005; Fonseca, 2005; Hakimpour, 2003; Jones & Taylor, 2004; Pinto et al., 2003), among which, most of them are technical studies. But without investigating all technical and non-technical issues together within a single framework, effective spatial data integration cannot be achieved.

This chapter aims to elaborate on the background and concepts of multi-source spatial data integration. The chapter also presents the major incentives and drivers of multi-source spatial data integration. It then discusses the potential issues of and barriers to effective data integration in various literature and presents the initiatives and products that have been developed to overcome these barriers.

2.3. Multi-source Spatial Data and Spatial Applications

Spatial data is playing a significant part in many applications that provide services to decision-makers. Much research (Ryttersgaard, 2001; Klinkenberg, 2003) has highlighted that about 80 per cent of all information utilised by decision-makers is spatial data or has spatial dimensions. The amount of spatial data circulated among stakeholders creates enormous opportunity for spatial services to serve more consumers at different jurisdictional scales. To achieve this aim spatial services utilise different spatial datasets ranging from fundamental to business-centric datasets.

Spatial data may contain very basic (fundamental) information or very rich (business-centric and value-added datasets) information (Buehler, 2003). Fundamental spatial datasets include very basic datasets and are utilised by private and public services. Fundamental datasets include some highly used spatial data that is mainly owned by governments (ANZLIC, 2004). Some examples of fundamental spatial datasets are topographical, vegetation, hydrography and administrative boundaries. Small to medium-scale fundamental data is produced at national (or state) level and is provided to other data users, while large-scale fundamental datasets are produced at local government level and combined at national level to form a national coverage map (National Research Council, 2003). The custodians of fundamental datasets are governmental agencies, therefore these datasets are more publicly available and most of the time for the cost of maintenance, so users can access and use them more easily. In many cases, as governments are obliged to provide data to a broad range of users, these datasets are accompanied by fewer restrictions. If governmental organisations commit to maintain metadata with these datasets, users capitalise on much useful information which is included in metadata. Many businesses build their information on fundamental datasets and create and maintain rich value-added datasets (Burrough & Masser, 1998). Due to the cost of value-added datasets these data sets are associated with more restrictions and are less available to the public. As they are business-oriented datasets, they are expensive and much specified to the needs of particular businesses (Figure 2.1).

Figure 2.1. Spatial data value chain

The business-centric datasets barely interact with other datasets and comply with the data model of the business they are targeted at. Conversely, the richer the datasets become, the more restrictions are applied to them. Table 2.1 summarises some of the characteristics of fundamental and value-added datasets.

Table 2.1. General characteristics of fundamental and value-added datasets (after Baker & Young, 2005)

	Fundamental	Value-added
Custodian	Mainly governments (public sector)	Mainly private sector
Scale	Small to medium (at national level) Large (at state and local levels)	Mainly medium to large
Format	Compatible with other governmental datasets	Business-specific
Availability	Publicly available	Restricted
Restrictions	Licensing	Licensing, distribution and manipulation
Data model	Compatible with other governmental datasets	Business-specific
Pricing	Maintenance and conveyance price	Cost recovery
Data sharing	Distributed	Silo-based
Metadata	Available and rich	Mostly no metadata
Coverage	National and state coverage	Interested area
Attribution	Most common attributes	Rich

As shown in Table 2.1, fundamental datasets are more compliant with standards and policies. They include compliancy with other governmental datasets on format, data model, metadata and attribution. The less restricted availability, pricing and data sharing arrangements for fundamental datasets create more incentives among users to utilise these datasets. Business-centric datasets are associated with more complexity and barriers. These datasets are less compliant with other standards, policies and datasets that are owned by other stakeholders. The restrictions on availability, pricing, sharing, metadata and attribute content of these datasets hinder the access and utilisation of the datasets, while many spatial services require these datasets for delivery of their objectives.

Spatial services utilise datasets to meet their objectives. According to the complexity and required outputs, utilised datasets may vary from very fundamental to value-added and integrated datasets. As an example, emergency management services utilise a broad range of spatial datasets (Montoya & Masser, 2004; CRC.SI, 2004; Edwards & Simpson, 2002). In the case of Emergency Information Coordination Unit (EICU) in NSW, Australia, a broad range of different spatial data including fundamental spatial data, locational data and infrastructure and socio-economic spatial data, is utilised. The EICU provides service to many agencies and organisations who are involved in emergency management activities including Fire Authorities (Colless, 2005). Figure 2.2 shows various datasets that are used for emergency management services.

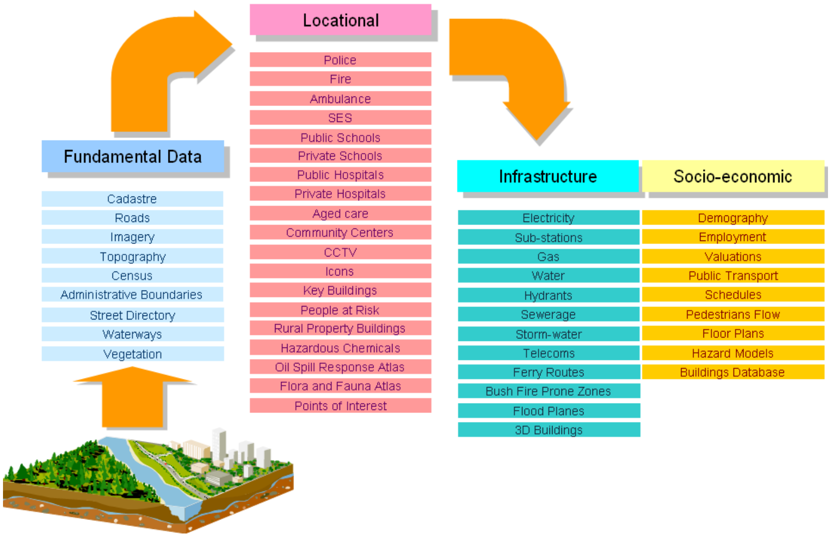


Figure 2.2. Spatial data required for emergency management (Colless, 2005)

These data sets are owned, coordinated and maintained by different organisations. Fundamental data sets are mostly owned and managed by governments, while infrastructure and socio-economic data are more business-oriented data and owned and used by businesses. The diversity and sensitivity of each data group for public and business sector cause different considerations on data management. Fundamental data sets are more publicly available and more interoperable data has richer metadata, more consistent data models, clear pricing and privacy policies and fewer access points, which make it easier for the users to discover and find data (ANZLIC, 2004). Business-type data, including infrastructures, utilities, and socio-economic data, is less publicly available and each business sets its own data standards and frameworks. These data sets have poor or no metadata, and restrictions in use.

Conversely, spatial data is utilised to meet the objectives of projects at different scales. Spatial data at the local scale is mostly utilised by users who are interested in small areas with more details and accuracies. Many applications including local emergency management services, urban planning systems and catchments management serve such requirements.

At regional, national and state levels, vast areas are required to be covered by spatial data. Depending on the requirements of the services, accuracies and details of data may vary.
Therefore, depending on the area of coverage, the richness of data content and the number of datasets involved, the complexity and the issues of coordination and matching datasets vary. Figure 2.3 illustrates that with any increase in area, data richness (value-added data) and number of datasets, the number of issues and barriers in coordination of the datasets and matching datasets will increase dramatically.

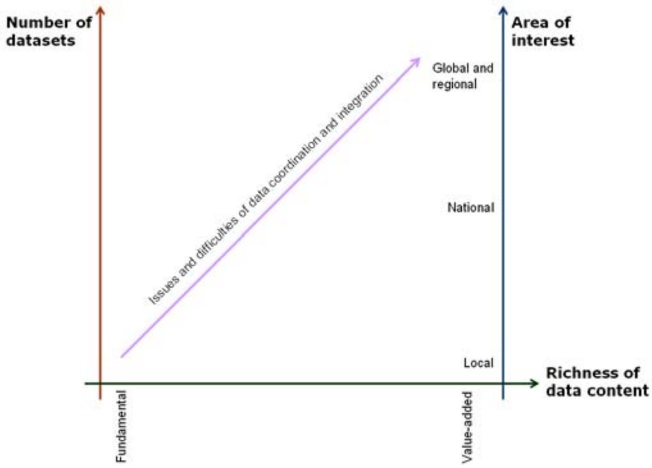


Figure 2.3. The increase of issues and difficulties of data coordination and integration in relation to the number of involved datasets, richness of data content and area of interest (adopted from Van Loenen, 2006)

Value-added datasets are produced mostly by integrating different datasets and forming much richer data. Integration adds value to the source datasets by providing more information from different sources (Longley et al., 2001; Van Loenen, 2006).

Longley et al. (2001) has stated that the value of a dataset rests on its "coverage, the strengths of its representation of diversity, its truth within a constrained definition of that word, and on its availability". The users decide whether or not to use the dataset on the basis of its technical and non-technical characteristics although each user may have specific needs from spatial datasets (Van Loenen, 2006).

The diversity of users' requirements from spatial data for the delivery of the objectives requires the combination of different spatial data. Spatial datasets are mostly coordinated by different custodians. Therefore users need to utilise an integration of multi-source datasets to achieve their aims.

2.4. The Integration of Multi-source Spatial Data

Due to the ever-increasing complexity in environmental analysis and high cost of data production, sharing and integrating spatial data are necessary. Therefore, the data integration and interchange between organisations have become an essential issue

(Pinto et al., 2003). Organisations with interests in spatial information are no longer able to produce and maintain necessary spatial data. Therefore, multi-source spatial data integration has become the fundamental motivator throughout the organisations.

Many studies have been devoted to investigate different aspects of spatial data integration. Different researchers have different perceptions of data integration and expect different results from it.

Some researchers have defined multi-source spatial data integration with emphasis on the establishment of the relationship between different objects of datasets. Uitermark (2001) defines spatial data integration as the establishment process of relationships among corresponding object instances in different, autonomously produced, geographic datasets of the same geographic space. Usery et al. (2005) have investigated spatial data integration and have defined it as the process of matching different datasets geometrically, topologically and establishing correspondence of attributes. Uitermark (2001) has defined spatial data integration as “the process of establishing links between corresponding object instances in different, autonomously produced, spatial datasets of the same geographic space”.

Spatial data integration in some studies has been described as a tool for value-adding to original datasets. In this context, spatial data integration is the practice of bringing source spatial datasets together to create a new product that is richer in content than the original sources (National Technology Alliance, 2005).

In some literature spatial data integration has been defined as a process. Samadzadegan (2004) describes data integration as a number of processes including the acquisition, processing and synergistic combination of multi-source datasets. Ronsdorf (2005) believes that in all applications the vital part is to fit multi-source data spatially in order to facilitate the joint use and analysis.

Besides technical combination and collation of multi-source spatial datasets, researchers address a number of non-technical processes that should be established for effective data integration.

Some researchers believe that an institutional component is a vital component to be considered in order to fulfil effective data integration. Ryttersgaard (2001) highlights the institutional component of data integration alongside technological developments as the most important issues for efficient use of spatial datasets. Masser (2006a) and Williamson et al. (2003) address institutional structure as a facilitator for spatial data integration. Van Loenen and De Jong (2007) identify institutional settings that can greatly affect and facilitate data usage. Capacity building and raising awareness have been highlighted also in some literature as a social activity to assist spatial data integration (Alexiadou & Rajabifard, 2006; Williamson et al., 2003). Some studies (Burrough & Masser, 1998; Van Loenen, 2003; Onsrud et al., 2004) consider the legal component as important as the technical component for effective spatial data integration. Van Loenen (2003) has identified some policy components including pricing and access, which should be considered for comprehensive spatial data integration.

All the definitions outlined above, represent a particular aspect of spatial data integration. This thesis capitalises on the research and studies in the area of spatial data integration and introduces a more comprehensive definition that overarches complex components of spatial data integration.

Multi-source spatial data integration is not only the establishment of the process of matching different datasets geometrically, topologically and having correspondence of attributes (Usery et al., 2005), and synergistic combination of multi-source spatial datasets (Samadzadegan, 2004) and enriching multi-source spatial data with the establishment of relations between them (National Technology Alliance, 2005), but also is the establishment of necessary institutional, legal, social and policy frameworks.

Figure 2.4 illustrates the technical core of spatial data integration and surrounding non-technical components.

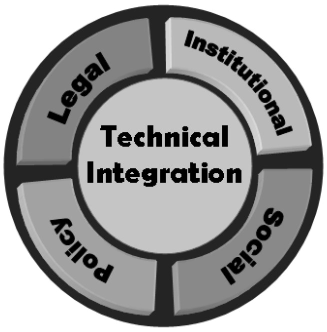


Figure 2.4. Spatial data integration and its technical and surrounding non-technical components

The complexity of spatial data integration is also impacted by the complexity and the level of analysis that the users expect from integrated product. Spatial data users require spatial data integration at different levels including visual and analytical levels.

2.5. Spatial Data Integration Drivers

As discussed in previous sections, many spatial applications gather datasets from different sources and put them together. Spatial stakeholders expect some results from spatial data integration. Some of major advantages of spatial data integration are:

2.5.1. Spatial Datasets Control through Comparison with Other Datasets

Spatial data integration provides a qualitative and quantitative comparison of underlying spatial datasets and it can strengthen the confidence in the underlying data (Edwards & Simpson, 2002). It is particularly important for the reuse of updates from one dataset to another dataset (Uitermark, 2001). Spatial data integration puts datasets beside others and produces a realistic benchmark to assess datasets against others. Spatial datasets with better quality or richer content can be utilised as a measure to evaluate other datasets and improve the internal consistency of the participant datasets (Usery et al., 2005).

2.5.2. More Complete Spatial Data Product

Spatial datasets represent some aspects of the real world and assist practitioners to model the environment and measure and analyse the impacts of environmental phenomena (Rajabifard & Williamson, 2004b). The more complete the model, the better the result from spatial services. There is no single source of spatial data that perfectly represents the environment for all spatial services. Therefore, this pluralistic reality suggests the integration of data from various sources obtains the unique advantages of each source to enrich the final product of integration (Edwards & Simpson, 2002).

The integrated data elements reinforce, augment or contradict others; therefore, the integration of spatial data generates new information. Different practitioners interpret the integrated datasets differently, which can provide a wide range of new information (Edwards & Simpson, 2002).

2.5.3. Horizontal and Vertical Integration

There are two distinct types of integration: horizontal and vertical (Finn et al., 2004). In order to produce a seamless, vast area coverage (regionwide, nationwide or statewide), spatial datasets are integrated horizontally (Usery et al., 2005). Horizontal integration provides seamless data comprising patches of datasets. From a horizontal integration viewpoint, two maps must share the exact attribution so an extension of a feature from one horizontal partition to another remains the same feature with the same attributes (Usery et al., 2005).

Conversely, different levels of government and the private sector conduct the vertical integration of spatial datasets, which is bound to the borders of the area of interest (Figure 2.5)..

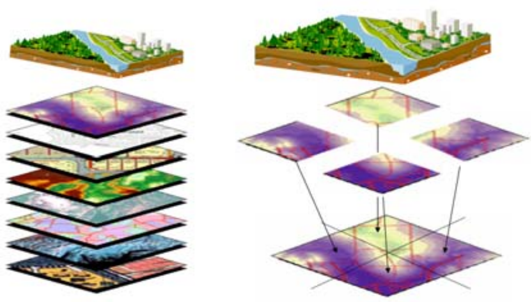


Figure 2.5. Vertical and horizontal integration

Vertical integration will result in the production of the best available scale spatial data product that can be utilised to derive smaller scale mapping needs. It also avoids the duplication of effort in governments (federal, regional, state, local and municipal governments) and the private sector (Burgess, 1999) and provides internal consistency among different themes (Usery et al., 2005)

2.5.4. Unified Data Collection

An ultimate aim of spatial systems is to provide a platform in which different datasets can be easily integrated and used. In order to achieve this aim some organisations have developed an integrated collection of necessary datasets. This contains spatial data from different sources that have been amended based on the requirements and standards that have been adopted for the collection.

This can be a logical (virtual) collection of different and independent databases in a single unified database (Buch, 2002). The collection of data can be treated as one entity and viewed through a single user interface. Within this collection, data stakeholders share the data logically through the communication media. This approach prevents data inconsistency and redundancy (Iqbal, 2007).

2.5.5. Facilitating Multi-disciplinary Initiatives

An emerging trend in Spatial Systems (GIS) is the use of multiple systems and diverse datasets in a single study. "In practice, sharing spatial data among several systems is difficult because of incompatibilities in spatial data formats and limitations within existing GIS" (Piwowar & LeDrew, 1990). Spatial data integration can get different disciplines together and produce multi-disciplinary knowledge. Cowen and Craig (2003) have highlighted the importance of integrating multi-source spatial data sets for building multi-purpose cadastre. Another example of multi-disciplinary applications of spatial data integration is biodiversity. A vast quantity of information has been collected and compiled on the earth's living organisms. These include biodiversity data on the distribution of plants, animals and microbes etc. (De Santos et al., 2002). In order to link and analyse these datasets, spatial location of these phenomena is of high significance, which dramatically increases the reliance on partnerships and commercially available data (Usery et al., 2005).

The above-mentioned items introduce some of the incentives and advantages of spatial data integration. There is a potential in facilitated spatial data integration to assist spatial services and users; however, the diversity of approaches, arrangements and standards utilised by source organisations lead to a heterogeneity among datasets. In order to implement spatial data integration efficiently, associated barriers and issues should be investigated and identified. Many of the issues and obstacles of data integration have been studied and identified by researchers. The following section discusses the issues of and barriers to spatial data integration from the perspective of other researchers. Chapters Four and Five will discuss the issues and considerations for effective spatial data integration in more depth by conducting some case studies.

2.6. Issues and Obstacles of Multi-source Spatial Data Integration

Spatial data integration within a single organisation is not so problematic (Usery et al., 2005); however, between organisations and agencies, it faces many problems. Data produced and collected by a variety of agencies and companies is in theory interesting and valuable (Fonseca et al., 2002). The diversity and number of mapping organisations and data providers are the most significant barriers for effective spatial data integration (Clarke et al., 2002). Spatial data providers create and maintain spatial data for their own and their users' needs. Therefore, the datasets properly suit the requirements of the target users. Each of the data categories sit in a separate silo. Each of the silos complies with a certain set of policies and considerations that may differ from other silos (Williamson et al., 2003).

Different organisations utilise different specifications, tools, standards, policy frameworks and institutional arrangements. Also legal considerations and social background of organisations especially within different jurisdictions differ a great deal. These arrangements cause many technical and non-technical issues (Piwowar & LeDrew, 1990). At the same time, many spatial applications manage some aspects of the environment that do not necessarily sit within the borders of any particular jurisdiction. Many of the spatial applications require spatial data from different areas. These applications provide additional value to the spatial data by integrating them.

The diversity of the requirements of spatial applications together with the diversity of standards, policies and approaches utilised by different spatial data custodians cause many issues that hinder effective data integration.

The technical and non-technical characteristics of source datasets add up to the characteristics of the integrated dataset. This includes technical inconsistencies and many institutional, policy, legal and social issues. Many studies have been conducted to investigate the technical and non-technical issues associated with data integration.

2.6.1. Technical Issues

Technical issues can be categorised into different classes. Some technical issues are relevant to the characteristics of the dataset itself. These characteristics are not much related to external factors, like access policies (Van Loenen, 2003). Some of the technical issues are sourced from standards (Sen, 2005) and some others are because of the access and distribution channels that are utilised by corresponding organisations (Usery et al., 2005).

Many researchers have studied spatial data integration and associated issues in the context of distinctive projects and initiatives (Usery et al., 2005; Pinto et al., 2003; Backe & Edwards, 2005). The technical integration of spatial datasets is the biggest concern of these initiatives. For example, Usery et al. (2005) identified a number of technical issues of spatial data integration including boundaries, resolution and positional accuracy, and believe that effective data integration improves the internal consistency of the involving datasets. Pinto et al. (2003) have studied data models and highlight the heterogeneous data structures utilised by the different data models that represent the graphical and non-graphical components of spatial phenomena as the greatest barrier to spatial data integration.

Some of the major technical obstacles of spatial data integration are as follows:

- Different data sources and accuracies (Finn et al., 2004)
- Data quality (Krek, 2002; Van Loenen, 2003; EUROGI, 1997)
- Multiple raster and vector formats (Pinto et al., 2003; Usery et al., 2005)
- Variety of spatial resolutions (Edwards & Simpson, 2002; Usery et al., 2005; Van Loenen, 2003; Finn et al., 2004)
- Temporal resolution (Ulubay & Altan, 2002; Edwards & Simpson, 2002; Usery et al., 2005; Van Loenen, 2003; Finn et al., 2004)
- Different scales (Usery et al., 2005; Van Loenen, 2003; Ulubay & Altan, 2002)
- Metadata concerns (Jones & Taylor, 2004; Sheth, 1998)
- Interoperability problems (Ulubay & Altan, 2002)
- Different semantics and representations (Ulubay & Altan, 2002; Pinto et al., 2003; Iqbal, 2007)
- Compilation standards (Edwards & Simpson, 2002)
- Differences in datum, projections, coordinate systems (Finn et al., 2004; Ulubay & Altan, 2002)
- Data models (Finn et al., 2004; Usery et al., 2005, Ulubay & Altan, 2002; Iqbal, 2007)
- Precision and accuracy (Finn et al., 2004; Usery et al., 2005, Ulubay & Altan, 2002)

- Different purposes of datasets (Ulubay & Altan, 2002)
- Different base maps (Ordinance Survey, 2003)

2.6.2. Non-technical Issues

Non-technical obstacles of data integration can be caused by institutional, policy, legal and social issues and considerations of spatial stakeholders (Williamson et al., 2006; Mohammadi et al, 2006; Burrough & Masser, 1998; Van Loenen, 2005; Chrisman and Niemann, 1985).

Institutional Issues

With the involvement of different organisations, it is difficult to provide discoverable and available data in a structured way (Piwowar & LeDrew, 1990). Many researchers have investigated institutional obstacles of spatial data integration. Some key findings are as follows:

- Inter- and cross-organisational access, retrieval and display arrangements (Zaslavsky et al., 2004; Baker, 2005; EUROGI, 1997),
- Sharing data among organisations (Weaver, 2004; Baker, 2005),
- Different coordination and maintenance arrangements (Ordnance Survey, 2003),
- High degree of duplication (Baker, 2005; Burgess, 1999),
- Weak collaboration (Baker, 2005),
- Uncoordinated specifications and standards across spatial stakeholders (Baker, 2005),
- Lack of central access gateway (single point of access) (Baker, 2005); and
- Building awareness and capacity (Clausen et al., 2006).

Policy Issues

From a policy perspective, the diversity of involved organisations with different policy drivers and priorities affect the integration. Some of the key issues are listed below:

- Access policies (Donker & Van Loenen, 2006): Concerning use requirements, users require both transparency of information policies and consistency in the access to policies throughout government. Differences in pricing, use restrictions, and liability regimes may result in confusion and ultimately limited use of the datasets. (Donker & Van Loenen, 2006)
- Pricing models (Donker & Van Loenen, 2006): As a consequence, it is time-consuming to explore a potential avenue to cost-recovery, among other things (Donker & Van Loenen, 2006).
- Use restrictions (Meixner & Frank, 1997; Donker & Van Loenen, 2006)

Legal Issues

The integration of spatial datasets raises a number of legal issues. It is obviously necessary to clarify the nature of datasets and the stakeholders and their particular rights in data (Burrough & Masser, 1998). In 1995, the European Umbrella Organization for Geographic Information (EUROGI) commissioned RAVI, the Netherlands Council for Geographic Information, to conduct a survey on the legal problems:

- Different licence conditions (Donker & Van Loenen, 2006; EUROGI, 1997)
- IP and licensing (Baker, 2005; Donker & Van Loenen, 2006)
- Liability regimes (Donker & Van Loenen, 2006)

Social Issues

Many institutional and policy issues are caused by the social behaviours of jurisdiction. Building collaboration among organisations is an example. However, some explicit social behaviours directly hinder effective spatial data integration as follows:

- Silo mentality without effective mergers among silos (Ordinance Survey, 2003)
- Aversion to data sharing and integration: People often resist sharing data across organisational boundaries due to loss of control, power and independence (Clausen et al., 2006)

In order to have a holistic study of technical and non-technical obstacles of data integration, Table 2.2 has summarised the above-mentioned issues and obstacles identified.

Table 2.2. Integration issues

Technical issues	**Non-technical issues**			
	Institutional issues	Policy issues	Legal issues	Social issues
▪ Different data sources and accuracies ▪ Data quality ▪ Multiple raster and vector formats ▪ Variety of spatial resolutions ▪ Temporal resolutions ▪ Different scales ▪ Metadata concerns ▪ Interoperability problems ▪ Different semantics and representations ▪ Compilation standards ▪ Differences in data, projections, coordinate systems ▪ Data models ▪ Precision and accuracy ▪ Different purposes of datasets ▪ Different base maps	▪ Inter- and cross-organisational access, retrieval and display arrangements ▪ Sharing data among organisations ▪ Different coordination and maintenance arrangements ▪ High degree of duplication ▪ Weak collaboration ▪ Uncoordinated specifications and standards across spatial stakeholders ▪ Lack of central access gateway (single point of access) ▪ Building awareness	▪ Pricing models ▪ Access policies ▪ Use restrictions	▪ Different licence conditions ▪ IP and licensing ▪ Liability regimes	▪ Silo mentality without effective mergers among silos ▪ Aversion to data sharing and integration

Many of these issues are tightly connected to others. For example, the aversion to data sharing and integration hinders the establishment of effective cross-organisational access, retrieval and display mechanisms. Restricted pricing and access policies hinder maximum use and cause duplication of efforts. Diversity in coordination and maintenance arrangements result in different data characteristics including data models, quality, metadata content and coordinate systems etc. Therefore, without considering all the issues (Figure 2.6) within a single holistic framework, effective spatial data integration cannot be achieved.

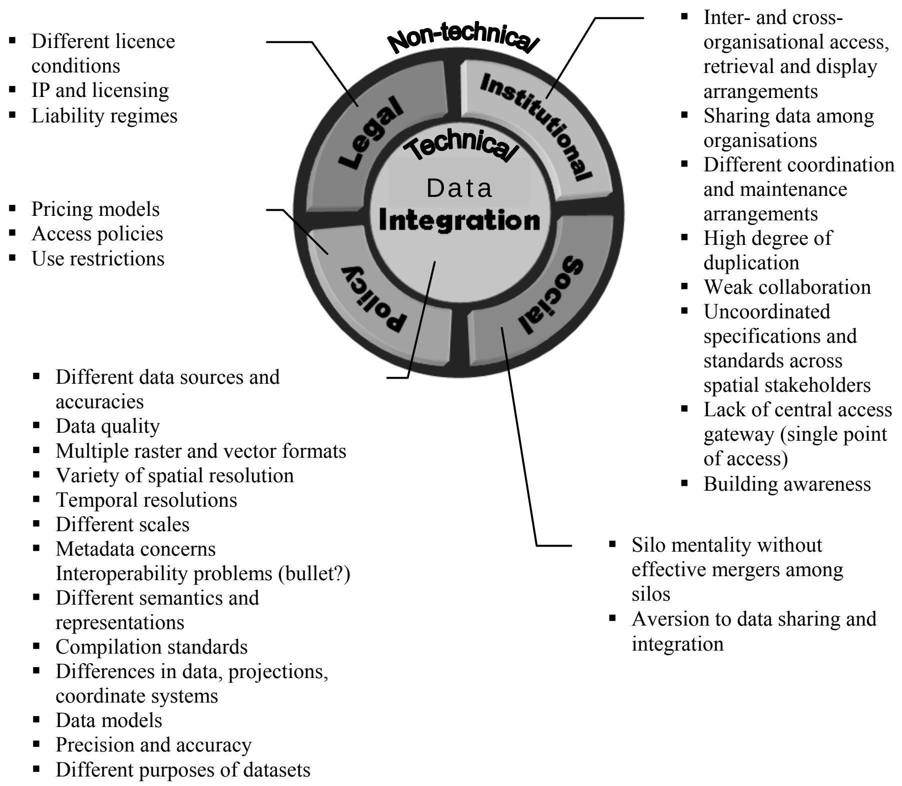


Figure 2.6. Technical integration and associated non-technical considerations

In order to overcome these issues many approaches have been taken by practitioners. Some approaches are based on a single gateway that obtains datasets with certain characteristics. Federated database and generalisation approaches are based on this concept. Some other approaches, including mediators and ontology, are utilised to integrate multi-source spatial data in general.

2.7. Spatial Data Integration Approaches

Due to the necessity of spatial data integration within multi-source and multi-disciplinary projects, responsible parties may take a short-term and project-specific approach. Called the project-based approach, specific datasets are integrated. Within the framework of the project necessary institutional arrangements, policy considerations, legal arrangements and social hindrances are dealt with and overcome. Some major approaches are generalisation, federated databases, ontology-driven data integration, spatial mediators, feature manipulation engines and spatial interoperability; these are elaborated on in the following section.

2.7.1. Generalisation

A single database that holds the best available datasets is another substitute for huge heterogeneous databases (VSC, 2004). This single database is not necessarily centralised, but points to the best available data. Other datasets with less accuracy and smaller scales can be generalised and extracted from the main dataset (Figure 2.7).

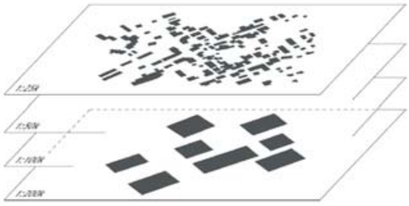


Figure 2.7. Generalisation for smaller scales (Cecconi, 2003)

Generalisation is an important issue in spatial data integration. It refers to a reduction in the volume and appearance of data as we move from very large scales to small scales (Burgess, 1999). Cecconi (2003) has defined generalisation as the process of reducing the amount of data and adjusting the information to a smaller scale and less detailed theme. Independent of the data form, the detail and content of data are strongly dependent on the scale of the data. It is evident that data in small scales contains less detailed information than large-scale data from the same region.

Brassel and Weibel (1988) have designed one of the most detailed conceptual frameworks for generalisation which consists of five major steps: structure recognition, process recognition, process modelling, process execution and data display.

According to Muller (1991), generalisation is the application of spatial data transformation which is driven not only by a reduction of scale representation but also is initiated by four main requirements including economic requirements, data robustness requirements, multi-purpose requirements and display and communication requirements.

2.7.2. Federated Databases

A federated database is a logical (virtual) collection of diverse stand-alone databases in a single unified database (Buch, 2002) or in simple terms it is a collection of databases that are treated as one entity and viewed through a single user interface (PC Magazine, 2008). Federated databases aim to access and integrate the data and specialised computational capabilities of a wide range of data sources (Haas & Lin, 2002) as shown in Figure 2.8.

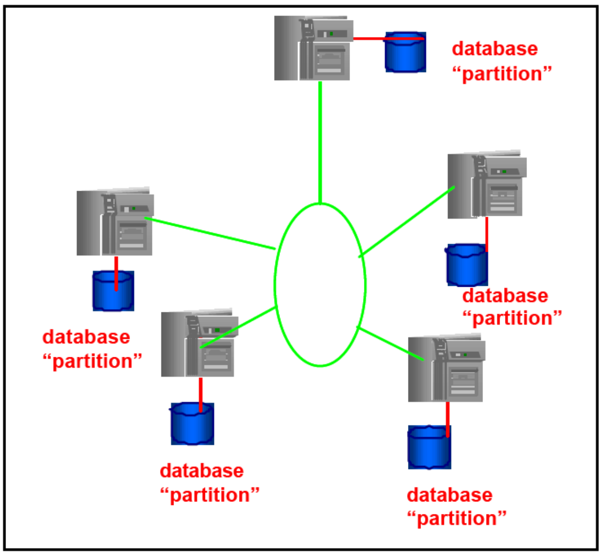


Figure 2.8. Federated databases (Buch, 2002)

The federated database system offers a powerful facility for combining different types of datasets from multiple data sources. Federated database is transparent to end users since it makes the set of federated data sources look like a single system (Muilu et al., 2007) and from the user's viewpoint, it looks like there is only one database (Sundaresan & Hu, 2005). In the federated database, heterogeneity emerges as the degree of differentiation in the various data sources. Sources can differ from different aspects. Different hardware, use of different network protocols, and having different software to manage the data stores are a few of many heterogeneity items (Haas & Lin, 2002).

A key characteristic of a federation is the cooperation among independent systems and organisations. This is reflected by a controlled and sometimes limited integration of the autonomous components. This kind of cooperation is often referred to as interoperability. To this extent, a federated database provides an explicit interface to its database components (Hammer & McLeod, 1993). Hence, federated model standards are developed jointly by the participating parties (Burgess, 1999).

2.7.3. Ontology-driven Spatial Data Integration

People, organisations and technical tools need to communicate effectively. However, due to the diversity of the requirements, purposes and backgrounds, there are widely varying understandings and assumptions among them. This leads to the difficulty in identifying requirements and in the definition of the specifications. The way to address these problems is to create a unifying framework for different viewpoints through the basic description of the entities (Uschold & Gruninger, 1996).

A very basic description of the real entities of the world, the description of truth is called *ontology* (Fonseca, 2001). Hence, ontologies are theoretical definitions that use a specific vocabulary to describe entities, their properties, and functions related to a certain view of the world (Fonseca et al., 2002). Ontologies provide ways to represent knowledge and resolve much heterogeneity (Harvey, 2002).

Many studies have been conducted to apply ontology to spatial data integration. Fonseca (2001) has introduced a framework for integration of multi-source spatial data based on ontologies. He created a mechanism that integrates spatial datasets based on their meanings. Uitermark (2001) has studied the ontology for update propagation and introduced the ontology-driven spatial data integration as a necessary condition for this purpose. Hakimpour (2003) has utilised ontologies to handle the semantic heterogeneity among diverse databases.

In another study that utilizes ontology for spatial data integration, Visser et al. (2002) highlight the involvement of various datasets from different areas ("e.g. data regarding the river, adjacent waste dumps, ground water flow etc.") in an environmental domain ("e.g. where does the high sulfate concentration in the river come from?"). This study addresses the problems of data integration including distribution and diversity in format.

2.7.4. Spatial Mediation/Mediators

Mediation has been developed as a means of providing the tools for the integration of multi-source heterogeneous datasets. Spatial mediation has been studied as a means of making use of these techniques to integrate spatial data (Miller & Nussar, 2003). Mediation has been utilised as one of the approaches for database integration. It also provides data access for distributed and heterogeneous data sources. The mediation technique capitalises on one global component (a mediator) and multiple local components (wrappers) (Pinto et al., 2003). The mediator accesses the information shared by the wrappers, which in turn manage the access to the local data sources. This architecture provides the binding of heterogeneous data to its domain application. Hence, for identification of the corresponding datasets and their appropriate domain application, each dataset is associated with a metadata (Gupta et al., 1999). This approach promotes data integration and mechanisms to translate data requests across ontologies as illustrated in Figure 2.9.

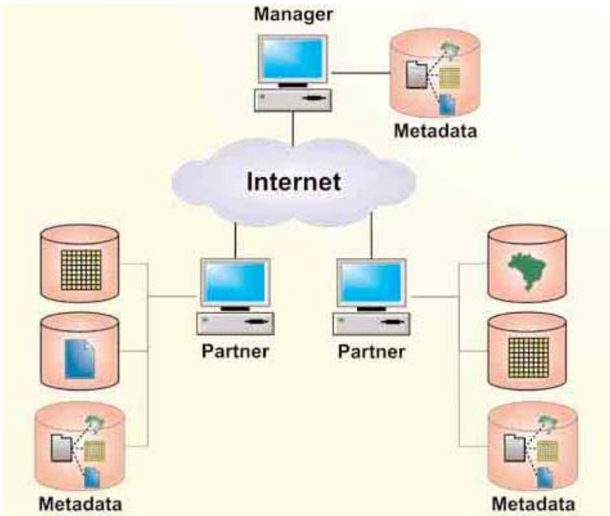


Figure 2.9. Mediator structor (Pinto et al., 2003)

One of the studies on spatial mediation that has been conducted by the San Diego Supercomputer Center (SDSC) has been very active in the development of spatial mediators (Miller & Nussar, 2003). This project extends to spatial information integration. For this purpose, spatial data and services are wrapped in a module that converts them to an XML document. The mediator dispatches user queries to multi-source heterogeneous data sources, and assembles query results. The structure of the system consists of three main components including (1) a network of "XML-wrapped" geographic sources and services; (2) the information mediation middleware that supports global resource discovery and query; and (3) an end-user interface capable of map rendering on a web client (Zaslavsky et al., 2000).

In another study Gupta et al. (1999) introduce a three-level architecture for mediator systems that includes a foundation level of wrapper, a mediation level and an application level with user interface (Figure 2.10).

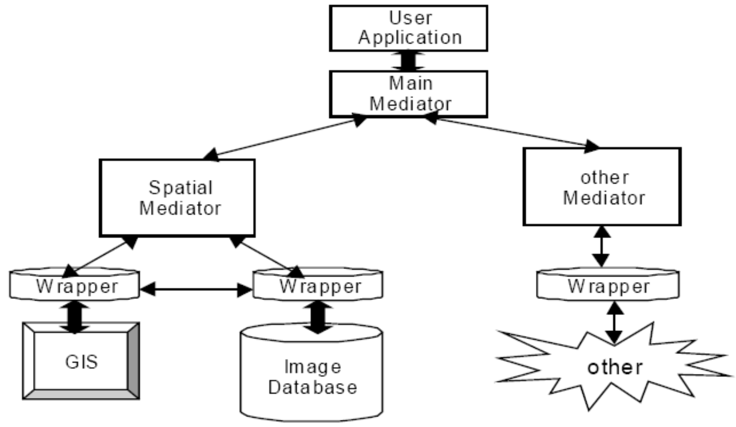


Figure 2.10. Mediator model proposed by Gupta et al. (1999)

2.7.5. Feature Manipulation Engine

Some organisations utilise spatial Extract, Transform and Load (ETL) tools to overcome the inconsistency among different data formats, data models and standards. ETL tools accommodate the spatial data transform tools.

The Feature Manipulation Engine (FME) is an integrated collection of spatial ETL tools. The FME has been introduced as a complete interoperability solution and eliminates barriers to spatial data integration and transforms and enhances data-sharing capabilities between diverse applications (Axmann, 2008). The FME works entirely on generic features and concentrates on providing building blocks that enable users to manipulate the data into the desired representation. The FME also allows the integration of different datasets, possibly of different types and different coordinate systems, into one logical dataset. Although this capability is most often used to merge data from adjacent datasets into a single dataset, it is also used to integrate data from several different sources through semantic translators (Visser et al., 2002).

Gerasimtchouk and Moyaert (2007) have applied the FME to integrate multi-source spatial data for the management of multi-source airport spatial data. This project has been initiated to bring a number of inconsistent datasets together and build a consistent integrated database.

2.7.6. Spatial Interoperability

The solution to many challenges of spatial data integration can be addressed by interoperability (Ouksel & Sheth, 1999; Goodchild et al., 1998), which has been backed by the definition of interoperability offered by Bishr (1998). Bishr defines interoperability as “the ability of a system or components of a system, to provide information portability and inter-application cooperative process control”.

In simple terms, interoperability is the ability of software and hardware on different machines from different vendors to share data (Webopedia, 2008). To be interoperable, one should actively be engaged in the ongoing process of ensuring that the systems, procedures and culture of an organisation are managed in such a way as to maximise opportunities for exchange and re-use of information, whether internally or externally (Miller, 2006).

As the central issue of geographic information science, spatial data interoperability has received considerable attention in recent years. Numerous spatial interoperability research projects and initiatives have addressed the heterogeneity among multi-source spatial datasets. The focus of interoperability research is moving towards the models and technologies that facilitate spatial data integration (Zaslavsky et al., 2005), hence the interoperability can be assumed as a part of integration (Sen, 2005).

Peedell et al., (2005), De Vries (2005), Goodchild et al. (2005) and Brodeur et al. (2003) have defined three levels of interoperability including syntactical, structural and semantic levels:

- **Syntactical** (related to different encoding, e.g. format conversion of databases –the output schema can remain the same) (Peedell et al., 2005). Syntactic interoperability helps to overcome the challenge of easy use of information including the integration of datasets for visualisation, query and analysis. The need for syntactic interoperability in multi-source and multi-vendor distributed software architecture is addressed with standards-based web services and vector formats like Geographical Markup Language (GML). But this is only the first step toward real information integration. There is also the structural and semantic heterogeneity aspect to solve (De Vries, 2005).
- **Structural** (related to differences in schemas, e.g. differences in attributes of two schemata). Structural interoperability provides a basic level of conversion from one schema to another (Peedell et al., 2005). Any spatial data is structured according to a certain conceptual view of the corresponding phenomena. This structure is strongly affected by the relevant organisation's vision; hence it differs from one dataset to another. As a consequence, the data model (database schema) is influenced. Structural interoperability deals with the inconsistency such as names, specifications, attribute names, granularity (many object types with few attributes, or few object types with many attributes), domain values, etc. (de Vries, 2005).
- **Semantic** (related to the differences in intended meaning of terms in specific contexts). Semantic interoperability provides the basis for more complex conversions (Peedell et al., 2005). Semantic interoperability is the ability of one user/system to understand the *meaning* of data from another user/system (Goodchild et al., 2005). Apart from differences in data structure, differences in information semantics will also stand in the way of unproblematic multi-source data integration. It is the specification, definitions and the meaning of the terms within a domain (Miller, 2006). The users within the source organisation understand the context, so they are implicit and seldom published. The major problem emerges when the data is reused by other organisations and this "inside knowledge" is not there (De Vries, 2005).

The structural and semantic heterogeneity of the data sources, in the worst case, can lead to inappropriate interpretation and unsuccessful integration (De Vries, 2005).

Sen (2005) has addressed different levels of interoperability to achieve effective integration. He proposes the fulfilment of three levels of interoperability as an ideal situation (Figure 2.11).

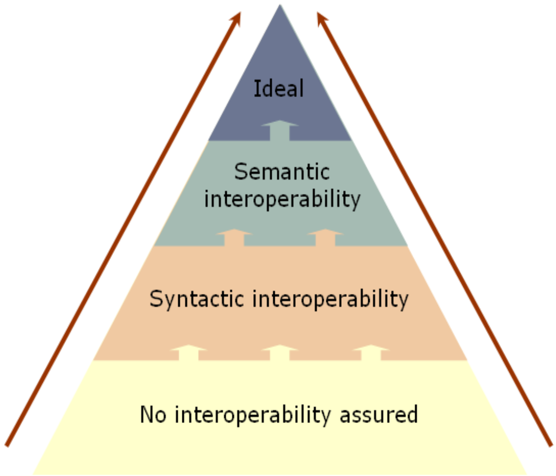


Figure 2.11. Value of interoperability (Sen, 2005)

In many ways interoperability at the technical level (data and services) is the most straightforward aspect of maintaining interoperability. Consideration of technical issues includes ensuring an involvement in the continued development of communication, transport, storage and representation standards such as Z39.50, the work of the World Wide Web Consortium (W3C), Open GIS Consortium (OGC), International Standards Organization (ISO), etc.

There is far more to ensuring interoperability than using compatible data, software and hardware. Rather, assurance of effective interoperability will require often radical changes to the ways in which organisations work and especially in their attitudes to information (Miller, 2006). Hence, besides interoperability at the technical level, interoperability at institutional, policy and social levels (Figure 2.12) needs to be established to ensure effective integration (Sen, 2005) and interoperation among the organisation's data and systems (Mohammadi et al., 2007).

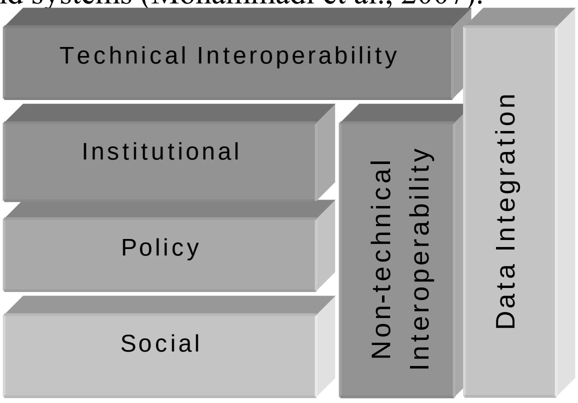


Figure 2.12. Technical and non-technical interoperability for effective spatial data integration (Mohammadi et al., 2007)

Researchers have tried to tackle the barriers to multi-source spatial data integration through the development of different techniques and tools. These approaches have been focused to overcome the inconsistency among datasets. For example, data interoperability introduces a set of standards and guidelines that prepares the compliant and consistent data. These approaches discuss data integration at data level, while the process level of data integration requires more studies and development of not only technical tools but also non-technical arrangements and mechanisms.

2.8. Data Level and Process Level Data Integration

The above-mentioned data integration approaches have been made to facilitate the integration of multi-source spatial data. Most initiatives have proposed solutions for available data. They have not investigated any potential barrier of data acquisition, availability and sharing. They try to overcome the issues of and barriers to integration at data level (Figure 2.13) and not at process level.

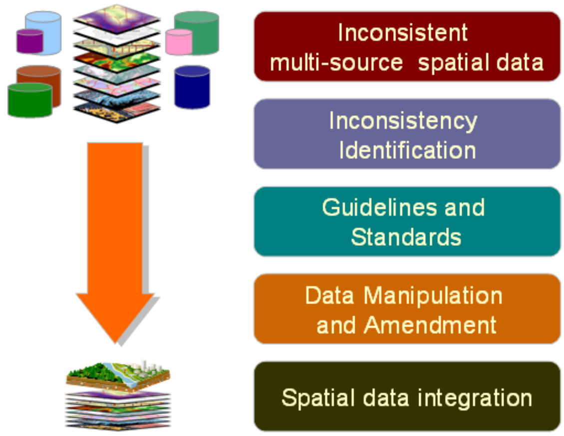


Figure 2.13. Data level integration

However, there are many technical and non-technical issues in the process of data integration which are related to the stages prior to the conflation of data. Therefore, two levels of data integration can be defined namely, data level and process level integration.

At data level integration, Kolbe (2006) has divided the integration of spatial data into two levels: model and visualisation levels. If the objective is met with integrated visualisation only, it is adequate to combine the graphical representation of spatial datasets. At this level, the combination of the spatial data content (spatially and aspatially) is not necessary. Kolbe also identifies a higher degree of integration that is done on actual data, which is model level integration. At this level, the entities and data model are integrated.

Ulubay and Altan (2002) have introduced a two-level model for spatial data integration in another study. In this model multi-source spatial datasets can be integrated at visual integration or analytical integration levels. Visual integration comprises the processes of:

- Matching coordinate systems and projections
- Compatible scale
- Compatible formats
- Visualisation.

The analytical level comprises:

- Visual integration
- Attribute matching.

Data model matching is another stage in the analytical level which is quite essential for organisations and businesses that maintain and coordinate integrated products (Geiger, 1998). For some organisations (e.g. PSMA Australia) that integrate spatial data at state and national levels, dealing with diverse data models is of great significance. In such cases different data models that are utilised by source agencies require being integrated and matched.

However, there has not been a holistic study at process level integration (Rajabifard & Williamson, 2004b). At a process level, technical and non-technical issues associated with any stage of spatial data integration should be addressed. Drawing from Ostensen (2006), who has identified major spatial activities, this researcher has identified six major integration tasks (Figure 2.14) as follows.

- Data exploration (discovery)
- Data assessment
- Communication with data providers
- Access to data and data acquisition
- Geo-processing
- Data collation.

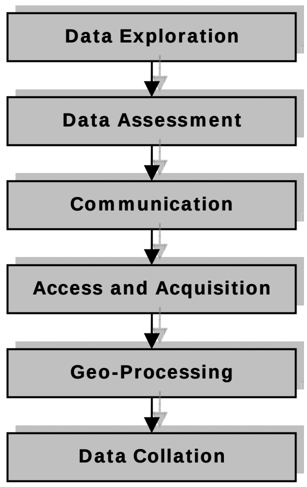


Figure 2.14. Spatial data integration tasks

2.8.1. Data Exploration

Appropriate data exploration leads to finding the data and data source that suit user needs. It requires good knowledge of data characteristics and the place where data sits. This knowledge can be achieved if information on available datasets is consistent and publicly available.

Comprehensive and consistent metadata, which provides information on data description, characteristics and providers together with an effective data dictionary as a single metadata access channel, is necessary for effective data exploration. Implementing consistent and comprehensive metadata and a data dictionary requires a concrete custodianship framework be in place to identify data custodians and data sources. Populating data and offering awareness on available products (Paull, 2006) are also essential in implementing an effective data exploration service.

2.8.2. Data Assessment

Data exploration is by no means an end in itself as explored data needs to be assessed from different perspectives including fitness-for-purpose, legal issues surrounding data, pricing and data quality. To evaluate datasets for integration, they need to be assessed against characteristics to identify their suitability for integration.

Metadata can play a significant role in providing users with information on assessing data. Product description documents are also helpful for this purpose.

2.8.3. Communication with Data Providers

Data exploration is important and in some cases leads to data acquisition, but in most cases, where users need accurate, detailed, business datasets, they need to find a communication channel with data providers. In some cases users know where data lies but they cannot find any channel to communicate and collect data. Easy access to an effective communication channel requires the provision of tools to link users to data providers including data dictionaries.

2.8.4. Data Access and Acquisition

Access to data and data acquisition can also be challenging if data access and acquisition tools are not available. These tools comprise not only technical tools including web services and single point of access, but also provide non-technical mechanisms including legal, social, policy and institutional considerations to facilitate data access and acquisition.

2.8.5. Geo-processing

Applications that rely on more than one data source mostly encounter inconsistency of datasets. Data providers follow their own policies, standards and specifications in creating and maintaining data that may differ from other data providers. Inconsistencies in format, data models, projections, attributes, standards and specifications are some of them.

To deal with these issues, users need to convert some characteristics of data to match other datasets. Most of these matching processes are technical including converting format and datum, manipulating attributes and data structure. However, there are also non-technical issues including legal considerations involved in this process. One of the most important legal issues is the way in which privacy issues of original datasets and IP issues affect final integrated products.

2.8.6. Data Collation

This stage is the same as the data level integration. Data collation is the final step in data integration and if done properly can lead to an effective decision or analysis. Data collation can be done at different levels based on user requirements. Kolbe (2006) identifies two levels (visualisation and data model) for data integration (Figure 2.15). If superimposition of data meets user needs, geometrical integration of data is then the ultimate goal. In some cases users need more than geometrical collation of datasets. For example, they may want to establish a link and relation between the features of different datasets. Hence, the more complex the data integration (higher integration level), the more sophisticated the tools and the effort required to collate datasets.

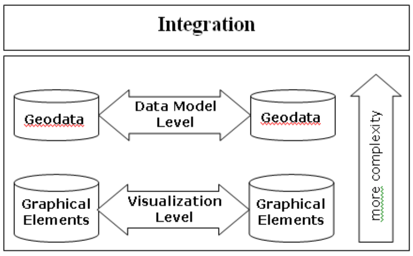


Figure 2.15. Visualisation and data model integration levels (adopted from Kolbe, 2006)

Due to the high amount of time and cost needed to deal with inconsistencies, data collation is problematic, time consuming and an expensive process at higher integration levels. Also metadata should provide information on data collation including data model and attributes.

To address a process level of spatial data integration a holistic framework is required which possesses the necessary technical and non-technical mechanisms to overcome data integration barriers (Figure 2.16).

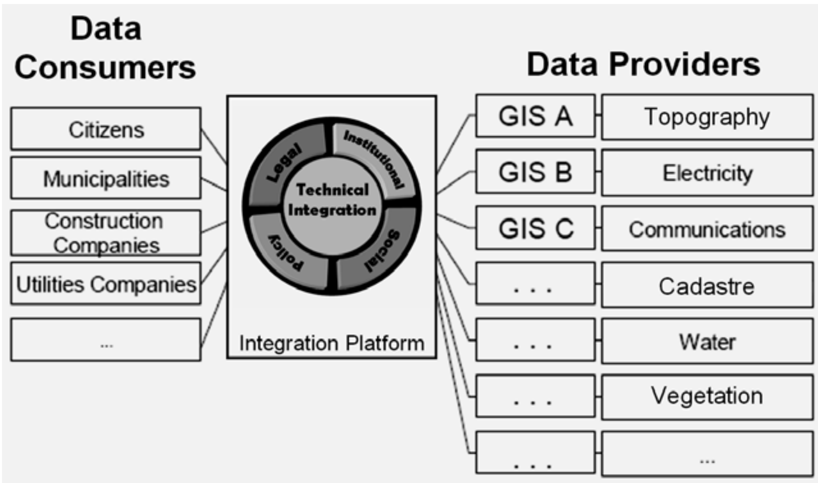


Figure 2.16. The holistic integration platform

This platform should have a number of characteristics to be able to accommodate effectively the spatial data integration initiatives. Spatial data and its effective relation to the provider organisations and consumers should be established. Standards and specifications should be developed in accordance with the requirements of the stakeholders and the advancement of the technical tools. Policies and institutional arrangements are not distinct parts, therefore the proposed framework should accommodate policy making and institutional arrangements along with special attention to the social behaviour and issues of the stakeholders involved.

Rajabifard and Williamson (2004a) and Morales (2006) believe that Spatial Data Infrastructure (SDI) possesses the above-mentioned characteristics. They highlight the needs for effective data integration and propose SDI for this aim where one of its key objectives is facilitating access to and the integration of a wide range of spatial data from various custodians and agencies and different jurisdictional levels. Sen (2005) also highlights the comprehensiveness of SDI in data sharing and integration from technical and non-technical aspects.

Over the last decade, the establishment of SDI was seen as a means of overcoming data access and sharing impediments. Indeed a key SDI objective is to facilitate the integration of multi-source datasets (Rajabifard & Williamson, 2004a). The SDI solution is described in many publications and international declarations including the Bogor declaration (FIG, 1996), which urges the establishment of national SDIs to ensure uniform data integration. Despite research into models and development, the design of SDIs capable of integrating multi-source spatial data remains problematic (Parker & Enemark, 2005). Many jurisdictions face difficulties in integrating spatial data (UNRCC-AP, 2003). The future progress of SDI establishment now depends on more thorough investigation of impediments to the integration of multi-source spatial datasets (UNRCC-AP, 1997).

Individual countries are also looking to highlight the importance of study on spatial data integration in the context of SDIs. The Australian Spatial Information Council (ANZLIC) has identified the integratability of multi-source spatial datasets as a priority area for implementation of the Australian SDI (ANZLIC, 2003a). The Canadian Geospatial Data Infrastructure (CGDI, 2001) indicated the integration of disparate geospatial information as a major component of CGDI architecture. In Europe, the Infrastructure for Spatial Information in Europe initiative (INSPIRE) recognised spatial data integration (combination) as one of its principles (INSPIRE, 2006). Therefore, the INSPIRE implementation will gradually harmonise data and information services, eventually allowing the seamless integration of systems and datasets at different levels into a coherent European SDI (INSPIRE, 2002).

An SDI is a platform that facilitates the interaction between people and data by providing required access channels, policies and standards (Rajabifard & Williamson, 2001; Nebert, 2004; Masser, 2006b) as illustrated in Figure 2.17.

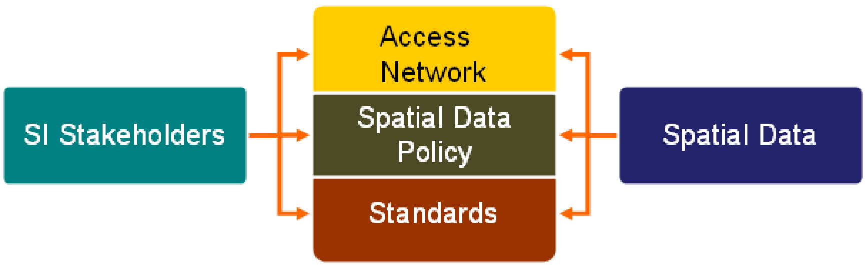


Figure 2.17. SDI and its components (Rajabifard and Williamson, 2001)

2.9. Chapter Summary

Multi-source spatial data integration is essential for many spatial applications for delivery of their objectives. At the same time, the diversity of stakeholders involved in the coordination and distribution of datasets causes much inconsistency among datasets.

This chapter discussed the importance of data integration for spatial stakeholders and applications. The chapter then listed a number of drivers for spatial data integration including controlling datasets, producing a more complete dataset, raising the analytical level of datasets, horizontal/vertical integration, complying with the uniformity of spatial services and providing service to multi-disciplinary science. The chapter highlighted issues and problems associated with effective data integration. Issues of data integration have been discussed in two main categories: technical and non-technical issues.

In order to study the previous initiatives and studies in the area of spatial data integration, major initiatives including generalisation, federated databases, ontology-driven data integration, spatial mediators, FMEs and interoperability have been discussed which led to the discussion of a holistic framework to manage diverse technical and non-technical aspects of spatial data integration.

This chapter then introduced and generally highlighted the characteristics of SDIs to meet the objectives of effective data integration and identified the need for research to investigate issues of multi-source data integration within SDI initiatives. The primary goals were to open the discussion of data integration within the context of SDIs and to design and present a methodological framework to facilitate the integration of multi-source data sets. This framework then can serve multi-disciplinary applications and can advance knowledge of SDI capacity by facilitating data integration, which will be discussed in Chapter Three.

Chapter Three (Concepts of SDIs) will give an insight to SDIs and the strengths that SDIs offer for effective spatial data integration. The chapter discusses the SDI hierarchy and SDI components. It also highlights the role of SDI components to address and facilitate issues and problems of multi-source spatial data integration.

Chapter 3

Spatial Data Infrastructures: Concept and Theory

3.1. Chapter Aims and Objectives

This chapter aims to present the concept of Spatial Data Infrastructures (SDIs) as an enabling platform to accommodate the necessary technical tools and non-technical guidelines, standards and instructions for effective spatial data integration. In this regard the continuum from information to information infrastructure and spatial data infrastructure is presented. It is accompanied by the diverse definitions of SDIs and SDI components within different communities. Hence, it discusses Spatial data sharing as one of the most significant components of SDIs. It also addresses the integration of spatial data as a compelling reason for data sharing. Spatial data integration as one of the success factors of SDIs has been discussed.

3.2. Introduction

Practitioners use information to build a realistic model of the world. This model is used for informative decision making. In order to achieve this aim it is necessary that the information can be shared and accessed effectively. Information infrastructure allows effective information transmission and sharing. Information infrastructure consists of facilities, processes and standards to provide information to the users. Information infrastructure also creates new information from existing knowledge. Further, it enables information users to share knowledge with others, through information sharing and exchange.

Spatial information is a specific form of information and is a key commodity to the present day. Spatial information differs from other information forms. Spatial information is any information that can be linked to a geographic location. It is necessary to understand that spatial information is complex and while it can be included and managed in databases alongside other information, its coordination and maintenance require special skills. Spatial information enables the delivery of good governance and efficient business. As a result, spatial information must be accessible for analysis and use by decision-makers. SDI aims to facilitate the sharing, exchange and integration of multi-source spatial information through the provision of standards, policy framework, access and the establishment of partnerships and collaborations among spatial stakeholders.

SDI is a combination of technology, policies, standard and human resources that is necessary to facilitate and coordinate the exchange and sharing of spatial information between stakeholders of the spatial community. Data integration is a compelling reason for sharing data. Data integration is the ability to share access to data sources or access common databases. In order to achieve this aim, different components have been identified for SDIs. These include spatial information, people, institutional arrangements, standards, metadata, access network, partnerships, governance and capacity building.

The next significant advancement in SDI development is the spatial enablement of society and governments, in which location and spatial information are regarded as common goods that leverage effective business and efficient governance.

3.3. From Spatial Data to Spatial Data Infrastructure

The term *infrastructure* typically represents the facilities such as roads, sewer lines, electric lines, airport and similar physical structures or networks, for whose construction and support the government is mainly responsible. Therefore the term *information infrastructure* (II) allows more effective transmission and sharing of information in support of the public good. It also consists of facilities, processes and standards by which essential information is made available to citizens, business, scientists and other governmental agencies and sectors (Onsrud et al., 2004).

The evolution of societies from industrial form to information is derived from data and Information and Communication Technology (ICT) advancements. One of the characteristics of modern societies is the focus on publicising digital data and building information infrastructures. The publicity of information causes the diversity and heterogeneity among multi-source information, hence the effectiveness of information infrastructures greatly depends on integrated systems and information. This wave requires the ability of organisations and systems to cooperate and liaise, which means information becomes interrelated (Muggenhuber, 2003).

3.3.1. Information Infrastructure

Information allows practitioners to model a virtual world for orientation and decision making. Therefore information provides the basis for measuring and controlling the world through observations. In order to make this mechanism work, an information infrastructure is required (Muggenhuber, 2003). The information infrastructure helps the maximum use of information by more users and also facilitates the trade of digital information as a commodity (Warnest, 2005).

The term information infrastructure has been widely used during the last few years. It gains its rhetorical thrust from certain so-called visions. These visions were initiated by the Clinton (Executive Office of the President, 1993) and Gore (Gore, 1998) plans and followed up by the European Union's plan for Pan-European II (Hanseth & Monteiro, 1998).

In September 1993, the US President's Administration announced the construction of a National Information Infrastructure (NII). NII aimed at some critical principles and objectives (Executive Office of the President, 1993), including:

- To ensure that information resources are available to all at affordable prices
- Promote seamless, interactive, user-driven operation of the NII
- Ensure information security and network reliability
- Protect intellectual property rights
- Coordinate with other levels of government and with other nations
- Provide access to government information and improve government procurement.

NII was mainly aimed at changing the way people live, work, and interact with each other. In the context of the European Union, the building of a European II would satisfy the following expectations:

- A more caring European society with a significantly higher quality of life, services and entertainment
- New ways to exercise creativity for content providers
- More efficient, transparent and responsive public services

- Enterprises to gain more effective management and organisation, improved access to training and other services
- Europe's telecommunications operators to gain the capacity to supply an ever wider range of new high value-added services
- New and strongly growing markets for the computer and consumer electronics industries.

Information infrastructure refers to the communications networks and associated software that support interaction among people and organisations (Clarke, 2004). The internet was a primary driver for the recognition of information as an infrastructure (McDougall, 2006). It is a widespread information infrastructure, the initial prototype often called the Global Information Infrastructure (Hanseth & Monteiro, 1998).

Building an information infrastructure does not just facilitate the availability of information (searchability, discoverability and accessibility) but it also provides links as a bridge from existing knowledge to new information. Information exchange as one of the major processes within information infrastructure enables information users to share knowledge with others. In order to overcome the barriers to data exchange and data sharing, it is not sufficient to deliver data. New knowledge can only be generated when existing information can be linked to existing knowledge. Therefore, a different view on information results in a different knowledge, which also depends on existing knowledge.

The distinction between concepts like data, information, knowledge and wisdom is not discrete and is a continuum (McDougall, 2006). Clark (2004) has considered data as the facts created through research, gathering or discovery (Figure 3.1).

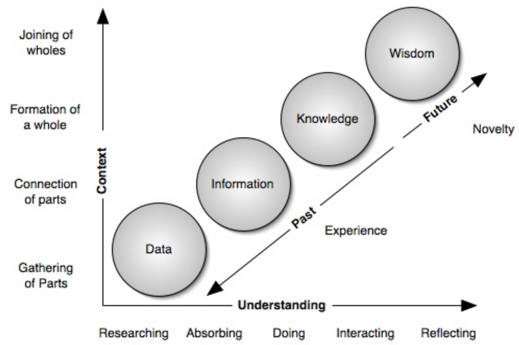


Figure 3.1. Continuum of data to wisdom (Clark, 2004)

By organising it, data is turned into information, so that we can easily draw conclusions. Therefore, information has context. Data can also be turned into information by "presenting", such as making it visual or auditory (Cleveland, 1982). Knowledge has the complexity of experience, which comes about by seeing it from different perspectives. This is why training and education are difficult – one cannot count on one person's knowledge transferring to another. Knowledge is built from scratch by the learner through experience. Information is static, but knowledge is dynamic as it lives within us. Wisdom is the ultimate level of understanding. As with knowledge, wisdom operates within persons. One can share experiences that create the building blocks for wisdom; however, it needs to be communicated with even more understanding of the personal contexts of our audience than with knowledge sharing.

Often, the distinctions between the data, information, knowledge, and wisdom continuum are not very discrete, thus the distinctions between each term often seem more like shades of grey, rather than black and white (Shedroff, 2001).

It was identified that a list of resources (including human resources, natural resources, utilities, spatial data infrastructure) which contribute to the development of infrastructures for a modern society highlight that most of these infrastructures have a spatial component.

3.3.2. Spatial Information

Spatial information is another form of information. About 80 per cent of all information utilised by decision-makers is spatial information (Ryttersgaard, 2001; Klinkenberg, 2003). Spatial information (also known as geographic information) is any information that can be geographically referenced (WALIS, 2008), i.e. describing a location or any information that can be linked to a location (ANZLIC, 2008).

Spatial information can now be stored and coordinated in databases, but the specific characteristics of spatial information make it a different form of information. Spatial information is scale-dependent. It is dependent on data models that are diverse and have many dimensions (Williamson, 2006). The size of spatial information and the need for management of spatial and attribute information require a specific set of tools and arrangements (Egenhofer, 1993). Hence, understanding the collection, management, manipulation, integration, use, presentation and querying of spatial information requires special skill sets.

The term *spatial information* is used almost interchangeably with *spatial data*, *geospatial information* or *data* and *geographic information* (Warnest, 2005). However, data and information have been approached differently in the continuum of data to wisdom. The amount of spatial data circulated among stakeholders creates enormous opportunity for spatial services to serve more consumers. Spatial information is a key and integral component for the delivery of good governance, defence, promoting efficiency in business and supporting sustainable development. It provides enabling technology for modern societies. It is also recognised as fundamental for wealth creation and good decision making.

As a result, spatial information is becoming the inseparable component of many existing and new/emerging services including emergency management, health services and census. Therefore, policy-makers and managers have begun to realise the value of spatial data for their business. They consider spatial data as a resource and also a part of fundamental infrastructure that needs to be coordinated and managed effectively (Ryttersgaard, 2001).

Spatial information plays a significant role in many social, economic and political decisions. Governments, business and the general public rely heavily on spatial information for their daily decision making (Onsrud & Rushton, 1995). The utilisation of spatial information and spatial services is a suitable means to optimise the sustainable management of resources (Muggenhuber, 2003). Former American Vice President, Al Gore (1998), took this vision one step further and insisted on geo-referenced information as the building block of a national information infrastructure. The Australian Government also pushed spatial information infrastructure to the fore, initiating spatially enabled government (Williamson, 2006).

3.4. Spatial Data Infrastructure

Over the last few years governments at state and national levels have given much attention to spatial data infrastructure (SDI), which is driven by business needs and technological developments to support both the government and the rapidly expanding spatial information industry (Williamson et al., 1998).

SDIs facilitate the sharing of data, by avoiding duplication associated with generation and maintenance of data and integration with other data sets (Williamson et al., 2003). SDIs aim to provide a framework based on spatial data with which spatial data stakeholders can interact easily.

Within SDIs spatial data can be used to its maximum potential by facilitating sharing, access and the integration of datasets. The integration of spatial data adds more value to the data; however, it is a time-consuming and costly process. Therefore, facilitating the integration of spatial data will save time and cost for spatial data users.

SDIs are more than a single dataset or database, and include geographic data and attributes, sufficient supporting documents including metadata and standards, a number of tools to facilitate discovery, visualisation, and evaluation of datasets together with methods to provide access to the data. To make an SDI operational, organisational agreements and capacity building are also required (Ryttersgaard, 2001).

3.4.1. SDI's Birth

Every nation undertakes the development of strategic national mapping and spatial data activities to meet their national planning and management needs. Bringing these activities together over time has resulted in the identification of key linkages between organisational and technical aspects which are quite similar in many respects to other forms of infrastructure.

The concept of SDI was first introduced in the mid 1980s around the need for cooperation and sharing of spatially related information across a nation. In the US, discussion about the national SDI initiative initially began primarily in the academic communities around 1989 and soon after in the government. These discussions progressed especially rapidly when in the early 1990s, the National Research Council's (NRC) Mapping Science Committee articulated the way that spatial information needed to be handled from an institutional perspective (Onsrud et al., 2004) and after the executive order from the President's office was issued in 1994 (Executive Office of the President of the US, 2002).

The recognition of the importance of SDI for the governments was accompanied by the formation of the Federal Geographic Data Committee (FGDC) in 1990 (McDougall, 2006). Since then, the FGDC attempted to develop a coordination framework, standards and the documentation of best practices in accordance with the national SDI objectives in building a national digital spatial data resource. In Australia, national land-related information initiatives commenced with a government conference in 1984, which eventually led to the formation of a committee responsible for SDI development (Williamson et al., 2003).

These initiatives, then followed by a number of other initiatives, characterised the first wave of SDI development. From 1999 to 2005, the Canadian Federal Government put $60 million in funding towards a national partnership initiative to make Canada's geospatial information accessible on the internet, while provincial and territorial governments and the private sector are investing over $50 million in funding (GeoConnections, 2008). GeoConnections developed the policies, standards, technologies, and partnerships needed to build the Canadian Geospatial Data Infrastructure (CGDI).

In Europe, the Commission of the European Communities (2004) submitted a proposal to the European Parliament and the Council of the European Union. The proposal aimed to make interoperable spatial information readily available in support of both national and community policy and to enable the public to access to this information. This was a major milestone for the use of spatial information in Europe as a contribution to environmental policy and sustainable development. It was the first step in a co-decision procedure that led to the formal adoption of the pan-European SDI (INSPIRE, 2007).

In the mid 1990s, Global SDI was formed with a special focus on promoting international cooperation and collaboration in support of local, national and international spatial data infrastructure developments that will allow nations to better address social, economic, and environmental issues of pressing importance.

GSDI aims at providing a point of contact and an effective voice for jurisdictions in the global community involved in developing, implementing and advancing spatial data infrastructure concepts to foster spatial data infrastructures that support sustainable social, economic, and environmental systems integrated from local to global scales, and to promote the informed and responsible use of geographic information and spatial technologies for the benefit of society (GSDI, 2008).

Also, many of the countries around the world are developing SDI at different jurisdictional levels. Each jurisdiction has its own definition of SDI that springs from jurisdictional backgrounds and requirements.

3.4.2. SDI Definition

Masser (1999) addresses the diversity of approaches to SDI definition and development and believes that SDI has developed in all shapes and sizes. Within the SDI community there are differences in the understanding of SDI and its potential benefits (Grus et al., 2007), therefore SDI is viewed, defined and interpreted differently by different practitioners. The European Commission (2006) highlights this as one of the most challenging obstacles for SDI assessment and development. It also argues that there is much confusion resulting from the lack of an agreed definition of SDI, its components and the relationships between them.

Table 3.1. shows a number of SDI definitions and perspectives.

Table 3.1. SDI definitions by different communities

Source	Definition
Executive Office of the President (1994)	SDI means the technology, policies, standards and human resources necessary to acquire process, store, distribute, and improve the utilisation of geospatial data.
Brand (1998)	A Global Spatial Data Infrastructure is one that encompasses the policies, organisational remits, data technologies, standards, delivery mechanisms and financial and human resources necessary to ensure that those working at the global or regional scale are not impeded in meeting their objectives.
Coleman & McLaughlin (1998)	SDI encompasses the policies, technologies, standards and human resources necessary for the effective collection, management, access, delivery and utilisation of geospatial data in a global community.
ANZLIC (2003b)	SDI is a framework for linking users with providers of spatial information. SDI comprises the people, policies and technologies necessary to enable the use of spatially referenced data through all levels of government, the private sector, non-profit organisations and academia.
Wikipedia (2008)	An SDI is a framework of spatial data, metadata, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way. Another definition is the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data.
Groot & McLaughlin (2000)	SDI encompasses the networked geospatial databases and data handling facilities, the complex of institutional, organisational, technological, human and economic resources which interact with one another and underpin the design, implementation and maintenance of mechanisms facilitating the sharing, access to, and responsible use of geospatial data at an affordable cost for a specific application domain or enterprise.
Rajabifard et al. (2002)	SDI is fundamentally about facilitating and coordinating the exchange and sharing of spatial data between stakeholders in the spatial community.
Nebert (2004)	SDI is a collection of technologies, policies and institutional arrangements that facilitates the availability of and access to spatial data.
GSDI (2005)	SDI supports effective access to geographic information. This is achieved through the coordinated actions of nations and organisations that promote awareness and implementation of complementary policies, common standards and effective mechanisms for the development and availability of interoperable digital geographic data and technologies to support decision making at all scales for multiple purposes.

The above table encompasses most of the SDI definitions. These definitions mainly emphasise the facilitation of data access, sharing and use. They also urge on the interaction between spatial data stakeholders and spatial data through a number of technical and non-technical components including people, fundamental data, technology, metadata, standards, policies, institutional arrangements and financial resources. The next section will articulate the components of SDI within different SDI communities.

3.4.3. SDI Components

The core elements that comprise an SDI have been defined differently by different communities. Rajabifard and Williamson (2001) have proposed a five-tier SDI structure. This model proposes that the fundamental interaction between spatial data and the stakeholders (people) is governed by the dynamic technological components of SDI including access networks, policies and standards as shown in Figure 3.2.

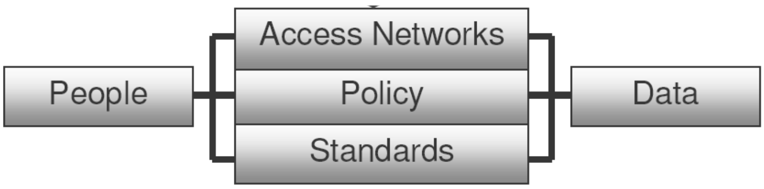


Figure 3.2. The SDI model (Rajabifard and Williamson, 2001)

ANZLIC (2003b) has also named three key components for Australian SDI, which are spatial data directory, standards and metadata.

The Executive Office of the President of the United States (2002) has introduced five components for US national SDI. The components of the SDI are fundamental data themes, metadata, the National Spatial Data Clearinghouse, standards, and partnerships (Figure 3.3).

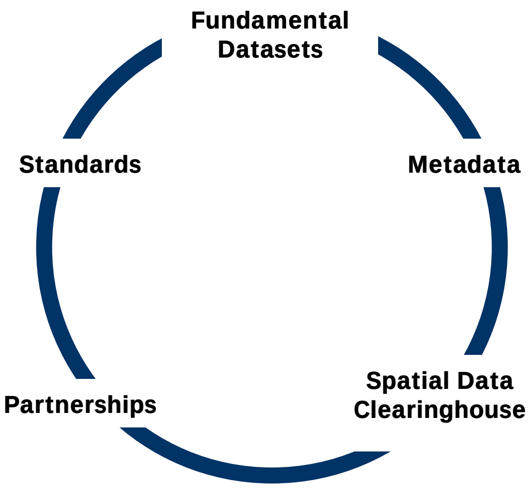


Figure 3.3. SDI components (Executive Office of the President of the US, 2002)

In Canada, the CGDI has identified five main components for Canadian SDI, including technology, policy, framework, standards and access network as illustrated in Figure 3.4 (GeoConnections, 2008).

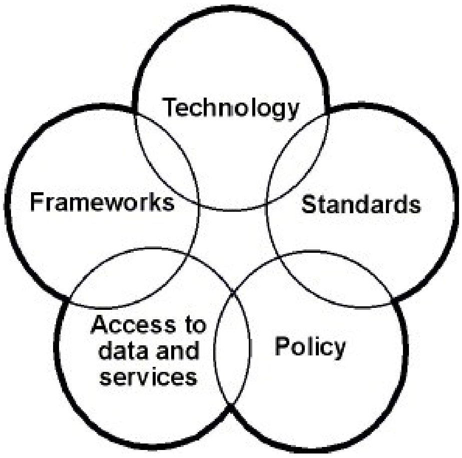


Figure 3.4. SDI components identified by GeoConnections (2008)

In recent years, as the concept and the development of the SDI framework have matured, the role of some other elements has been greatly realised. In particular, capacity building, spatial data sharing, partnership and governance have been recognised to have a great impact on the effectiveness and success of SDIs.

Table 3.2. summarises the most important components of SDI.

Table 3.2. Major SDI components

Data	Data themes are electronic records and coordinates for a topic or subject, such as elevation or vegetation. Themes providing the core, most commonly used set of base data are known as framework data, specifically geodetic control, orthoimagery, elevation and bathymetry, transportation, hydrography, cadastral, and governmental units.
People	Includes stakeholders who use, provide, value-add, manage or own the data. Users can be corporate, small and large business or individuals, public and private sectors.
Institutional Framework/Policy	Includes the administration, coordination, policy and legislative components of an SDI. The institutional framework is reliant on successful communication and interaction between stakeholders within and across jurisdictions.
Standards	Standards are common and repeated rules, conditions, guidelines or characteristics for data, and related processes, technology and organisation. To broaden the global use of federal data and services, international standards and protocols must be used.
Metadata	Metadata, commonly defined as “data about data”, is a structured summary of information that describes data (SEDAC, 2006). Metadata contains information about data and/or geospatial services, such as content, source, vintage, spatial scale, accuracy, projection, responsible party, contact phone number, method of collection, and other descriptions. Metadata is critical to document, preserve and protect agencies’ spatial data assets.
Access Network	Includes access and distribution networks, clearinghouse and other mechanisms for getting spatial information and data to the stakeholders
Partnership	Partnerships are critical components of SDI development, which can be inter- or cross-jurisdictional (Williamson et al., 2003).

	Building an effective SDI will require a well-coordinated partnership among federal, state, local government, and academic institutions, as well as a broad array of private sector geographic, statistical, demographic, and other business information providers and users.
Data Sharing	Spatial Data Sharing (SDS) is defined as transactions in which individuals, organisations or parts of organisations obtain access from other individuals, organisations or parts of organisations to spatial data (Omran, 2007a).
Governance	It is necessary to go beyond establishing the machinery for SDI coordination and give top priority to the creation of appropriate SDI governance structures that are both understood and accepted.
Capacity Building	SDIs are likely to be successful when they maximise the use made of local, national and global GI assets in situations where the capacity exists to exploit their potential. The creation and maintenance of SDIs are also a process of organisational change management. Capacity building is important in less developed countries where the implementation of SDI initiatives is often dependent upon a limited number of staff with the necessary GI management skills. Further, there is still a great deal to be done to develop GIS capabilities in many more developed countries, particularly at local level.

3.4.4. Hierarchical Nature of SDI

Many countries are recognising more inclusive models of SDI governance to meet the requirements of a multi-level multi-stakeholder SDI. Therefore, it seems necessary to go beyond establishing the machinery for SDI coordination and prioritise the creation of appropriate SDI governance structures. Obviously, it will often not be possible to bring all stakeholders together for decision-making purposes, and structures must be devised for keeping all informed and giving them an opportunity to have their opinions heard. The simplest solution is to create hierarchical structures at national, state and local levels.

Rajabifard et al. (2002) has proposed a hierarchical system for SDIs that should be viewed from a perspective in which each level of development supports the higher level of development (Figure 3.5).

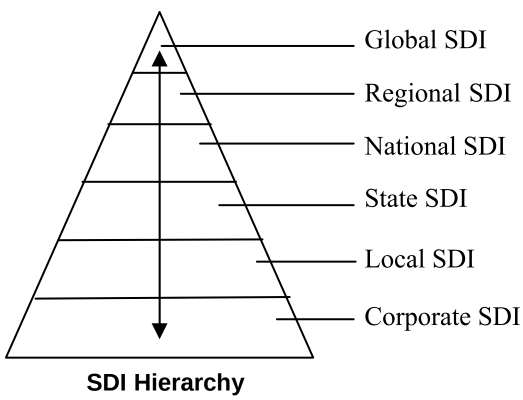


Figure 3.5. SDI hierarchy (Rajabifard et al., 2002)

The main reason that a hierarchy concept is applicable to SDIs is that all properties and reasons for developing a hierarchical structure are applicable to the SDI concept (Williamson et al., 2003). An SDI at a high level, like a global level, consists of one or more SDIs from a lower level. The hierarchical structure is common in modelling many manmade and natural systems (McDougall, 2006). In hierarchy, any element of the higher level consists of one or more elements from the lower level (part-whole property). Any element at any hierarchical level also has two faces. One face looks towards wholes in a higher level and the other looks towards parts in a lower level. This property is called the Janus Effect. Another property of a hierarchical structure is decomposability. It represents the nesting of the system within larger sub-systems. It also states that interaction between various systems decreases in strength with distance from other systems. Masser (2005) believes that although the properties and characteristics of the hierarchical system might be essential for the development of a consistent data structure, the absence of a strict hierarchical structure does not necessarily inhibit the implementation of SDI initiatives. For example, in the case of the US, the FGDC work directly with local governments without reference to the state level. This kind of structure is already operational to some extent in some countries including Australia and is implicit in the proposals for a 50-state initiative in the US (Masser, 2006b).

3.5. Spatial Data Sharing

SDIs are thought to have a dynamic structure. This is addressed by both change in the nature of SDIs and the external environment including the advancement in technology (Rajabifard & Williamson, 2003). The dynamic environment of SDI presents uncertainty for the organizations involved (Omran, 2007b), which leads them to focus on cooperative relationships. One of these relationships is data sharing. Spatial data sharing has been defined as transactions in which individuals, organisations or parts of organisations obtain access from other individuals, organisations or parts of organisations to spatial data (Omran, 2007a). A coordinated approach to sharing spatial information will result in a number of benefits to participants, including:

- reduction in the duplication of datasets, systems and processes
- sharing of investment costs (Montalvo, 2003)
- higher quality datasets
- improved access to spatial data
- development of partnerships across the entire spatial sector (public, private and academia).

Other benefits of this approach include certainty and security for custodians who make their data available through such networks; interoperability; adoption of common standards for data; and broader coverage of data across multiple jurisdictions and sectors (Victorian Spatial Council, 2005).

The capacity to meet such user needs and deliver services and tools within the spatial information community has gone far beyond the ability of single organisations (Rajabifard et al., 2005a). There is now a wide range of products and services available for a wide range of information technology applications, and hence the development of an enabling platform can facilitate access to data and sharing resources and tools among different practitioners (Omran, 2007b).

The sharing of spatial data involves more than simple data exchange. In order to facilitate the spatial data sharing, spatial stakeholders need to deal with many issues including the technical and non-technical aspects of data integration (Onsrud, 1995). The appropriate focus for sharing data is data integration. Integrating data in a spatial system increases its effectiveness and creates opportunities for wider enterprise benefits that accrue to entire organisations and consistencies. Data integration facilitates the ability to share access to data sources or access common databases (Montalvo, 2003).

3.6. SDI and Data Integration

The value of a spatial dataset rests on its "coverage, the strengths of its representation of diversity, its truth within a constrained definition of that word, and on its availability" (Longley et al., 2001) and the integrability with other datasets (Rajabifard & Williamson, 2004b).

Backx (2003) has categorised spatial datasets into three major classes – in terms of usability – of known (recognisable/findable), reachable (available/payable) and usable (clear/handleable/reliable) datasets (Figure 3.6).

Figure 3.6. Spatial data usability model (Backx, 2003)

Moreover, there is urgent demand for harmonisation and integration services that harmonise data for optimised common use. In many cases, it is often difficult or even impossible for users to sensibly combine data from different sources (Ryttersgaard, 2001). Muggenhuber (2003) identifies the continuum of spatial data management within the context of SDIs. He explains the progress made and highlights the current challenges. In this progress, GIS attracted the focus on implementation and tuning of stand-alone tools. The next wave began with the availability of large amounts of spatial data in the form of digital data. In the last decade, the stakeholders realised the urgent need for sharing spatial data with other stakeholders. This aim was highly dependent on institutional arrangements, which requires cooperation and partnership. The current demand is to provide integrable and harmonised spatial data for broad and maximum use (Figure 3.7).

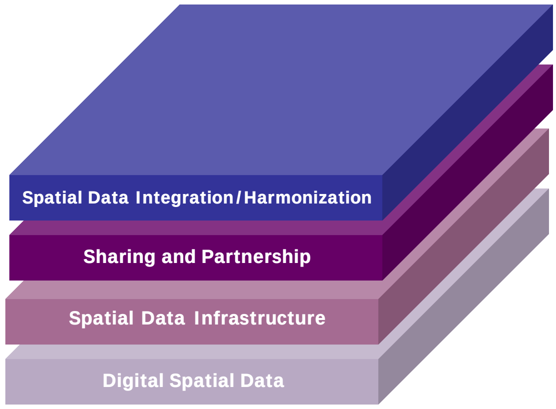


Figure 3.7. Spatial data management continuum (after Muggenhuber, 2003)

The issues and obstacles associated with multi-source spatial data integration had been recognized for many years (National Research Council, 1980; Chrisman and Niemann, 1985), however there have not been a holistic approach to deal with these issues before the introduction of SDIs. One of the major international challenges of building SDIs is linking distributed heterogeneous spatial information resources from different data providers in an application-oriented and user-oriented way (Donaubauer, 2005). The importance of investigating the integration of multi-source spatial datasets in the context of SDI initiatives has been highlighted in different publications, declarations and resolutions, in particular, UN resolutions. Rajabifard and Williamson (2004a) identified the integration of multi-source spatial data sets as a major concern in the success of national SDI initiatives. Resolution 15 of the 14th UN Regional Cartographic Conference for Asia-Pacific (UNRCC-AP) called for the investigation of issues in the integration of multi-source spatial data sets (UNRCC-AP, 1997). The UN Bogor Declaration (FIG, 1996) urged the creation of national SDIs to ensure integration and highlighted the homogeneity of the multi-source data sets to achieve maximum integration potential. Resolution 4 of the 17th UNRCC-AP (2006) emphasised the development of SDI principles to support the integration of multi-source spatial data.

It is also in line with the goal of SDI to reduce duplication of effort among agencies, improve quality and reduce costs related to geographic information, to make geographic data more accessible to the public, to increase the benefits of using available data, and to establish key partnership with states, counties, cities, tribal nations, academia and the private sector to increase data availability (National Research Council, 2003).

Effective spatial data integration can also be an identification and measure to show the success of an SDI. The assessment of SDI is difficult due to its complex, dynamic and ever-evolving nature. The perceptions of SDIs are also not the same in different parts of the world and many political, economic and cultural circumstances affect these perceptions (Grus et al., 2006).

Spatial data integration is claimed as one the most important aims and future directions of SDIs. Hence the degree of success in providing effective spatial data integration measures the success of SDIs.

3.7. Current Challenges of SDIs

In order to assist the spatial community, spatial information resources should be used widely by a broad range of citizens through SDI initiatives. In this regard priority SDI research should be directed at knowledge advancement which helps stakeholders, including policy-makers, scientists, business leaders, and ordinary people to better utilise spatial information and understand the relationship between government information policies, spatial information resources, products and services (Onsrud, 2004).

SDI has been in the spatial community for less than two decades and there are still many gaps in SDI advancement, which should be filled through conducting research and study. Onsrud (2004) highlights social and institutional issues as the most outstanding issues to focus on in future developments of SDI. He also recommends a number of specific projects that might be undertaken within the context of SDIs including:

- Real-time case studies to measure the effects of different legal, economic, and information policy choices on the development of spatial information infrastructures
- Evaluate the costs, benefits, effectiveness, and efficiencies of current government information policies
- Explore and develop a range of institutional and legal arrangements for accessing geographic resources
- Capacity building in spatial information resource management through the development of curricula, educational programs, and professional training
- Strategy development for increasing public access to government information
- Examine the role that pricing and cost recovery practices play in public access and commercial uses of data
- Compare local, state, and national government dissemination policies for allocating public and private funds to sustain government investments in an SDI
- Develop guidelines for increasing public participation in the identification, creation, use, and exchange of spatial information resources to inform community decision making
- Experiment with collaborative projects that are based on local knowledge and incorporate information to support public awareness and enhance decision-making processes
- Model the components and dimensions of an expanded view of the SDI focusing on technology and institutional developments and how they are embedded in other processes and media.

Masser et al. (2007) believe that the next significant step in SDI development is the spatial enablement of the government. They also urge that the future of SDIs is reliant on the ever-increasing involvement of the government in SDI development. There are many parallels between concepts based on which SDIs are developed and the vision of spatially enabling the governments.

One of the major achievements that moves governments and societies towards spatially enablement is the concept of virtual jurisdictions (Rajabifard et al., 2005b). Virtual jurisdictions represent an entity (such as a government) representing a defined territory (such as the State of Victoria) operating in an electronic medium, principally the internet (Robertson, 2004). The development of such a virtual system requires a set of concepts and principles to enable the design of an enabling platform that facilitates interoperability and interaction of functional entities within a heterogeneous environment and SDI has taken a lead role in attempting to meet these objectives. It also provides a foundation for identifying best practice and key performance indicators of SDIs in terms of their policy, technology and institutional frameworks.

Spatial data integration and harmonisation have been identified as a major challenge for next generation of SDIs (Rajabifard et al., 2005c; Muggenhuber, 2003). Among others, the integration of multi-source spatial data is within the scope of this thesis; however, the integration of spatial datasets is also a requirement in order to achieve spatially enabled government (Mohammadi, 2007 and virtual jurisdictions (Rajabifard et al., 2005b).

3.8. SDI and Spatially Enabled Governments

Nowadays, spatial data is framed within strategies that primarily aim to work towards a better government and improved living standards for society (Blakemore, 2004). Spatial data is also utilised within governance initiatives including e-government, e-society and e-democracy. Hence, spatial data and services evolve into a kind of nervous system for our planet throughout government and society. Spatial data and services will be able to take the pulse of the earth (Dangermond, 2006).

Wallace et al. (2006) articulate the Spatially Enabled Government (SEG) and state that a government can be named spatially enabled "where location and spatial information are regarded as common goods made available to citizens and businesses to encourage". SEG and the society are part of a broader picture of e-government/e-society and e-democracy, and aim at building a better relationship between citizens and governments (VSC, 2007). The vision of SEG identifies the necessity to make data, information and related spatial business services ubiquitous in the daily conduct of government agency business and in the efficient and effective delivery of government services (Gordon, 2007).

The implementation of the SEG vision leads to a number of significant achievements including informed and improved decision making, reduced cost of administration, consistent whole-of-government (WOG) outcomes and enhanced industry development opportunities (Rajabifard, 2007) together with effective interaction between citizens and government and better living standards for citizens (VSC, 2007).

The development of SEG was also a key outcome of the 17th United Nations Cartographic Conference for Asia and the Pacific (UNRCC-AP) and the 12th meeting of the UN-sponsored Permanent Committee for GIS infrastructure for Asia and the Pacific (PCGIAP) where SEG was defined as an environment where "data, information and related business services with spatial content become ubiquitous in the daily conduct of government agency business and in the efficient and effective delivery of government services ". The achievement of SEG vision requires collaboration at national level and the involvement of the private sector (Williamson, 2007).

SEG builds on SDI initiatives that are an important and integral part of a country's infrastructure. SDIs aim at developing an enabling platform, including institutional, legal, governance and political arrangements. In simple terms SEG is about using SDI to improve the operation and processes of government, and deliver better policy implementation and decision making by extending the use of spatial information to the whole of government and society. SEG is also an important part of countries' ICT, e-government and information-sharing strategies as a key activity that fosters innovation. The focus of SEG is on the use of spatial information to achieve government policy objectives, though SDI is essential to achieving SEG outcomes (Williamson et al., 2007).

3.9. Chapter Summary

The evolution of societies from industrial societies to information societies necessitates new types of commodities and infrastructures. Information and consequently information infrastructure are urgent demands of the new era, where data and information play a significant role in life and business. Eighty per cent of the information utilised by citizens and businesses are spatial or have spatial dimensions. Therefore spatial data and spatial data infrastructure seem crucial to satisfy the requirement of people and businesses and assist them with informed decision making. SDIs are interpreted differently – with different meanings and components – by different communities. However, some critical objectives and components look the same. Facilitating the use, exchange, sharing, access and distribution of spatial data is the most important task of SDI, while components like fundamental data, spatial data stakeholders, policy framework, standards, access networks, partnerships, governance and capacity building have been highlighted as the most crucial components required to fulfil these tasks.

SDIs also emerged as inseparable components of the vision of spatially enabled governments and societies. Not only this vision, but also the spatial community requires the integration of spatial data sources, which has not been achieved fully. The integration of spatial data is also a measure of the success of any SDI, as it was claimed as one of the most important objectives of SDIs.

The following two chapters investigate the challenges and issues involved in the integration of multi-source spatial data. This is done through a number of national (Australian) and international case studies. The chapters introduce study methodologies and at the end the results and findings are detailed, which provide the basis for the identification of necessary data integration toolbox components in Chapter Six.

Chapter 4

Data Integration:
International Case Studies

4.1. Chapter Aim and Objectives

In order to have effective data integration, a framework within the context of SDIs is required. The integration framework should highlight and articulate the components and tools that are necessary to assist practitioners to overcome the challenges and hindrances of spatial data integration.

This chapter discusses the results that have been achieved through a number of case studies that will be used for identifying the components of the framework. The case studies include seven countries and have been conducted within the Asia-Pacific region. The case studies have been selected from a range of different jurisdictions with different sizes, population and political structures. The major objective was to identify the issues and challenges within each jurisdiction, among which the most common and challenging issues are identified. The impact of the identified issues is also addressed in this chapter; however, the impact of all issues that have been identified through international and national case studies will be elaborated in the sixth chapter. This chapter presents a methodology and a roadmap to identify the issues and barriers. This chapter and Chapter Five (Australian Case Studies) will assist in the identification of necessary approaches and tools that act as enablers to overcome the obstacles of effective data integration.

4.2. Introduction

Different jurisdictions encounter a variety of issues and challenges when attempting to integrate multi-source spatial data (Williamson et al., 2003; Van Loenen & De Jong, 2007). Data integration depends on many factors including the institutional structure, the maturity of spatial data coordination within and across agencies, the attitude and awareness of stakeholders, the existence of well-developed and effective standards and tools, and so on. In order to thoroughly investigate the effective factors within different jurisdictions, a number of case study countries have been selected. The selection was based on the accessibility of the participating countries information.

The case study investigation was carried out at two different stages within a number of countries. The methodology of investigating the case studies is illustrated in Figure 4.1.

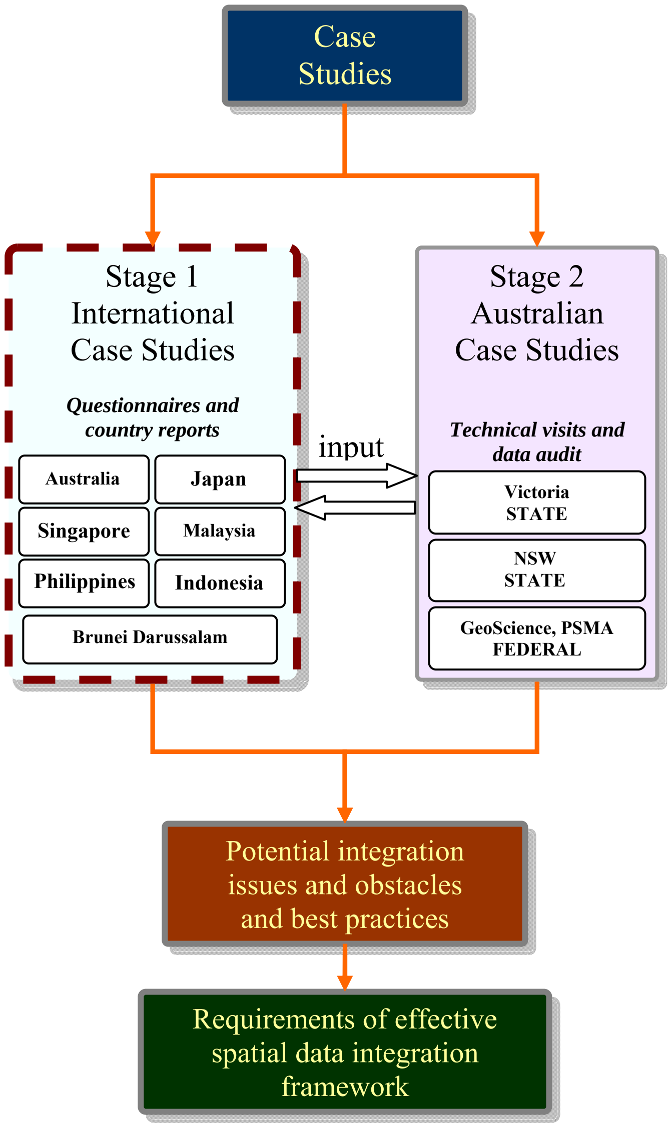


Figure 4.1. Case study investigation methodology

In the first stage, seven countries (Japan, Singapore, Australia, Brunei Darussalam, Indonesia, Malaysia and the Philippines) in the Asia-Pacific region were selected through the channel and support of UN-sponsored Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP). The case studies aimed to investigate the issues of and barriers to spatial data integration in the countries with different social, governmental and geographical characteristics.

The research was given the permission by PCGIAP-WG3 to utilise the documents and information that have been provided by the member countries (Appendix 2).

Seven case study countries participated to provide input as reports on their activities and current practices of multi-source spatial data integration (UNRCC-AP, 2006). The reports aimed to identify and highlight the common potential issues that hinder effective spatial data integration in the case study countries and also to provide inputs to form an SDI model with a focus on spatial data integration.

In this regard, an integration template (Appendix 3) was designed as a standardised generic proforma to enable the discovery of information, including matters concerning member countries' spatial information policies, laws and regulations, infrastructure implementation, institutional arrangements, technology, and integration issues as well as human resources and capacity building for spatial data integration.

In order to identify and investigate the technical and non-technical issues, a second stage of case studies (which was at a state level with interaction at national level) was conducted (datasets audit and technical visits) within two states of NSW and Victoria, Australia (Mohammadi et al., 2007). The agencies in Australia have been the research project partners who have expressed much interest in the outcome of the project.

This chapter is devoted to developing the case studies at first stage with international jurisdictions as described above. Australian case studies (stage 2) will be discussed in Chapter Five. The methodology for each stage is also articulated in detail. Each stage is accompanied by analysis of the outcomes and findings of case studies. The analysis of the outcomes and observations will provide input for the identification of the required mechanisms and tools for an effective spatial data integration framework.

4.3. International Case Study - Design and Results

The first stage of case studies aimed to provide holistic observation and findings on the challenges, obstacles and best practices of multi-source spatial data integration. Stage 1 collects information on the integration of multi-source spatial data within a number of countries. The information contains the situation of data integration and an insight into the issues and challenges of different jurisdictions. These observations will be compared in order to identify the major challenges.

Seven countries including Japan, Singapore, Australia, Brunei Darussalam, Indonesia, Malaysia and the Philippines participated through the PCGIAP channel to provide input as reports on their activities and current practices of multi-source spatial data integration (UNRCC-AP, 2006). They were also asked to give presentations in a workshop on integration (PCGIAP, 2006). An integration template was designed as a standardised generic proforma to enable the discovery of information, including matters concerning member countries' spatial information policies, laws and regulations, infrastructure implementation, institutional arrangements, technology, and integration issues as well as human resources and capacity building for spatial data integration.

The questionnaire and country report came in a single document. The country report was followed by an introduction to the integration project and its aims and objectives. In the country report sections, case study countries were asked to provide explanatory information on the following topics:

- Country Context including Geographical Context, Historical Context and Current Political and Administrative Structures
- National SDI Context including History and Status of National SDI Initiative, Historical Outline of Cadastral and Topographic Data Development and Current Administration of Cadastral and Topographic Data
- Institutional Framework for Integration – Data Provider Perspective
- Institutional Framework for Integration – Data User Perspective
- Issues in the Integration of Built and Natural Environmental Datasets including Need for Integration, Major Issues and Barriers

Second, a questionnaire aimed at identifying the main issues of integration in five major categories of technical, institutional, policy, legal and social collected information on individual items. The reports and questionnaire show the importance of solving non-technical issues along with technical issues for effective data integration.

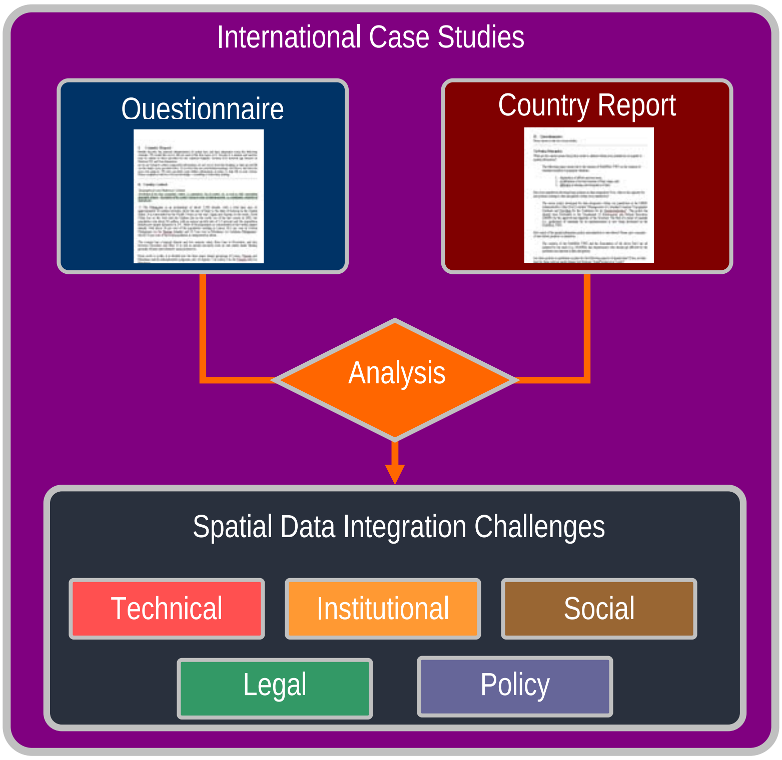


Figure 4.2. International case study design

The country report briefly describes the basic geographic context, that is, population, size of country, and so on, as well as other outstanding geographic features. It also describes the country's history in terms of relevant periods, for example, colonisation, and political development. This is followed by a description of the current political and administrative structures, such as the political system, number of states or provinces, and so on and how this may affect efforts to integrate spatial data. The country report also illustrates the national SDI context with history and status of national SDI initiative, the historical outline of multi-source spatial data development and the current administration of spatial data. The institutional framework for spatial data integration is described from both the data provider perspective and data user perspective. The final section of the country report is devoted to indicating the issues and challenges of spatial data integration with emphasis on the need for integration, major issues of data integration and the initiatives in case study countries to integrate the multi-source spatial data and the outcomes of initiatives.

The following data is gathered utilising case study country reports in the above-described sections. The information is first described for every single case study country and then a comparison is made and the results summarised and concluded.

4.3.1. Case Study Countries' Context

The Philippines

The Philippines is an archipelago of about 7,100 islands, with a total land area of approximately 30 million hectares (300,000 sq km). It is surrounded by the Pacific Ocean on the east, Japan and Taiwan on the north, South China Sea on the west and the Celebes Sea on the south. As of the last census in 2008, the population was about 96 million with an annual growth rate of 1.99 per cent and the population density per square kilometre is 271. It is rich in natural resources such as vast arable lands, fishing grounds, forests and extensive mineral reserves. Under the 1987 Constitution, the Philippines was declared a democratic republican state whose system of government is the presidential form. The national government has three branches: the *executive branch* headed by the President, the *legislative branch* and the *judicial branch*. For the purpose of administration and development planning, the Philippines is divided into 16 administrative regions. The territorial and political subdivisions of the Philippines are as follows: province, cities, municipalities and barangays (the smallest political unit into which cities and municipalities are divided; they consist of less than 1,000 inhabitants). Each territorial or political subdivision enjoys local autonomy. The President exercises supervision over local governments.

Figure 4.3. Map of the Philippines

The Philippines initiated the development of its national SDI with the participation of major stakeholder agencies and also the orientation of technical working groups and participant agencies. The driver for SDI development has been the resource and hazard management. This is also reflected in the initiatives that drive spatial data integration development in the Philippines. In this regars, one of the Philippines' SDI working groups has the mission to create the standard for seamless topographic data.

Brunei Darussalam

Brunei Darussalam is situated on the north-west of the island of Borneo, between eastern longitudes 114° 04' and 11° 23' and northern latitudes 4° 00' and 5° 05'. It has a total area of 5,765sq km with a coastline of about 161 km along the South China Sea. It is bounded on the north by the South China Sea and on all the other sides by the Malaysian State of Sarawak. The population of Brunei Darussalam in mid-year 2008 is estimated at 380,000 persons. Brunei is an independent sovereign sultanate that is governed on the basis of a written constitution. The 1959 Constitution provides for the Sultan as the Head of State with full executive authority. The Sultan is assisted and advised by five councils – the Religious Council, the Privy Council, the Council of Ministers (the Cabinet), the Legislative Council and the Council of Succession. The 1959 Constitution established the Chief Minister as the highest official. Brunei's administrative system is centred on the Prime Minister's Office, which has provided the thrust behind His Majesty's aim to introduce greater efficiency in the government.

Figure 4.4. Map of Brunei Darussalam

Brunei Darussalam has recently focused on the SDI development for the delivery of e-government services and hazard management. These drivers have been also identified as the main incentives for national spatial data integration.

Japan

Japan is an island nation stretching arch-wise in north-east Asia between the North Pacific and the Sea of Japan. The land consists of four main islands and over 6,800 islands with a total area of approximately 380,000 sq km. The population is approximately 127 million with a negative growth rate. The imperial structure has continued since the Japanese nation was founded. Under the present Constitution of Japan, the Emperor serves as the symbol of the state while the Cabinet has power over all aspects of the nation. The Diet, consisting of the House of Representatives and House of Councillors elected by direct vote of the people, designates the Prime Minister from among its members. The Prime Minister then appoints the other ministers and organises the Cabinet. The government bureaucracy was reorganised in 2001 into 10 ministries, two agencies and the Cabinet Office.

Figure 4.5. Map of Japan

Japan formed an inter-agency GIS liaison committee to initiate the national SDI. Despite the abolishment of this committee, its missions for establishing the collaborative initiatives to facilitate the use and exchange of spatial data among a broad range of users, have been a national priority and recently backed by an enacted law in 2007 (Kazuhiko, 2007). One of the major objectives of the missions is to provide integrated datasets to mitigate and manage hazards including earthquakes and tsunamis.

Singapore

With a population of 4 million living over an area of 700 sq km, Singapore is one of the most densely populated islands (countries) in the world. Singapore was under the British Administration in the early days. It gained its independence in 1965. It is a republic state headed by a President and its government is elected every five years. Singapore is a democratic country. It is not divided into states. However, for political administration, it is divided into 14 group representatives. As for public administration, Singapore has 15 ministries overseeing all the public services.

Figure 4.6. Map of Singapore

Singapore is moving from an exploratory stage to development stage of national SDI. Singapore's national priorities are land (natural resources), administration and national security and hazard management. In order to respond to these priorities, spatial data integration initiatives should be addressed effectively. The major barrier in this regard is the agency-specific standards and specifications that differ among various agencies.

Australia

Australia is the largest island continent in the world, with a total area of over 7,600,000 sq km, lying south of the Equator between the Indian and South Pacific Oceans. The population is approximately 19.5 million, with a growth rate of about 1 per cent. The majority of the population (85%) resides in urban areas along the east and south-east coastline and fertile plains. Much of the interior of the country is flat, barren and sparsely populated. Australia also lays claim to the third largest marine jurisdiction in the world and has a coastline extending more than 36,700 km. Australia has been an independent member of the British Commonwealth since 1901 when it became a Federation of States. A referendum to change Australia's status from a Commonwealth headed by the British monarch to a republic was defeated in 1999 and hence Queen Elizabeth II of England remains the Head of State. The constitution vests in the Governor-General, representing the Head of State exercised by tradition on behalf of the elected government. The government is based on a bicameral Federal Parliament headed by an elected Prime Minister consisting of a Senate, which has proportional representation among the states. The Federal Government has powers over defence, foreign affairs, trade and commerce, taxation, customs and excise duties, pensions, immigration and postal services. Other powers are left with the eight state and territory governments, such as health, education, state transport networks, town and rural planning and land administration (cadastral system, land registration).

Figure 4.7. Map of Australia

Australia has one of the best practices in terms of SDI development with the ANZLIC Australian governments represented in ANZLIC (Australia and New Zealand Land Information Council), which aims to develop and implement the Australian SDI.

Australia responded to the needs of governments and businesses for seamless nationwide spatial datasets through the establishment of Public Sector Mapping Agency (PSMA). PSMA is responsible for developing necessary standards and policies for spatial data integration. However, different states and territories have tried to develop their own integrated datasets, but at a national level, these datasets are not consistent.

Indonesia

Indonesia is an archipelagic country with about 17,502 islands and has more than 100,000 km length of coastline, the second longest coastline in the world after Canada. The total area of Indonesia is over 5.3 million sq km, where 63.7 per cent of its area consists of water. The population of Indonesia is about 2437 million. Twenty-two per cent of the population live on the coast and rely on marine activities for their living sources, especially in coastal areas, which occupy about 80 per cent of the activities. In the last two years, several natural disasters hit some parts of the coastal area of Indonesia, causing the death of hundreds of thousands of people, while those who survived lost their properties. A devastating earthquake and tsunami hit Aceh and its surrounding area in December 2004, followed by the Nias earthquake in March 2005. The last earthquake and tsunami hit the southern coast of Java Island in July 2006, causing more than 500 people to die and thousands of survivors to lose their properties. Indonesia is the world's largest archipelagic state. Current issues include alleviating poverty, preventing terrorism, consolidating democracy after four decades of authoritarianism, implementing financial sector reforms, stemming corruption, holding the military and police accountable for human rights' violations, and controlling avian influenza. In 2005 Indonesia reached a historic peace agreement with armed separatists in Aceh, which led to democratic elections in December 2006.

Figure 4.8. Map of Indonesia

The Indonesian National Coordination Meeting on Surveys and Mapping initiated the development of Indonesian NSDI. Natural disaster management has been the most significant challenge for Indonesia. Because Indonesia consists of numerous islands, the seamless management of the integration of respective datasets is a challenging issue.

Malaysia

Malaysia covers a land area of about 329,758 sq km, consisting of 13 states in Peninsular Malaysia, two states in the island of Borneo (Sabah and Sarawak) and three federal territories. Peninsular Malaysia, covering 131,598 sq km, borders Thailand to the north, the Straits of Malacca to the west, the South China Sea to the east, and the island of Singapore to the south. The states in Borneo covering 198,160 sq km border the territory of Indonesia's Kalimantan to the South, Brunei and South

China Sea to the north, and the Sulu Sea and Celebes Sea to the east. Malaysia lies close to the equator between latitudes 1° and 7° north and longitudes 100° and 119° east. The population is approximately 24.92 million, with a growth rate of about 1.7 per cent. Malaysia is a multi-racial country and the majority of the population resides along the west coast of Peninsular Malaysia.

Malaysia's government is modelled on the British system, somewhat modified because Malaysia's federal structure incorporates 13 states and three federal territories. Nine of those states have rulers or sultans and they elect a monarch, the supreme ruler, every five years. The government is based on a parliamentary system, headed by an elected Prime Minister. The Federal Government has powers over such areas as external affairs, defence, internal security, civil and criminal law, federal citizenship and naturalisation, finance, trade, commerce and industry, taxation, customs and excise duties, shipping, navigation and fisheries, communications and transport, federal works and power, education, medicine and health, social security and tourism. The states' powers include those over land and its administration, Islamic law, Malay customs, permits and licences for mines prospecting, agriculture, forests, local government, states works and water, and riverside fishing. Peninsular Malaysia is a federation of states, each of which is responsible for its own land matters. All states operate a Torrens system of land registration, administered by the State Land Offices and coordinated by the Department of Land and Mines. Cadastral surveys are controlled by the Department of Survey and Mapping, Malaysia (JUPEM) which is a federal department. JUPEM is responsible for undertaking cadastral survey work within Peninsular Malaysia and topographic mapping throughout the country.

Figure 4.9. Map of Malaysia

The development of SDI in Malaysia was first initiated by the development of a National Infrastructure for Land Information System (NaLIS), which later was renamed Malaysian Geospatial Data Infrastructure (MyGDI). Malaysian Centre for Geospatial Data Infrastructure (MaCGDI) is responsible for implementing MyGDI with support from the various technical committees. Malaysia is one of the fastest developing countries; therefore the management of land resources and infrastructure is a priority for Malaysian governments. In this regard the integration of spatial datasets has been identified as a critical issue that can greatly support national priorities.

4.3.2. Country Context

The diversity of the case study countries ensures the identification of diverse issues and challenges of spatial data integration. The country contexts including the geographical characteristics of countries together with the history, size, population and political structure define the priorities and drive the attempts. Countries like the Philippines and Japan struggle with the management of numerous islands. The Philippines, Japan and Indonesia possess about 7,100, 6,800 and 17,500 islands respectively. Singapore, Malaysia and Indonesia have water boundaries with neighbours, which makes the definition of precise boundaries a must. This makes the effective management of water bodies and water boundaries and resources a substantial priority for these countries. As a consequence creating the availability and the sharing of water-related datasets with rich content on resources and boundaries is quite crucial for these countries. The density of populated areas defines the priority of countries in effective management of land and resources in countries like Brunei Darussalam and Singapore. The political structure also greatly influences the approach and behaviour of spatial stakeholders, especially governmental stakeholders. Some of these countries (Brunei Darussalam, Indonesia, Malaysia and the Philippines) are developing countries (at different stages of development) with average population growth rate over 1.5 pe rcent annually and also a huge human–environment development. Countries like Singapore, the Philippines, Japan and Indonesia are densely populated and the impact of man-made developments is huge. Thus, another major issue that comes as a priority for such countries is monitoring and assessing the impact of human activities on natural resources. This necessitates the integration of human-related and environmental datasets to create a realistic model of the environment to better manage the impact of human activities on the environment. The political structures of the case study countries differ. The sultanate (central) government of Brunei Darussalam, imperial and central structure of Japan, central government of Singapore, federated structure of the Philippines, Australia and Malaysia reflect the diversity of governance bodies, standardisation organisation and policy-making agencies. In most central governments a single organisation is responsible for data coordination, policy-making and standard development, while in federated countries it is the breakdown among states to develop their own standards and policies and data management schemes. Table 4.1 summarises the geography, size, population and political structure of the case study countries.

Table 4.1. Geographic, demographic and political characteristics of international case study countries

Case study country	Geography	Size	Population* (Annual growth rate)	Development	Political Structure
Philippines	archipelago of about 7,100 islands	~ 300,000 sq km	~ 96 million (1.99)	Developing	democratic republican state with 16 administrative regions
Brunei Darussalam	land and sea boundaries (~ 161 km coastline)	~ 5,765 sq km	~ 380,000 (1.78)	Developing	independent sovereign sultanate government based on written constitution centered on the Prime Minister's Office
Japan	island nation stretching arch-wise (4 main islands and over 6,800 islands)	~ 380,000 sq km	~ 127 million (-0.14)	Developed	imperial structure (the Emperor serves as the symbol of the state while

					the Cabinet has power over all aspects of the nation)
Singapore	one of the most densely populated islands (countries) in the world	~ 700 sq km	~ 4.6 million (1.135)	Developed	democratic country not divided into states 14 group representative constituencies
Australia	the biggest island in the world	~ 7,680,000 sq km	~ 20 million (1)	Developed	federated states with 8 states and territories
Indonesia	an archipelagic country with about 17,502 islands and more than 100,000 km coastline the second longest coastline in the world	~ 5,300,000 sq km (~1,900,000 sq km of land)	~ 237 million (1.175)	Developing	republic with 30 provinces
Malaysia	Peninsula with two main lands	~ 330,000 sq km	~ 25 million (1.7)	Developing	federal structure

* as in 2008

4.3.3. SDI Development in Case Study Countries

As a part of the country reports, information on different aspects of SDIs including the initiation of SDI (SDI driver), the level of accommodating agencies, the priorities of SDI development in terms of technical working groups, and utilised standards was collected. In terms of SDI development, the case study countries are at different stages ranging from the well-developed and effective SDI of Australia and Malaysia to the exploratory stage in Singapore. The major difference is the driver behind the initiation of SDIs in case study countries. The national priority can be identified as the horsepower of SDI initiative. Taking the Philippines as an example, SDI was initiated with the participation of seven member agencies forming the Inter-agency Task Force on Geographic Information (IATFGI). This includes the following agencies: Housing and Land Use Regulatory Board (HLURB), National Statistics Office (NSO), National Computer Center (NCC), Department of Public Works and Highways (DPWH), Department of Science and Technology (DOST) through the Philippine Institute of Vulcanology and Seismology (PHIVOLCS), Bureau of Soils and Water Management (BSWM), and National Economic and Development Authority (NEDA) (Crisostomo, 2007). It was also accompanied by the formation of technical working groups including:

- Agriculture, Environment and Natural Resources
- Lands and Survey
- Infrastructure and Utilities
- Socio-economics
- Research, Training and Technology

The orientation of the technical working groups and participant agencies shows the importance and priority of the resource and environmental managements together with hazard management. This is also reflected in the initiatives that drive spatial data integration development in the Philippines.

Brunei Darussalam is in the early stages of SDI development with special emphasis on collaborations to leverage e-government services. The development of SDI is supported by the Department of Survey and in partnership with the Land and Town and Country Planning, which highlights that resource and land management for planning purposes with major focus on the hazard management is one of the priorities of Brunei (Kamis, 2007).

Japan started the development of SDI in 1995 by forming an inter-agency GIS liaison committee. This committee was abolished in 2005 but the mission that the committee was meant to deliver has been followed. The mission was the establishment of a collaborative initiative to promote the use and sharing of spatial data among a broad range of users (Kazushige, 2007). This includes the adoption of the long-term plan for building a national geospatial data framework and promoting the use of GIS; fundamental development such as standardisations relating to the national geospatial data framework; completion of the national geospatial data framework and the dissemination, adoption of standards and development plan for national geospatial data framework; the adoption of standards for national geospatial data framework, adoption of national geospatial data framework development plan, measures for future development and promotion of GIS; the formulation of government policies to emphasise the digitisation and provision of its database; planning of measures for the effective utilisation of GIS by the government; agreement of guidelines for the provision of government geographic information; and guidelines for disseminating government geographic information to the general public. The mission was backed by a law enacted in May 2007 (Kazuhiko, 2007). The purpose of the law was to support the mission through the advancement of policies concerning the advancement of utilising geospatial information in a comprehensive and well-planned manner.

Singapore has managed land very well since 1989. In this regard the electronic submission of plans, job data storage and digital imaging were launched in 2005 (Khoo, 2007), but Singapore is still in the exploratory stage of SDI development. There is no SDI steward in Singapore to initiate working groups and policy and guideline development.

Australia most probably has one of the best practices in terms of SDI development with the ANZLIC (Australia and New Zealand Land Information Council) as a national council to develop and endorse spatial data-related standards, guidelines and policies. Australian governments have a good liaison with ANZLIC with representatives sitting in ANZLIC. ANZLIC has the support of federal and state governments; however, the adoption and implementation of endorsements of ANZLIC are not compulsory in all states. A national SDI is being coordinated in Australia through ANZLIC. The council's role is to facilitate easy and cost-effective access to the wealth of spatial data and services provided by a wide range of organisations in the public and private sectors. This is done through the development of the Australian SDI (ASDI), which comprises the people, policies and technologies necessary to enable the use of spatially referenced data through all levels of government, the private sector, non-profit organisations and academia (ANZLIC, 2006). ANZLIC's main role is to establish national leadership through the creation of national policies and guidelines for data custodianship, data access and metadata. It collaborates with the eight states and territory governments of Australia to implement the ASDI on a cooperative basis – the implementation of the ASDI is not legislated. The ASDI is actually made up of eight state and territory SDIs.

Due to Australia's federated system of government, the state and territory governments have power over land administration and state spatial information, which has seen each state government build its own SDI. Each state creates its own policies in regards to data access and discovery, with harmonisation and cooperation between states made possible through ANZLIC. ANZLIC is made up of a representative of each of the state governments and the Federal Government, which enables promotion of nationally consistent standards and policies to be implemented within each jurisdiction. The development of the ASDI is not supported by any legislation; therefore ANZLIC's strength lies in the creation of partnerships and collaborations among spatial stakeholders. This is a very important factor in the ability to implement an SDI across all jurisdictional levels in Australia.

The development of National Spatial Data Infrastructure in Indonesia was implemented when recommended by the National Coordination Meeting on Surveys and Mapping (Rakornas Surta) in 2000. Bakosurtanal is the coordinating agency for the development of Indonesian NSDI (Matindas & Purnawan, 2004). The development covers five aspects of an Indonesian NSDI:

- Institutional aspects
- Legal aspects
- Fundamental data set aspects
- Technology research and development
- Human resources aspects.

The need for geospatial information to assist in the planning and development of land resources in Malaysia had been felt since the early 1970s. As a consequence much effort was made by various government agencies in establishing computerised land information systems. However, they existed as stand-alone systems in "islands of information systems", thereby making it difficult for users to obtain access to them. As a consequence there was an urgent need for the sharing of information and for the said systems to be linked. National Infrastructure for Land Information System (NaLIS) was consequently formed in 1997. Basically, NaLIS was established as an initiative of the government to promote and facilitate the sharing, exchange, dissemination and use of geospatial information among land agencies and other users. This would be through providing online access to the information that resides in the various agencies, over the internet. At the beginning of 2002, NaLIS was renamed MyGDI (Malaysian Geospatial Data Infrastructure) following the restructuring of the NaLIS Secretariat to a division known as Malaysian Centre for Geospatial Data Infrastructure (MaCGDI). These changes were performed to increase the efficiency of SDI implementation. Major SDI activities occur at the national level and they are being carried out by MaCGDI, with support from the various technical committees under MyGDI, including framework, standards and clearinghouse technical committees. Data custodianship guidelines, geospatial information policy guidelines and policy on data sharing have been developed within MyGDI (Nordin, 2007).

Some SDI aspects of case study countries have been summarised in Table 4.2.

Table 4.2. SDI development in international case study countries

Case study country	SDI driver activities / players	Accommodating agency/level	Technical Working Groups (TWG)	Capacity building
Philippines	1993 Inter-agency Task Force on Geographic Information (IATFGI)	National Mapping and Resource Information Authority (NAMRIA)/ **national level**	1. Agriculture, Environment and Natural Resources 2. Lands and Survey 3. Infrastructure and Utilities 4. Socio-economics 5. Research, Training and Technology	-
Brunei Darussalam	2005 e-government Collaboration-based	Department of Survey in partnership with the Land and Town and Country Planning/ **national level**	-	1. officers are sent for short and long-term overseas training 2. improvement of technologies used 3. continuous awareness program
Japan	1995 GIS Liaison Committee of Ministries and Agencies to establish NSDI	Assistant Chief Cabinet Secretary/**national level**	GIS Liaison Committee of Ministries and Agencies	on-the-job training (OJT)
Singapore	1989 Land Data Hub (LDH) for land-related datasets **full NSDI is still in the exploratory stage**	in exploratory stage	no TWG yet	not available
Australia	Australian SDI for effective data sharing and coordination	Australia and New Zealand Land Information Council (ANZLIC) and state SDIs /**national and state levels**	1. Standing Committee on Land Administration 2. All-Hazards (Emergency Management, Counter Terrorism, CIP) Standing Committee 3. Intergovernmental Committee on Surveying and Mapping	regular meetings, forums and conferences
Indonesia	2000 National Coordination Meeting on Surveys and Mapping (Rakornas Surta)	Bakosurtanal/**national level**	1. Institutional aspects 2. Legal aspects 3. Fundamental data set aspects 4. Technology research and development 5. Human resources aspects	intent collaboration and regular meetings
Malaysia	1997 NaLIS (National Infrastructure for Land Information System) and in 2002 renamed as MyGDI (Malaysian Geospatial Data Infrastructure)	Malaysian Centre for Geospatial Data Infrastructure (MaCGDI) /**national level**	1. Framework Technical Committee 2. Standards Technical Committee 3. Clearinghouse Technical Committee	training

4.3.4. SDI Components

Some major SDI components that strongly influence the effectiveness of spatial data integration have also been discussed in country reports. SDI components including access and sharing mechanisms and arrangements, pricing policies, metadata and metadata standards, spatial data models and data management standards, play significant roles in the integration of multi-source spatial data. Easy access and sharing of spatial data that is owned by different agencies are a requirement for spatial data integration. Spatial data integration is meaningless without accessing data. It is also essential to have a reasonable price and easy pricing policy. Restricted access and pricing policies severely hinder the use and integration of spatial data. Easy access and flexible pricing policy can attract more users to utilise and integrate spatial data. Rich metadata content can also facilitate the integration of spatial data as it provides information on different technical and non-technical aspects of data. This includes information on the content, scale, quality (spatial and aspatial accuracy and logical consistency), completeness, currency, access channel, geographical extent, spatial reference system (datum and projection system), restriction of data use, jurisdiction, owner and pricing policies. The metadata content not only can assist users in the assessment of spatial data, but also can impact the decision of users in order to use datasets. Metadata standards can also be a significant issue, as a common standard can provide the basis for comparison and can facilitate the integration of metadata content to create the metadata for integrated data. Table 4.3 summarises the above-mentioned aspects in the case study countries.

Table 4.3. SDI components in the international case study countries

Case study country	Access mechanism	Pricing policy	Metadata	Data model
Philippines	data browser and web-based GIS	free data transfer cost maintenance cost	based on ISO TC211	based on Standards on Data Classification (IATFGI Resolution No. 1, Series of 1995)
Brunei Darussalam	no publicly available clearinghouse inter-governmental internet-based link	free data transfer cost	no metadata and metadata standard	-
Japan	clearinghouse, Web-GIS, Digital data (CD distribution)	free data transfer cost	Japanese Metadata Profile (JMP) based on ISO 19115	ISO 19100 data encoded in XML
Singapore	open and free access for public and private agencies dissemination using CDs on a quarterly basis LandNet an online system for Land Data Hub	data transfer cost	no metadata and metadata standard	no data model
Australia	state-level portal, value-added resellers (VARs)	state-dependent in most cases full cost recovery	ANZLIC profile based on ISO TC211	state-level data model at national level - harmonised data model
Indonesia	spatial data warehouse	-	-	-
Malaysia	clearinghouse	governmental data is free subsidised cost recovery	ISO TC211	spatial and aspatial data stored separately in different data models

Most countries have realised the importance of easy access and sharing of spatial data through effective web services with less restrictions. The Philippines has developed a data browser and web-based tool to provide spatial datasets to users at little or no cost. Japan is also moving in the same direction with the provision of many sophisticated access tools including clearinghouse and Web-GIS. Digital data is also distributed through CD media to users. Datasets are available for use free of charge or for the cost of transfer. Brunei Darussalam has adopted the same pricing arrangement, but as it has been recently developing its SDI, there is still no publicly available clearinghouse for data access and dissemination. However, governmental organisations can share and have access to other governmental datasets through an inter-governmental internet-based link. Singapore also provides open and free access to its spatial datasets for public and private agencies in the form of CD dissemination on a quarterly basis. However, land-related data is available online through LandNet, an online system for Singaporean Land Data Hub.

Australian states most probably have the toughest spatial data access restrictions in terms of pricing arrangements. Most Australian states charge spatial data users for the cost recovery price. This greatly hinders the use of spatial datasets for private users. However, Australian states coordinate and maintain quite sophisticated access channels. These include spatial data directories and clearinghouses. Malaysia possesses a well-developed access tool in the form of a clearinghouse that provides most of the spatial datasets to the users. In terms of pricing, governmental data is free of charge and other non-governmental and value-added datasets are available at a subsidised cost recovery price.

In terms of metadata maintenance, all case study jurisdictions except Singapore and Brunei Darussalam maintain metadata. Singapore and Brunei are also in the exploratory stages of metadata development, though there is little structured metadata available in these countries. All other case study countries maintain metadata and all have adopted the metadata profile of ISO/TC211.

Data modelling is one of the biggest challenges as each jurisdiction has adopted or developed a particular different data model. This issue greatly hinders the integration of spatial data as it produces different data structures, data categorisation and spatial and aspatial content. This will be discussed in depth in Chapters Five and Six.

4.3.5. Spatial Data Integration Drivers and Challenges

The priorities and major challenges of each case study country drive spatial data integration initiatives. The applications and services that are developed revolve around these drivers and in response to national and societal priorities.

The maturity of spatial data coordination and the advancement of initiatives in the area of spatial data integration also strongly affect the complexity and nature of the challenges of data integration. Apart from sustainable development, this has been mentioned as the priority of case study countries and as one of the major drivers behind spatial data integration. Sustainable development aims to meet environmental protection and society cohesion objectives alongside economical growth. This requires a realistic model of the environment to monitor the impact of human activities on the environment.

The Philippines has created a technical working group specifically for the creation of standard seamless topographic data. The main aim of the working group is to integrate multi-source spatial data related to topography in order to create a national coverage. The Philippines has encountered a number of different challenges and problems in this regard. One of the major issues that hinders this task is the difference in scales and accuracies of multi-source data including topographic and cadastral data. In some cases there has been no digital data available for integration with other datasets. The Philippines also has found it quite difficult to come up with a standardised platform for data and processes to facilitate the access, exchange and integration of datasets; the lack of technical standards at national level in particular makes it much more difficult. Duplication of efforts in data creation and management also wastes resources. This has led to the proliferation of several versions of base maps in the Philippines. Spatial data integration and sharing have been mentioned as one of the most difficult tasks to be effectively addressed within the national SDI of the Philippines.

Brunei Darussalam requires integrating multi-source spatial data to respond to many GIS applications which have been developed to meet the national priorities including sustainable development and natural hazard mitigation. This is hindered by the lack of a reliable access channel for a broad range of spatial data users as there is no web-based service available to deliver data to the public. Brunei Darussalam also suffers from lack of metadata and metadata standard to enrich its spatial data content. This hinders effective discovery, use, assessment and integration of datasets. Technical tools are also not sophisticated enough to implement multi-source spatial data integration effectively. Lack of standardisation in spatial data domain has also caused heterogeneity and incompliancy among different spatial datasets. Custodianship agreement is also a big issue in Brunei Darussalam as a robust custodianship agreement ensures the quality of data content and the provision of rich metadata content. Most spatial data is created by data provider agencies to respond their own needs with little consideration of other user needs so most datasets do not fully fit the purposes of other users. Legal issues are also hindering effective data use, sharing and integration as there is lack of policy of data issuance, confidentiality and security of data, and copyright laws. The lack of effective legal mechanisms and consequent confusion hinder spatial data sharing and distribution.

Japan's national priority is the management of hazards like earthquakes and tsunamis. In this regard spatial data from different sources needed to be integrated effectively and with minimum effort. Otherwise in the situation of hazard in which ready-to-use data may save lives and resources, the data is worthless. In the deliverance of this task, Japan encounters some problems. A major problem is the unavailability of cadastral data through the national clearinghouse, so there is no intention among users to utilise cadastral maps. The data model is also a huge problem that one of its major consequences is the lack of metadata level relations between cadastral and topographic datasets.

Singapore, as the most populated country in the world, has prioritised land (natural resource) administration alongside other critical issues such as national security and hazard management. These national priorities are also drivers of spatial data integration initiatives in Singapore. However, there are challenges that should be addressed for an effective delivery of data integration. These include challenges such as lack of metadata and metadata standards that hinders discovery, use and integration of data. Lack of collaboration and support from various public agencies has also been a hindrance to effective data coordination and sharing. Lack of legislative support also threatens the effectiveness of spatial data initiatives. Without legislative support, stakeholders are not obliged to provide necessary data and processes. In many cases the policy frameworks are agency-specific which may differ among different agencies.

Australia experienced one of the best practices in the region with the establishment of Public Sector Mapping Agency, which is responsible for creating seamless integrated

Australia-wide datasets. PSMA's establishment was to respond to the need of governmental and public sector agencies for the best available nationwide spatial data. PSMA sets policies and rules and standards for spatial data integration. It has also established collaboration between different stakeholders to obtain, share and disseminate spatial data. Despite the fantastic achievements of PSMA, there are still problems and hindrances. Australia is a federated state country comprising eight different states and territories. States form independent jurisdictions in which all policies and standards are made in-house. The federated system of governance provides significant difficulties in attempting to create a national overview of spatial information for use by different communities of practice such as environmental management, emergency management and counter-terrorism operations. This also causes the development of different data access policies within different states. Another issue that hinders spatial data use and sharing and impacts the decision of users not to utilise spatial data is the data price, which is in most cases for full-cost recovery. Apart from non-technical issues such as licensing and privacy, some substantial technical aspects of spatial data are also non-compliant among different Australian states. These include diverse data models and data specifications. This especially hinders effective data integration as different themes of data are categorised and have structures that differ from others.

Natural disaster management including earthquakes and tsunamis is one of the biggest challenges of Indonesia and is the driver behind many spatial initiatives including Indonesian SDI. Another national priority for Indonesia are the numerous islands that require effective integrated land–sea management. In this regard the integration of land and sea spatial datasets is a necessity for the robust management of the resources and environment. However, the difference in land and sea management in Indonesia causes many problems including different spatial reference systems, storage formats, scales, feature and object definitions, data modelling and data quality. Effective data management and integration also suffer from lack of consistent standards and policies, as policies are agency-driven. Most of the data is still hardcopy or stored in formats that are not compatible across stakeholders, thus hindering data sharing and exchange as well as data integration. Each institution or organisation has a different understanding and knowledge about NSDI. More information and socialisation about NSDI are required to have a better understanding of common interests and goals. No regulation has been implemented to enforce all spatial data providers to become involved in and contribute to the development of NSDI. Such regulation is still in the draft version and being processed by the Ministry of Law. Most of the spatial data providers do not publish enough information (spatial metadata) to enable users to discover the spatial data easily.

Malaysia has identified land development and infrastructure development as the national priority. Malaysia is in dramatic transition from a developing to a developed country and resources and infrastructure play a significant role in the country's development. Therefore, the integration of multi-source data is an essential task that can ensure appropriate management of land and resources. Despite this, data integration is a challenging and problematic task. Data access and availability are significant issues, while different stakeholders have conflicting interests. This makes the usability and distribution of data problematic. Data quality is also weak and so does not meet the expected objectives. A few other problems of data integration, which have been identified by Malaysia, are diversity of reference systems and diversity of data structures and models.

A significant issue that has been highlighted by Malaysia is the weakness of GIS environments in terms of database and software components in data integration. GIS systems are not capable enough to overcome multi-source spatial data integration including database level integration and data model level integration. It also makes the lack of topological relations between classes and inconsistent attribute data. It is more problematic because the spatial and aspatial data are stored separately in different data models in Malaysia. Table 4.4 summarises the spatial data integration drivers and also the challenges in the case study countries.

Table 4.4. Spatial data integration drivers and challenges in international case study countries

Case study country	Data integration drivers	Data integration challenges
Philippines	the creation of a Technical Working Group for the creation of a standard seamless topographic database	1. difference in scales and accuracies between the Philippine topographic and cadastral maps 2. not all maps in digital form 3. lack of standardisation of data and processes to facilitate access, exchange and integration of data 4. lack of technical standards 5. duplication of efforts and resources 6. proliferation of several versions of base maps 7. difficulty in sharing and integration of data
Brunei Darussalam	natural hazards GIS applications	1. no data access tool (e.g. clearinghouse) for the public 2. no metadata and metadata standard 3. lack of technical capability 4. lack of standardisation of data 5. custodianship issues 6. policy of data issuance 7. confidentiality of data 8. security of data (sharing) 9. lack of copyright law 10. no user-driven datasets
Japan	hazard mitigation	1. cadastral maps are not registered in clearinghouse and there is no intention to utilise cadastral map in the public 2. no data model level relation between cadastral and topographic datasets
Singapore	land administration national security hazard management	1. no metadata and metadata standard 2. lack of support from various public agencies 3. lack of legislation support 4. lack of standards 5. collaborative support for NSDI 6. data access policy is agency-specific 7. no user-oriented data
Australia	Pan-Australian integrated datasets (PSMA) and through Harmonized Data Model (HDM)	1. The federated system of governance provides significant difficulties in attempting to create a national overview of spatial information for use in different communities of practice such as environmental management, emergency management and counter-terrorism operations 2. different data access policies 3. full cost recovery 4. different data models 5. different data specifications
Indonesia	natural disaster (mainly earthquake and	1. land and marine are managed differently 2. differences in spatial reference system (horizontal datum, vertical datum, coordinate

	tsunami) land-marine integrated management	system) 3. differences in storage format 4. differences in scale of data source (map scale) 5. differences in feature or object definition 6. differences in spatial data quality due to the differences of resolution or data acquisition methods 7. the differences in spatial data modelling (geometry, features name, attributes, field types, topology, etc.) 8. lack of standards 9. each institution or organization has different policies and rules on managing spatial data 10. most of the data are still in hardcopy format or stored using formats that are not compatible for data sharing or exchange as well as integration 11. each institution or organisation has a different understanding and knowledge about NSDI; more information and socialisation about NSDI are required to have a better understanding 12. no regulation has been implemented to enforce that all spatial data providers should be involved in and contribute to the development of NSDI; such regulation is still in the draft version and being processed by ministry of law 13. most spatial data providers do not publish enough information (spatial metadata) to enable users to find spatial data easily
Malaysia	land development infrastructure development	1. data availability 2. conflicting interests 3. data quality in terms of accuracy and consistency 4. reference system 5. data structure 6. GIS environment (database and software requirements) 7. lack of topology and relationship between classes, and inconsistent attribute data 8. spatial and aspatial data stored separately in different data models

The following section discusses the structure of the questionnaire and the analysis of data gathered through the questionnaire within case study countries. Six out of seven initial case study countries managed to respond and provide information via the questionnaire. These include Japan, Australia, Singapore, Brunei Darussalam, Malaysia and the Philippines.

As a part of the country report a questionnaire was also developed. The questionnaire was aimed at gathering specific information on different aspects of policy, institutional, technical, legal and social principles of case study countries. The questionnaire was also designed to gather as much comparable information as possible. In this regard, the policy principle section comprised descriptive information to address the general policy requirements for spatial information management, policies arrangements regarding the spatial data integration, capacity and policies relating data integration especially from the user perspective. The policy section also contains a section with a table of different policy aspects including existing policies at national, state/provincial and local levels for information management, data modelling, metadata, custodianship, pricing, access, distribution, privacy, security and procurement.

The next section asks the participant countries about the importance of institutional principles including the funding model, collaboration, awareness of data existence, licensing and data access in hindering effective spatial data integration. This section also studies the different access methods in the case study countries. It includes access mechanisms including paper maps, digital access such as CDROM and other portable media, email, online data catalogues, Local Area Network (LAN), Wide Area Network (WAN) and so on.

A part of this section identifies whether spatial data is managed in a centralised or distributed manner within different stakeholders especially governmental organisations. The funding arrangement for spatial data coordination is also identified in this section highlighting the contribution of governments, private sector and public/private partnership in funding or taking a cost-recovery approach.

The technical section of the questionnaire tries to identify the major issues that hinder effective spatial data integration. It includes the importance of some technical issues including computational heterogeneity, vertical topology between multi-source datasets, reference system, data quality, existence and richness of metadata, and data format. The international organisations and bodies for spatial data including ISO, W3C, jurisdictional standard bodies and OGC are also identified in this section. The importance of legal issues hindering data integration in case study countries is also studied in next section. It includes some important legal issues including copyright, intellectual property, data access, privacy and data licensing.

The major users of the two main spatial data themes (topography and cadastre) in terms of the technicians, managers, private sector, academia or military are also identified in the questionnaire. The participant countries are asked to provide information on any capacity building initiatives that have been underway within the jurisdiction in regards to spatial data development. The cost of available data is also gathered in terms of free and open access, cost of data transfer and full cost recovery. This issue is quite important as restricted data access and high costs greatly hinder the use and integration of datasets. Table 4.5 summarises the issues of spatial data integration in case study countries which have been identified through the questionnaire.

Table 4.5. Findings of spatial data integration questionnaire

Issues			Japan	Australia	Singapore	Brunei Darussalam	Malaysia	Philippines
Policy principles	Existence of policies	Management	national	national/ state/local	national	local	national/ state/local	-
Policy principles	Existence of policies	Data Model	national	national	-	local	-	-
Policy principles	Existence of policies	Metadata	national	national/ state	-	-	national/ state/local	-
Policy principles	Existence of policies	Custodianship	national	state	national	local	national/ state/local	-
Policy principles	Existence of policies	Pricing	national	state	-	-	national/ state/local	-
Policy principles	Existence of policies	Access	national	national/ state/local	national	local	national/ state/local	national
Policy principles	Existence of policies	Distribution	national	national/ state/local	national	-	national/ state/local	national
Policy principles	Existence of policies	Privacy	national	national/ state/local	-	-	-	-
Policy principles	Existence of policies	Security	national	national/ state/local	-	local	national/ state/local	-
Policy principles	Existence of policies	Procurement	national	national/ state/local	national	-	national/ state/local	national
Institutional principles	Importance of issue	Funding	I	I	I	I	NVI	I
Institutional principles	Importance of issue	Collaboration	I	I	I	I	NVI	I
Institutional principles	Importance of issue	Awareness of data existence	NVI	I	VI	I	NVI	NVI
Institutional principles	Importance of issue	Licensing	VI	I	I	I	NVI	NVI
Institutional principles	Importance of issue	Data access	N	I	I	I	NVI	I
Institutional principles	Access mechanism	Paper maps	secondary	-	-	primary	primary	primary
Institutional principles	Access mechanism	Directory	-	-	primary	not used	primary	primary
Institutional principles	Access mechanism	CDROM or other digital medium	-	primary	-	not used	primary	secondary
Institutional principles	Access mechanism	Email	-	secondary	-	secondary	not used	secondary
Institutional principles	Access mechanism	Internet	primary	primary	primary	secondary	primary	secondary
Institutional principles	Access mechanism	LAN and WAN	-	-	-	primary	primary	secondary

	SI management (1)		centralised	decentralised	centralized	centralized	centralized and decentralized	decentralized/ federated
	Funding body (2)		government	government/cost recovery/ public–private	cost recovery	government	government/cost recovery	government/cost recovery/ public–private
Technical principles	Importance of issue	Computational heterogeneity	NVI (3)	NVI	VI	I	VI	VI
		Vertical topology	NVI	I	VI	I	VI	I
		Reference system	NVI	NVI	N	I	VI	VI
		Data quality	NVI	I	VI	I	VI	I
		Metadata	NVI	NVI	VI	I	NVI	I
	Standards	International Standards Organization ISO, Technical Committee for Geographic Information/ Geomatics – TC 211	√	√	√	√	√	√
		National Standards Committee or Body	√	√	-	-	√	-
		Open GIS Consortium OGC	√	√	-	-	-	-
		World Wide Web Consortium W3C	-	√	-	-	-	-
		Other	-	-	-	-	-	-
Legal principles	Importance of issue	Copyright	VI	VI	I	I	NVI	-
		Intellectual property	I	VI	N	I	NVI	-
		Data access	NVI	I	I	I	NVI	I
		Privacy	I	VI	I	I	NVI	I
		Data licensing	VI	I	N	I	NVI	-

(1) Centralised (National Government) or Decentralised (State/Provincial/Local)
(2) Government (public sector)/Private sector/Public-private partnership
(3) VI: Very Important, I: Important, N: Neither, NVI: Not Very Important, NI: Not Important

As illustrated in this table, technological issues are not major issues, while institutional arrangements and policy framework developments have been mentioned as major concern of the case study countries. The case study also concludes the role of government is maintenance and sharing of data. This implies the custodianship of government on spatial data with pertaining governmental policies and arrangements. In the institutional category, some outliers have also been seen. It includes the importance of funding and access policies.

The questionnaire goes into a deeper level of detail in the case study countries. This is divided into four major categories of institutional, technical, legal and policy classes. From a policy perspective, the availability of the policies at three national, state and local levels has been studied. This includes policy for different technical and non-technical issues such as management, custodianship, pricing, access, distribution, privacy, security, procurement, data model and metadata. The study shows that the political structure and the SDI development progress have substantial impact on the development and the jurisdictional level of policies. Japan has the data management and SDI policies in place at the national level. Australia shows one of the most mature policy frameworks with some nationally consistent policies including metadata, privacy, access and data models. Australia has policies at state level with policies for the areas in which state is responsible including datasets custodianship, pricing, distribution and metadata policies. Singapore is in the SDI exploratory stage, hence only the general policies in data management, access, distribution, and custodianship have been developed, while more SDI-oriented policies, including technical data model and metadata standard, privacy and pricing, have not yet been put in place. Brunei Darussalam is also in the recent stages of SDI development and there are no well-developed national-level policies especially for some SDI-relevant policies including metadata, pricing and privacy. Local authorities are driving the policies within the boundaries of their jurisdictions. Malaysia also possesses a fairly well-developed SDI in which both national and local policies are developed in major areas such as data management, metadata, custodianship, distribution and access policies. The Philippines is not very well progressed in spatial data policies. However, a few areas have national-level policies in place including procurement, distribution and access.

The institutional arrangements cover four main areas namely, *importance of institutional issues* including funding, collaboration, awareness, licensing and data access from the data integration perspective, *access mechanisms*, *spatial information management* approach in terms of centralised and distributed approaches and *funding bodies*. Most case study countries have highlighted funding arrangements and collaboration as important issues for effective data integration. Awareness of data is addressed as a *not very important* issue in Japan, Malaysia and the Philippines while it is an *important* issue for Australia and Brunei and *very important* issue for Singapore. Japan has claimed that licensing is very important issue within the SDI context, while it is not very important issue for Malaysia and the Philippines. Licensing and data access are mentioned as important issues for Australia, Brunei and Singapore as it avoids misuse of data. Generally speaking, the setting and development of policies are an important issue as they provide common and agreed platforms and mechanisms for coordination of different aspects of spatial data and processes.

In terms of centralised SI management, it closely follows the political structure of the jurisdiction and also the degree of SDI development. SDI development supports the distributed architecture of connected data silos. Countries with central governmental structure and in the early stages of SDI development have centralised the management of SI. This includes Japan, Singapore and Brunei. Other case study countries especially countries with federal structure and well-developed SDIs have decentralised spatial data management. This includes countries like Australia, Malaysia and the Philippines.

Governments are the main funding body in all case study countries. In countries with established data sharing and data distribution mechanisms among stakeholders, the public–private sector collaboration and also data and service provision cover a part of SDI development. To name a few, the Philippines, Australia and Malaysia are among the countries in which both government and service provision through SDI contribute to cover the costs.

Generally speaking, technical issues including computational homogeneity, topology, consistent reference system, data quality and metadata content are important enablers that assist effective data integration, as addressed by the case study countries. All case study countries have adopted ISO standards fully as their primary data coordination and maintenance standard or partially to handle some aspect of data management such as metadata content. It shows the significance of common standard framework for spatial data management. Many of these standards including metadata, quality and spatial reference standards facilitate the integration of spatial data, as they provide a common basis for different technical characteristics of the spatial data and processes. Among case study countries Australia has adopted OGC, W3C standards as well as the standard guidelines developed by national standard bodies. Japan and Malaysia have developed their required standards, while Japan has developed its required standards including Japanese Metadata Profile (JMP).

Generally, all case study countries also believe that legal issues and mechanisms are important enablers for data integration. Case study countries have mentioned copyright, data access and privacy as the most important legal principles that should be considered in the development of an effective spatial data integration framework.

4.3.6. Summary of International Case Studies

The international case studies have raised a number of technical and non-technical challenges and consequently suggested enablers and facilitators to respond to the challenges. They also highlighted the importance of the issues for their national integration initiatives. Taking into account the political and development conditions of case study countries together with the development of SDIs, the issues can be prioritised in order to achieve an effective platform for spatial data integration. The issues can be categorised into five major classes (technical, institutional, policy, legal and social) based on their impact and nature, as summarised in Table 4.6.; however, more investigation has been conducted through Australian case studies which together with the outcomes of this chapter will provide inputs to the development of an integration toolbox.

Table 4.6. Potential spatial data integration challenges in international case study countries

Technical Challenges/Issues	Non-technical			
	Institutional	Policy	Legal	Social
difference in scales and accuracies not all maps in digital form lack of standardisation of data and processes lack of efficient tools for data integration (database and software) lack of metadata lack of consistent metadata standard lack of data model relation between datasets (geometry, features name, attributes, field type, topology, etc.) lack of link between data specifications differences in spatial reference system differences in data structure and storage format differences in scale of data source (map scale) differences in feature or object definition (specifications) diversity of data quality (accuracy, logic and consistency) lack of interoperability lack of topology and relationship between classes, and inconsistent attribute data	duplication of efforts and resources (different versions of data) unclear custodianship lack of legislation support federated system of governance with little link between jurisdictions fragmented management of different data themes (e.g. land and marine) conflicting interests	data provider-driven datasets agency-specific policies restricted cost recovery	restricted data security (confidentiality of data) complex copyright law	lack of support from various public agencies lack of collaborative support for NSDI silo mentality different understanding and knowledge about NSDI and its missions

The above-mentioned non-technical issues such as conflicting interests, fragmented data management, restricted data security and different understandings and knowledge hinder spatial data integration directly and also impact or cause some technical and non-technical issues. As an example fragmented data management causes a diversity of collaboration arrangements, data sharing approaches and funding models that are utilised to coordinate different aspects of spatial data and services. It also may lead to different data model adoption, diverse coordinate systems, different data quality and different formats that are consequent and indirect technical problems. In cases where there is a well-established SDI with the contribution and agreement among all stakeholders, there are less issues and challenges. In these cases there are also effective conversion mechanisms in place to overcome different technical and non-technical issues.

The non-technical challenges also emerge as the major source of problem and highlighted similarly in different case studies. Some of the technical issues are also the result of non-technical issues. It includes data provider-driven policies which restricts the interoperability and consistency. Silo mentality, lack of collaborative SDI support and complexity of access and distribution policies have also been the source of data integration challenges.

The focus on collaborative SDI support at national level, building capacity and awareness among spatial data stakeholder and also the development of necessary policy and legislative mechanisms could effectively address and overcome these issues.

The investigation of the impact of mentioned challenges requires more study through a methodological approach. This chapter provides some essential components to be investigated in next chapter including:

- a number of technical and non-technical issues
- possible technical and non-technical enablers to overcome challenges
- some contextual details that need to be gathered through technical visits
- technical methodology components
- the consideration of country context and SDI progress in designing an integration toolbox and their impact on the issues and approaches
- the consideration of SDI context as a holistic platform that overarches the integration toolbox components.

Last two items show the mutual impact of *SDI development* and *development within SDI context* for the success of the integration toolbox. Spatial data integration is more straightforward within a well-developed SDI, while the SDI context provides a comprehensive platform to investigate and identify the requirements of effective data integration.

4.4. Chapter Summary

International case studies show the diversity of issues that hinder spatial data integration both technically and non-technically. The issues of and barriers to data integration including institutional, policy, legal and social issues are major concerns, as highlighted in most country reports. In this regard, the establishment and maturity of SDIs, the participation of stakeholders in spatial initiatives and also the country context such as political structure have a high impact on the effectiveness of coordination and management of data and as a consequence on the effectiveness of multi-source spatial data integration. Countries with central governmental systems have more consistent national policies and standards as the national government is in charge of the development and endorsement of policies and standards. Conversely, countries with a federal structure have fragmented spatial data coordination and maintenance among states, which are responsible for their own policy and standard developments.

The case study countries have also identified a number of specific technical and non-technical issues. From a technical perspective, the major technical barriers to effective spatial data integration are difference in scales and accuracies, hardcopy maps, lack of standardisation of data and processes, lack of efficient tools for data integration (database and software), lack of metadata or consistent metadata standard, lack of data model relation between datasets (geometry, features name, attributes, field type, topology, etc.), lack of a link between data specifications, differences in spatial reference system, differences in data structure and storage format, differences in scale of data source (map scale), differences in feature or object definition (specifications), diversity of data quality (accuracy, logic and consistency), lack of interoperability, lack of topology and relationship between classes, and inconsistent attribute data .

There are also a number of issues that have been identified by case study countries that emanate from institutional arrangements. These include duplication of efforts and resources (different versions of data), unclear custodianship, lack of legislation support, federated system of governance with little link between jurisdictions, fragmented management of different data themes (e.g. land and marine), and conflicting interests. It is obvious that some of these issues are valid in specific countries. For example, the federated system of governance is not applicable to countries that are ruled by a central government.

Some barriers are also of a policy nature. Obstacles arise from the policy priorities and arrangements of respective jurisdictions. The adoption of agency-specific policies, restricted cost-recovery policies, and agency-driven dataset generation are the major policy barriers identified in the case studies. Case study reports also highlighted legal issues such as hindrances of data integration. Restricted data security (confidentiality of data) and complex copyright laws are two major legal barriers, as identified by the case study countries.

Some of the issues raised by countries have social roots and emanate from societal behaviours. Therefore, they are not valid for all situations as they are context-specific. This includes possible social issues including lack of support from various public agencies, lack of collaborative support for NSDI, silo mentality, and different understanding and knowledge about NSDI and its missions.

The more non-technical issues identified, the more the necessity for a holistic framework of SDI becomes obvious. Technical issues can be treated only if they are not the result of a non-technical mechanism. Many technical issues including the diversity of format and reference system have appropriate solutions; however, there are many others that require technical advancements and sophisticated technologies and approaches including the development of data validation tools. Therefore, effective multi-source data integration requires the provision of both technical and non-technical mechanisms to overcome the barriers. Effective data integration is not achievable easily unless these prerequisites are identified and coordinated under the holistic framework of SDI. This is addressed in Chapter Five where Australian case studies are investigated based on a methodological approach. This includes a technical case study part of which has been conducted to audit the multi-source datasets and study the potential technical barriers and also possible solutions. A series of technical visits have also been done to investigate the institutional, policy, legal and societal barriers to data integration. However, a number of technical obstacles and likely solutions have been identified through the technical visits. Based on the outcomes of Chapters Four and Five, in Chapter Six an integration toolbox is proposed with technical and non-technical enablers to assist practitioners for effective spatial data integration.

Chapter 5

Data Integration:
Australian Case Studies

5.1. Chapter Aims and Objectives

Chapter Five aims to elaborate on the potential barriers to effective spatial data integration and possible solutions through conducting a number of technical case study visits and data audits in Australia. This chapter also capitalises on the findings and observations of the previous chapter (Chapter Four) where some technical and non-technical challenges of spatial data integration were identified through a number of international case studies. Chapter Four provided a methodological approach (comprising questionnaire and integration template) within which the major technical, institutional, legal, policy and social issues are investigated together with the impact of the issues on effective data integration in the international case study countries.

This chapter comprises two main sections. The first section presents the observations and results of a technical case study investigation. The technical case study has been conducted utilising spatial datasets for four local governments in Australia. This section of the case study aimed to identify the technical obstacles of data integration through the auditing of actual spatial data from different federal and state agencies. This includes the study of data themes, data specifications, data models, metadata, spatial and non-spatial contents and so on. The second section discusses the outcomes of a series of visits to a number of state and federal agencies and also the investigation of multi-source spatial data integration in different organisations. This included investigation of the activities in mapping, maintenance and coordination parties. Meetings and discussions have also been conducted with technical people who have been involved in different inter- and cross-agency spatial initiatives and also with policy-makers who have been responsible for setting up policies, arrangements, standards and collaborations within and across agencies. This chapter summarises major potential technical and non-technical problems of integration.

This chapter is followed by Chapter Six in which a spatial data integration toolbox is presented. The spatial data integration toolbox discusses the necessary components (in terms of technical and non-technical enablers) for effective multi-source spatial data integration.

5.2. Introduction

Chapter Four has highlighted that any jurisdiction may encounter a variety of issues and challenges in the integration of multi-source spatial data. This depends on many factors including the institutional arrangements and political structure, the maturity of spatial data coordination infrastructures (especially SDIs) within and across agencies, the attitude and awareness of stakeholders, the existence of well-developed and effective standards and technical tools and so on. In this regard, Chapter Four conducted a number of case studies in seven countries in the Asia-Pacific region. Based on the findings of Chapter Four, Chapter Five utilises a methodological approach to audit multi-source spatial data and also studies diverse institutional, legal, policy and social issues and possible obstacles of spatial data integration in Australia. The methodology of investigating the case studies and the aim of each case study together with the relation between international and Australian case studies have already been illustrated in Figure 4.1. Figure 5.1 highlights the Australian case study stage of the methodology.

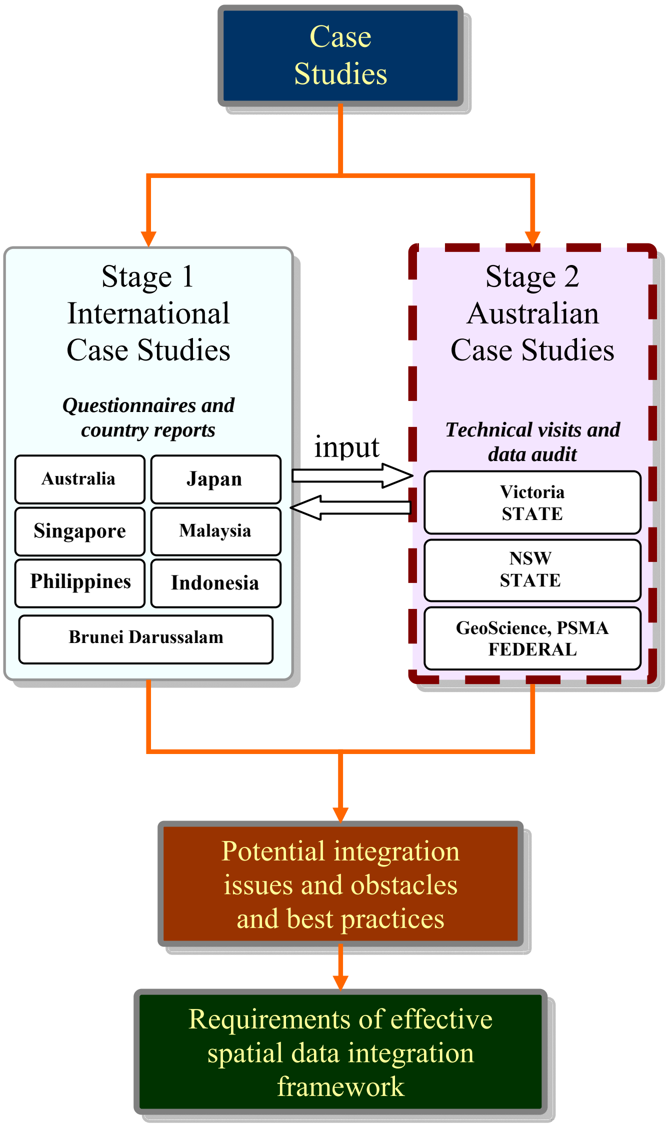


Figure 5.1. Case study investigation methodology (as in Figure 4.1)

The Australian case studies consist of two major sections:

- a technical evaluation (data audit) of a number of Australian state and federal spatial datasets for four local councils
- a series of technical visits at state and federal levels and visits to respective agencies.

In the technical evaluation (data audit), the states of Victoria and New South Wales (NSW) were asked to provide spatial data in different themes including topology, cadastre, administrative boundaries, imagery and so on. Department of Lands, New South Wales (NSW) and Department of Sustainability and Environment, Victoria, which are custodians of the most fundamental datasets within their respective states, manage and disseminate state spatial data. These agencies have been project partners that have supported the research project. These agencies provided datasets within the boundaries of two local councils. GeoScience Australia is the federal mapping agency and provided national datasets for the same local councils. Therefore, spatial data at state and national levels for four councils has been collected and investigated based on a methodology that has been designed to investigate the integration issues among datasets. The methodology tries to investigate datasets based on a structured approach in which different technical issues (identified in Chapter 4) can be studied in detail. There is also a template designed to document the issues and challenges (Appendix 3). The second phase of Australian case study investigation provides information on the history and status of SDI initiatives, capacity for and policies relating to data integration and institutional support for and barriers against integration, based on jurisdictional case studies of the Federal Government (including GeoScience Australia and the Public Sector Mapping Agency), Victoria and NSW. Overall findings and conclusions based on national synergies for data integration, including recommendations, are provided in the final section of this chapter.

Through the investigation and data assessments, visits and meetings, common potential issues that hinder effective spatial data integration within the case study jurisdictions have been identified (Figure 5.2) and have provided inputs to the spatial data integration toolbox (Chapter Six) with focus on facilitating the multi-source spatial data integration process.

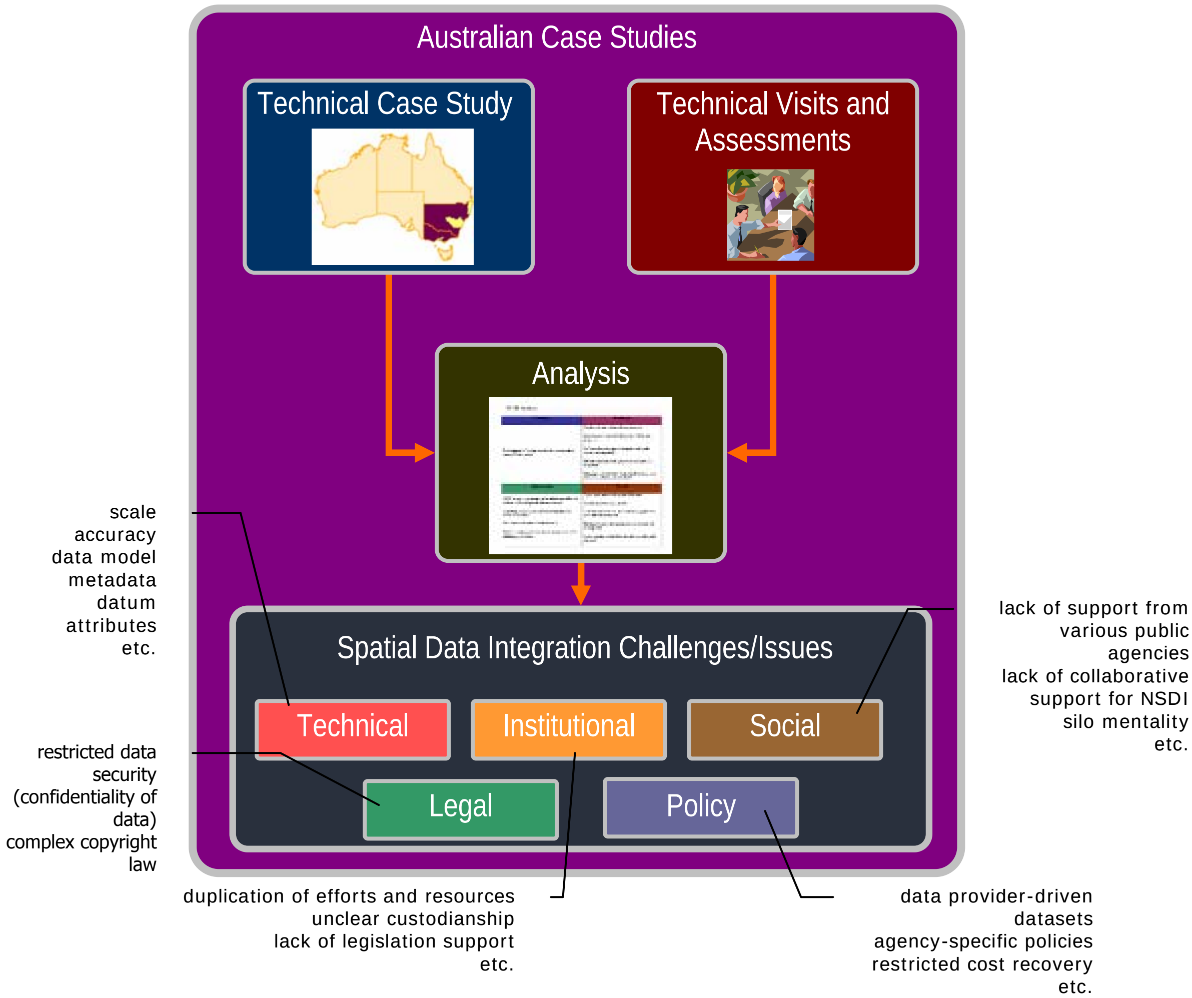


Figure 5.2. Australian case study design and outcomes

Two phases of Australian case study investigations assisted in recognising the major barriers and obstacles of multi-source spatial data integration and consequently the identification of necessary tools and mechanisms to overcome the issues.

5.3. Technical Investigation

The technical data assessment and investigation was carried out on two municipalities in the states of Victoria: Casey and Yarra Ranges and two municipalities in NSW: Blue Mountain and Coffs Harbor. The investigations have been conducted based on a methodological approach with consideration of the issues raised and outcomes concluded in Chapter Four. The issues have also been documented based on a template. The template consists of problem description, involving datasets, snapshot of occurrence, frequency of the problem, cause and possible impact of the problem on the effective integration of datasets.

Yarra Ranges Shire is located in Melbourne's outer east – between 30 and 110 km east of Melbourne's city centre. The Shire covers an area of almost 2,500 sq km, the largest area of any metropolitan council and is home to more than 143,000 people. The Shire offers a mixture of urban and rural communities. Around 70 per cent of the Shire's population lives in the urban areas of the Shire that represents approximately 3 per cent of its landmass. The remaining population is distributed throughout rural areas. The City of Casey is located 35 km from the Melbourne city centre, in Melbourne's south-east. Casey has a total area of around 400 sq km, and its population is estimated at 221,058. Currently, approximately 55 families move into the area each week, totalling 8,700 each year. It is expected the population will grow dramatically, making it the largest and fastest-growing municipality in Victoria, and the third-fastest growing city in Australia behind Brisbane and Gold Coast City Councils. Coffs Harbour is a coastal city located on the north coast of New South Wales. The region has a population of nearly 70,000. Popular with people wanting to relocate from big cities to small towns on the coast or in rural areas, Coffs Harbour continues to grow at an exceptional rate, with a population projection of 80,014 by the year 2016. The City of Blue Mountains is a local government area of New South Wales, Australia, governed by the Blue Mountains City Council. The city is located in the Blue Mountains range west of Sydney.

5.3.1. Methodology Design and Investigation

There is a methodology designed for the technical investigation phase. This methodology capitalised on the technical outcomes of the previous chapter including the differences in data models, formats, attributions and so on (Table 5.1). It also tried to address and investigate these issues in next phase of jurisdictional investigation through visits and meetings.

Table 5.1. Technical challenges identified in Chapter Four

Technical Issues (identified through international case studies)
difference in scales and accuracies
not all maps in digital form
lack of standardisation of data and processes
lack of efficient tools for data integration (database and software)
lack of metadata
lack of consistent metadata standard
lack of data model relation between datasets (geometry, features names, attributes, field types, topology, etc.)

lack of link between data specifications
differences in spatial reference system
differences in data structure and storage format
differences in scale of data source (map scale)
differences in feature or object definition (specifications)
diversity of data quality (accuracy, logic and consistency)
lack of interoperability
lack of topology and relationship between classes, and inconsistent attribute data

The methodology starts at the highest level of data themes (e.g. property and transport) and breaks down to the feature level (points, lines, polygons) and investigates the available datasets for possible challenges at theme to feature levels. As an example, at the highest level: the transport network *theme*, roads and ferry routes *layers* etc. and finally individual segments of the roads *features* from one data provider have been investigated and compared to the same level from the other data provider.

It also investigates the spatial, aspatial and metadata content and documents the potential issues and challenges. The datasets are compared against datasets of other jurisdictions and also datasets of same jurisdiction. As illustrated in Figure 5.3, the datasets are then investigated for potential issues and barriers including any discrepancy, incompliancy and heterogeneity.

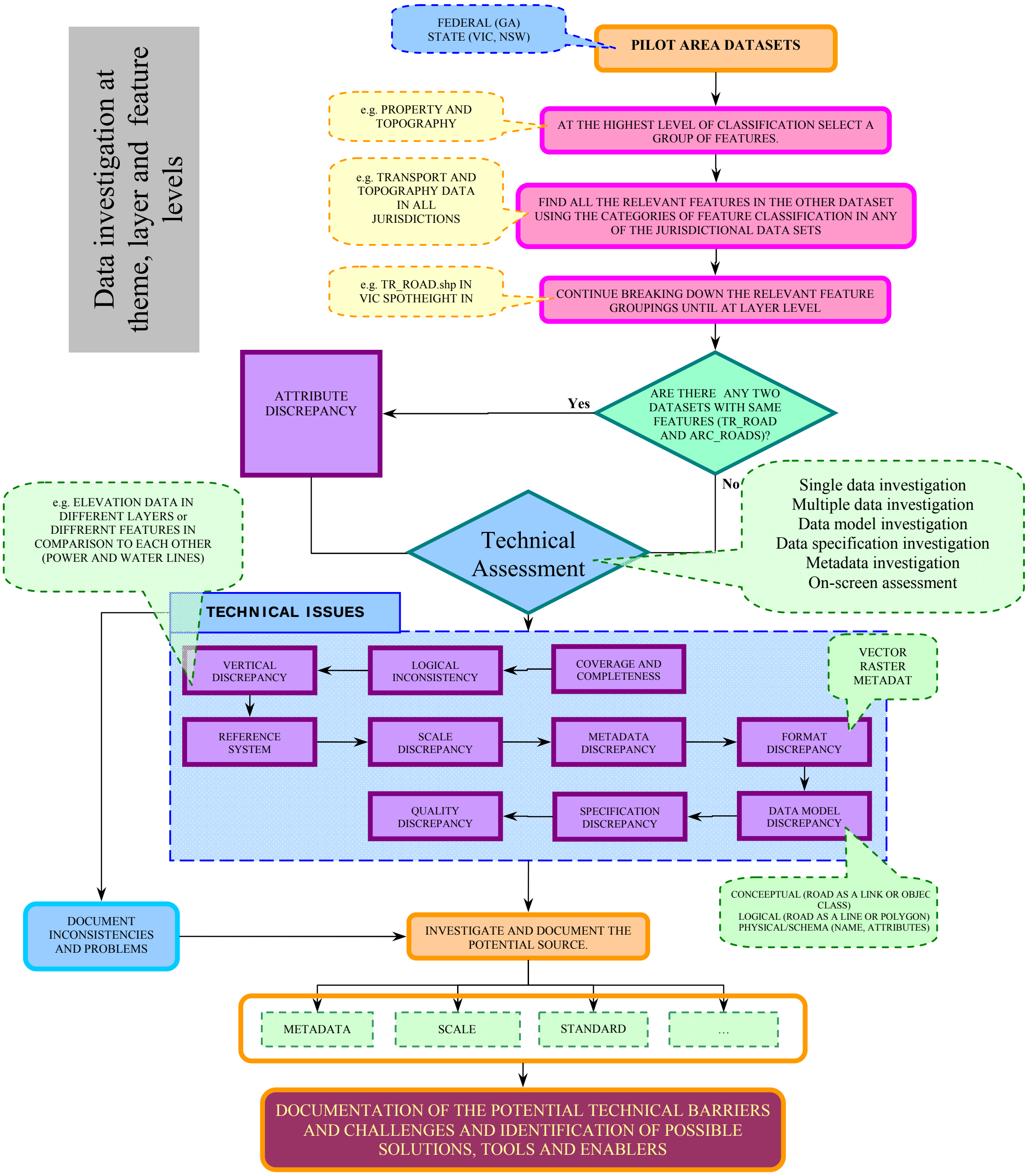


Figure 5.3. Technical investigation methodology for Australian case studies

The technical assessment model comprises a number of approaches in which spatial data layers and features have been checked for any issue and incompliancy. It includes the investigation of all data layers individually to identify their characteristics, the investigation of multiple data layers to compare the characteristics data layers, data model investigation to identify the data classes and relationships, data specification document investigation to identify the restriction, descriptions and detailed specification of datasets, metadata investigation for further data characteristic extraction; and on-screen data comparing. Investigating the data content from theme to layer and finally features provides an appropriate and detailed study to identify the issues among involving datasets. At each level datasets from the same themes or different themes have been investigated. In this regard eight themes with around 70 data layers from Victoria's datasets, six themes with around 100 layers from NSW and 20 themes with around 90 layers from Geoscience Australia have been involved in the technical assessment.

The technical assessment also included comparing data from either the same state (also federal level) or two states or states and Commonwealth. Major investigations and items which have been utilised have been shown in Table 5.2:

Table 5.2. Investigation activities and utilised items for technical assessment case study

Investigation activity	Items utilised
investigating and comparing data specifications (through metadata, data models and data specification documents)	Metadata, data models, data specification documents
investigating and comparing data specifications (through data content, file names, layer names and aspatial content)	Data layers, data model, classification, names and spatial and aspatial content
spatial data content investigation (content and classifications)	Data model and data layers
aspatial data content investigation (types, names and content)	Aspatial tables
conceptual data model investigation (and with regards to specifications)	Data models and data specification document
physical data models investigation (in terms of feature definition and relation between features)	Data models
data quality investigation in terms of currency, spatial and aspatial accuracies (through metadata and data specifications)	Data layers, metadata and data specification document
visual investigation of data quality (completeness, coverage, logical consistency, data redundancy and scale)	Data layers
metadata consistency investigation (content and profile)	Metadata and metadata DTD (Document Type Definition) document
metadata suitability investigation (measurable content and machine-readability),	Metadata
database management approach (database design and storage forms)	Data structure and geodatabase
appropriateness of validation and integration tool (from database, data model and vertical topology aspects)	Database, data model and actual data layers
vertical relation between features of different themes	Data models and data layers
spatial reference system (SRS)	Data layers and metadata

Mainly spatial and aspatial investigations have been conducted in an ArcGIS environment. Metadata, metadata profiles, specifications and available data models have been also investigated based on the provided documents. ArcGIS is one of the most reliable GIS software available and provides many different functions in order to investigate, assess and validate data. In order to document the findings in a structured manner a template has been designed which not only contains the problem and its description, but also elaborates the frequency, cause of the issue and potential impact of the effective data integration (Figure 5.4).

Name of the Problem:
Problem Description:
Theme/Layer/Feature/Document Used:
Source Jurisdiction(s):
Snapshot of the Problem: <usually a picture or sample of the problem illustrating the problem> 
Frequency (Uniqueness of the Problem):
Possible Cause:
Possible Impact:

Figure 5.4. Technical case study investigation template

The problem title assigns a name to the identified problem in order to categorise and represent the problem. The problem description contains the observation of the occurrence of the problem and the instance of the problem that has been identified. The template also contains the evaluated data or documents together with the source jurisdiction(s). It also illustrates the observation of the problem with a snapshot of the problem. Frequency, possible cause and impact on effective data use and integration have also been detailed in the template. An effort was made to investigate the major technical issues. The following section presents the main findings in the various categories of data specification, data model investigation, data quality investigation, metadata investigation, database management approach, spatial reference systems and validation and integration tools.

Data specification

In order to identify the issues and challenges of data specifications, investigations have been made through two different sources: accompanying documents of datasets including metadata, data models and data specifications document; and data content, file names, layer names and aspatial content. Each of the three jurisdictions (states of Victoria and NSW and Geoscience Australia) has specific data models and data specifications that are different in many aspects from other jurisdictions. Differences include:

- fundamental datasets
- layer and feature classification
- classes of features
- detail level of features
- aspatial content
- naming system.

Victoria maintains VicMap fundamental datasets that are a combination of about 70 layers within eight themes (SII, 2008). NSW coordinates around 100 layers within six main themes (ASDD, 2007) and GA has around 90 layers within 20 themes (ASDD, 2007). As identified by case study investigation, layers and feature classifications are also differently defined within these jurisdictions. There is also a different number of layers in similar themes within case study jurisdictions. These three jurisdictions maintain different levels of detail for some features such as road-side infrastructures. This is also applicable to aspatial content, where there is a different number of attributes with different naming and types in the same layers. The naming system is another inconsistency among the case studies. Table 5.3 presents some examples of the above-mentioned data specification issues for three case study jurisdictions.

Table 5.3. Examples of identified data specification issues

	Victoria	NSW	Geoscience Australia
Fundamental datasets	8 themes with around 70 layers*	6 themes with around 100 layers*	20 themes with around 90 layers**
Layer and feature classification	No separate layer for water tanks and it is a part of **HY-WATER_STRUCT_POINT** with watering places and swimming pools in **Hydro** theme	**TankPoint** layer in **TopoFD** theme	**Point_watertank** layer in **Drainage** theme
Classes of features (e.g. administration theme)	23 layers in **Vic_Admin** theme	3 layers in **Admin Boundaries** theme	4 layers in **Administration** theme
Detail level of features (e.g. road-side facilities)	Road-side facilities kept in **Road** theme	No road-side facilities kept in **Road** theme	Road-side facilities kept in **Road** theme
Aspatial content (e.g. administrative boundary layer)	12 attributes for administrative boundaries	8 attributes for administrative boundaries	16 attributes for administrative boundaries
Naming system	**AD_LOCALITY_AREA_POLYGON** layer in **Vic_Admin** theme	**Administrative Areas** layer in **Admin Boundaries** theme	**Polygon_Administrative** layer in **Administration** theme

* for 1:25,000 datasets
** for 1:250,000 datasets

The data specifications documents contain fairly detailed conceptual definitions of data features, and metadata in case study jurisdictions is based on ANZLIC profile and is common across case study jurisdictions (both federal and state). Metadata contains information on different aspects of data including custodian, source of data, technical characteristics (accuracy, scale and datum etc.), restrictions and access information. According to the findings through case studies, different specifications and differently defined fundamental datasets with inconsistent conceptual data models are the major cause of the issues presented in Table 5.3. Data specification documents are differently developed in terms of structure and content within case study jurisdictions. GA and NSW's data specifications contain a brief description of layers and features and also attributes which mainly presents the content rather than the concept and characteristics, while Victoria specifies the feature based on their characteristics. For example, Table 5.4 presents the definition of "freeway" features in Victoria and NSW.

Table 5.4. The description of "freeway" features in Victoria and NSW's data specifications

Victoria	NSW
Freeway is a hard surface formation, high volume, high speed roads declared as "Freeway"; comprising dual carriageway and full access control and grade separated intersections; i.e. no direct access from adjoining properties or side roads and all crossings are by means of overpass or underpass bridges with traffic entering or leaving carriageways by means of ramps. Single carriageway sections forming part of declared freeways may be included within this category.	Freeway is a part of **DualCarriageWay. DualCarriageWay** is a subtype of **roadsegment**. **Roadsegment** is a line feature class representing the centreline of a section of road having common attributes and terminating at its physical end or at an intersection with another road at the same grade (same level). An integer attribute called **ClassSubtype** singles out the **dualcarriageway** within **roadsegment** class by quantity of **3**.

The definition of freeway feature in Victoria's data specification document is the combination of a number of aspatial and spatial properties and also in relation to other features which can be assigned to a feature (in data model) by attributes (hard surface, high volume, high speed), geometry (line/multi-line), restrictions (no direct access from other roads), rules (separated intersections) and relations (no direct access from features of property theme), while the description of a freeway in NSW is not clear and it is more a description from a database design perspective. Consequently, this causes a time-consuming process of integration of relevant features that requires extracting features from different datasets. Therefore, the integration of diverse datasets especially at data model level is not possible. Creating a new integrated data model requires reclassifying the features, which can be done based on the conceptual definition, spatial and aspatial specifications, restrictions and relations.

The lack of a conceptual and comprehensive definition of features that contain the above-mentioned information is a common problem in all participating agencies. A well-established agreement among spatial data custodians to provide this information can assist to overcome this issue.

Data model investigation

Each of the case study jurisdictions use their own data model that is different from the data models utilised by other jurisdictions. This is different at both conceptual and physical levels. At the conceptual level there are different conceptual definitions for spatial entities, classifications and relations to other features. This influences the physical data model design, which mainly concerns storage and database management issues. As an example, transport network (road) classes are differently defined within case study jurisdictions. Figures 5.5a, 5.5b and 5.5c illustrate the road data models for NSW, Victoria and GA respectively.

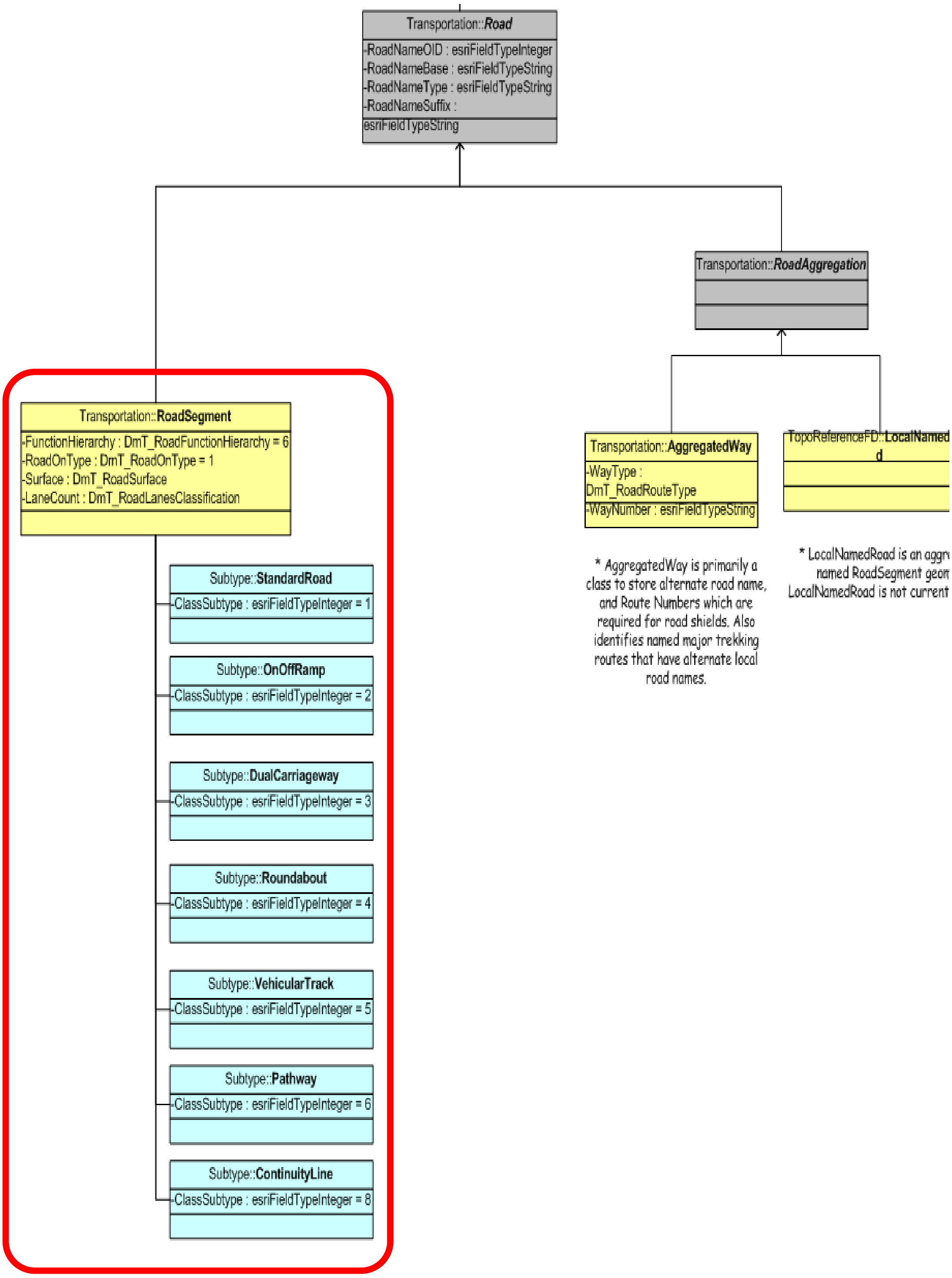
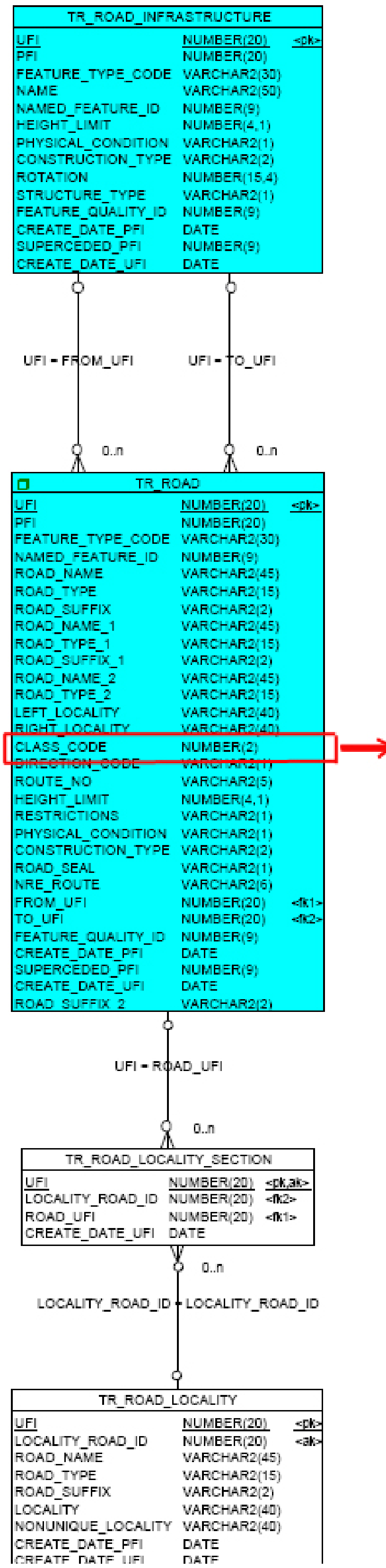


Figure 5.5.a. Road conceptual data model of NSW



LOOK UP TABLE CODELISTS APPLICABLE:
Class_Code

Code	Description
0	Freeway
1	Highway
2	Arterial
3	Sub-Arterial
4	Collector
5	Local
6	2wd
7	4wd
8	Unknown
9	Proposed
11	Walking Track
12	Bicycle Track

Figure 5.5.b. Road conceptual data model of Victoria

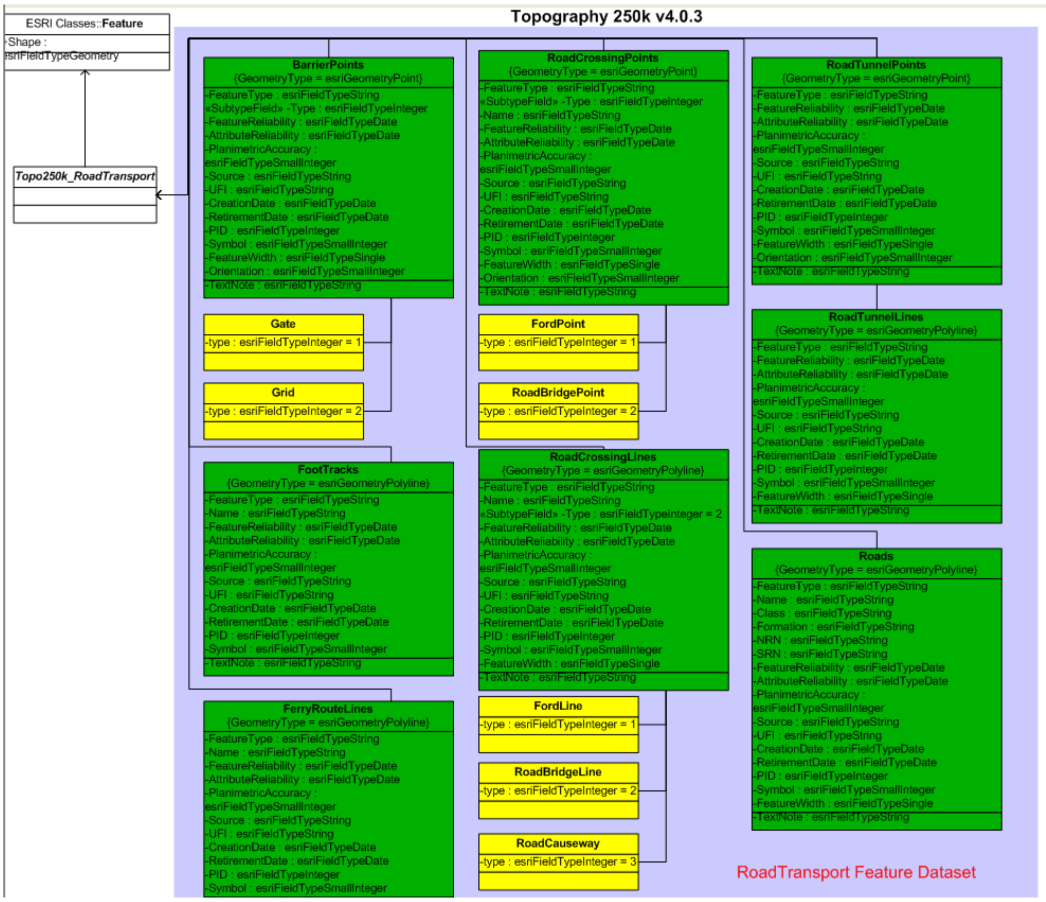


Figure 5.5.c. Road conceptual data model of GA

The State of NSW maintains seven classes of roads that are all within a single class of "roadsegment" and are singled out by an attribute (subType). Victoria's road theme consists of 12 road types that are stored within a single class of "TR-ROAD" and are determined by an attribute called Class-code. GA has taken another approach and has separated some road types including foot tracks and ferry routs from other road types. Therefore, in GA's road data model there are three major road classes. These classes are not the only distinction between data models. Almost all features have been classified and modelled differently, and also the relation between different classes has been defined differently. Taking the same example, road infrastructures in Victoria, including bridges and facilities are kept in a single class of road infrastructure (TR_ROAD_INFRASTRUCTURE). GA has different classes for road infrastructure including barriers, crossings, bridges, tunnels and so on. Victoria's road class has direct relation (association) with two single classes of road infrastructures and localities, while GA's road class has an association with different road infrastructure classes.

This would be a huge problem for spatial data integration, especially when datasets are required to be integrated at (conceptual, logical and physical) data model level(s). The reason is that similar spatial features have been differently categorised and classified. For data model integration, features from different data models should be extracted and encapsulated in a new data model structure. This is not easily practical, unless a common mapping method between diverse data models is available, which identifies and extracts the same features. An approach to this process is ontology-based data modeling, which is based on feature ontology and conceptual specification. The main objective of ontology is the sharing and reuse of knowledge by different disciplines (Lin et al., 2001). Ontology defines the vocabulary based on which queries and assertions are exchanged among entities. This can be used to develop a common vocabulary for multi-source spatial features. This approach will be elaborated in the next chapter.

Data quality investigation

Data quality can be reflected in a number of data characteristics including:

- spatial and aspatial accuracies
- currency
- coverage
- completeness
- logical consistency
- no data redundancy
- scale (Goodchild & Gopal, 1989).

The major source of data quality information is actual data and metadata. In order to derive data quality items in the case studies a number of sources have been investigated. Data contains some explicit and implicit information on quality items. Scale can be easily derived from data within almost all GIS software, while coverage, completeness, logical consistency and redundancy can be derived from geometrical (spatial) content of data. By examining data against other datasets, it can be ascertained whether data is covering the expected area or not. It also shows whether data is complete, logically consistent and not redundant. Metadata also contains information of the quality items.

In the case studies, both sources of information (actual data and metadata) were investigated. In some cases there was some inconsistency observed frequently, in terms of completeness, logical consistency and redundancy. In other cases no inconsistency was observed. Figures 5.6a, 5.6b and 5.6c illustrate some snapshots of the inconsistencies observed.

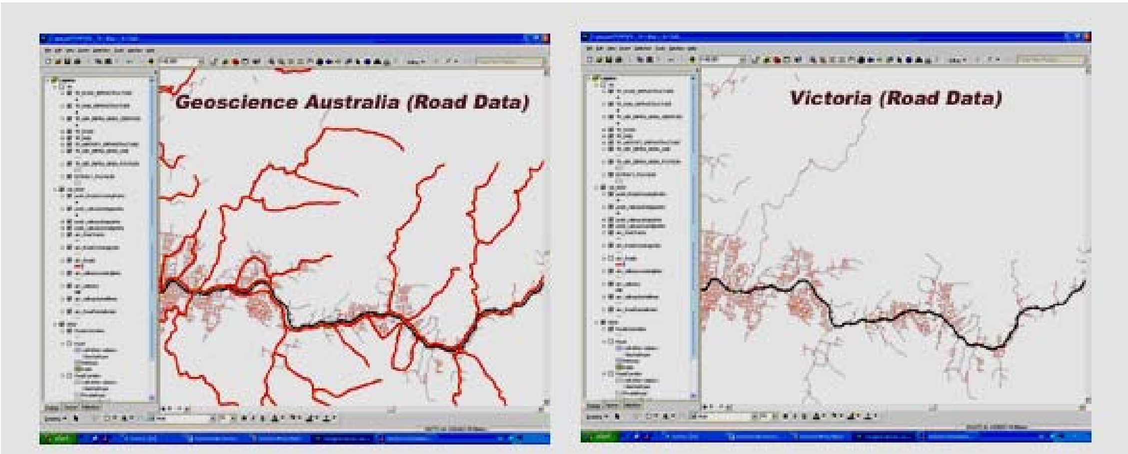


Figure 5.6.a. Data completeness using same area but different datasets

The completeness of the data refers to the area that a specific data layer covers. If the data layer is not complete, the integration and analysis cannot be done as the information does not exist for that particular area.

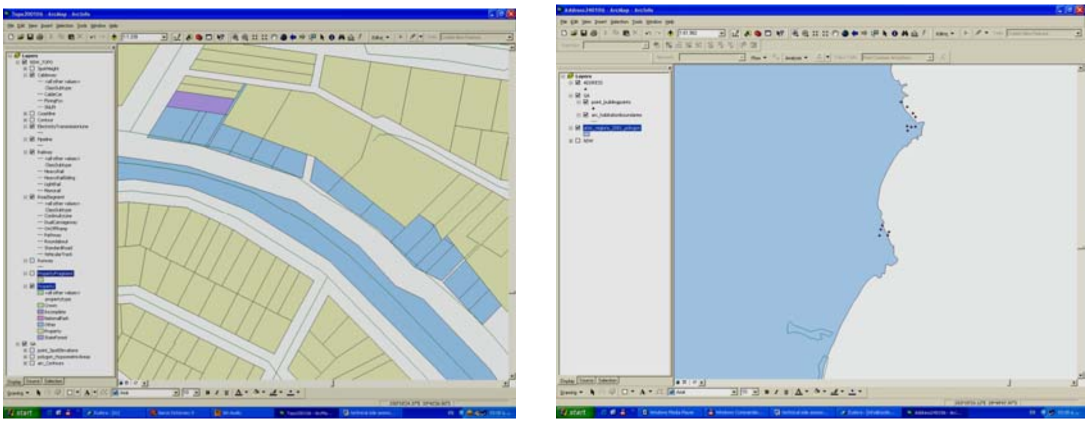


Figure 5.6.b. Logical inconsistency: NSW roads crosses buildings and building points are out of maritime boundaries

Logical consistency also impacts the data analysis. In some cases it was observed that the data layers do not comply with the logical consistency rules. For example, the property features are located within the water bodies.

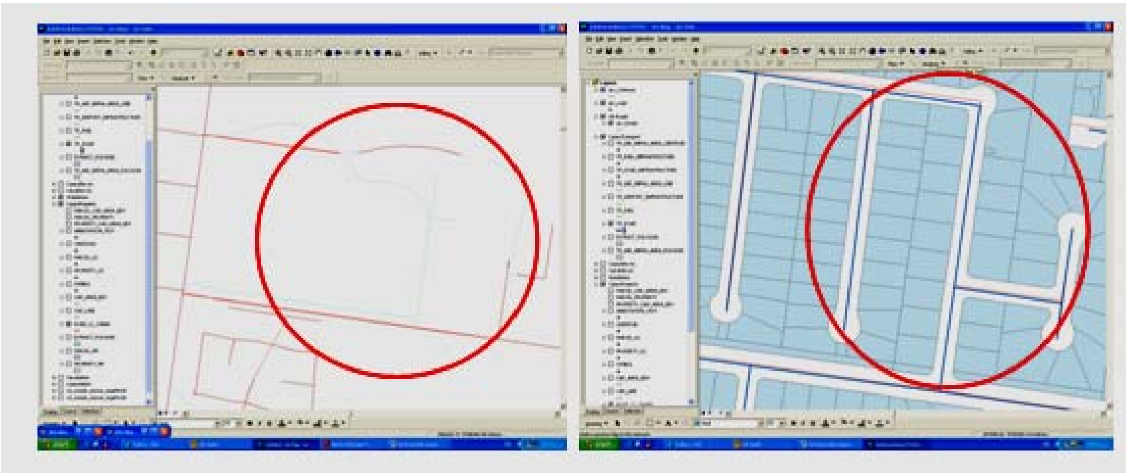


Figure 5.6.c. Data redundancy in Victoria's road datasets (two distinct road layers)

Data redundancy was also observed in some cases. As an example, Victoria has a number of road layers that are kept within different themes (road network and property themes etc.), which are geometrically and aspatially different.

Metadata investigation

Metadata contains invaluable information about different aspects of datasets. It includes information on source, custodian, quality; and access channels. Metadata follows a structure that is defined by metadata profile. In the case of Australia, ANZLIC's metadata profile has been adopted as the nationally consistent metadata profile by Australian jurisdictions. The ANZLIC guidelines (ANZLIC, 2006) have been developed to promote a consistent standard to describe a number of core metadata elements that are generally common for all types of data and designed to indicate the data, its content, geographic extent, how useful it might be for other purposes and where more information about the data can be obtained. The purpose is to make information about all available data freely available so that existing data can be reused for other purposes if it is suitable.

From an integration viewpoint, a number of metadata problems in Australian case studies have been identified through metadata investigation, which includes:

- the richness and currency of metadata content
- no machine-interpretable content
- lack of measurable content.

The metadata standard is consistent across Australian states and territories, while in some cases the information content of metadata is outdated. For example, there is no metadata for PSMA's products (ASDD, 2007). There is outdated metadata in the ASDD (Australian Spatial Data Directory) including outdated contact information for some of NSW's data (*for example email address <name@dtm.gov.au> no longer exists*). Some data themes, including building datasets, also suffer from lack of metadata. This is a hindrance for spatial data integration where users require updated and rich information on data quality, access channel, restrictions on data and so forth.

Another problem is the format, structure and content of metadata to be used for automation of information extraction. Metadata is mostly maintained in Word document, PDF (Portable Document Format) and html format. In some cases (like GA), metadata is also maintained in eXtensible Markup Language (XML) form, which is a structured and machine-readable form. XML provides a better structure for information content, which can be interpreted by computer programs (Walsh, 2008).

Database management approach

Spatial databases have been structured and designed utilising different tools and approaches within case study jurisdictions, which differs from other jurisdictions. Storage form and distribution of spatial data have also been coordinated differently. Victoria mainly provides and distributes the datasets in ESRI shapefile format. GA provides quite a few different formats including MapInfo and ESRI shapefile. NSW coordinates datasets within ESRI Geodatabase, which is a completely new and different approach of database design and management. Within Geodatabases integrated collection of datasets are managed. This includes the definitions, integrity rules, and behaviour of datasets (Arctur & Zeiler, 2004), while a shapefile or MapInfo file is a stand-alone and autonomous file. This makes the integration of different datasets problematic as the storage and management tools are different for the above-mentioned formats.

Spatial Reference Systems

The spatial reference system (SRS) is one of the geographical components of spatial datasets that define the locational metrics for datasets. Many GIS tools are capable of converting reference systems; therefore it is not a big issue. The only inconsistency that has been observed through the case studies was the diversity of definitions used by the three case study jurisdictions for a single datum. Australia's standard datum is GDA94, which has been specified variously as:

- Map Grid of Australia Zone 55 (GDA 94) for Victoria
- GCS_Assumed_Geographic_1 for NSW
- GSC_GDA-1994 for GA and some NSW's datasets.

Validation and integration tool (database, data model and vertical topology aspects)

The entire case study investigation developed the idea of having an effective data validation/assessment and integration tool as a part of any sharing platform in an SDI initiative. This was also backed by the necessity for an automated process that prevents the time-consuming and difficult process of data assessment. The automated process of data integration and validation provides a computer application with a consistent approach to evaluate and integrate multi-source datasets. It also provides a structured approach to evaluate and integrate the datasets. In an ideal situation, the tool can have two major roles. First, it can assess and evaluate the datasets and second, it can provide a gateway for an SDI database as a single access, integration and sharing point. The integration capability at different levels including data model, topology and attribution makes it an effective tool for data integration. The availability of such a tool has been discussed and probed in case study jurisdictions during the visits, which is detailed in the next section (technical visits) of case study investigation in this chapter.

5.3.2. Analysis and Conclusion

The case studies suggest a number of key issues and challenges exist, which should be considered and addressed for effective spatial data integration and also the development of an integration toolbox that aims to facilitate the preparation, validation and integration of multi-source datasets. Many of these issues could not be identified unless a detailed investigation of data contents and comparison of data themes and features have been done. In this regard a number of key issues have been identified. Table 5.5. summarises the identified issues through technical evaluation of the case study datasets.

Table 5.5. Technical issues identified through Australian case studies

	Identified Issue
specification	Diverse fundamental datasets
	Different naming and classifications of data themes
	Inconsistency in non-spatial content (attribute, name, type)
	Diversity in theme content
	Diversity in detail level of datasets
data model	Different conceptual data models
	Different physical data models
data quality	Completeness
	Coverage
	Currency
	Logical consistency
	Data redundancy
	Scale
	Accuracy (spatial and aspatial)

metadata	Outdated metadata content
	Lack of metadata content suitable for data integration
database	Diverse database designs
	Diverse storage formats
data integration tool	Lack of tools for data validation and integration (database, data model and vertical topology)
vertical integration	Different vertical (model level) link
SRS	Diverse reference system [definition]

The information and analysis that have been gained through the investigation of datasets identified and highlighted a number of major challenges and issues for effective data integration. This could not be achieved unless the actual datasets and accompanying documents were assessed and investigated. Some items that emerged as the most significant issues include a data validation and integration tool, integration data model, metadata content and structure and standardised data specification.

The methodology of the technical investigation of the case studies was developed to cover the potential issues and problems. It showed that the process of data assessment could be a quite time- and resource-consuming task. Therefore, one of the major hindrances to data analysis was the lack of a validation tool that could be configured to evaluate the datasets against a number of criteria.

The integration data model, which can bring different data models together, is another issue that has been highlighted during the case studies. This helps build a consistent data model in which the same features can be grouped and the relation between features can be effectively established.

Metadata content plays a significant role in the data assessment process as it provides invaluable information on different aspects of data including quality, access channel, data restrictions and so on. This information can assist practitioners to evaluate datasets. Metadata content can be better utilised if the content is structured and machine-interpretable. Some forms like XML provide appropriate structure and form based on which information content can be extracted automatically.

Data specifications also contain very detailed information on data origin, conceptual descriptions, data structure and content. This is a rich source of information for the conceptual design of the integrated data model. It also provides significant information at feature level, which can lead to the extraction of appropriate features for data integration.

5.4. Australian Federal Agencies Case Studies: Technical Visits and Investigation

The overarching policy at Australian national level concerning spatial information is the development of the Australian Spatial Data Infrastructure (ASDI). Many of the early advancements in SDI development were pioneered at the national level due to the overarching federal structure of government and the historical nature of surveying and mapping organisations that produced natural environmental data. At a national level along with ANZLIC, Geoscience Australia and PSMA play important roles in the development of the ASDI through the creation, maintenance and use of diverse spatial datasets. The states' major area of involvement was through the creation of digital cadastral data used to create an efficient and effective land market. The creation of the ASDI enables these major initiative areas to be brought together.

The case study investigation has been conducted to cover the above-mentioned agencies. Figure 5.7 below outlines the methodology for visits to the organisation of case study jurisdictions .

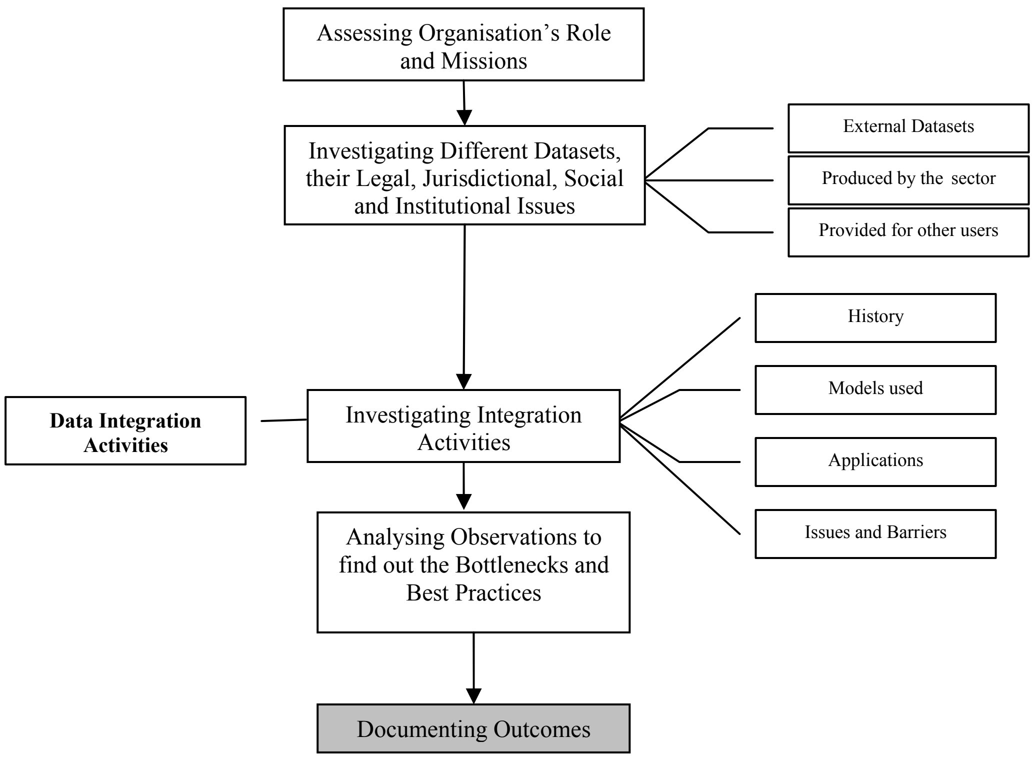


Figure 5.7. Australian case study investigation methodology

The methodology provides a number of steps for the investigation and technical visit conduction. First, the role and mission of the organisation was investigated in relation to the creation and management of spatial data. The management of different datasets was then investigated and data integration activities documented. The observations seen throughout all organisations were then analysed to find the bottlenecks and best practice areas in data integration.

5.4.1. Federal Government Systems in Australia

Australia operates as a Federation of States as an independent member of the British Commonwealth. The Federal Government has powers over defence, foreign affairs, trade and commerce, taxation, customs and excise duties, pensions, immigration and postal services. Other powers including health, education, state transport networks, town planning and land administration (cadastral and land registration) are the responsibility of the state and territory governments (Dalrymple et al., 2003).

Australia's parliamentary system, not unlike other democratic nations, is complex and while considered rigorous, comes with much duplication of powers, effort and expense to the community (Hodgins, 1989). In federated systems particularly, the replication of the parliamentary system across two levels of government equates to immense duplication. Although justified for the most part with separated constitutional powers, roles and responsibilities, much duplication remains. The determinations of matters that are of state, national or of shared interest are central to the confusion and debate (Warnest, 2005).

This style of government system has direct implications for the creation of a national SDI. For many years, spatial information has played a vital role in the land administration process, enabling activities such as taxation, planning and so on to be undertaken to a greater degree of accuracy and at greater speed. This has meant that over time, states have built up their use of spatial information and spatial information tools in many ways and with various degrees of success.

There is also a different culture in the way that the Federal Government and the state and territory governments operate which appears to significantly impact on the decision-making processes of each government. While mapping and spatial information are not considered responsibilities of the states (nor the Commonwealth), the states do require maps and spatial information to manage their jurisdictions sustainably. The Federal Government plays a national leadership role and usually provides significant funding to facilitate activities such as health, higher education and agriculture, but there is no such funding for mapping and spatial information (Williamson, 2004). The investment in mapping and spatial information activities has been an operational business decision that was made on an agency-by-agency or project-by-project basis. This has seen the creation of an SDI for Australia that has tended to be fragmented throughout the states in line with the development of Land Administration systems, which has only recently been filled by ANZLIC at a federal level, Office of Spatial Data Management at a federal level, and Land Administration-based agencies (in general) at a state level. This is an important step in reducing duplication of effort and in producing tools and functions for cross-jurisdictional outcomes that benefit multiple agencies and stakeholders.

5.4.2. Spatial Data Coordination at Australian Federal Agencies

Small-scale spatial information underpins land use planning and management, mining, agriculture and forestry, environmental, infrastructure, defence and emergency and counter-terrorism services activities across the country. This makes national coordination and administration of spatial information a major task for federal and national government agencies such as ANZLIC – the Spatial Information Council (including the Inter-governmental Committee on Surveying and Mapping (ICSM)), the Public Sector Mapping Agency (PSMA) and the Office of Spatial Data Management (OSDM). Figure 5.8 illustrates these agencies.

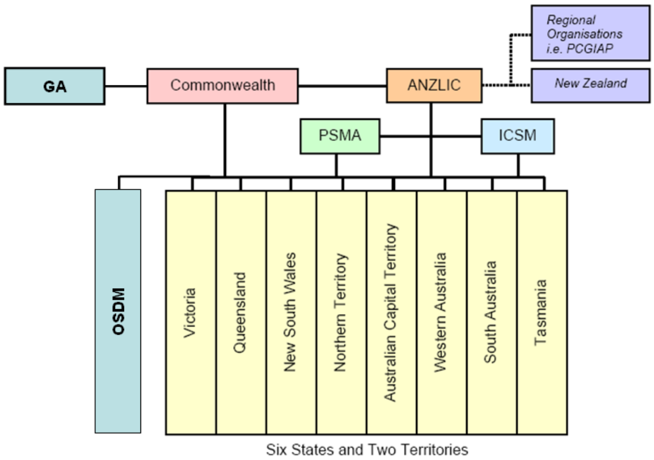


Figure 5.8. Organisation chart for Australian federal and state case study agencies (adopted from Warnest, 2005)

Geoscience Australia

Geoscience Australia (GA) is a federal government organisation responsible for producing maps and spatial datasets at national level. GA's focus is more on natural environmental datasets and it provides nationally consistent data for Australia, utilising datasets from states and also in-house productions. GA is also responsible for national-level hazard management services including the tsunami mitigation service (GA, 2008). GA both produces and uses data which forces it to utilise new approaches and methods to provide more useful and integrable data, especially with recently defined tasks falling to GA in the areas of utility, emergency and hazard management. The National Mapping and Geo-hazards Divisions of Geoscience Australia were merged in 2005 to reflect a global trend in dealing with an increasing range of real world phenomena. Under the new Geospatial and Earth Monitoring Division (GEMD), the maintenance of standard small-scale national topographic mapping remains a major responsibility.

GA plays a critical role in producing first-class geo-scientific information and knowledge. This can enable the government and the community to make informed decisions about the exploration of resources, the management of the environment, the safety of critical infrastructure and the resultant wellbeing of all Australians. GA produces a range of spatial data including Topographic Maps (paper and digital), Reference and Thematic Maps (paper), Topographic (GIS) Data, Thematic (GIS) Data, Digital Elevation Data, Satellite image data, Aerial photography, Airborne Geophysical Products, Geophysical Maps/Images, Digital Geology, Geological Maps, Research Publications, Marine Data, Petroleum Data (Repository), Online National Geoscience Datasets, Geodetic and Geodesy Products, Educational materials.

GA, on behalf of the Commonwealth Government, is interested in datasets required by policy drivers at the commonwealth level. These drivers are emergency management, counter-terrorism, tsunami and so on. In this regard, GA seeks collaboration with each state and territory to complete the jurisdictions' map products.

GA is recognised by the government, industry and the community as a world leader in geoscience and geographic information. GA is not directly involved in the research and development of the ASDI, but as a commonwealth research agency, it is a keen supporter and participant of research projects related to the ASDI. In addition, GA manages the gateway to the Australian Spatial Data Directory (ASDD); provides a range of national fundamental datasets through its National Mapping Division and is the custodian of commonwealth data licensed to PSMA Australia. GA is taking great steps towards utilising a well-developed national SDI and as a consequence utilising integrable and interoperable data.

Office of Spatial Data Management

The Australian Government Office of Spatial Data Management (OSDM) was established to promote spatial information in the jurisdiction of the Australia Government. It is responsible for the development of Australian Federal Government's spatial policies and standards and generally speaking, OSDM is responsible for development of the Australian Government's Spatial Data Integration (AGSDI). This includes implementing the Australian Government Policies such as the Policy on Spatial Data Access and Pricing.

The role of OSDM (OSDM, 2007) is to:

- provide administrative support to the Spatial Data Policy Executive (SDPE) and the Spatial Data Management Group (SDMG)
- implement the work plan and manage the working groups established by SDMG
- facilitate sharing of experience and expertise between Australian Government agencies
- provide technical advice to the SDMG
- promote efficient use of Australian Government spatial data assets
- represent the Australian Government's interests in spatial data coordination and access arrangements with the states and territories
- foster the development of a private sector spatial information industry.

OSDM operates across the whole of the Australian Government. GA provides administrative support including accommodation, personnel services, web hosting, IT and technical support to OSDM. OSDM (OSDM, 2007) has developed three strategic documents including:

- Australian Government Policy on Spatial Data Access and Pricing
- Australian Government Metadata Profile based on ISO 19115
- Australian Government Custodianship Guidelines.

The work of the OSDM also strongly encourages government agencies to adopt international open system standards to aid in the development of tools, functions and data that are integrable and interoperable. OSDM's operations have not been mandated, hence other agencies are not obliged to adopt standards. Consequently this will limit the effectiveness and ability of OSDM's initiatives to integrate cross-jurisdictional data. There is currently a strong political priority and resources allocated to national cross-jurisdictional issues such as emergency management/counter terrorism, animal health and emergency and natural resource management; to be effective, all of these initiatives require access to data and services that are integrable. While this is a step in the right direction, the development of a framework and tools for spatial data integration that serves cross-jurisdictional purposes will aid in facilitating solutions to all of these issues at the one time, as opposed to the creation of single-purpose solutions to each of the national issues.

Public Sector Mapping Agency

Public Sector Mapping Agency (PSMA) is a government-owned company which functions as a "clearinghouse" within the ANZLIC model for the ASDI. Its main purpose is to coordinate, assemble and deliver standard and compliant national datasets for government, industry and community use. Hence the main activity of PSMA is to facilitate the integration of multi-source spatial data from Australian states and territories.

PSMA combines reliable spatial data from Australia's governments with leading-edge technology to create national spatial information datasets. PSMA presents the data in meaningful and useful ways for a wide range of industry, government and community uses that deliver economic, environmental and social benefits to Australia (PSMA, 2006).

PSMA does not produce spatial data in the traditional data collection, analysis and creation method. Its data is gained from each of the state and territory jurisdictions and integrated to form complete Australia-wide datasets. The majority of PSMA's activities in this regard are non-technical activities concerned with establishing partnerships and cooperation arrangements with the various jurisdictions, along with the creation of the legal and policy basis for data integration. As PSMA gathers data from different jurisdictions (each with their own standards, policies and spatial information arrangements), PSMA encounters inconsistencies in the datasets. This highlights the need for a common policy and standard framework to be adopted by Australian governments.

There are five datasets currently licensed by PSMA Australia with several others in various stages of development. The five licensed datasets are discussed below.

- Geocoded National Address File (G-NAF): G-NAF is the authoritative address index for Australia. It contains the State, Suburb, Street, Number and coordinate reference or Geocode for street addresses in Australia. Names are not part of G-NAF, nor does G-NAF contain any personal information. G-NAF started as an empty database and it uses existing and recognised address sources including the state and territory government land records, as well as address data from Australia Post and the Australian Electoral Commission. Through a rigorous process involving textual address comparison, matching and geospatial validation, both national consistency and national coverage are achieved at levels not previously obtainable.
- CadLite: CadLite is a digital representation of all cadastral boundaries excluding easements and road/drainage casements for Australia. CadLite, comprising digital cadastral boundaries and their legal identifiers, is derived from cadastral data custodians of each of the Australian state/territory jurisdictions.
- Points of Interest (POI): PSMA Australia Points of Interest include over 175,000 cultural points with feature code and name attribution. Features include: Accommodation, Community Services Centres, Cultural points, Defence areas, Education & Training, Facilities, Finance facilities, Gaols,

Government buildings, Grounds, Homesteads, Locality/Suburb, Medical facilities, Mines & Quarries, Mountains & Hills, Places of Worship, Post Offices, Relief Feature Names, Transport, Utilities, Waste Disposal sites, and Water bodies.

- Administrative Boundaries: PSMA Australia administrative boundaries dataset is a digital representation of suburb/localities, local councils, state electoral districts, commonwealth electoral districts and state boundaries for Australia. The data also includes Australian Bureau of Statistics (ABS) collector districts, statistical local areas and urban centre localities for the 2001 census taken from state-based data.
- Transport and Topography: The Transport and Topography dataset is underpinned by a road centreline layer of over one million kilometres of roads, together with more than 30 feature types within transport, hydrology and green space themes. The transport component of this dataset encompasses the roads, rail, rail stations and airport infrastructure networks across the entire nation of Australia. The roads layer includes more than 1,000,000 km of named roads. The rail and rail station layers depict the national rail network (including tram lines). The airports layer also includes landing grounds. The topography component of this dataset is made up of two themes – hydrology and green space. Two layers of hydrology are made up of water bodies, major rivers, minor waters and oceans. The two green space layers are urban parks plus national parks and other reserves.

5.4.3. Current Integration Initiatives at Australian Federal Level

The importance of spatial data integration leads federal agencies to implement initiatives to facilitate the integration of multi-source spatial data. This includes initiatives in data sharing, integrated data modelling, integrated data repositories and so on.

Geoscience Australia

GA has started to implement initiatives that will aid the process of multi-source data integration. One of the significant initiatives is the development of a new data management and sharing platform entitled Single Point of Truth (SPOT). SPOT is a virtual repository of the best available datasets and aims to be the most reliable channel to obtain data for Australia. SPOT aims to gather reliable datasets with the best accuracy and detail from other Australian jurisdictions, so that less accurate and detailed datasets can be derived from original datasets. This is called scaleless database, which contains the best available scale data sourced from state, territory or commonwealth data. Smaller scale maps are produced with generalisation of the best available scales.

This approach relies on both spatial and aspatial content of datasets. The attribution remains intact at different scales, and geometry will change during the generalisation process. To generalise the features, some features should be flagged and ascertained for a certain large scale to depict generalised features at smaller scales (as illustrated in Figure 5.9).

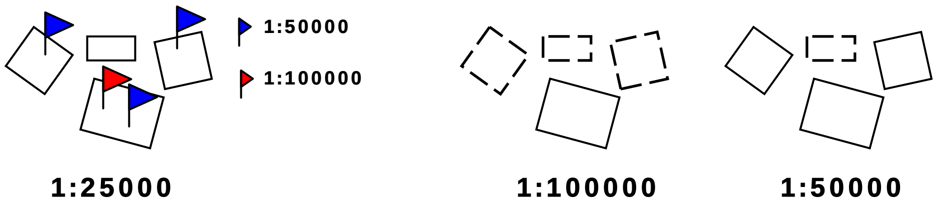


Figure 5.9. Generalisation of features (SPOT's approach)

Data models also will be built based on the best available data. GA is developing data models that are compatible across different scales. GA tends to utilise a common methodology for different scales. Generalisation in small scales should be considered in data models.

PSMA

PSMA has developed sophisticated initiatives such as LYNX (spatial data warehouse) and also has utilised an integrated data model in order to help streamline data access and integration practices. PSMA has implemented a spatial data warehouse called LYNX (Position, 2006) in order to manage, register and distribute spatial datasets. LYNX improves data logistics. It facilitates the data transfer between data suppliers or custodians (government agencies), data managers (outsourced data integrators) and PSMA Australia's data distribution channel of Value Added Resellers (VARs). It is a significant step towards automating the complicated integrated data delivery and also is a concrete step in the development of the Australian Spatial Data Infrastructure (Scott, 2007). Specifically, LYNX delivers:

- an integrated database that combines PSMA Australia's suite of datasets into a single data model and single database
- a data transfer facility comprising a website and associated "back-end" that enables the upload and download of data by suppliers, data managers and VARs, and the monitoring of these data movements by PSMA Australia. The back-end incorporates a series of quality assurance functions to enable online acceptance testing
- a services-orientated environment to support the automated extraction, processing and delivery of data using a variety of web services
- a suite of enhanced and documented business processes covering data supply, data update, data access and distribution, data quality assurance, and customer administration.

PSMA has also utilised a consistent Integrated Data Model (IDM) for its spatial datasets. The IDM is utilised to accommodate spatial features that are collected from Australian states and territories. In the past, PSMA Australia's distinct datasets have been updated at either annual or six-monthly intervals. However, G-NAF precipitated a need to review the PSMA Australia data management environment with respect to integrating the various data models and aligning the update regimes for PSMA Australia datasets. Since the G-NAF maintenance processes rely on many of the other PSMA Australia datasets, there is a high degree of interaction between them. This interaction leads to inefficiencies when common data elements are duplicated across datasets. Integration of the data will remove the duplication and any chance of a mismatch between duplicated information. Furthermore, integration of the data models facilitates a greater degree of cross-dataset analysis enhancing the maintenance process for all datasets. PSMA Australia's datasets currently overlap in some areas, that is, common entities are captured in different ways in different datasets. The purpose of the IDM is to remove duplication between the data models and any chance of a mismatch between duplicated data.

PSMA Australia receives data from multiple sources in multiple formats, from across Australia. The data is then integrated to produce datasets that provide seamless and consistent national coverage on a 90-day update cycle. Data quality forms a vital part of this procedure and 1Spatial's Radius Studio (1Spatial, 2008) will provide an automated solution to ensuring a high level of data quality throughout the integration process.

Radius Studio provides automated conformance checking of spatial data against a pre-defined set of business rules certifying the quality of, first, the source data, and second, the integrated dataset, which will be delivered to the customer. The benefit of this rules-based approach is the assurance of the overall data quality and the automation enables increased efficiency and productivity, which ultimately leads to savings in resources. In the longer term, Radius Studio will also form part of PSMA Australia's LYNX infrastructure. LYNX provides enhanced mechanisms to perform data integration and delivery between PSMA Australia and the various data providers (custodians) and their clients (Vector1media, 2008).

5.4.4. Issues of and Barriers to Data Integration in Federal Case Studies

The following section outlines the issues and barriers that are hindering the ability to integrate multi-source spatial datasets in the federal case studies (GA and PSMA). The issues and barriers are described in relation to experiences and practical applications. In this regard, technical visits were conducted to investigate the integration activities within case study agencies. The investigation of integration models, applications, and collaborative activities has been conducted. The ability to integrate multi-source spatial datasets relies heavily on having necessary technical, as well as institutional policies, legal arrangements, policy mechanisms and practices in place. The case study visit to Australian federal agencies facilitated the investigation of issues and limitations that currently hinder effective spatial data integration within respective agencies. These are listed in the section below.

- **Inter-organisational collaboration**

  In many cases policies are developed by policy-makers and implemented by technicians. As an example the technical feasibility of implementing of single access point to multi-organizational data can not be achieved unless the technical and access policy issues are resolved. Without considering the technical requirements, many of the policies fail to be implemented. In order to overcome this failure, an effective collaboration between policy-makers and technicians should be established.

- **Pricing policy**

  In some cases organisations including GA allow free access to some of their datasets, while states take different approaches including full-cost recovery and maintenance-cost recovery. Integrating different datasets with different pricing policies hinders the integration of respective datasets, as the product of integration need to comply with a combination or either of the policies.

- **Data model**

GA gathers data from different states and converts it to its own data model. The data model used by GA acts as a repository and there is no link between features and it just maintains the logical consistency.

- **Lack of a single channel for data access**

There is no single channel for data access within states that facilitates access to all major state datasets. Therefore, obtaining data needs communication with different agencies, which is a time-consuming and inefficient process. This causes a huge workload of communication and interaction among different data providers at different institutional levels including states, councils and utility companies.

- **Inability to successfully implement national policies**

States are not obliged to adopt and apply national spatial information policies developed by agencies like ICSM and ANZLIC. Therefore, the policies that meet the priorities and major concerns of states are adopted.

- **Metadata content**

Metadata in its present form is not complete, detailed or accurate enough for efficient and effective data integration. Metadata also does not contain detailed information on attributes and topology so datasets can be linked and integrated.

- **Custodianship**

There are a number of datasets including building data which do not have a custodian at any level (especially at a national level) so there is a lack of effective management in order to capture, update and provide access for these datasets. For example, bathymetry is an important dataset that is essential for many applications including tsunami mitigation services, but there is no national custodian to look after this significant dataset.

- **Inconsistent data models**

Each of Australia's states and territories has developed its own data model to meet its own needs. These data models do not comply with other states' data models. This makes the integration of data model problematic.

- **Restrictions on data**

Datasets can be restricted in different ways. Some data custodians may restrict the access, distribution and manipulation of datasets. Tough restrictions are huge barriers for data use and integrations. Most users consider the restriction before utilising the data and restrictions can limit the use and sharing of data.

5.4.5. Summary of Federal Case Studies (Geoscience Australia and PSMA)

GA is the national mapping agency of Australia and focuses on the creation of small-scale natural environmental data including topography, imagery and aerial photos. GA creates seamless nationwide datasets across Australia. In order to implement this task, GA gathers data from different states and maps it to its own data model. The data model used by GA is a silo-based data model and there is no link between features.

To meet users' need, GA is developing a distribution data model and considering links between features in the new data model. This data model will create a customised and theme-based data distribution of data based on users' needs instead of providing all features and datasets. GA is also developing a new paradigm called SPOT. SPOT is a virtual repository (clearing house) that provides users with the best available datasets. SPOT ensures that there is no better data.

GA coordinates some national-level initiatives including utility and emergency management. These initiatives require large-scale data. In this regard, GA collects data from various local government councils throughout the states. The integration of multi-source datasets is a significant task that requires dealing with different technical and non-technical issues. GA manages the data flow and exchange internally and across states.

There are also essential datasets for initiatives including emergency management such as bathymetry and buildings data without national coverage. There is also no custodian for the above-mentioned datasets that is responsible for coordination and maintenance of them.

PSMA is the peak body at Australian national level that is integrating the best available data from different Australian states and territories to create nationwide data. PSMA has no rivals in this area and does not compete with any other organisations. PSMA is also focusing on business data including G-NAF. Business data needs to be accurate and application-oriented, which requires more effort in removing integration barriers. In this regard, legal and institutional issues seem to be more problematic than technical issues.

PSMA communicates with different states to collect required data. PSMA mostly liaises with states, but there are also some datasets that are not under states' mapping agencies' control. In such cases, PSMA needs to communicate with other organisations and governments including local councils. Diversity of communication channels and the process of finding the channels is a time-consuming process. In this regard and in order to facilitate the interaction, PSMA has developed a web-based tool called LYNX. LYNX provides sophisticated tools to perform data integration and delivery between PSMA and different data providers and clients.

Each jurisdiction in Australia has developed its own specifications that hinder proper integration of datasets and data models at a national level. Technically PSMA needs to map jurisdictions' specifications to its model. To produce nationwide datasets, PSMA needs to map different models used by jurisdictions to a single model.

5.5. The State of Victoria: **Technical Visits and Investigation**

Victoria is a state located in south-eastern corner of Australia with 87,884 sq mi (227,620 sq km). It is bounded by the Indian Ocean, Bass Strait, and the Tasman Sea. Melbourne is its capital. Victoria is Australia's second smallest state, though it is the most densely populated state of Australia (Figure 5.10).

Figure 5.10. State of Victoria

Victoria is a major innovator in the administration and coordination of spatial information in Australia. Led by the Department of Sustainability and Environment, Victoria has pioneered initiatives such as the development of a digital online titling system (Victorian Online Titling System) and arguably has the most up-to-date and accurate fundamental datasets of any of the case studies. The development of policies and the establishment of partnerships among spatial stakeholders have been realised by the Victorian Government Spatial Committee (VGSC) and Victorian Spatial Council (VSC). Victoria also boasts the best example of facilitated coordination in Australia between the state and local governments through the development of the Property Information Project – recognised as best practice, which other states are trying to emulate.

5.5.1. Spatial Data Coordination in the State of Victoria

Spatial information affairs in Victoria are conducted mainly by the Department of Sustainability and Environment (DSE) through the Spatial Information Infrastructure (SII) group and Land Victoria, in which the state's digital cadastral map, land registry and title office are embedded. In 2005 the push for wider, whole-of-government and whole-of-industry strategies across the spatial information sector drove the creation of the VGSC and VSC, which act as consultation and coordination mechanisms for spatial information across the state. Additionally a range of government departments and agencies listed below are sharing information together:

- Department of Sustainability and Environment (DSE)
- Department of Primary Industries (DPI)
- Department of Infrastructure
- Department of Justice
- State Revenue Office

- Sustainable Energy Authority
- Parks Victoria
- Local Councils
- Utilities.

VSC has also developed the Victorian Spatial Information Strategy (VSIS). VSIS is the mechanism that the government has used to promote a whole-of-government strategy towards access and use of spatial information. This document is refreshed every three to four years, and is concerned with all aspects of Victoria's spatial information industry, considering the roles and requirements of the public and private sectors and academia in advancing Victoria's social, economic and environmental goals through the provision and application of spatial information (VSIS, 2005).

Major built and natural environmental datasets within Victoria form part of a suite of products called VicMap. This range of spatially related data products is made up of individual datasets and is the underlying foundation to Victoria's primary mapping and geographic information systems. VicMap products are produced and managed within DSE and include:

- VicMap Geodesy
- VicMap Address
- VicMap Property
- VicMap Transport
- VicMap Administrative Boundaries
- VicMap Elevation
- VicMap Hydrology
- VicMap Vegetation
- VicMap Planning
- VicMap Imagery.

Maintenance of these datasets is outsourced to private agencies. These agencies are responsible for amending and maintaining data based on the requirements of councils (local governments), state governmental agencies and private sector stakeholders.

Geodesy, Transport, Elevation, Hydrology, Vegetation and Imagery are maintained directly under supervision of DSE on behalf of the state, while Addresses, Property, Planning and Administrative Boundaries maintenance are driven by local governments and the final result is provided to DSE.

The range of VicMap products is managed through the VSIS with policies setting out maintenance cycles, custodianship, stewardship responsibilities and so on and these are mandated across the datasets. There is a range of other built and natural environmental datasets, however, that falls outside the VicMap suite that must also be considered within the focus of data integration. These datasets (examples listed below) are used by a variety of users, often outside the domain of spatial professionals through areas such as Crown Land Management, Parks and Forests and Land and Catchment Authorities.

- Crown Lands

- Tree Cover
- Flora and Fauna
- Pest Information
- Wetlands
- Bio-sites
- Ecological Vegetation Classes (EVC)
- Infrastructures, etc.

It must also be noted, that data relating to utilities, finance and other information that are often used in conjunction with built and natural environmental datasets is held within agencies outside of DSE. There are generally no authoritative custodians for these datasets.

5.5.2. Current Data Integration Initiatives in the State of Victoria

A number of data integration activities have been initiated in Victoria to respond to the need of local councils, business and industry to utilise integrated products. The development of partnership initiatives such as the Property Information Project (PIP) along with the Victorian Spatial Information Strategy (VSIS), VSC and VGSC have laid the foundation for more efficient and effective multi-source data integration. This is being complemented by several other initiatives such as the seamless database and development of vertical topology.

Data validation tool

A data validation tool in Victoria has been developed to evaluate data against a set of predefined measure and rules. These measures are mostly performed on aspatial content of datasets in the form of queries. The tool is more focused on the data quality assessment for the datasets that have been collected from data providers (Figure 5.11).

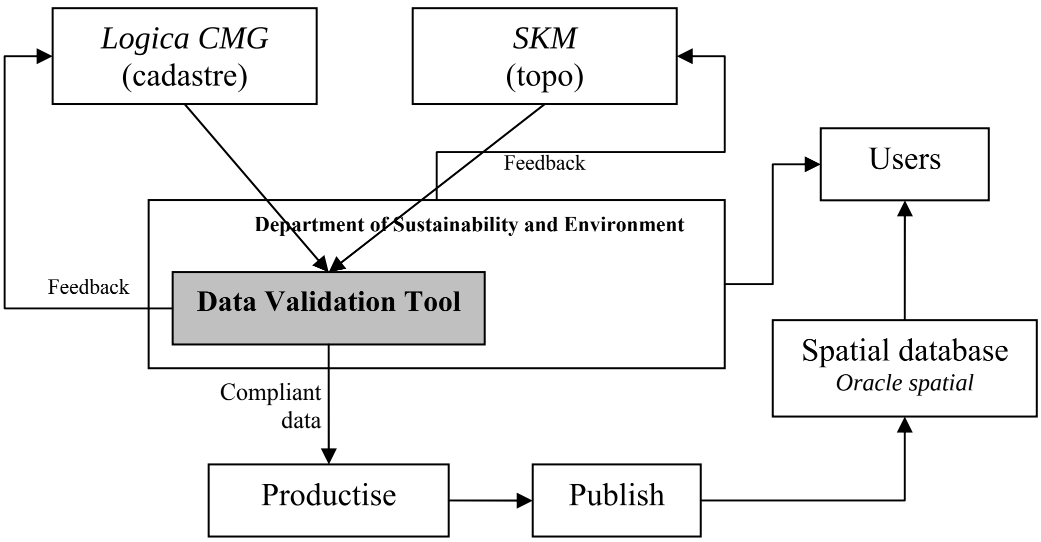


Figure 5.11. Victoria's data validation tool

Some of the characteristics of this tool are as follows:

- Java and ESRI Development Environment
- ability to work with locally stored data
- measures aimed for data quality assessment
- results provided in forms of reports and tables.

VicMap: Integrated Spatial Database

As mentioned before, SII maintains fundamental datasets that form a suite of products called VicMap. These datasets are stored in separate layers without any relation (topology). These datasets are separately maintained. Consequently, VicMap datasets are geometrically integrated, but there are no attributes links and topology between them. At model level, there are links between related features such as localities and roads.

Seamless VicMap Database

SII has been maintaining a tiled-based spatial database called Cooperative Geographic Data Library (CGDL). Within CGDL spatial datasets have been coordinated in separate patches and linked through a master grid. Recently, SII has been migrating from CGDL to a seamless database called SDE. In this approach datasets are seamlessly managed. This also helps effective maintenance and integration of datasets.

Property Information Project

Victoria has initiated a program – Property Information Project (PIP) – between local governments and state. This program has focused on collaboration between state and local governments to develop a framework for maintaining integrated property information including property, address, transport network and administrative boundaries.

Acknowledging that accurate land records underpin the State Valuation of Land Act, the Local Government Act and the recently introduced Road Management Act amongst others, it is essential that the land records used to execute these Acts meet a certifiable standard. SII currently administers the PIP through 78 individual agreements with councils participating in the program.

5.5.3. Issues of and Barriers to Data Integration: Victoria

The following section outlines the issues and barriers that are hindering the ability to integrate multi-source spatial datasets. The issues and barriers are firstly described in relation to technical, institutional, social, legal and policy experiences, with the major barriers and issues summarised at the end of the section.

- **Inconsistent data specifications and terminology**

  Data specifications and terminologies play a significant role in the integration and linking of datasets, as they contain conceptual definitions and specification of data and features. They also form the basis for data generation, database design and implementation. Inconsistency of data specifications leads to inconsistency of data and consequently hinders effective data integration.

- **Institutional issues** include:
    - o organisations are changing without reflecting these changes in capacities, tasks and so on
    - o lack of raising awareness among spatial data/service users of data and its source
    - o many datasets contain rich information but it has been poorly managed
    - o minimal awareness of the importance of integration among senior managers.
- **Lack of awareness of data existence and access**

Spatial experts within the spatial community are mostly aware of the main stakeholders and data access channels. But people from outside the spatial community are not aware of these arrangements, existence of data and data access channels. This causes difficulty in the use of spatial data by a broad range of users. Spatial data clearinghouses and directories can play a significant role in raising awareness of data and its access points. In the case of Australia, access to metadata is possible through Australian Spatial Data Directory (ASDD). However, ASDD does not feed the needs of non-spatial practitioners in locating and accessing built and natural environmental data, as it contains technical and professional information on data.

- **Silo-based management mentality among data providers**

Crown land has its own registry and is managed separately to private land. Adding other registries to this list, including the water registry, there is a huge amount of issues to overcome in integrating different registries to form a single one. Different assets including water, mines, private and public lands are not effectively managed if they are separately registered.

- **Insular mentality**

People basically insist on doing their own tasks and nothing more. This is a hindrance against initiatives that are based on agreement at policy-makers level and consequently causes aversion against data sharing.

- **Lack of local government-level cooperation**

Local councils individually accomplish their own activities and operations and develop their own initiatives, in isolation and with little communication with other local councils. PIP created a good channel for state and councils to talk to each other but councils do not talk to each other (Figure 5.12). Greater emphasis is needed to create channels for local councils to communicate and collaborate on developing spatial information and tools.

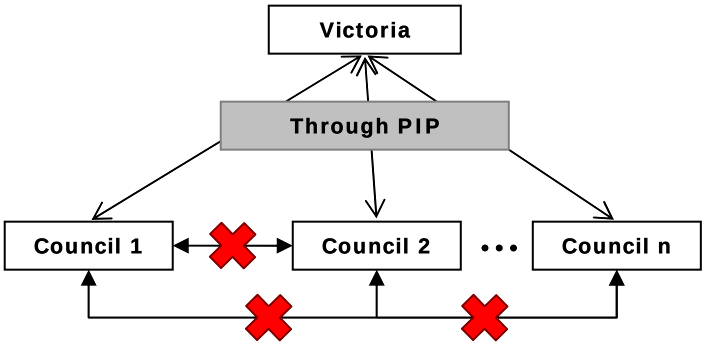


Figure 5.12. Lack of local government communication

- **Project-driven data integration**

Multi-source spatial data integration mostly is initiated within project boundaries and is targeted to meet the need of the project. Few initiatives (e.g. PSMA's datasets) are conducted to integrated datasets for a general purpose; hence this is limited to the scope of the project. It includes the funding, collaboration, technical standards and so on.

- **Custodianship arrangements**

Custodians of spatial data play a very significant role in coordination and maintenance of actual data and its accompanying documents including data specifications and metadata. A robust custodianship arrangement results in a reliable network of agencies that coordinate datasets, metadata and data specification. Currently a number of datasets including habitat data have not been assigned to any custodian. In some cases, there are different agencies interested in a single theme (Figure 5.13). The proposed custodianship program covers both governmental and utility datasets.

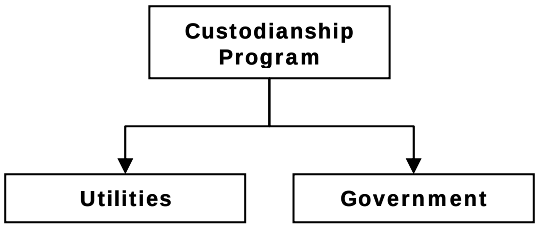


Figure 5.13. Custodianship program in the case of Victoria

5.5.4. Summary of the State of Victoria Case Study

DSE initiates most of the spatial data coordination activities in the State of Victoria including the creation and maintenance of framework datasets, liaison with councils to update and maintain large-scale datasets and the provision of state policies. Having most of the spatial information initiatives under the umbrella of one agency helps the provision of a solid platform for facilitating data integration.

Within DSE, however, there are still some fragmented institutions with interests in land management and environmental interest with the main spatial strategy body being Spatial Information Infrastructure (SII). There are also other departments in

Victoria that create or provide datasets including Department of Primary Industries (DPI), Department of Infrastructure and so on. The fragmentation of institutions leads to a diversity and inconsistency in strategies and policies and less interoperability between different sectors; for example, interests in land are managed by different authorities including crown lands management, Land Victoria, water sector group, forests and parks and so on.

There is no single point of access to all state datasets, while different organisations create and maintain built and natural environmental datasets. Datasets within VicMap have been integrated geometrically; however, there is no vertical topology between datasets. VicMap products are integrated and they are logically consistent, while there is no link to relate different features to others.

Besides DSE there are two central state level bodies, which are VSC and VGSC, who are responsible for the implementation of SDI initiatives and strategy development. Almost all governmental SI stakeholders within Victoria are involved in spatial activities through VSC and VGSC. The VSC has the specific role of driving and supporting the Victorian spatial industry through the initiation and development of spatial information policy, maintaining a focus on the development and use of spatial information, and the establishment of a mechanism for communication and cooperation across all spatial industry sectors. The Council has representation from local, Victorian and Australian Government, industry, academia, and the professional associations. The VSC has initiated activities that can aid in data integration, including a custodianship program.

The VGSC has the role of setting the strategic direction for spatial information policy and decision making. This includes promoting a coordinated and consistent approach to the planning and allocation of resources for the development, management and use of spatial information, the development of a whole of Victorian government registry of spatial information, and the promotion of spatial information best practice.

Victoria has initiated a best practice program to link councils with the state government, called PIP. This program is one of the best practices of local–state collaboration in Australia. The PIP represents a collaborative initiative between the two levels of government that has not been achieved in any other state of Australia. However, it is finely balanced between achieving outstanding success or alternatively potential fragmentation of Council support if the currency and reliability of the data cannot quickly be improved to the satisfaction of its end-users.

The development of a successful PIP will aid in the ability to integrate built and natural environmental data, as it is the key to cooperation between local councils and the state government. Such cooperation is vital, if information is to be kept accurate, up to date and useable.

5.6. The State of New South Wales: Technical Visits and Investigation

New South Wales (NSW) is Australia's oldest state, located in the south-east of the country with 809,444 sq Km area to north of Victoria and south of Queensland (Figure 5.14).

Figure 5.14. State of New South Wales (NSW)

NSW is the pioneer of built and natural environmental data creation within Australia. Full coverage of the state's Digital Cadastral Data Base was completed in 1990, leading to a thriving land and property sector. There are, however, often disadvantages in being a pioneer of LIS (Land Information System). NSW does not currently have any overarching spatial information policy/strategy to effectively lead development. Combined with the decentralisation of the Lands Department, there has been a decline in the development of spatially enabled integration initiatives to more effectively manage NSW built and natural environmental datasets. There has been, however, a strong emphasis on Natural Resource Management not seen in other states.

The highly populated east coast and low-populated west region also create issues for the management of built and natural environmental data and information in the state as they must cater for both areas in a different way to other states such as WA and Queensland due to the fact that the east coast is the most densely populated area in Australia and the western area one of the lowest.

5.6.1. Spatial Data Coordination in New South Wales

Administration and coordination of spatial information and in particular built environmental information in NSW are led by the Board of Surveying and Spatial Information (BOSSI) which is charged with the development and promotion of spatial information initiatives within NSW. The Board has developed a vision for spatial information in NSW being:

> To provide NSW with the skills and resources for economic growth, social and environmental development through the application of Best Practice and Standards in the areas of Surveying and Spatial Information. (BOSSI, 2006)

The Board sits within the Department of Lands, one of several mega-departments within the NSW Government. The department was decentralised in the 1970s with the

Central Mapping Agency being moved to Bathurst, 250 km west of Sydney. This has created somewhat of a division within the land department between the spatial section located in Bathurst (dominated by surveyors) and the more textual components of the department (land register, land titles etc.) located in Sydney. The Department has, however, been a leader in responding to the mounting concerns of terrorism and community safety as the first department in Australia to establish an Emergency Information Coordination Unit (EICU).

The Department of Lands consists of Land and Property Information (titling, valuation, surveying, and other spatial information); Crown Lands administration and management (land leases and licences, reserves and State Parks and land uses from cemeteries to iconic development/business sites to tourist and recreation areas); Native Title and Aboriginal Land Claims; Soil Conservation Service (soil conservation earthworks and consultancy services); Land Boards as well as the previously mentioned Emergency Information Coordination Unit (spatial data needs for counter-terrorism and emergency services planning, research and consequence management).

Other built and natural environmental data is coordinated by a range of authorities including the Department of Lands, Department of Infrastructure, Planning and Natural Resources (DIPNR) and Department of Environment and Conservation. One of the major developments in terms of natural environmental data coordination in NSW has been in developing access mechanisms to natural resource information from across government agencies through the Community Access to Natural Resource Information (CANRI) project.

NSW was the first state to establish complete statewide coverage of the Digital Cadastre DataBase (DCDB) (Warnest, 2005). It is based on a collection of data from various sources ranging from digitised 1:100,000 base maps to survey accurate data. This varying data capture policy was necessary due to the large area of low-density population west of the Dividing Range and the highly populated coastal plains. The result of the data capture means that the cadastre is a graphical best fit of data with accuracy ranging from less than 0.33 m to less than 46.2 m in the less densely populated areas. Data for the DCDB was also gained from various authorities within NSW, including the Albury Council, Hunter Water Board and Sydney Water. This has also contributed to the varying degree of accuracy.

The Digital Topography DataBase (DTDB) is managed and maintained separately to the DCDB and provides topographic map data for a range of services including tourism, areas of interest, for mapping key localities and communities and for emergency services situations. There is currently a project underway to integrate road centre lines between the two datasets, creating a statewide dataset for LGA, for emergency management and helping to implement the state's responsibilities for the Geo-coded National Address File. This process is time consuming with each cadastral update related to aerial photography and the road centre line updated on the topographic database. There are currently no other projects looking at the integration of the two datasets; however, integration is seen as a future area of need for the two datasets.

The DCDB and DTDB were originally in the form of a map tile-based system within the Hewett Packard Genomap environment. This type of system requires an enormous amount of effort to maintain and update, as the updating of data for a road for example requires amending every affected tile or map sheet. The system has undergone a process of being converted from Genomap to a seamless ESRI environment that lends itself to an online incremental update system. This allows users to have a direct feed into the DCDB with incremental updates available as they occur.

5.6.2. New South Wales Spatial Data Infrastructure

As discussed above, NSW's Department of Lands ("Lands") is responsible for spatial data coordination in NSW and takes on the role of building the state's SDI. The development of NSW's SDI is primarily driven by individual initiatives occurring in isolation of each other. However, the success of some of these projects, such as CANRI, in fostering cooperation between organisations and promoting the use of spatial information has became a benchmark for other jurisdictions (CANRI, 2006). "Lands" has decentralised key spatial data organisations across the state. Sydney, Bathurst, Newcastle and Parramatta (through DIPNR), sequentially have been appointed for land registry and titling, data coordination, soil data coordination, and natural resource management, in order to distribute resources across the state. This approach has caused some problems in communication and conveyancing of resources.

At the same time NSW's approach to the production and maintenance of data is an in-house approach and Lands has all the tools and equipment to produce and maintain the data. Lands department has all tools and resources to produce (even sophisticated camera and photogrammetry equipments), manipulate, store, disseminate and print maps. The development of built data is done in cooperation with local councils; however, there is no policy or relationship in place to efficiently deal with councils as a whole and this creates gaps in data coordination in NSW. In order to support the coordination of spatial data infrastructure and establish effective relations among spatial data stakeholders, the Board of Surveying and Spatial Information (BOSSI) has been established.

BOSSI plays a significant role in managing the policies and relations between agencies. In this regard, BOSSI convenes different organisations under a number of groups called clusters. These clusters comprise organisations with similar spatial interests including natural resources, utilities and education. These clusters help spatial stakeholders to achieve understanding and trust amongst stakeholders with same interests. BOSSI (2006) has been developed to overarch all clusters as seen in Figure 5.15.

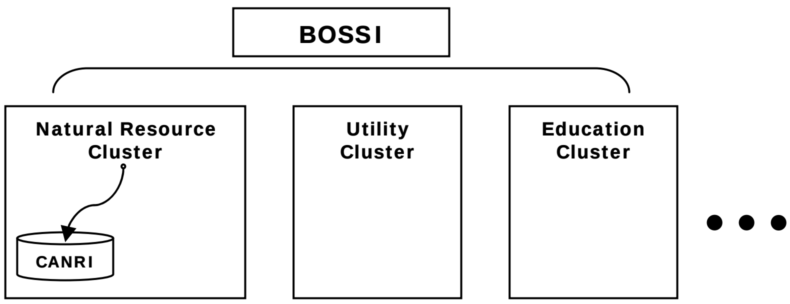


Figure 5.15. BOSSI's cluster structure

In order to provide a dissemination channel, Department of Lands has developed a Geospatial Portal. The geospatial portal is the communities' gateway to spatial data and spatial data services across NSW. The portal is a multi-service interface enabling integration, viewing and searching of spatial data. The portal also provides access to topographic maps and aerial photography. The geospatial portal provides vector data and the system is based on open standards and XML. There are 31 options within the portal to locate features (e.g. plan number, towns, addresses, utility facilities, roads, etc.) with the core cadastre, topographic and imagery datasets used as the base for displaying data (Figure 5.16).

The portal is not as well developed as in some other jurisdictions, but there is a strong emphasis on the development of services within the NSW portal (still under development) which will assist in making spatial information useable to the wider community in a more integrated fashion, once fully developed.

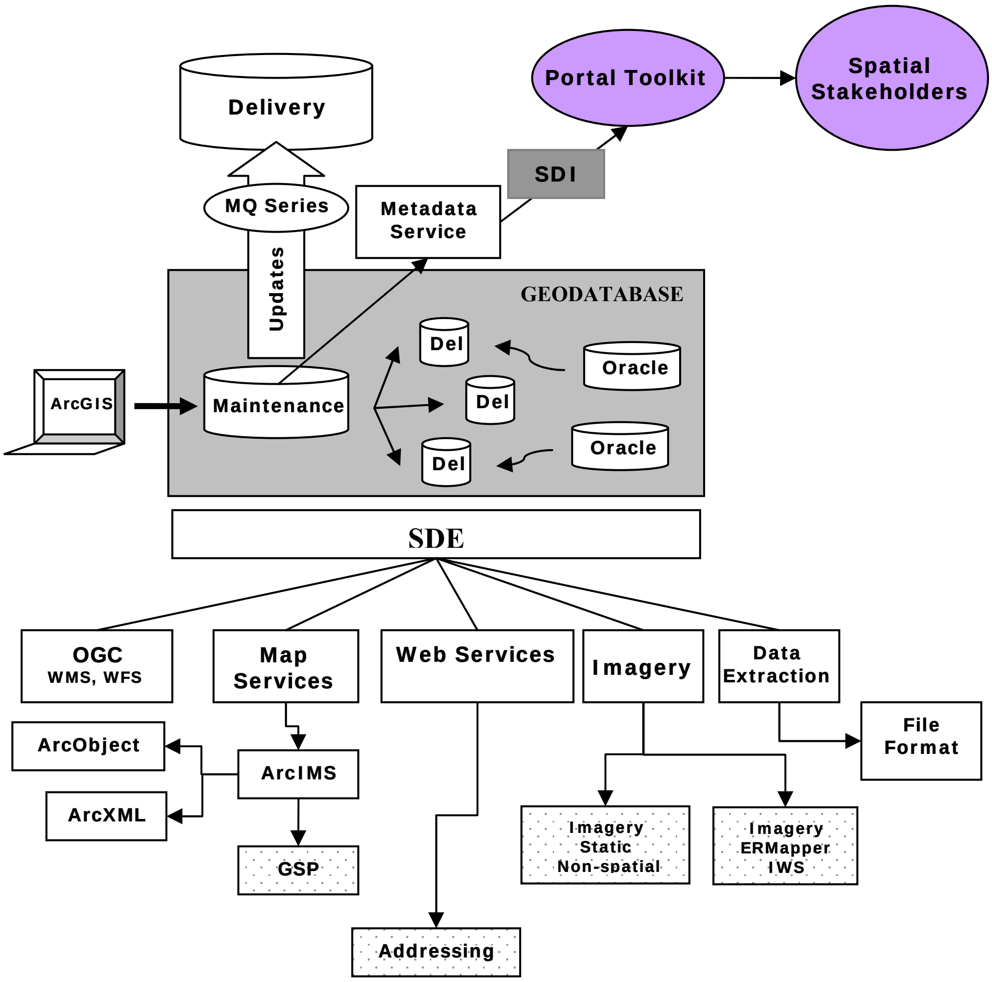


Figure 5.16. Department of Land's geospatial portal architecture

Land eXchange is another service to provide spatial information to users in NSW. The main focus of Land eXchange is imagery and satellite images; however, it also serves vector data. The launch of the new Lands' Spatial Information eXchange website highlights DoL's commitment to the "connect.nsw" direction of NSW Government, endorsed by the Cabinet on 5 August 1998. Key objectives addressed by this site include:

- improved access to government information and services

- alignment of government services to customer needs
- improved choice for the delivery of government services
- integrated delivery of government services
- development of business and the community for the betterment of NSW
- Lands' SDI Metadata.

Department of Lands has developed a new service recently to store and provide metadata to the public. The Lands' SDI Portal represents the first steps in the publishing of a spatial data search and discovery engine for Lands' spatial repositories. Built using the Portal ToolKit from ESRI, initial datasets to be published relate to the imagery datasets and vector data used in the Spatial Information eXchange.

The Natural Resource Atlas is the NSW portal to maps and data for environmental management, planning, research and education. It aims to be a comprehensive catalogue of authoritative, significant natural resource databases and geographic information held by the NSW Government, as well as providing links to significant data holdings in local and federal government and other sites. This site is a part of the natural environmental data CANRI initiative.

There are also some other thematic services including SPADE (http://spade.dlwc.nsw.gov.au/) for soil maps and State of Environment (SoEdirect at http://soedirect.nsw.gov.au/app/index.jsp). There is no link between these systems, however, which inhibits the ability to effectively access integrated data.

5.6.3. Current Integration Initiatives

Recent data integration activities in NSW have focused on large projects that aim to bring together information in order to contribute to cross-agency communities of practice such as emergency management and environmental management. This demonstrates NSW's commitment to creating integrated information and services that can be utilised across government.

Emergency Information Coordination Unit

Emergency Information Coordination Unit (EICU) ensures that Emergency Service Organisations (ESOs) have the best spatial and related spatial data available to deal with multi-agency emergencies. EICU is a new unit set up by the Director General Lands, DoL, as a counter-terrorism initiative. There is, however, a commonality in the data required by ESOs for bushfires, floods, earthquakes, storms, and criminal activities. The EICU aims to implement and maintain a collaborative data-sharing system on behalf of ESOs. The main aim of the unit is to provide seamless delivery and a consistent supply of data to the emergency management organisations through the implementation of a collaborative data-sharing system on behalf of ESOs.

The EICU plans to have single authoritative datasets that can be accessed by the EICU in the case of an emergency. These datasets will initially be required to be co-located on a single database so that integration and interoperability are assured as well as a seamless set of common attributed data. Once all datasets are finalised the original custodian will take on the role of maintaining them. Access to the datasets will be strictly monitored through the development of protocols as some of the information which the EICU will have access to will be of a personal nature, creating privacy and commercial issues for agencies. A demonstrator has been built and the EICU are using this to show various agencies the benefits of cooperating effectively with the EICU.

Community Access to Natural Resource Information

Community Access to Natural Resource Information (CANRI) is a collaborative initiative involving all of NSW natural resource agencies led by the Department of Infrastructure, Planning and Natural Resources. CANRI provides information and products tailored for community-based local and regional environmental management in NSW. It was the first program of its kind to offer integrated access to maps and other data held at various sites by various agencies and stakeholders to the community, enabling the natural resource agenda to be moved forward through engagement of the community. The program is a whole-of-government initiative involving all organisations and agencies with natural resource management information.

CANRI is built on an open technology framework of applications, catalogues, operators and data repositories in a distributed fashion, all connected over the internet enabling maps from various websites to be accessed and operated on the one system as shown in Figure 5.17. These components are managed by various government agencies and other organisations with components connected via published industry-standard interfaces (CANRI, 2006). The framework supports access to a wide framework of remote data servers over the web.

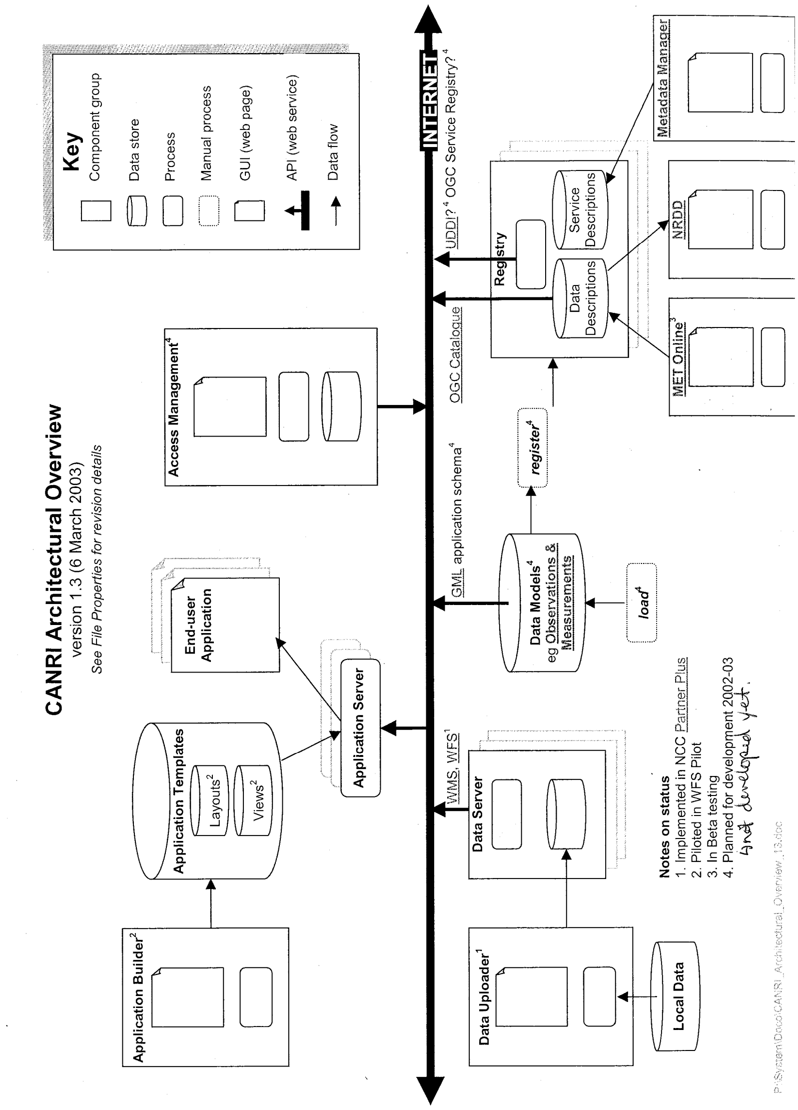


Figure 5.17. CANRI architecture overview (CANRI, 2001)

The CANRI program addresses six themes including coordination, data, systems, quality and standards, products and services, and communications (Warnest, 2005). This includes 150 datasets belonging to 10 custodian agencies. CANRI itself is recognised worldwide as a leading model for SDI development based on a distributed network (ANZLIC, 2004).

CANRI is the only initiative to keep track of different datasets and metadata on datasets in NSW. It is also the only initiative that develops spatial data guidelines and an SDI framework in NSW; however, it is more environmental-oriented.

Single Land Cadastre

The development of a single land cadastre in most states is taken for granted with this fundamental dataset the major domain of government. In NSW, however, due to the state obtaining data from various sources in the creation of the DCDB as explained above, many authorities maintain an updated duplicate cadastral database. The major cadastre maintained outside of the DCDB is that of Sydney Water (approximately 2 million titles in the greater Sydney area are affected). A project in 2001 attempted to create a common land cadastre between Sydney Water and Land and Property Information (LPI) but was abandoned after 15 months. A new project is currently underway between the two, which aims to finish the job started in 2001. A memorandum of understanding has been developed to undertake the project as there are some different aspects incorporated in each of the cadastres. For example, the Sydney Water cadastre has a utilities layer – LPI does not; easements are recorded within the cadastre differently by the two; and there are differences in strata needs – with Sydney water needing billing-metre attachments, which differs from LPI. The Single Land Cadastre (SLC) contains all parcels with titles and boundaries. Cadastral data provided for SLC by LGAs is inconsistent from different perspectives:

- *Content:* local councils capture data contents based on their priorities, which may differ from other local councils
- *Accuracy*
- *Completeness*
- *Currency*

The development of single land cadastre for NSW that creates up-to-date and accurate information is of great importance. Other spatial information and services can be integrated, assessed and shared based on this service.

Geo-coded Urban and Rural Addressing System

In NSW, addresses were being stored in database systems. The new approach is to move existing addresses to a new system and attach them geometrically. This system is called Geo-coded Urban and Rural Addressing System (GURAS) and it aims to serve a broader range of users with integrated address data. To achieve this aim GURAS utilises data from different areas, such as valuation, cadastre, topography and councils' addresses.

ValMap and ValNet – Integrated Land-based Data

GURAS integrates address data from councils, roads data from DTDB (entry point of property), cadastre from DCDB, and valuation information from a text-based system called ValNet. It then provides property layer to another service called ValMap. ValMap is a spatially enabled service to valuating properties. ValMap also supplies data to ValNet. This information contains new valuation data based on integrated information within ValMap. As ValNet has been customised to meet valuing needs, it is used to a greater extent by valuers. ValMap, however, contains spatial information and is used by spatial information experts. With new advancements and developments in spatial data management and new and emerging applications, the merging of these two systems into a single system has been targeted, which facilitates the greater and more effective service to users (Figure 5.18).

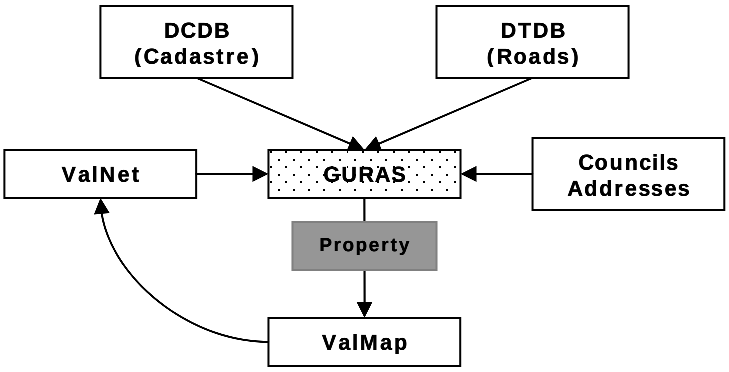


Figure 5.18. Information linkages within GURAS

5.6.4. Issues of and Barriers to Data Integration

As a result of the assessment, the following section outlines the issues and barriers that are hindering the ability to use and integrate multi-source spatial datasets in the State of NSW. The issues and barriers are firstly described in relation to technical, institutional, social, legal and policy experiences, with the major barriers and issues summarised at the end of the section.

- **Policy involved in spatial data sharing**

Because of the importance and sensitivity of spatial data, politicians tend to control spatial data and its activities (production, maintenance, management) and also they want to control and limit its distribution. Thus, the politics involved in data coordination limits effective data coordination including data exchange and access to data.

- **Data supply and users' demands**

Often, user needs are not reflected in spatial data policies and standards. Especially for business-oriented datasets, this issue is more intense. Providers and managers of the data seldom interact and communicate with actual users of data to consider their needs in data coordination.

- **Social barriers**

Addressing social issues are as important as technical and institutional issues for effective data integration. It also influences the development of spatial policies, but has received little attention so far. Historical background, culture and social behaviour are the most prominent issues. The establishment of collaboration and communications plays a significant role in effective data integration and generally SDI development. Therefore, considering the culture and social behaviour of the stakeholders is quite important in order to develop a well-established collaboration within the SDI framework. Some social hindrances are as follows:

- background of organisation including the source of funding and institutional structure, this deprives organizations from keeping pace with collaborative initiatives and organizational changes
- social characteristics of a particular region (for example peoples may be resistent against new/emerging technologies)

- resistant to data sharing
- Conservative officers still exist and attend meetings and forums and express intension to limit the use of data (including pushing full-cost recovery pricing policy).

Generally speaking, effective spatial data integration needs the social preparedness of stakeholders. Social issues affect the success of the integration. Resistance of stakeholders to share data hinders effective data integration. Preparation of integratable data is also possible if data providers accept the importance and necessity of data integration.

5.6.5. Summary of the State of NSW Case Study

In NSW the development of an overall state-level authority responsible for SDI development has been implemented through the development of BOSSI. BOSSI advises government on matters relating to any aspect of spatial information industry and overarches scattered spatial activities in NSW. Most strategies come from the Department of Lands (DoL), which is the highest-level organisation in NSW and coordinates spatial data within the state. Most mapping activities in NSW are also done by DoL. However, there is no sector to specifically look after SDI development.

In the area of natural environmental data, Department of Infrastructure, Planning and Natural Resources (DIPNR) is leading with a superb initiative called Community Access to Natural Resource Information (CANRI). These two organisations are more data provider than data consumer. CANRI has created a single point of access to natural environmental datasets and provided an effective SDI platform in the area of natural data, though it does not cover all state datasets. There is no complete metadata service in NSW; even ASDD does not supply a complete set of metadata for NSW's datasets.

NSW is an old state in terms of land management, with a culture of change being hard to implement. The silo mentality and resistance against data sharing also hinder proper data coordination and integration.

5.7. Chapter Summary

Despite the significance of data integration in decision making, many jurisdictions still have fragmented institutional arrangements and data custodianship. The fragmentation of institutions leads to a heterogeneity of approaches and policies in data coordination and turns the integration of multi-source spatial data to a costly and time-consuming process. Fragmentation of jurisdictions is more explicit in federated state countries including Australia where less interoperability of institutions occurs. Consequently, despite SDI development and sophisticated technological progress in Australia, the lack of a holistic approach to coordinate these activities within a single framework has hampered many of the applications to access, integrate and use spatial data efficiently and easily on a national level in Australia. Each data provider also creates and maintains the datasets in a manner that responds to its own requirements without considering other users' needs, the reuse or application of its data to other areas.

At a national level, there is generally good coordination of initiatives including utility and emergency management by Geoscience Australia (GA) and some effective integration initiatives being implemented by PSMA. However, these are generally on a small scale and fail to take into account the large scale, people-relevant data that is often more important to stakeholders and businesses. At Australia's state level, each state has its own arrangements for spatial data activities. Two examples are the states of Victoria and NSW. The Department of Sustainability and Environment (DSE) in Victoria initiates most of the spatial data coordination activities in the state, while most mapping activities in NSW are done by Department of Lands. Within each state, councils are also major providers and consumers of large-scale spatial data. They liaise closely with states to capture and maintain spatial data for their use; however, they follow their own policies with little consideration of states' policy frameworks.

In the case of Victoria, the Victorian Spatial Council (VSC) has been set up to coordinate spatial information development in Victoria and led the implementation of the Victorian Spatial Information Strategy. The CANRI program provides information products tailored for community-based local and regional environmental management in NSW.

The fragmentation of different-level institutions causes complexity of interaction and communication and inconsistency of policies and coordination approaches among institutions and jurisdictions. These inconsistencies should be considered in institutional spatial policies especially for overarching policies at a national level. A dynamic environment is needed to facilitate the interoperability of not only datasets and services, but also policies, institutional arrangements, legal systems and social behaviours.

From a technical point of view these inconsistencies includes heterogeneity of data models, attributes, metadata, data specifications, data quality, integration tools and so on. Data models do not comply with a single framework and there are even some national-level data models including PSMA's IDM, each state follows its own constructions and specifications. The story is the same for metadata, attribute set and data hierarchy. One state may use 10 categories of road data, while another uses six road categories. Metadata that is utilised in Australia is not suitable for data integration purposes. Integration needs feature-level metadata that includes attributes, and detailed information on data. There is also a lack of effective tools for coordinating different aspects of data integration including data validation, database integration, data model and attribution integration.

Vertical collaboration from local councils through state and federal levels is not well established. Local councils follow their own approaches and there is no effective collaboration model to link councils with state governments. However, the PIP program in Victoria is an exception. Local councils are at different levels of maturity. Maturity in tools, data creating approaches, personnel and strategies are at different levels across councils. As most large-scale dataset sources come from local councils, diversity of maturity levels results in inconsistency of data, policies and coordination approaches.

Business data users are the focus of most data providers. Victoria is moving towards providing business data such as navigation data. PSMA is also developing new products based on business requirements such as ARIA. Business data is used in applications that mostly rely on more than one resource of data; therefore business data should be more integrable. Overcoming integration issues at both technical and non-technical stages is therefore a necessity.

The use of a national SDI as the catalyst for data integration would enable users to reduce duplication of effort and expense in integrating data. For this to occur effectively however, socio-technical issues such as immature institutional arrangements, inconsistencies and incomplete knowledge about the availability and quality of data along with technical issues need to be resolved. The re-engineering of Australia's SDI must take these issues into account if the integration of built and natural environmental data on a national level is to be achieved. In this regard, Chapter Six discusses the components of a spatial data integration toolbox. The major findings and outcomes of this chapter form the components of an effective toolbox that facilitates the integration of multi-source spatial datasets. It includes the following components:

- comprehensive data integration guideline
- structured approach for data validation/integration based on the guidelines including necessary technical tools
- Integrated data model
- consistent and conceptual spatial (geographical and attributes) data specifications
- reliable, machine readable and consistent metadata standard and content
- custodianship arrangements, etc.

Chapter 6

Spatial Data Integration Toolbox: Necessity and Components

6.1. Chapter Aims and Objectives

Effective multi-source spatial data integration requires a number of technical tools together with institutional and policy arrangements within an SDI platform. SDIs as enabling platforms for spatial data sharing could enable users to reduce duplication of effort and expense in integrating data. Spatial data sharing is a major and ultimate goal of SDIs. Within an SDI framework, multi-source spatial data is delivered to users through a number of technical and non-technical components (Rajabifard & Williamson, 2001). The usability of multi-source datasets is highly dependent on their integrability (Rajabifard & Williamson, 2004b). Therefore, the components and tools that facilitate spatial data integration should be addressed and developed within the context of SDIs.

This chapter aims to introduce the data integration toolbox and elaborate on the necessary components of a spatial data integration toolbox within the context of an SDI platform. In this regard, this chapter draws on the findings and outcomes of international and Australian case studies. The international and Australian data integration case studies have identified a number of key components for a spatial data integration toolbox. This includes a comprehensive methodological guideline that addresses the technical and non-technical barriers to multi-source spatial data integration.

Further, data assessment and validation is an issue that requires developments of computer application to automate this process. The automation of the process of data assessments and validation within an SDI context together with necessary extract, transform and load (ETL) capabilities is also a substantial contribution to the spatial data integration toolbox. In order to automate the process of informed integration of spatial datasets, reliable and machine-readable metadata content and data specification are also necessary. The spatial data integration data model is another technical component of the toolbox that plays a significant role in the harmonisation of multi-source spatial datasets.

This chapter also discusses the need for the essential components of the spatial data integration toolbox. The benefit of the development of each component is also discussed. This chapter is followed by Chapter Seven in which a design and development approach has been introduced for the spatial data integration toolbox.

6.2. Introduction

SDIs have been developed to facilitate the access and sharing of multi-source spatial datasets. In this regard, SDI can be considered as a channel through which spatial data stakeholders can share, access and collect data and services. In order to establish an effective data-sharing platform, datasets should be integrable against the measures that have been defined to well meet the purposes of the respective spatial community (Muggenhuber, 2003). Having said that, the spatial data integration toolbox can accommodate necessary components to achieve this aim.

As a result of the international case studies, a number of technical and non-technical requirements of the spatial data integration (including the data validation tool, consistent metadata, data specification and data integration guideline) toolbox have been identified. This was also elaborated and endorsed by the Australian case studies that studied the above-mentioned requirements within the context of a number of Australian jurisdictions. This includes the following key components (Figure 6.1):

- spatial data integration validation tool
- methodological spatial data integration guideline
- consistent, machine-readable metadata
- consistent, machine-readable data specification
- spatial data integration data model.

Figure 6.1. Spatial data integration toolbox components

However, there are also other components including data conversion tools and geometrical integration tools that are required for facilitating spatial data integration (Lanter, 1992). The tools that have been identified through the research project are important for data preparation and readiness assessment for spatial data integration within the SDI context.

The spatial data validation tool is a piece of software that evaluates spatial data against a number of measures including restrictions on data, projection systems, content limits (aspatial restrictions e.g. the attributes cannot accept null quantity). The compliancy of the datasets to the measure represents the fitness of datasets for data sharing and integration purposes. Compliancy to the measures also allows the data to be a part of the data collection component of the SDI platform.

The guideline provides a methodology for multi-source spatial data integration together with necessary tools to overcome potential technical and non-technical barriers. The guideline also details the technical tools together with non-technical mechanisms and approaches that can be utilised to overcome the barriers.

The spatial data integration data model also ensures the consistent integration of multi-source datasets at data-model level. An ontology-based data model provides the basis for integrating multi-source spatial datasets through the identification of logical connections or constraints between datasets and features. This leads to the design of a data model that integrates similar features in a single model through the conceptual definitions and specifications. In this regard, spatial data specification plays a significant role as it contains a conceptual description of features, logical connection between different features and also the constraints that exist between spatial features. A structured data specification that allows automatic information extraction will be a big step towards the effective integration of multi-source heterogeneous spatial datasets.

Metadata also contains invaluable information on different characteristics of datasets, therefore the consistent, rich and machine-readable metadata content assists in the information extraction and automation of data evaluation and integration.

These components form a suite of tools within the SDI platform that facilitates the integration of datasets.

6.3. Spatial Data Integration Toolbox in the SDI context

The spatial data integration toolbox is a suite of technical products and non-technical enablers and mechanisms that aim to facilitate the integration and sharing of the multi-source spatial datasets in the context of SDI (Van Loenen, 2003). Effective spatial data sharing is one of the major aims of SDIs. It ensures use and access of spatial data by a broad range of users (Rajabifard et. al., 2005c). In this regard, SDI aims to facilitate the sharing of spatial data. Spatial data sharing aims to provide usable multi-source datasets to a broader range of users (Figure 6.2). It includes the interaction between useable spatial data and the stakeholders through a number of technical tool and non-technical enablers. In this regard, spatial data integration should be facilitated for utilisation and use of multi-source spatial datasets to their maximum potential.

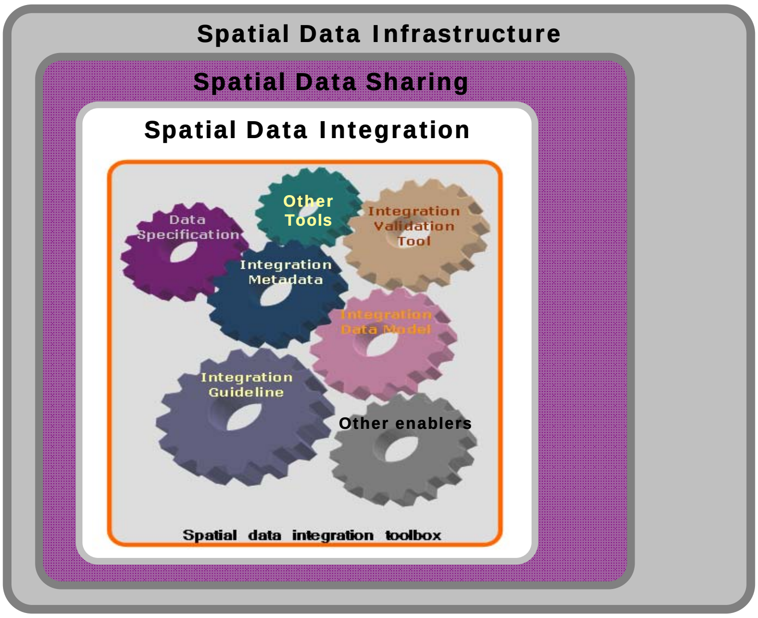


Figure 6.2. Spatial data integration in the context of SDI

An SDI has been developed based on the requirements of a particular jurisdiction. The development of SDIs greatly depends on the characteristics of the respective jurisdiction including the social and legal contexts, institutional arrangements, technical developments and technological infrastructures (Usery et al., 2005).

In this context, the spatial data integration toolbox can provide a gateway between data providers and data users. This gateway can reduce the time, cost and effort of data harmonisation. This also helps SDI developers to deliver consistent datasets to users (Figure 6.3).

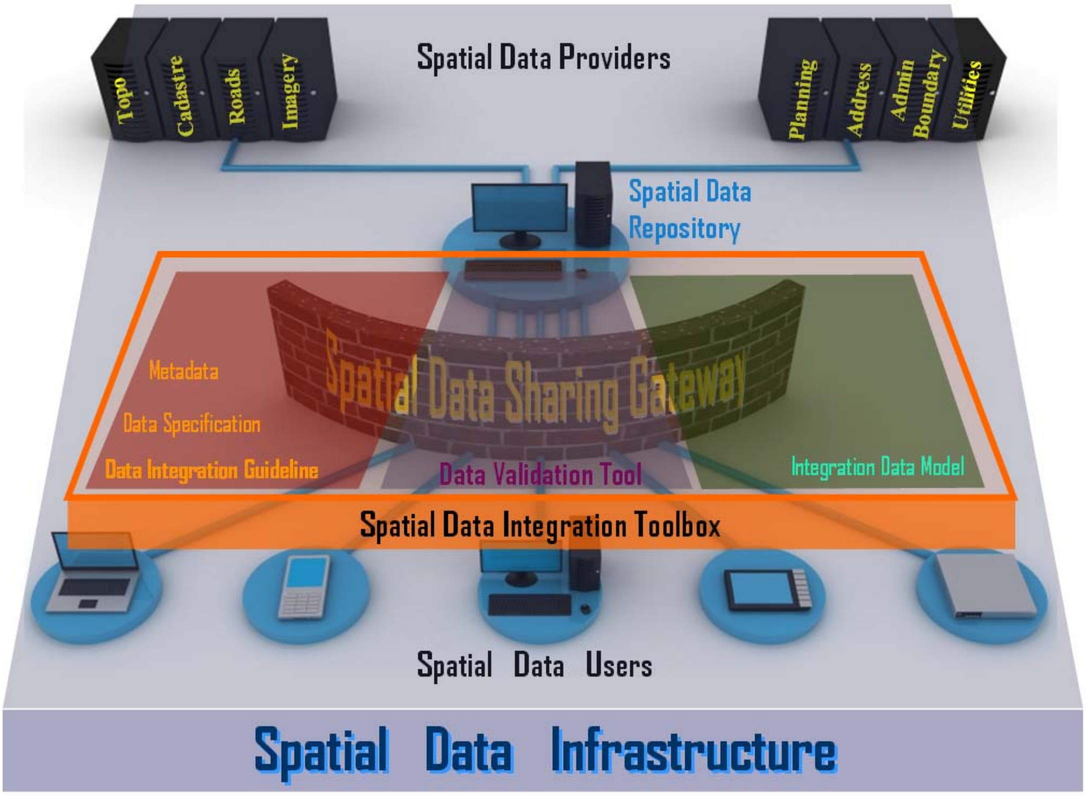


Figure 6.3. Spatial data integration toolbox in the context of SDI initiatives

In this regard, a number of technical tools are required to evaluate the readiness of datasets against a set of jurisdiction-defined measures and rules and also a number of tools to remove the barriers. A guideline can also help practitioners in regards to gaining knowledge of potential barriers and respective solutions. An effective data validation tool utilises the combination of other components of the data integration toolbox. The measures to evaluate the datasets and technical solutions can be derived from the data integration guideline. Consistent and machine-readable metadata provide necessary information to evaluate and integrate datasets and features. The integration data model also can form the data model part of the data validation data model (Figure 6.4).

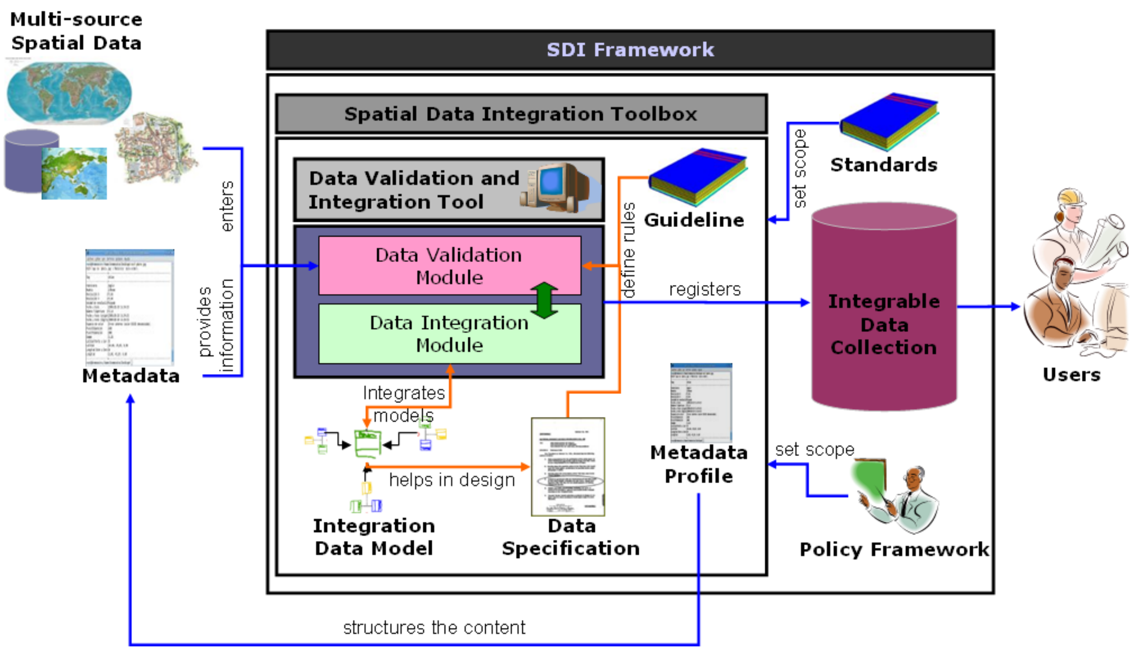


Figure 6.4. Spatial data integration toolbox components and their relation within SDI

Figure 6.4 illustrates the key components of the spatial data integration toolbox including the data integration validation tool and associated guidelines, integration data model, metadata and data specification and relations between them. Within the SDI framework, a number of general relationships can be defined among the components of the spatial data integration toolbox and SDI. This includes the impact of SDI components including policy framework and standards on the toolbox components and the association of the toolbox components on others. These relationships can be summarised as follows:

- Policy framework and standards in the SDI context specify scope of the toolbox components.
- Data enters the data validation and integration tool.
- Metadata provides information on data content.
- Data validation module evaluates datasets against measures and rules.
- Data integration module amends data to meet the rules.
- Guideline and data specification set rules and solutions.
- Metadata profile defines the structure and content of metadata.
- Data specification provides input in designing integration data model.
- integrable datasets are registered in data collection.
- Users access integrable data.

The relations among these components are detailed in next section. In this regard, key components of a spatial data integration toolbox including guidelines, a data validation tool, metadata and data specification together with an integration data model.

6.3.1. Spatial Data Validation and Integration Tools

Spatial data validation and integration tools are essential and integral components of any spatial data-sharing platform. Spatial data validation and integration tools facilitate the delivery and sharing of usable and integrable datasets among spatial data stakeholders.

The Need for the Spatial Data Validation and Integration Tools

Usable data has a number of characteristics that are defined within the context of respective jurisdictional SDI (Backx, 2003). Usable data should comply with the rules and measures that have been defined in the SDI initiatives. These include different technical and non-technical characteristics such as certain format(s), datum/data, metadata content, restrictions on data use, quality (spatial and aspatial accuracies, currency and coverage), pricing policy and so on. In order to examine and assess the compliancy of datasets against these measures, the data validation module evaluates datasets and the data integration module provides necessary functions to amend datasets based on the measures and integrates datasets.

Benefits of the Spatial Data Validation and Integration Tools

The spatial data validation and integration tools provide a number of functionalities in order to facilitate the assessment and amendment of multi-source spatial data for integration. These include:

- assessing spatial and aspatial content of datasets against measures and rules
- identifying the items of incompliancy among datasets
- amending spatial data based on data integration guidelines and integrable data collection rules
- extraction of metadata information content on data characteristics
- providing a structured and standardised approach in data evaluation
- saving time and effort of manual data evaluation.

The Data Validation and Integration Tools Components

There are different measures and rules that define whether or not a dataset is compliant and integrable with other datasets within an SDI initiative. These rules are specified within SDIs' data content specification based on the requirement of jurisdictions. This includes rules on geographical components of data including spatial quality and datum or attribute content including type and value of aspatial content. Non-technical rules including complying with privacy policies and restrictions on data use can also be defined within the data validation tool. This module compares data characteristics with specified rules. Data characteristics are extracted from different sources including actual data that contains different information on the datum, attribute content and scale, and metadata that contains information on source, quality, restrictions and jurisdiction of origin. Users of the system also can define some rules such as restrictions on aspatial content (e.g. NULL value is not accepted for attributes). The items of inconsistency are identified and if there is a technical solution within the tool, data can be amended based on the solutions provided in the guideline; otherwise the report on inconsistent items are provided to the users. The amendment functions are provided through the data integration tool. The data integration module provides necessary functionalities to overcome technical inconsistency with rules. This includes fine-tuning of spatial and aspatial content of data such as datum, attribute values and metadata content.

In summary, the major activities that form the functionalities of data validation and the integration tool are:

- Data validation module
    - spatial and aspatial measures and rules
    - non-technical rules
    - rule configuration
    - data information extraction
    - metadata content extraction
    - comparison with rules
    - incompliancy identification
    - incompliancy reporting
- Data integration module
    - spatial and non-spatial content amendment
    - reporting back to the data validation module
    - data registration (Figure 6.5)

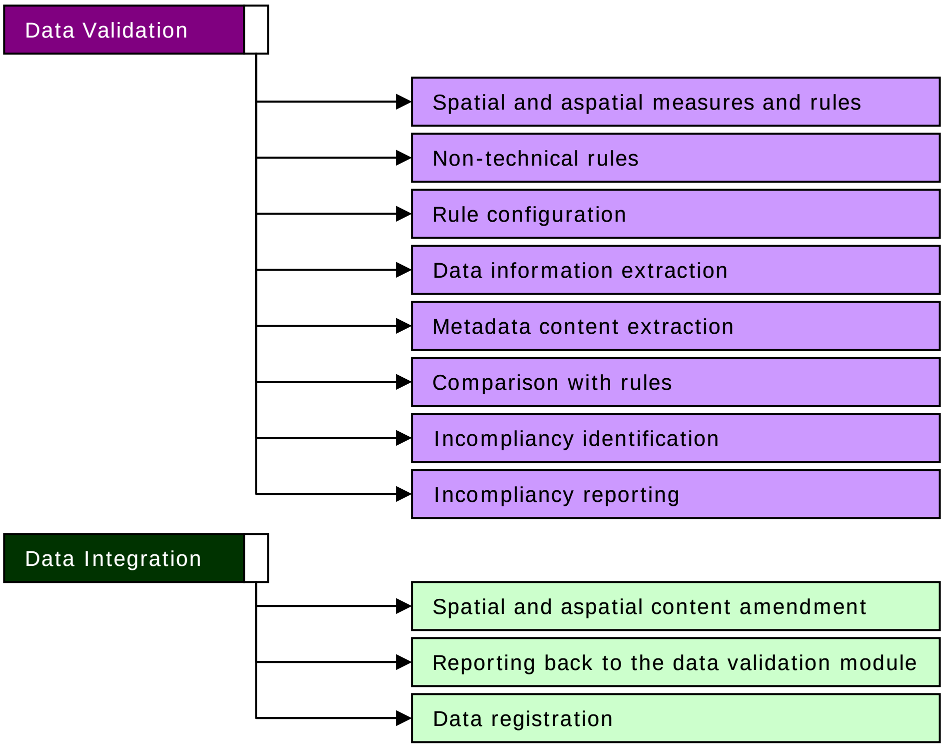


Figure 6.5. The major activities of the data validation and integration tools

The spatial data integration guideline provides the definition and configuration of the rules and measures and also provides the solutions and amendment approaches.

Machine-readable and consistent metadata is a critical component that facilitates the automation of the information extraction.

6.3.2. Spatial Data Integration Guidelines

The spatial data integration guideline is a document that details the spatial data integration steps. It also discusses potential barriers and proposes available technical solutions and non-technical enablers.

The Need for the Spatial Data Integration Guideline

Within SDI initiatives, there are policies and standards that have been developed to meet the data coordination and sharing requirements and objectives of the respective jurisdiction. There is still a lack of specific policies and guidelines for data integration, even within jurisdictions with well-developed SDIs.

Benefits of the Spatial Data Integration Guideline

The spatial data integration guideline details the key considerations for effective spatial data integration. The guideline mainly discusses the potential technical and non-technical barriers; available solutions are also provided in the guideline. An important part of the guideline is the methodological structure for data integration and evaluation. This structure defines the necessary steps to evaluate and integrate the multi-source datasets. The prototype is also designed to meet the stages proposed by this structure.

The Guidelines' Structure and Content

The data integration guideline consists of a number of major components. The key features of the proposed spatial data integration guideline are as follows:

- methodology for data evaluation and integration
- potential technical and non-technical barriers to spatial data integration
- jurisdiction-specific considerations for spatial data integration
- possible and available solutions for data integration barriers.

The methodology for data evaluation provides a structured stepwise approach for data evaluation. This includes the evaluation of actual data, attribution content and accompanying documents such as metadata and data specification. The data integration methodology also proposes a number of steps based on which multi-source spatial datasets can be integrated. The potential barriers to spatial data integration that have been defended through the case study investigations are summarised and accommodated in the methodologies. The guideline also provides possible and available solutions and methods in order to overcome the barriers. This document can be utilised separately for manual data preparation and evaluation.

Potential barriers to spatial data integration have been discussed in Chapters Four and Five. The key methods and solutions that have been addressed in the guideline are as follows:

- partnerships
- custodianship arrangement

- mapping intermediate tools
- metadata structure and content
- data specifications
- standards
- extract, transform and load (ETL) tools
- spatial data interoperability
- generalisation tools
- single point of truth
- data quality considerations
- privacy policies
- pricing policies, etc.

The guideline provides necessary information for practitioners in order to deal with the complexity of data integration. The guideline can be utilised as a part of the tool or as an individual document that helps identify potential barriers and possible enablers.

The structure and content of the guideline is also detailed in Chapter Seven as part of the design and implementation stage.

6.3.3. Spatial Data Integration Data Model

The data model defines the structure of data using spatial objects, relations between objects and attributes. This also defines the database design for spatial services.

The Need for an Integration Data Model

There are different data models utilised to define datasets. The integration of the data model into a single data model creates a consistent object structure and relationships, which facilitates database design and analysis within spatial services.

In this regard, an approach that integrates different data models can greatly facilitate the sharing of usable and integrable data among spatial data stakeholders.

Benefits of the Integration Data Model

The integration data model facilitates the integration of different data models; therefore it allows higher performance and ease in analysis and sharing of different datasets. The benefits of the integration data model can be summarised as follows:

- effective integrated analysis of datasets
- effective integrated data modelling
- integrated database design
- topological design and analysis of integrated features.

Integration Data Model Design Approach

In order to design an integrated data model a number of approaches including ontology-based data modelling have been investigated. Based on the conceptual and semantic differences among datasets, ontology-based data modelling has been selected as an appropriate method to design a data model. Ontology-based data modelling relies heavily on the conceptual definition of features and their content. In this regard, data specifications play a significant role in providing necessary information for data modelling.

6.3.4. Consistent Metadata

Metadata provides information on different technical and non-technical characteristics of spatial datasets. It includes information such as jurisdiction, custodian, data source, quality items, access channel and restrictions. An appropriate content of metadata can facilitate the integration of multi-source spatial datasets.

The Need for Consistent and Automated Metadata

Metadata contains a rich source of information on different characteristics of spatial datasets. It includes information like title, source, jurisdiction, spatial reference system, data quality, restrictions, access channel and so on. This rich and consistent content can greatly facilitate different spatial data use, evaluation, coordination and integration. Effective data integration requires data evaluation. Data integration also necessitates information on different characteristics of datasets.

Conversely, automation of spatial data validation and integration requires measurable and machine-readable content of metadata. Therefore, a suitable metadata for spatial data integration should have a number of key characteristics, including:

- consistent content
- rich and current content
- machine-readable content
- measurable content.

Benefits of Consistent and Automated Metadata

Consistent metadata content provides a homogeneous structure to store and maintain information on spatial data. The rich and current content of metadata that covers different aspects and the latest information of spatial data are also essential characteristics of suitable metadata for spatial data integration.

Machine-readable metadata content also facilitates the automated extraction of information from metadata. There are different forms that meet this objective. For structured, machine-readable and self-descriptive information management, the increasingly popular XML provides an appropriate form to store metadata.

Another important issue is the measurability of the content of metadata. For spatial data validation and integration, metadata content needs to be measurable and provide elements that can be measured and compared with others. However, many metadata items including quality are descriptive and more target the manual use of metadata rather an automated approach.

Metadata Investigation and Recommendations

This section investigates the content of these metadata profiles and studies whether the content of the above-mentioned metadata standards meet the objectives of automated data validation and integration. This section is also accompanied by recommendations on extra items that enrich metadata content for spatial data integration and validation purposes.

6.3.5. Consistent Data Specification

Data specification contains conceptual definitions of datasets and features. Data specifications provide the description of spatial phenomena in real world. Therefore, database design and data models can be derived from these definitions.

The Need for Consistent Data Specification

Data provider agencies maintain data specifications differently. In some cases it is a brief and general description of data and in other cases it may contain a rich content with detailed information on feature descriptions, categories, data models, feature-level metadata and links between features.

This turns a rich data specification to an invaluable source of information on spatial and aspatial content of data. Based on data specification, the conceptual data model can be designed independently from any existing data model. Feature-level information of data description also assists in developing data models in physical and logical levels. The conceptual definition of features can then be reflected in spatial characteristics, topology and attribute content of data.

Benefits of Consistent Data Specification

Consistent and detailed data specification can lead to the effective development of a number of key components for effective data coordination and integration. It includes the development of:

- conceptual data model
- logical data model
- physical data model
- topological relations between features
- attribute content of data
- geometrical characteristics of features.

This chapter aims to present the key components of the data integration toolbox that have been identified through the technical assessments and case study investigations. The components that have been identified include data validation and integration tool and associated guidelines, integration data model, integration metadata and integration data specification documents. The need for these components together with the benefits of their development to the spatial community have been discussed.

6.4. Chapter Summary

Sharing of multi-source heterogeneous spatial data is one of the ultimate aims of spatial communities. In this regard, it is crucial to harmonise and provide integrable spatial datasets to users. Multi-source spatial data integration has been identified as a problematic and time-consuming activity. There are a number of different barriers that hinder effective spatial data integration. These include technical barriers such as lack of effective technical tools for data validation and integration, heterogeneous spatial and aspatial data content and incompliant data models and also a number of non-technical issues including institutional arrangements, restrictions on data use in forms of tough privacy policies and restricted access.

In order to facilitate the integration of multi-source datasets, a number of technical and non-technical tools and guidelines are required. In this regard, this chapter has proposed an integration toolbox with a number of key technical and non-technical components. These components have been identified through a number of international and Australian case studies that have been conducted to investigate the actual spatial datasets and accompanying documents (data specification documents, privacy policies and metadata) and also the institutional arrangements, legal, social and policy issues within the case study jurisdictions.

The outcomes of case studies led to the identification of a number of components including data integration guidelines, integration validation tools, integration data models, metadata and data specifications.

A spatial data integration guideline has been proposed to provide a methodological approach for spatial data integration. It highlights the potential barriers and also possible and available solutions to the barriers. This helps practitioners to gain knowledge of the technical and non-technical problems that may occur in data integration. The spatial data validation tool responds to the need of the spatial community for a tool that can effectively assess and evaluate the readiness of spatial datasets for sharing and integration. The appropriate content and automation of metadata also assist the extraction of necessary information for data assessment. The development of measurable and machine-readable content of metadata is discussed in this regard.

Data specification documents also contain invaluable information on concepts of data features. Appropriate content of data specification documents can assist not only to evaluate datasets, but also facilitate the derivation of ontologies. Ontologies can be used to encapsulate features with similar characteristics in designated classes of data models. The integration data model capitalises on conceptual data specifications to develop an integration of concepts to build a conceptual data model for individual features. In this approach, the development of an ontology-based classification is discussed to extract similar classes of features to build an integrated data model.

Chapter 7

Data Integration Toolbox Design and Development

7.1. Chapter Aims and Objectives

This chapter discusses the design and development of key components of multi-source spatial data integration. The identification of the above-mentioned components has been made through a number of case studies that have been discussed in Chapters Four and Five. The result of the case studies highlighted the need for some technical tools including the validation and integration tool and integration data model together with a number of guidelines and documents including the data integration guideline, integration-oriented metadata and data specification. These components form a suite that has been called the data integration toolbox in the context of this thesis. The role of the data integration toolbox in facilitating spatial data sharing and integration within SDI initiatives together with the components of the data integration toolbox and their benefits have been addressed in Chapter Six.

This chapter aims to present the design and development of the components. It includes data integration guidelines and spatial data validation and integration tools. The design phase of the validation and integration tools has been presented based on common languages in the ICT area including Unified Modelling Language (UML). This communicates the presented concepts and ideas with practitioners and researchers from related disciplines. In the development phase of tool prototype, VBA codes have been programmed in the ESRI's ArcGIS environment based on the classes proposed by the UML model.

The guideline has proposed a methodology for data validation and integration. The guideline highlights the key issues of and barriers to spatial data integration and provides available solutions and enablers for respective barriers.

Integration data model development also has been presented based on the latest research in the area of data conceptual design for effective integration. In this regard, an ontology-based method has been proposed. Metadata and data specification components of the data integration toolbox have been also recommended based on best practice. The best practice can be used as effective approach. In some cases including metadata profiles, some recommendations have also been proposed.

7.2. Introduction

Within the context of an SDI, effective spatial data integration ensures effective sharing of usable spatial data among stakeholders. In this regard, if any dataset that becomes a part of SDI is evaluated and prepared against the integrability and interoperability rules and guidelines (which are defined within respective SDI and based on jurisdiction requirements), it facilitates the use of shared datasets with less time and effort.

In order to achieve this aim a spatial data integration toolbox has been proposed through a number of case studies. The international and Australian case studies endorsed a number of components for the toolbox. The components have been identified though the investigation of spatial datasets, the development of SDIs and the institutional arrangements of case study jurisdictions. The toolbox components comprise the spatial data integration guidelines, the spatial data validation and integration tools, integration data models, metadata and data specifications.

Spatial data integration guidelines propose a methodology for spatial data integration. They cover the potential technical and non-technical barriers to spatial data integration and also available solutions. The guidelines capitalise on the findings and outcomes of international and Australian cases studies to identify the barriers. They also propose solutions and instructions to overcome the barriers.

The spatial data validation and integration tools are a gateway within SDI that facilitates data sharing among spatial data stakeholders. Within SDIs there are a number of rules and measures that define usable and integrable datasets. A tool that automates the validation of datasets against these measures and rules can greatly facilitate the sharing and integration of spatial datasets. The proposed prototype tool has been designed and developed to meet these criteria. In the design phase of prototype, UML diagrams have been utilised to illustrate the architecture and components of the system. UML is a unifying language enabling IT professionals to model computer applications. In this regard the use-case diagram illustrates a unit of functionality provided by the validation and integration tools. The class diagram also has been designed to single out necessary classes for the prototype implementation.

The integration data model is also discussed in this chapter. The proposed approach for integration data modelling is based on ontology concepts. Ontology is a specification mechanism that defines the conceptualisation of entities (Gruber, 1992). The data model capitalises on conceptual definitions of spatial entities. It includes the definition of entity as phenomena in the real world that can be converted to spatial and aspatial definitions and also restrictions and relations to other entities. If the definition is comprehensive and covers every aspect of entities it can promise a holistic conceptual model. This model is independent of any model that is utilised by different stakeholders.

The chapter also discusses the appropriate content of spatial metadata and data specification documents that facilitates the integration of multi-source datasets. It includes the content and structure of them to provide reliable, measurable and machine-readable information on data content and its characteristics. This information can be utilised in the data validation tool to extract information on data and also for the integration data model. In this regard, a number of metadata and data specifications have been investigated and best practice with further recommendations have been discussed.

The chapter discusses the components of the multi-source spatial data integration toolbox with reference to and within the context of SDIs as a data-sharing platform. The chapter contains the deliverables and outcomes of the research that will be summarised in the next chapter.

7.3. Spatial Data Validation and Integration Prototype Tool - Design and Development

The spatial data validation and integration tools are essential and integral components of any spatial data-sharing platform. These tools facilitate the delivery and sharing of usable and integrable datasets among spatial data stakeholders. Any SDI defines a number of criteria for usable spatial data. It includes different technical and non-technical characteristics such as certain format(s), datum/data, metadata content, restrictions on data use, quality (spatial and aspatial accuracies, currency and coverage), pricing policy and so on. In order to examine and assess the datasets' compliancy with these criteria, the proposed tool can evaluate datasets and provide necessary functions to amend them based on the measures; it can then integrate the datasets.

The previous chapter identified a number of functionalities for the tool, including:

- assessing spatial and aspatial content of datasets against measures and rules
- identifying the items of incompliancy among datasets
- amending spatial data based on data integration guidelines and integrable data collection rules
- extraction of metadata information content on data characteristics
- providing a structured and standardised approach in data evaluation
- saving time and effort of manual data evaluation.

The main goal for building the prototype is to demonstrate the strength and effectiveness of data validation and integration tools to facilitate sharing and integration of multi-source spatial data within the context of SDIs. The tool aims to assist practitioners in preparing multi-source data for integration within a consistent framework.

In this regard, the tools comprise two modules: validation and integration. The validation module compares data characteristics (which are extracted from actual data and metadata) with specified rules. Users of the tool also can define some rules on spatial and aspatial content of data (e.g. NULL value is not accepted for attributes). The items of inconsistency are identified and if there is a technical solution within the tool, data can be amended based on the solutions provided in the guidelines; otherwise the report on inconsistency items is provided to users. The amendment functions are provided through the data integration tool. The data integration module provides necessary functionalities to overcome technical inconsistency with rules. This includes fine-tuning of spatial and aspatial content of data such as datum, attribute values and metadata content. Major activities within the above-mentioned tools can be summarised as follows:

- spatial and aspatial measures and rules
- non-technical rules
- rule configuration
- data information extraction
- metadata content extraction
- comparison with rules
- incompliancy recognition
- incompliancy reporting
- spatial and non-spatial content amendment
- reporting back to the data validation module
- data registration.

The tool capitalises on the instructions and mechanisms proposed through data integration guidelines. Spatial data integration guidelines provide the definition and configuration of the rules and measures and also provide the solutions and amendment approaches. Machine-readable and consistent metadata is a critical component that facilitates the automation of the information extraction. The content and structure of suitable metadata for data integration are also discussed in this chapter.

This section discusses the design and implementation of a prototype tool for spatial data validation and integration. The design stage of the prototype has utilised Unified Modeling Language (UML) in order to model the architecture, components and activities within the system. UML provides a unified model that acts independently from the development environment and allows developers to easily interpret the components and interactions between them (Bell, 2003).

7.3.1. Spatial Data Validation and Integration Prototype - Design

The design phase of the prototype utilises UML. UML was put together in order to define a standard notation for specifying, visualising, constructing, and documenting the artefacts of software systems, as well as for business modelling and other non-software systems (Keretho, 1999). UML is a graphical modelling language that is used to express designs (Artiso, 2008). This section utilises UML unified language to design the prototype in the form of a number of diagrams.

Data Integration Prototype Unified Modelling Language Design

As the strategic value of software increases, the industry looks for techniques to automate the software design and implementation and to improve quality and reduce uncertainty, cost and time. These techniques include component technology, visual programming, patterns and frameworks. Businesses also seek techniques to manage the complexity of systems as they increase in scope and scale. In particular, they recognise the need to solve recurring architectural problems, such as physical distribution, concurrency, replication, security, load balancing and fault tolerance. Additionally, the development for the World Wide Web has exacerbated these architectural problems. UML was initially designed to respond to these needs (Braun et al., 2000). In order to present different views of the system UML proposes a number of diagrams. UML diagrams represent three different views of a system model (Figure 7.1):

- Functional requirements view: Emphasises the functional requirements of the system from the user's point of view; and includes *use case diagrams*.
- Static structural view: Emphasises the static structure of the system using objects, attributes, operations and relationships; and includes *class diagrams* and composite structure diagrams.
- Dynamic behaviour view: Emphasises the dynamic behaviour of the system by showing collaborations among objects and changes to the internal states of objects; and includes *activity diagrams*, sequence diagrams and state machine diagrams.

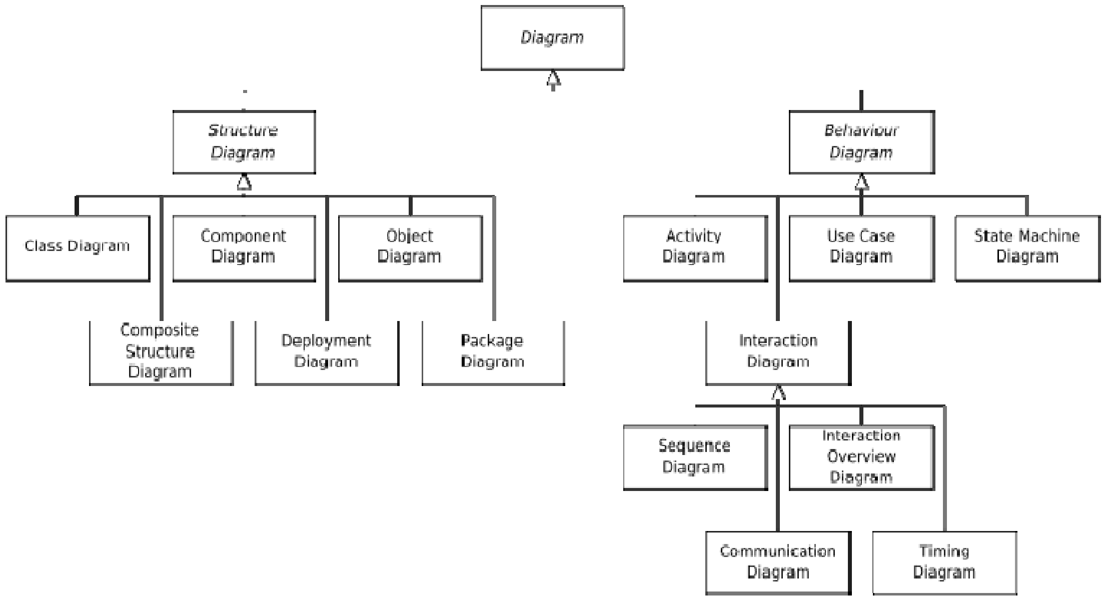


Figure 7.1. Hierarchical categorisation of UML diagrams (adopted from Wikipedia, 2007)

In many design processes, the use case diagram is the first that designers will work with when starting a project. This diagram allows for the specification of high-level user goals that the system must carry out. These goals are not necessarily tasks or actions, but can be a more general required functionality of the system (Artiso, 2008). More formally, a use case is made up of a set of scenarios. Each scenario is a sequence of steps that encompasses an interaction between a user and a system. The use case brings scenarios together that accomplish a specific goal of the user.

A use case can be specified by textually describing the steps required and any alternative actions at each step. For example, a very simple use case for validating spatial datasets might be shown as (the use case will be detailed in next section):

- defines evaluation measure
- evaluates dataset
- amends dataset
- registers dataset.

The use case diagram allows the designer to graphically show these use cases and the actors that use them (Figure 7.2). An actor is a role that a user plays in the system.

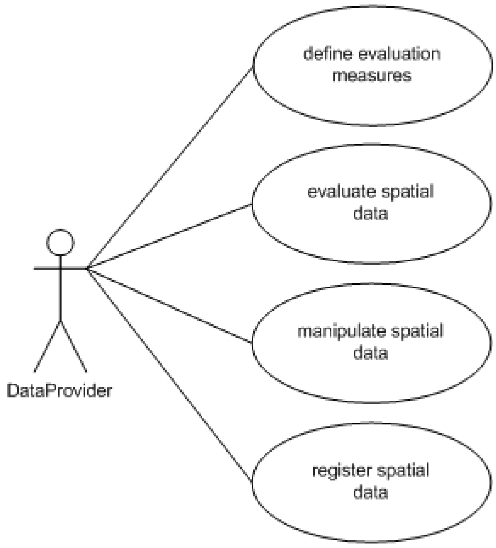


Figure 7.2. Use case diagram for data validation by data provider

A class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes. The purpose of a class diagram is to depict the classes within a model. In an object-oriented application, classes have attributes (member variables), operations (member functions) and relationships with other classes (Martin, 2008). The fundamental element of the class diagram is an icon that represents a class. For example, a class for a data provider can be shown as in Figure 7.3:

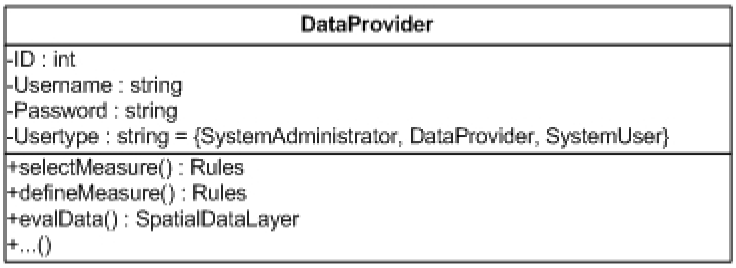


Figure 7.3. Data provider class diagram

Prototype UML Use Cases

The most difficult part of any design is the understanding of tasks and requirements. In most cases the problem statement for system requirements is vague (Stevens & Pooley, 2006). In order to make a solid design of the system and before agreeing on whether to tackle the design, a detailed analysis of the requirements of the system is needed.

In this regard, the tasks and requirements of the spatial data validation and integration tools have been assumed as follows:

> The spatial data validation and integration tool (the prototype) is used by data providers and SDI administrators (both have been categorised as data provider) to evaluate spatial datasets against a number of technical and non-technical rules and measures. These measures and rules are defined within the SDI context which best meet the requirement of the respective spatial community. This includes technical characteristics of data (such as datum,

metadata standard and accuracies etc.) and also a number of non-technical considerations including restrictions on data. In order to provide corresponding data characteristics, data provider supplies actual data and metadata. The prototype extracts information from data and metadata and evaluates the extracted information with the predefined measures and rules. The result of this stage is a report that indicates the evaluation outcomes and includes items of inconsistency. Available technical tools to amend data also can be provided by the prototype. If inconsistency has been removed and data has complied with the measures, it can be registered in the system. Other stakeholders of the system can also access the registered spatial datasets. (Give author, date and page number in brackets)

This scenario is illustrated in Figure 7.4.

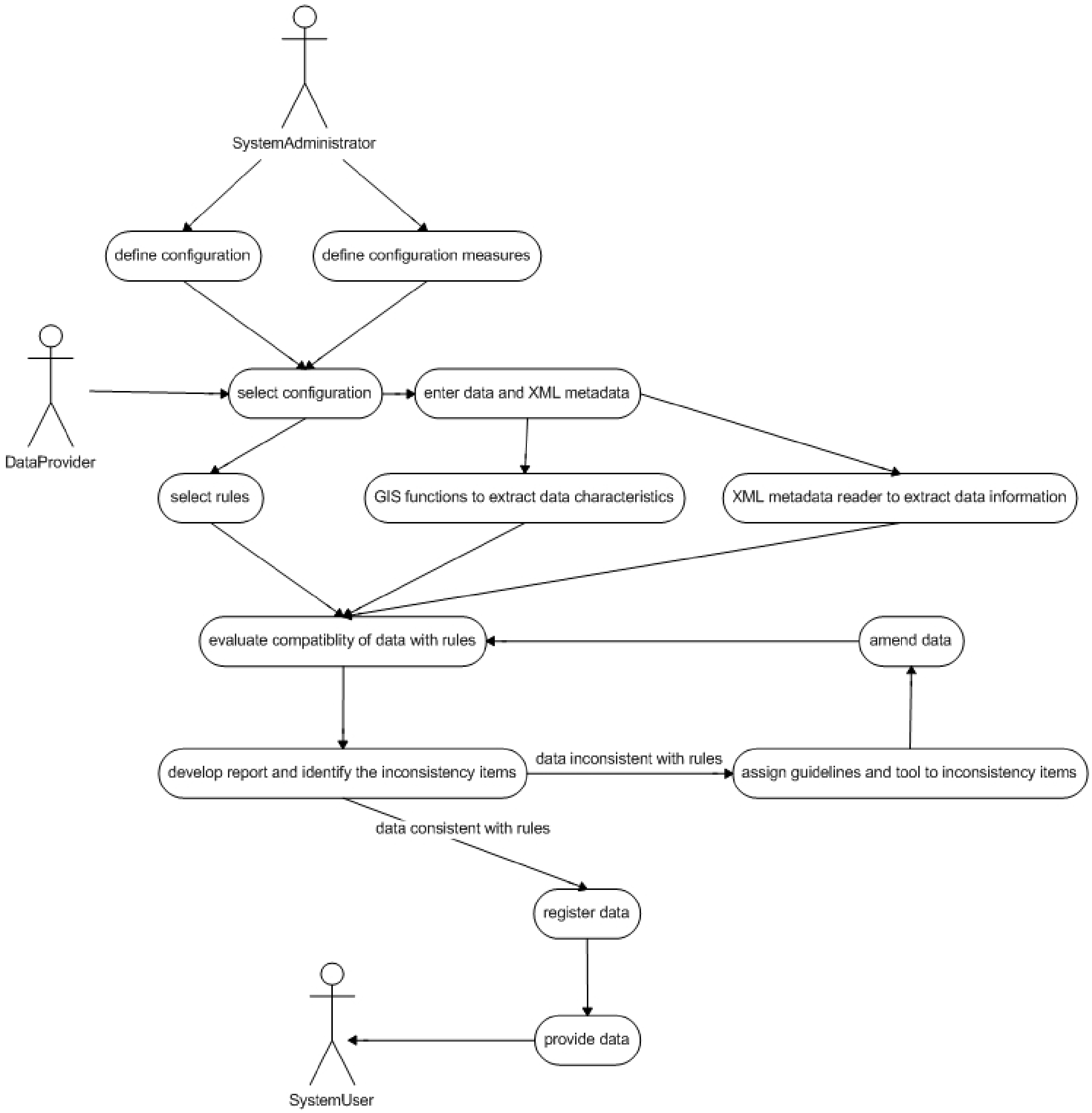


Figure 7.4. Spatial data validation and integration prototype scenario

After some careful investigation, the following facts emerge about the requirements that an ideal prototype tool would satisfy.

Dataset and metadata: the prototype investigates datasets and metadata to obtain information on different characteristics of datasets.

Rules and measures: system administrator defines rules and measures to evaluate datasets. The data providers are also able to define some rules on attributes.

Evaluation of datasets: the prototype should allow data providers to enter data and metadata as information sources and evaluate the extracted information with measures and rules.

Based on the requirement of the prototype, there are a number of associations between actors of the prototype system, and use cases can be identified. Three actors of the system are the system administrator (*SystemAdministrator*), who makes configurations based on the respective SDI and spatial community. For example, in the case of Australia, metadata should comply with the ANZLIC metadata profile, datasets should sit within the boundaries of Australia, certain pricing and privacy policies are applicable to datasets. The administrator also defines evaluation measures. This includes certain data accuracy, datum and restrictions on data and so on. Data providers (*DataProvider*) evaluate spatial data. This also entails the provision of spatial data and metadata. Then, a report is created which presents the result of the evaluation. Based on the evaluation results and in case of inconsistency with measures, available guidelines are assigned to the inconsistent items. This may include technical solutions or non-technical considerations to overcome the inconsistency. The data provider then can manipulate the data to meet the requirements of the prototype and if no consistency is recognised, data can be registered as compliant data. Another actor in the system is general system user (*SystemUser*), who can get integrable data from the data registry. The above-mentioned use cases are summarised as follows:

- SystemAdministrator makes configurations
- SystemAdministrator defines evaluation measures
- DataProvider evaluates spatial data
- DataProvider manipulates spatial data
- DataProvider registers spatial data
- SystemUser gets spatial data

The use case diagram, which shows these associations, has been illustrated in Figure 7.5.

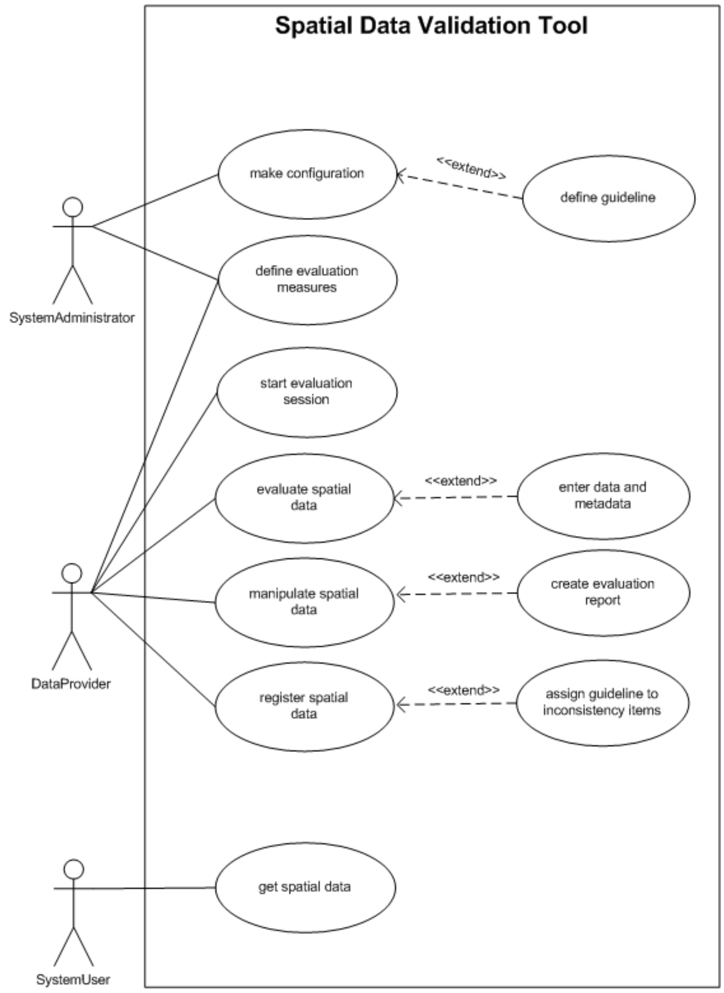


Figure 7.5. Prototype use case diagram

The use case diagram allows for the specification of high-level user goals that the prototype must carry out. These goals are not necessarily tasks or actions, but can be a more general required functionality of the prototype.

It also helps the identification of required objects and relationships between them in a class diagram. The class diagram describes the types of objects in the system and the static relationships between the objects. The next section discusses the objects and relationships of the prototype.

Prototype Classes: Class Diagram

Many users and use cases share some common ideas or concepts. These form the basis for classes that the prototype will need. It is also possible to clarify how these classes of objects are related. The links between them define what interactions are possible between different classes of objects (CSCI, 2007). In order to identify the classes and relationships between them, the entities that interact within the prototype have been singled out. They include the following classes:

- User with three subclasses of SystemAdministrator, SystemUser and DataProvider
- Measures: defined by SystemAdministrator
- Rules: defined by DataProvider

- DataRegSession: data registration session started by DataProvider
- SpatialDataLayer: spatial data layer that is evaluated
- Metadata: metadata for spatial data layer
- MetadataStandard: metadata standard that defines the structure and content of metadata
- DataSpec: data specifications
- Report: to present the results of the evaluation
- Inconsistency: items of inconsistency between data characteristics with measures and rules
- Guidelines: assigned available guidelines for items of inconsistency
- ManipulationTool: assigned applicable manipulation tool for technical inconsistency.

Figure 7.6 illustrates the classes and relationship between them.

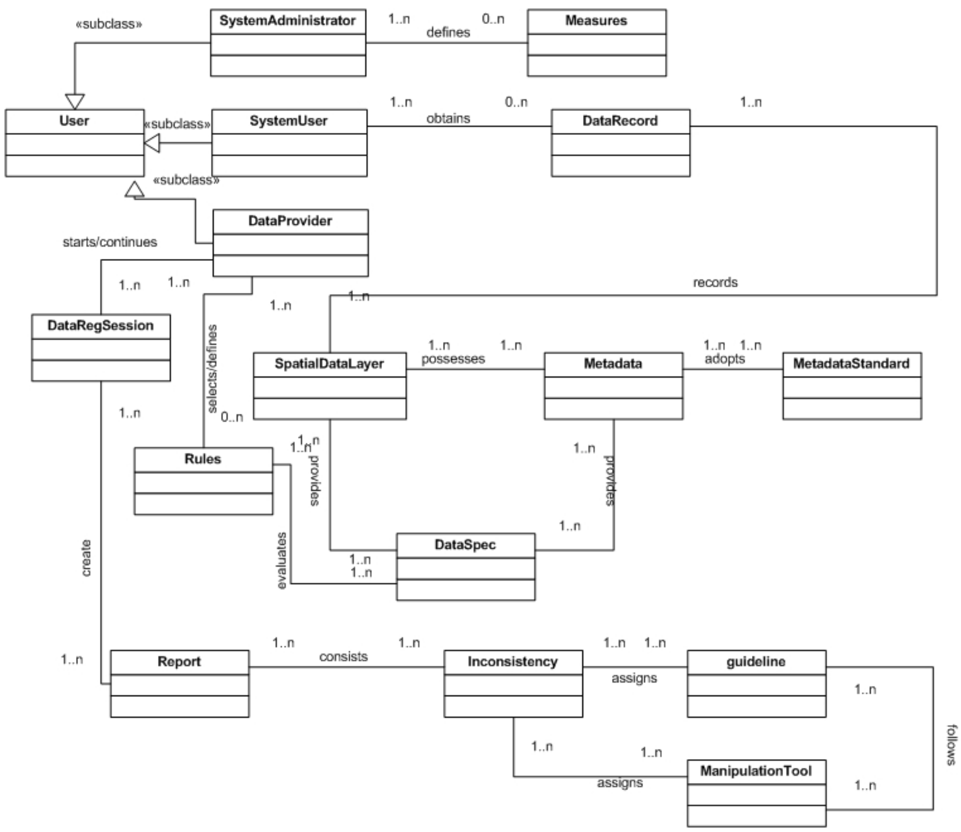


Figure 7.6. Integration validation and integration prototype class diagram

Class diagram shows implementation classes, which are the entities that programmers typically deal with (Bell, 2003). Therefore, the above-mentioned classes and relationships have been implemented within an object-oriented programming environment.

Object-oriented programming is a structure or design for computer programming languages. In an object-oriented programming language, you work with objects that have properties and behaviours/methods (Burke, 2003). In object-oriented programming, objects that share common behaviours can be grouped into distinct classes. Objects that work together to perform a task, communicate with each other by sending messages through methods (Arnow et al., 2003).

7.3.2. Spatial Data Validation and Integration Prototype - Development

ESRI ArcGIS is a general and powerful GIS environment that provides many GIS, data analysis and manipulation functionalities. ArcGIS is a suite of spatial tools for building complete GIS. ArcGIS provides an integrated GIS, combining object-oriented and traditional file-based data models with a set of tools to create and work with geographic data. ArcGIS supports VBA (MIT, 2006).

Capitalising on the strength of ArcGIS and VBA programming for ArcGIS, the prototype has been developed. ArcGIS provided the GIS engine and base functionality, while the customization and combination of the functions have been achieved through programming.The major components include:

- conversion of conceptual model classes to physical classes
- implementation of messaging between classes
- implementation of complex objects and tasks including metadata content extraction
- keeping the record of class properties (instances) in a database.

The major task in the implementation phase had been the conversion of classes that had been designed as conceptual models to physical classes in the VBA environment. This included the conversion of UML modelling language to VBA programming syntaxes.

The implementation of messaging and interaction between different components of the system had been the next significant task. In this regard, class methods and new procedures were required to carry out different tasks of the system. Some of the important tasks of the prototype that have been implemented are as follows:

- passing configurations and parameters between interfaces
- data and metadata content investigation
- assessment of measures with data characteristics through queries and comparisons
- establishment of an effective link with the database to keep a record of objects and also to retrieve existing records, etc.

The following sections detail the development process and provide detailed discussions on the objects, interfaces and database components of the system together with interaction between system components.

Prototype Implementation Environments: VBA Programming within ArcGIS

Visual Basic for Applications (VBA) is a simplified version of Visual Basic and is one of many object-oriented programming languages. The main difference between VBA and other object-oriented programming languages is that VBA has been designed to be embedded within different applications including ArcGIS (Burke, 2003).

VBA is an event-driven implementation of Microsoft's Visual Basic, a procedural programming language, and associated integrated development environment (IDE). By embedding the VBA IDE into applications, developers can build custom solutions using Microsoft Visual Basic. It has also been built into different applications including ArcGIS. It can be used to control almost all aspects of the host application, including manipulating user interface features, such as menus and toolbars, and working with custom user forms or dialog boxes. As its name suggests, VBA is closely related to Visual Basic, but can normally only run codes within a host application rather than as a stand-alone application (Wikipedia, 2000).

ArcObjects are the building blocks of ArcGIS. With ArcObjects, customised menus, tools, workflows, applications, and custom feature classes can be created for use with ArcGIS (Burke, 2003). ArcObjects are a set of computer objects specifically designed for programming ArcGIS. ArcObjects can be utilised within VBA to customise ArcGIS or develop new applications for it.

Prototype Database Design

Because of the system user diversity and also in order to retrieve the existing information on system classes, a database is also required. The database is used to keep a record of different components of the prototype. It includes the record of Users, DataRegSession, Measures, Guidelines, ManipulationTools and RegisteredDataLayers. In this regard a relational database has been designed in Microsoft Access that keeps the record of mentioned classes with their attributes and communicates the content of the database with VBA through Data Access Object (DAO), when necessary. DAO is a data access mechanism that provides an interface to the relational DataBase Management System (DBMS) functionality of Access (Brydon, 1997).

Figure 7.7 shows a part of a database that has been designed for this purpose.

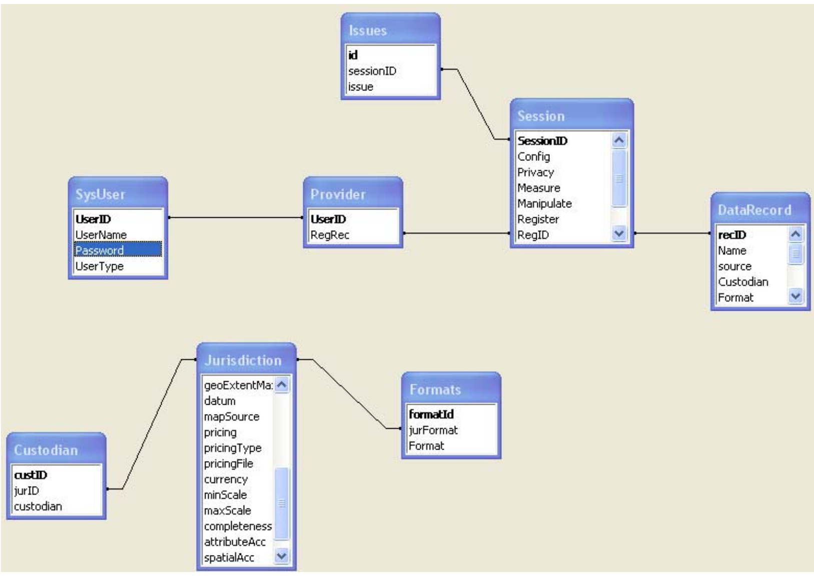


Figure 7.7. A part of a database designed for prototype

Every time a user tries to utilise the system, the access level and available functions for that particular access level are provided based on the user's records. Data providers also validate and register data through data registration sessions

(DataRegSession class). This helps the data provider not only start new sessions, but also resume previous sessions. Configurations and corresponding measures are also important components of the prototype, which should be kept as soon as a system administrator makes new configurations and defines measures. The database also stores a number of other critical and significant sets of information on registered datasets, inconsistency items and guidelines.

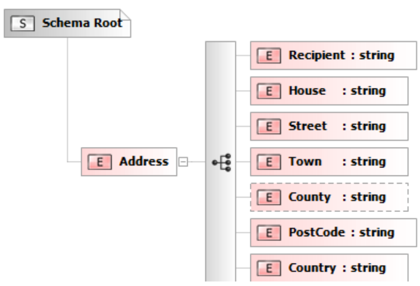
Some Remarks on Metadata Content Extraction

The content of metadata plays a very significant role in the validation process as it provides information on data characteristics. In this regard, the automation of parsing metadata content has been another challenging issue in the implementation of the prototype. The best approach to extract the content of metadata is to parse eXtensible Markup Language (XML) metadata files. XML files provide a structured, self-descriptive and machine-readable form to store information (Figure 7.8).

```
<?xml version="1.0" encoding="utf-8"?>
<Address xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="SimpleAddress.xsd">
  <Recipient>Mr. Walter C. Brown</Recipient>
  <House>49</House>
  <Street>Featherstone Street</Street>
  <Town>LONDON</Town>
  <PostCode>EC1Y 8SY</PostCode>
  <Country>UK</Country>
</Address>
```

Figure 7.8. A sample XML document

XML is a markup language for documents containing structured information. Structured information contains both content and some indication of what role that content plays (Walsh, 2008). In order to indicate the content of metadata, it uses element names that remain constant in different XML files and indicates the respective element. XML schema can be used to express schema as a set of rules to which an XML document must conform in order to be considered valid according to that schema. XML schema was also designed with the intent that determination of XML documents' validity would produce a collection of information adhering to specific data types (Figure 7.9). Such a structure would be useful in the development of XML processors (Wikipedia, 2003).




```
<?xml version="1.0" encoding="utf-8" ?>
<xs:schema elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="Address">
    <xs:complexType>
      <xs:sequence>
```

```
        <xs:element name="Recipient" type="xs:string" />
        <xs:element name="House" type="xs:string" />
        <xs:element name="Street" type="xs:string" />
        <xs:element name="Town" type="xs:string" />
        <xs:element minOccurs="0" name="County" type="xs:string"
/>
        <xs:element name="PostCode" type="xs:string" />
        <xs:element name="Country">
          <xs:simpleType>
            <xs:restriction base="xs:string">
              <xs:enumeration value="FR" />
              <xs:enumeration value="DE" />
              <xs:enumeration value="ES" />
              <xs:enumeration value="UK" />
              <xs:enumeration value="US" />
            </xs:restriction>
          </xs:simpleType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Figure 7.9. A sample schema for the XML document of Figure 7.8

Therefore, an XML processor can parse metadata into its components and extract the content based on the XML schema. Figure 7.10 illustrates the structure of ANZLIC metadata in an XML document. In the metadata XML file shown in Figure 7.10, for example, “<title>” is used to accommodate the title of data.

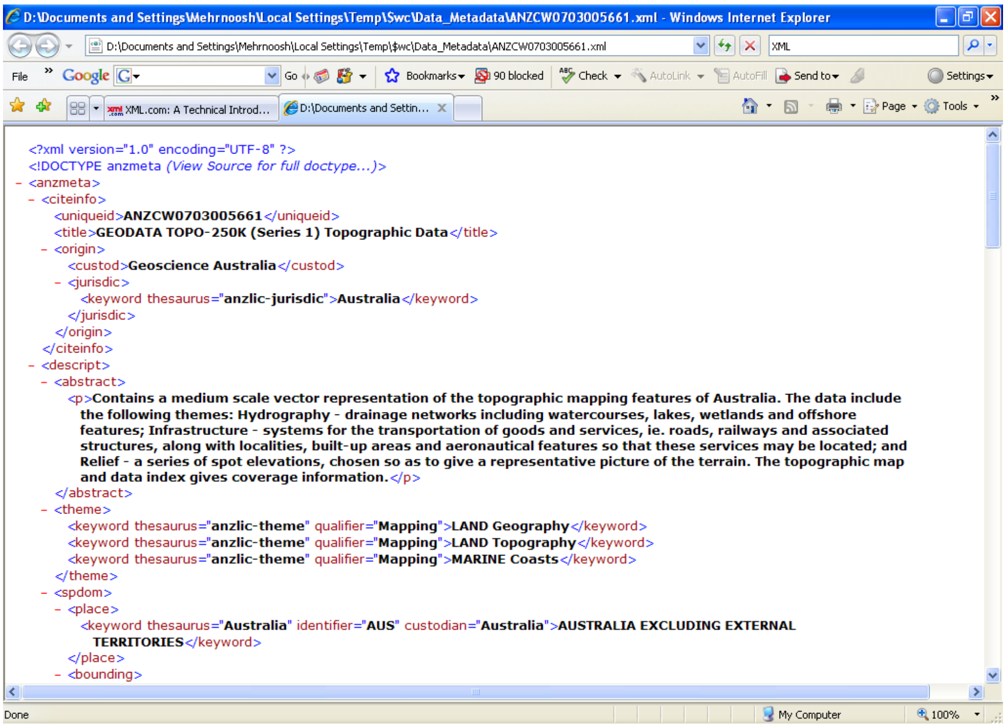


Figure 7.10. A sample of ANZLIC XML metadata

Many XML processor/parser applications have been developed to easily extract the content of XML files. The XML processor is used to read XML documents and provide access to their content and structure (Bray et al., 1998). For the prototype development, an XML processor that has been developed by Microsoft has been utilised. The XML processor is called MSXML (Microsoft, 2007) and provides capabilities to read XML metadata and schema and extract XML metadata content. ANZLIC XML metadata documents are interpreted based on Document Type Definitions (DTDs). DTD defines which elements appear in a document, which attributes can be assigned to an element, and which elements appear inside other elements. The successor to DTDs, XML schema, will allow designers to specify the acceptable data types for elements and attributes.

User Interfaces and their Components

Alongside the design and development of the objects and messages, user interfaces have also been developed. User interfaces provide communication channels with users of the system, whenever there is a need to get some input from users (such as data layers and metadata locations) or whenever a result should be communicated to users (such as a report on the result of validation).

In this regard, a number of interfaces have been developed to interact with different types of users. The prototype adds a new toolbar to ArcGIS. The toolbar provides five tools that help users to evaluate and manipulate datasets. The toolbar comprises the integration validation tool, coordinate converter, interoperability tool and attribute (table) manipulation tool (Figure 7.11).

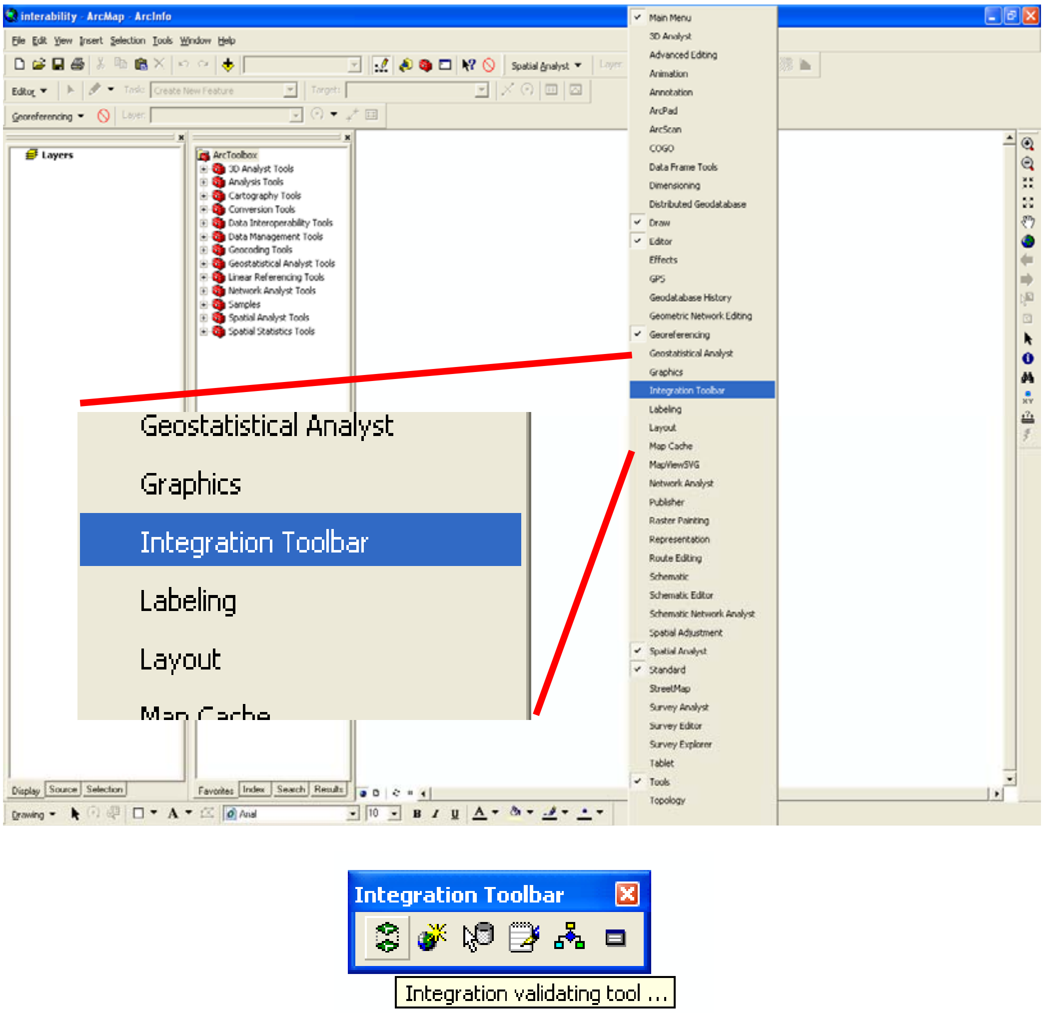


Figure 7.11. Integration toolbar within ArcGIS

The integration validation tool contains the core evaluation functionalities, while other tools help to manipulate data and their attributes.

There is a first general user interface that is used to determine the access level of the user including system administrator, data provider and system user. This interface (UserAccessInterface) obtains the username and password from the user. Based on the existing users' information in the database, the tool allows the user to follow the next steps. Other interfaces are developed to facilitate the use and interaction with the system for three above-mentioned user categories (Figure 7.12).

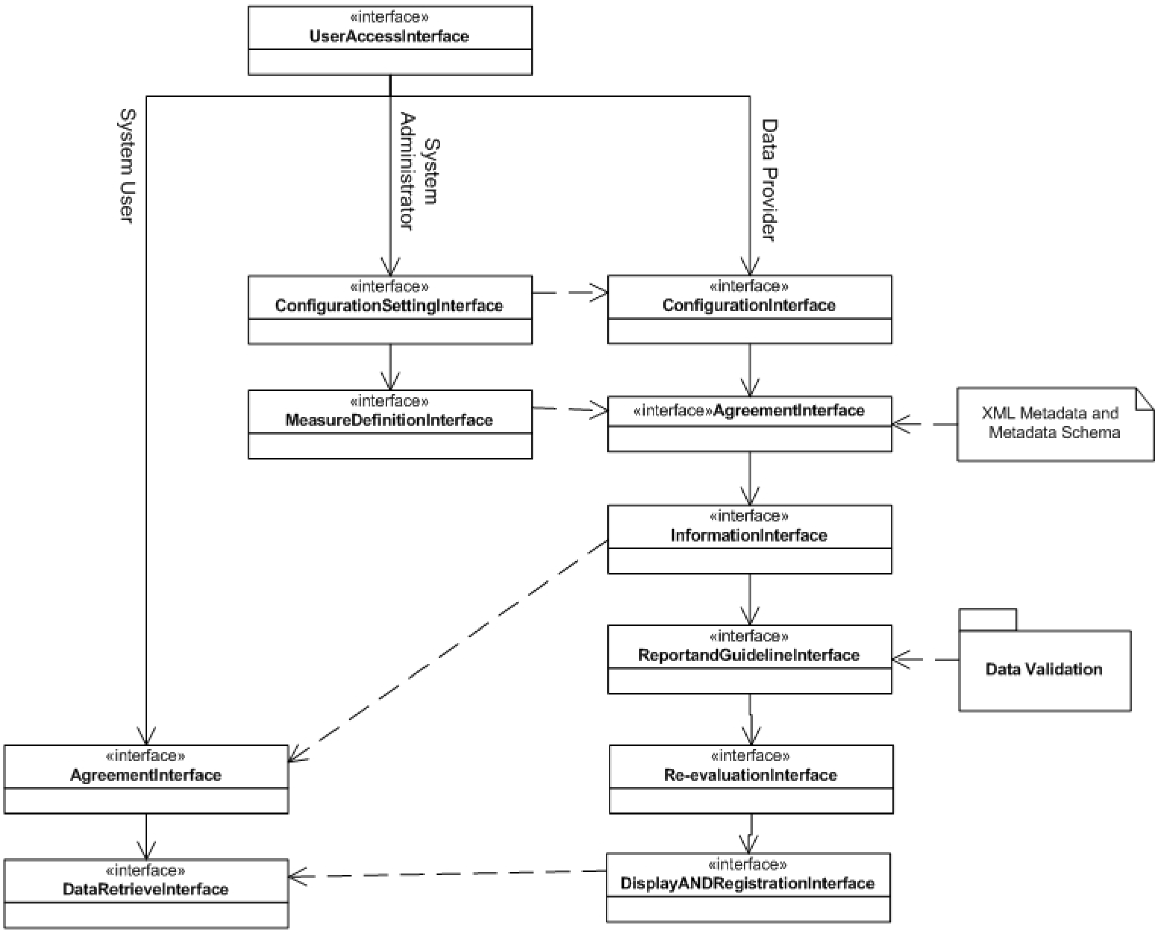


Figure 7.12. Prototype user interfaces and interaction between them

The following two interfaces are designed specifically for the system administrator. These interfaces allow the system administrator to define configurations and evaluation measures. They include the following two interfaces:

- ConfigurationSettingInterface
- MeasureDefinitionInterface

ConfigurationSettingInterface interface (Figure 7.13) allows the system administrator to define new configurations and related measures or manipulate existing configurations. There are a number of configurations that can be defined by the system administrators. They include:

- spatial reference system (datum and projection system)

- metadata standard (and respective schema)
- measure items from metadata
- acceptable geographical extent for data layers
- acceptable formats
- restrictions in forms of restrictions on use, manipulation and distribution
- privacy policy documents applicable to the datasets
- list of acceptable data custodians.

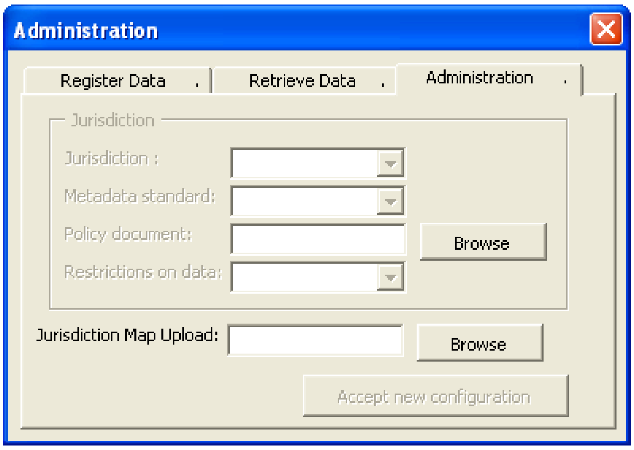


Figure 7.13. System administration interface

Apart from defining measure items that are defined to be extracted from metadata, the system administrator is able to define some rules in the form of restrictions of attribute values. For example, the system administrator can set a no-zero rule of a particular attribute that represents the unique identifier of features and can restrict attributes to values other than zero.

System users also can obtain the datasets that have been already been validated and registered in the system. These users need to agree with the restrictions and privacy policies that have been assigned to the datasets by the system administrator.

The major part of the prototype has been designed and implemented for data providers to evaluate and register compliant data in the system. For this purpose, six interfaces have been developed as follows:

- ConfigurationInterface
- AgreementInterface
- InformationInterface
- ReportGuidelineInterface
- Re-evaluationInterface
- RegisterationInterface.

Through ConfigurationInterface (Figure 7.14), data providers select corresponding contextual configurations that have already been defined by system administrators. This includes selection of jurisdiction, metadata standards, policies and restrictions.

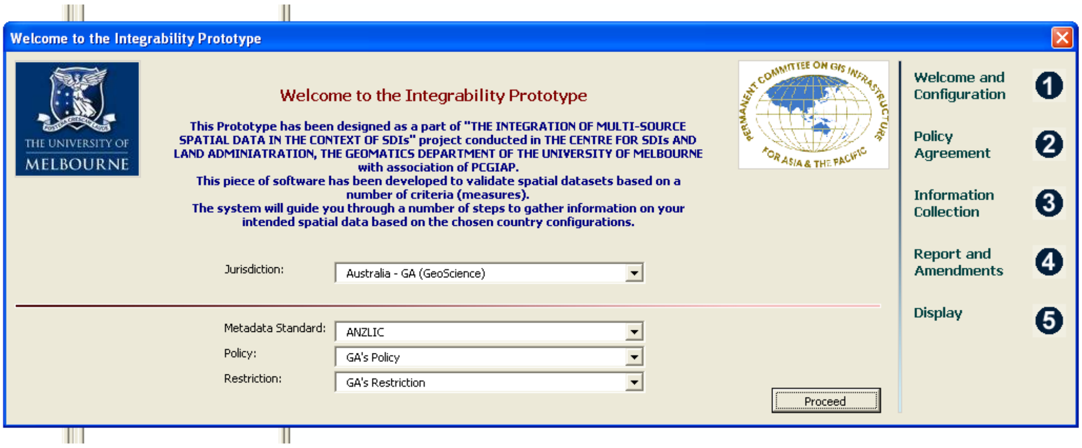


Figure 7.14. Configuration user interface

For any particular jurisdiction, available metadata standards, policies and restrictions are customised; therefore it ensures the selection of only items that are applicable for the corresponding jurisdiction. It also sets acceptable formats, spatial reference systems, restrictions, available attribution rules, geographical extent and metadata items (which should be extracted from the XML metadata file) for that particular session. For example, some of the items that can be extracted from ANZLIC metadata are as follows:

- data custodian
- pricing
- currency
- scale
- completeness
- attribute accuracy
- spatial accuracy
- logical consistency.

Next the interface (AgreementInterface) asks data providers if they agree with the policies and other restrictions that have been defined by system administrators (Figure 7.15).

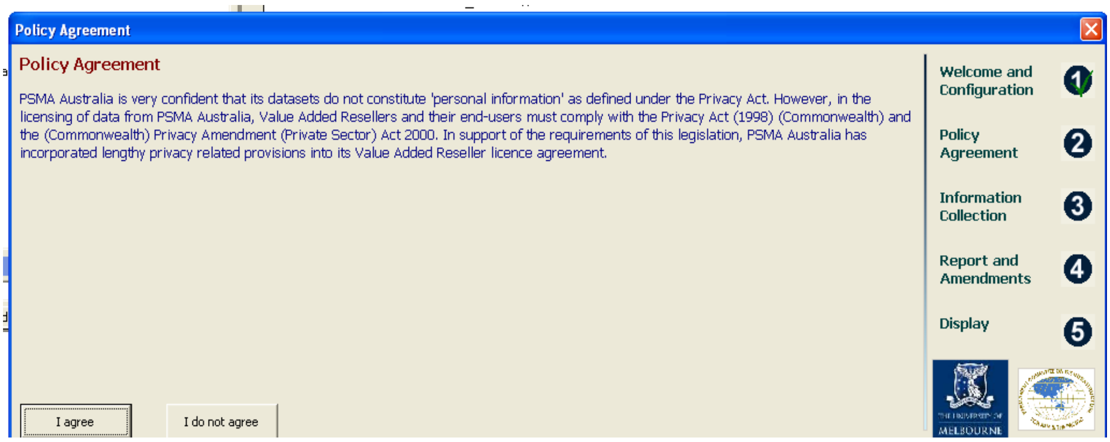


Figure 7.15. Policy agreement user interface

If data providers agree with the conditions, the interface then provides data providers with an interface (InformationInterface) that allows them to enter data and metadata locations as illustrated in Figures 7.16a and 7.16b. Based on the configuration that has been selected in the previous step, corresponding information on data characteristics including format, geographical extent and attribution rules and so on are extracted from actual data and metadata. The same items that represent counterpart values and quantities to the evaluation measures can be extracted from data and metadata. Actual data contains information on some of these values. It includes geographical extent, formats, attribution content and spatial reference systems. This interface also allows data providers to define more rules to restrict attribute content. It is a very useful functionality as it checks the quality of attributes.

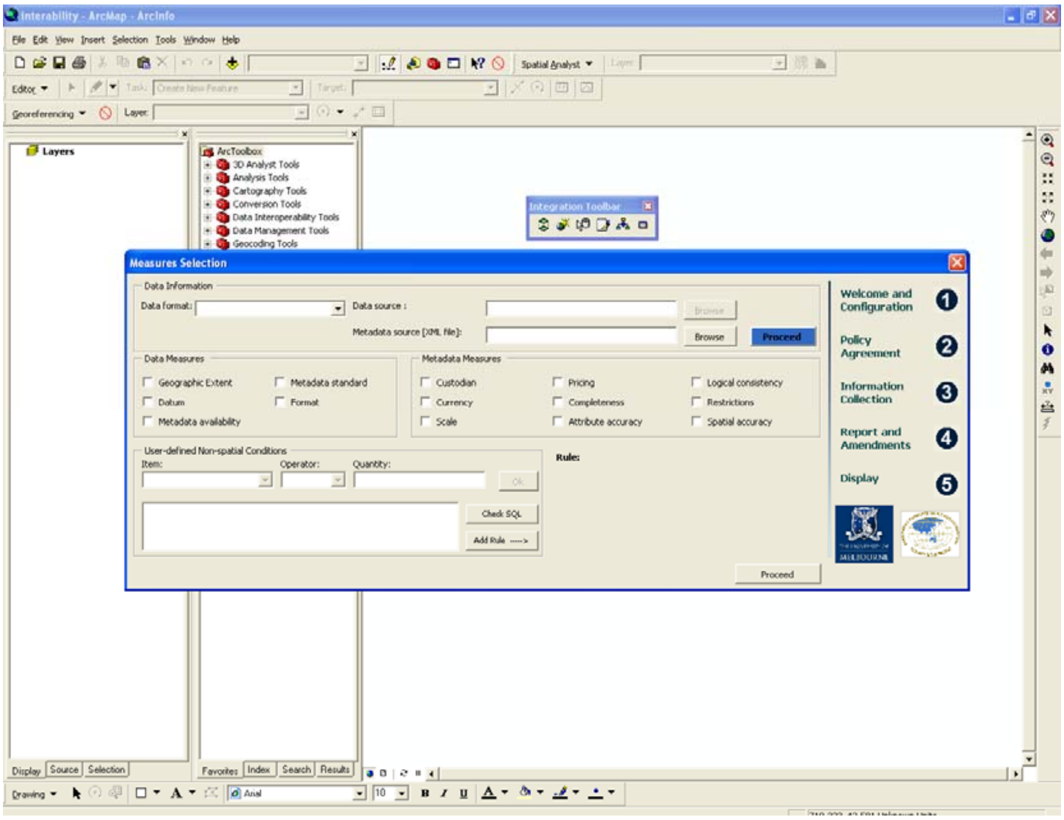


Figure 7.16a. Information user interface (before data and metadata entry)

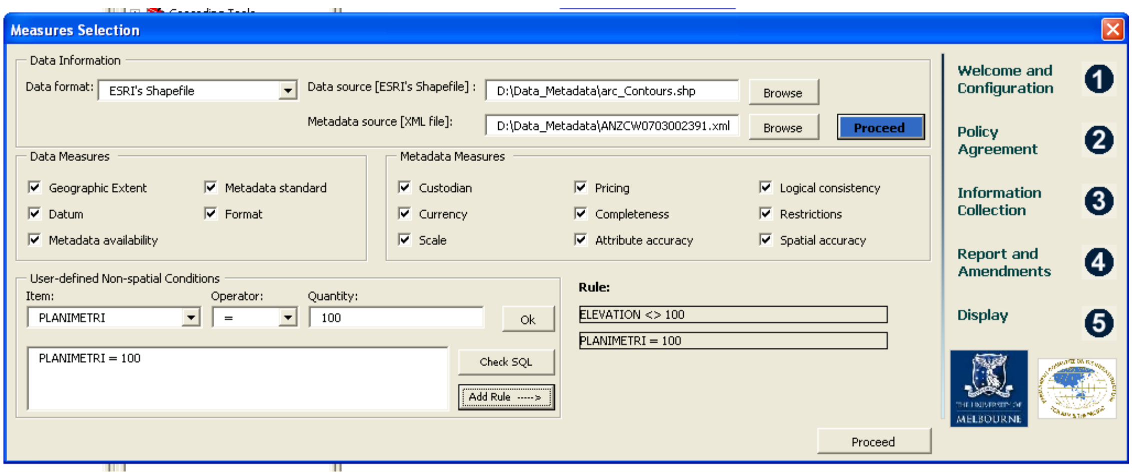


Figure 7.16b. Information user interface (after data and metadata entry)

ReportGuidelineInterface (Figure 7.17) presents the results of data validation that has been produced through comparing the measures with the information that has been extracted from data and metadata. The report also highlights the items of data inconsistency with measures. The report also assigns available and applicable guidelines to the inconsistency items. Availability of the guidelines refers to a link to guidelines in the guidelines table of the prototype database.

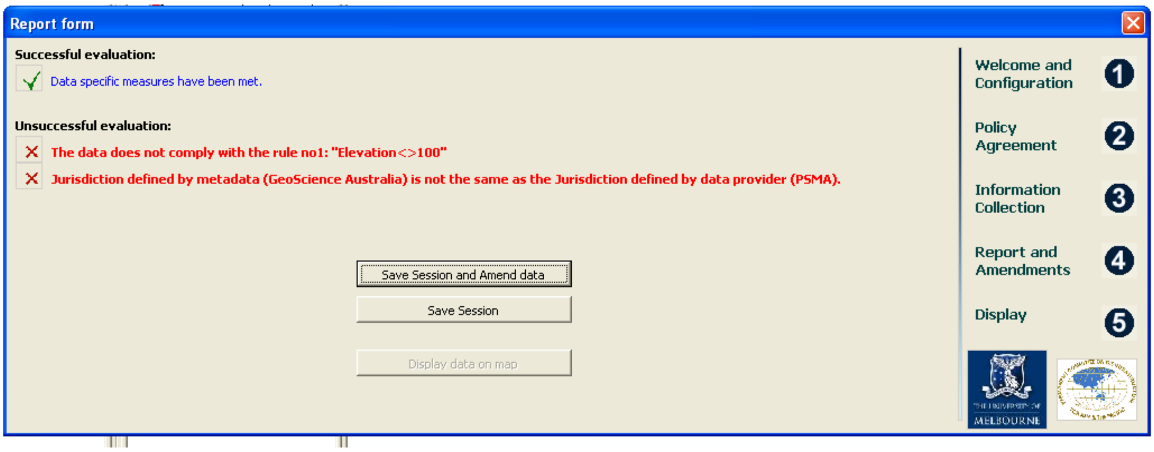


Figure 7.17. Report on the evaluation results

If the data layer has been amended and provided for re-evaluation, the same evaluation process is executed on data and if there is no more inconsistency (Figure 7.18) with measures, the data layer is displayed in the ArcGIS environment (Figure 7.19) and a data record is also added to the database.

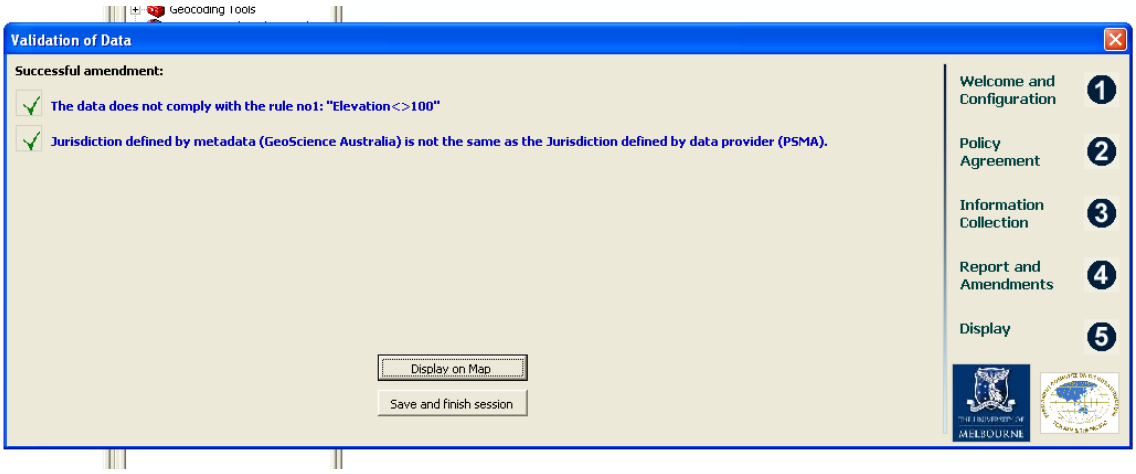


Figure 7.18. Re-evaluation after data amendment

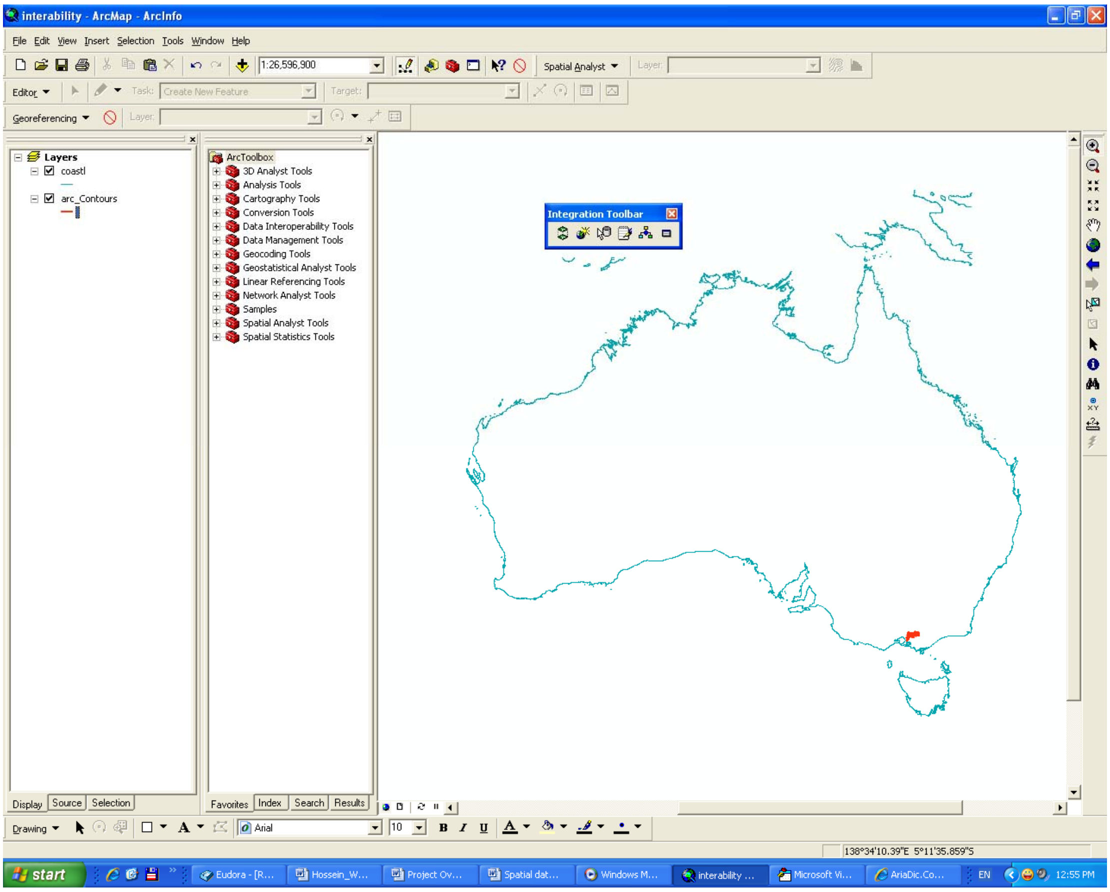


Figure 7.19. Data layer representation in the ArcGIS environment

Prototype Use Test

In order to conduct a realistic and practical example to show the capabilities of the prototype, the tools have been applied to an administrative boundary dataset in the State of Victoria, Australia. The data provider points to a dataset (local council administrative boundaries) and provides a metadata source based on the ANZLIC metadata profile. The system extracts the information from data and metadata and cross-checks the information with the jurisdictions (State of Victoria) specifications, which have been fed in as configurations of the prototype.

Based on the pre-defined specifications (for testing the tool), the datum of compliant data should be GDA94 (Geographic Datum of Australia); in the geographical extent of Victoria, the restriction on data was set to “complete restrictions” and a specific attribute’s value (jurisdiction) could not be “null”.

The dataset was tested and the outcome showed that the geographical extent is the same as the tools configuration, but the datum was WGS84, the restriction was “subject to licence” and a number of features had the attribute with null value. For the datum the specification GDA’s technical manual (ICSM, 2002) was proposed and the null values were reported. As the restriction was in a lower level of restriction, it was accepted. Table 7.1 shows the rules that were set as default rules and also the characteristics of the dataset. It also highlights the items of incompliancy with the defined rules.

Table 7.1. The result of the validation prototype use test

Measures	**Jurisdiction-defined**	**Dataset**	**Incompliancy**
Accepted Formats	ESRI shapefile, ESRI coverage, WFS/WMS, ERMapper ECW	MapInfo TAB	X
Metadata Standard and DTD/Schema	ANZLIC's profile, anzmeta1.3.dtd	ANZLIC's profile, anzmeta1.3.dtd	√
Policy Agreement	Privacy policy Restrictions	Subject to licence	√
Geographical Extent	Australia's boundaries	Australia, Victoria	√
Datum	GDA_1994	WGS86	X
Currency	Not older than one month	Two weeks old	√
Minimum and Maximum Scale	1:1,000 to 1:1,000,000	1:25000	√
Completeness	complete	Complete	√
And user-defined rules:	No "null" is accepted for attributes	Some "null" attributes	X

The use test showed that with a given list of measures and rules to evaluate the datasets, the users spend much more time investigating data content, running queries and extracting information from metadata and data manually. The use test also indicated that in some cases in the manual test, users take different approaches caused by the different interpretation of the rules. For example, the geographical extent can be defined by the maximum and minimum coordinates of the data features, but in some cases participants used the jurisdiction that the data belongs to, in order to specify the geographical extents. The use test also defined two major advantages, which include the time that is consumed for data evaluation and as a consequence the money spent, and also the comprehensiveness of the assessment.

The tool can be an integral component of any SDI to evaluate and assess the readiness of datasets to become a part of the SDI dataset. However, if users intend to use it as an individual tool for data evaluation, the data integration and validation tools can provide them with functionalities to evaluate the datasets based on pre-defined and customisable measures. Therefore, at least the tool can identify the compliancy of datasets upon user-defined measures. It is also necessary to clarify that this tool does not provide an environment in which two or more datasets are integrated geometrically and aspatially; however, there are many tools that have been developed to perform these activities.

Discussion on Prototype Development Findings

The tool test showed the automation of the process of data validation can be increased and facilitated with the tool. The tool decreased the manual, resource-intensive, time-consuming and cumbersome process of data investigation and validation.

Utilising a tool that contains a set of rules and guidelines, the process of data validation for integration can be automated much more easily. The standardised and routine process proposed by the tool also provides a consistent approach to evaluate different data sets.

The tool provides a data preparation platform, which assists data providers and users to evaluate the preparedness of spatial data to be integrated with other data sets and also provide functions to prepare the data on the basis of jurisdiction-defined rules and measures.

From an SDI data content perspective, once the tool is configured with the rules and specifications and necessary functions are developed, the data sets that are evaluated and verified by the tool, can ensure free-of-challenge integration with other data sets.

In the process of development, the availability of resources including appropriate rich metadata and practical instructions and specifications (including metadata schema integration) for data manipulation and incompliancy resolution remains a challenge in some jurisdictions. In some cases, some information including data on spatial and aspatial accuracies are not accessible and require a number of logical and statistical analyses on data that should be developed for the tool. Also, to be widely used by different jurisdictions, metadata schema for the jurisdiction (where it is not available) and the metadata schema conversion tool should be developed. In this regard, the data access module also requires more development to be able to access and obtain data from more formats including the raster format and services including Web Coverage Servers (WCS).

Many of technical and non-technical issues, including geographical extent and datum, can be measured utilising analysis and query tools. However, some of them including logical consistency and restrictions are not easily measurable, unless there is an indication of them in supporting documents including metadata. In order to automate the process of the evaluation of these items, the machine-readable documents are highly helpful.

Another issue raised during the implementation phase was the measurability of the metadata content, which helps a lot in the assessment process. Some metadata content including the accuracies, privacy policies and restrictions are kept in the form of a descriptive text content that is not easily comparable with another value, so the issue raised is the metadata content. If the metadata content is not only machine-readable (for example in the form of XML files) but also measurable, this helps many different analyses and assessments that require the evaluation and measurement of the metadata content.

The use test has shown that by utilising the tool the validation of spatial datasets can be highly facilitated through a structured approach. Some of the benefits of utilising the data validation tool are as follows:

- time and effort saving
- consistent approach
- provision of guidelines assists users to find the solution with minimum effort in minimum time
- can be used as an individual data validation tool.

From an institutional arrangement perspective, a rigorous custodianship agreement between data providers and owners that oblige them to provide a certain data content and accompanying documents will also assist effective data integration. Privacy, restrictions, metadata and pricing documents especially with measurable content (in the form of XML or other machine-readable structures) can greatly facilitate not only data evaluation but many other processes including data discovery, data use and sharing.

7.4. Spatial Data Integration Guidelines

Spatial data integration guidelines can form a comprehensive document that details the major integration activities. It outlines a methodology for data evaluation and integration, potential technical and non-technical barriers to spatial data integration, jurisdiction-specific considerations for spatial data integration and possible and available solutions for data integration barriers.

The guidelines' development is highly dependent on the needs and objectives of the respective jurisdiction and the context of the respective SDI. Each SDI has its own considerations and guidelines. It includes the roadmaps, standards, policies and agreements that are developed within SDI to facilitate the coordination of spatial datasets. In the case of data integration, the guidelines are specifically focused on facilitating the integration of multi-source spatial datasets. The guidelines can be utilised by practitioners to learn the issues and barriers that they may expect and also possible solutions for data integration. In this regard, the document should cover a number of necessary components including:

- spatial data validation steps
- potential barriers, challenges and issues (including the jurisdiction and SDI-specific considerations)
- possible solutions.

Saying that, a general guidelines document is presented here which tries to cover the above-mentioned components. The document presented here is supposed to support and demonstrate the idea of data integration guidelines. However, SDI coordinators can develop and expand the components and items of the guidelines.

7.4.1. Spatial Data Integration and Validation Steps

Chapter Two presented a two-level data integration: process and data-level integration. The combination of these two levels provides a holistic approach for data integration. The process-level data integration outlines a set of processes that should be performed for exploring, assessing, communicating with data custodians, acquiring data, data manipulation and collating datasets as illustrated in Figure 2.14, Chapter Two. The data-level integration discusses the activities that are related to the final step of process-level integration that is data collation (Figure 2.13). This includes inconsistency identification, utilising guidelines, data manipulation and amendments, and data integration.

Each of these levels has its own issues and considerations that have been discussed in detail in Chapter Two. The combination of these two data-integration levels provides a number of steps necessary for comprehensive data integration (Figure 7.20).

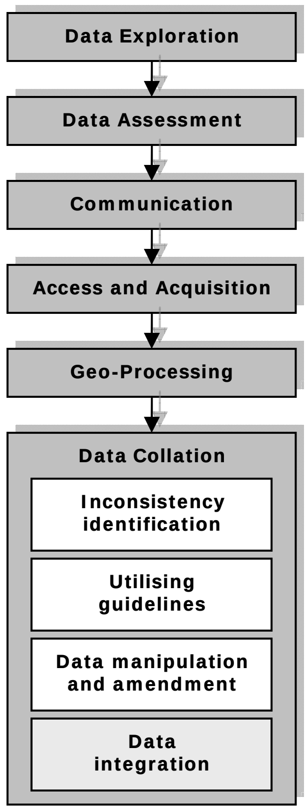


Figure 7.20. Steps for spatial data integration

7.4.2. Data Integration Challenges and Issues

The major challenges and issues that should be considered in general for effective data integration have been identified and listed in Chapters Four and Five.

SDI coordinators can select and include any of the items according to the needs of the accommodating jurisdiction and community. They may require adding more items according to the requirement of their jurisdiction. For example, the diversity of data models is an issue for Australian national SDI where states have adopted different data models (Chapter Five), but it is not a major issue for Indonesia, where data models are developed at national level and adopted by other authorities.

7.4.3. Possible Enablers and Solutions

The issues and challenges of spatial data integration require different enablers and solutions. Depending on the issues and their impact on data integration, effective solutions should be utilised. The solutions and enablers vary from technical developments to the establishment of collaborations and arrangements. Some of the key enablers that facilitate spatial data integration and overcome the barriers are as follows:

- custodianship arrangement
- conversion tools
- metadata structure and content
- data specifications

- standards
- extract, transform and load (ETL) tools
- spatial data interoperability
- generalization tools
- single point of truth/access
- data-quality considerations
- multi-agency privacy policies
- multi-agency pricing policies.

With referring back to the data integration steps and associated challenges and issues, each of above-mentioned enablers and solutions can facilitate one or many of the data integration steps.

As an example, effective custodianship arrangement and agreements assign dedicated custodians to coordinate particular spatial datasets. This can include updating the data and associated documents such as metadata and data specifications. Custodians can also take responsibility for distributing and managing the policies and restrictions on data. With such custodianship agreements in place, there are determined agencies that can be referred to to acquire the data and associated documents.

The development of technical tools such as conversion, generalisation and ETL helps users to amend and manipulate available datasets to meet the criteria of users' applications. Multi-agency collaborative initiatives within the SDI context such as the development of multi-agency standards, privacy and pricing policies also help to integrate the inconsistent policies and standards that are applicable to the integration of multi-source datasets.

7.5. Integration Data Model

The spatial data model represents the structure and integrity of spatial data features (Spyns et al., 2002). Data models are largely the domain of spatial applications and services for spatial data manipulation, structured queries and analysis (Teorey, 2006). The integrated data model adds great value to data analysis and queries that rely on multi-source spatial datasets.

Spatial communities have different perceptions of same spatial phenomena. The perception is more distinct in communities with diverse interests in spatial data. For example, the land parcel feature can be differently interpreted by surveyors, GIS users, environmental scientists and land lawyers (Figure 7.21).

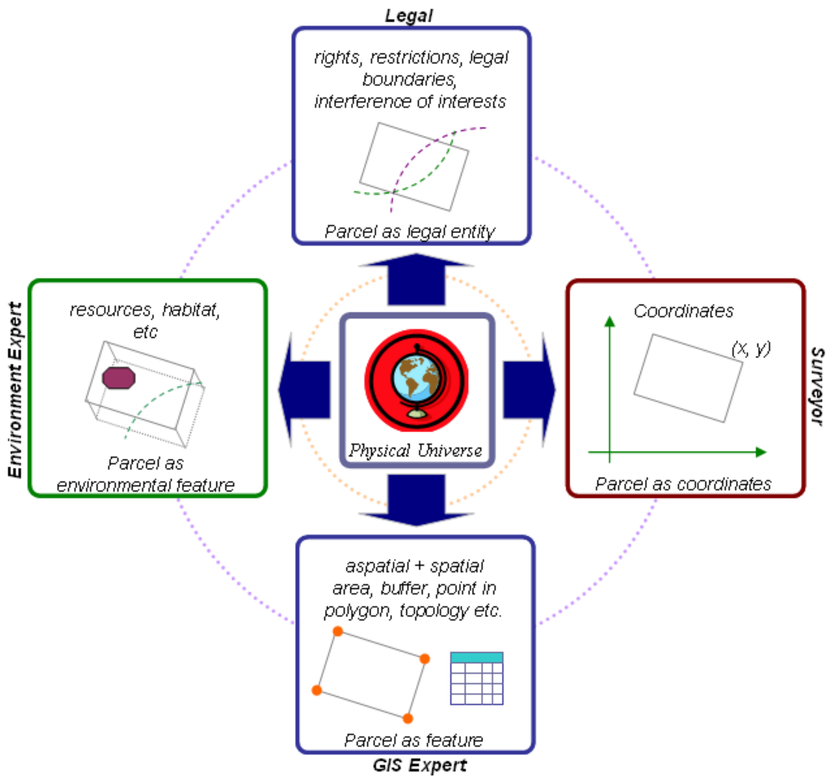


Figure 7.21. Different perception of spatial phenomena among different communities

A surveyor interprets a land parcel as a set of coordinates (or distances and angles). The interest of the surveyors' community in land is the quantities and measurements. GIS users require not only geometry and measures but also attributes and analysis of its different spatial and aspatial components. At the same time the environmental community studies the environmental features of the land. The land is resource and habitat of species for this community. The land lawyer has legal interests in the land. These include the rights, restrictions, legal boundaries and interference of interests on the land. Each of these communities requires a different view of a single phenomenon and as a consequence the tools to achieve the corresponding view could be different. Different views entail different modelling mechanisms.

Therefore, an effective integration data model should be developed based on the meanings of the geographical phenomena, so an effective approach is the ontology-driven data model (Uschold & Gruninger, 1996; Sheth, 1998; Fonseca, 2001). In this regard, the next section aims to present an ontology-based reclassification approach that can be used in the design of an integrated data model.

7.5.1. Ontology Definition

The main motivation of ontology is knowledge sharing and reuse (Lin et al., 2001). Ontology defines the vocabulary with which queries and assertions are exchanged among entities. In this regard, Najar (2006) has proposed that ontology can be placed on the level between the conceptual model and reality as illustrated in Figure 7.22.

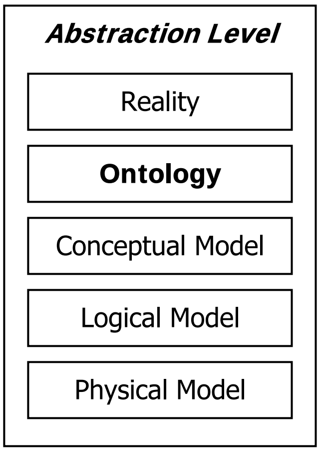


Figure 7.22. Level of model abstractions including ontology

Lin et al. (2001) propose a core data model approach based on ontologies, which deals with multi-source data. In this approach ontologies can be used to organise keywords that describe different aspects of data features and relationships between them. Ontologies are useful for data integration because they form a basis for integrating separate data features through identification of logical connection or constraints between information pieces.

Visser et al. (2002) adopted ontologies to extract features based on their specifications. In the method proposed by Visser et al., a new classification of spatial features has been proposed through the new concept term describing the information entity. In this method a new classification of features is presented that is independent from source classifications (Figure 7.23).

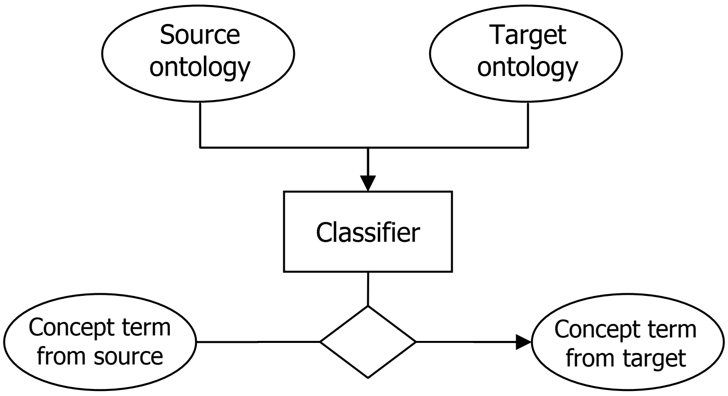


Figure 7.23. Reclassification of features through ontologies (Visser et al., 2002)

7.5.2. Feature Reclassification based on Ontologies

By adopting the above-presented method a new approach can be utilised for the reclassification of features into a new data model. This method defines a set of new rules for classification of spatial features that can result in a different classification from the source data model. The classification relies on actual data and its specification. If common vocabulary exists among source datasets, the new data model can be created through extraction of features with the rules defined for the new integration data model. This method is presented in Figure 7.24.

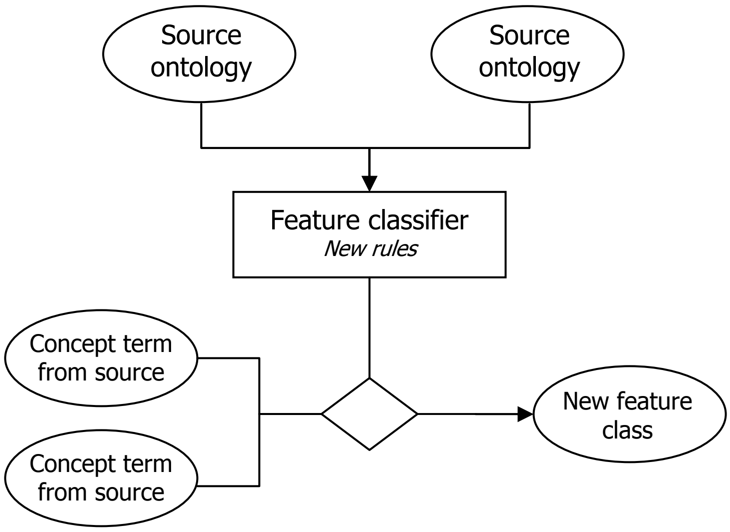


Figure 7.24. Reclassification of spatial features based on integration date model rules

In order to demonstrate the method, a reclassification of forest (as part of wooded-land class) has been presented utilising ontologies for vegetation datasets of the State of Victoria, Australia.

The wooded-land class is defined by Department of Agriculture, Fishery and Forestry (DAFF) although the wooded-land spatial data is coordinated by Department of Sustainability and Environment. In the definition presented by DAFF there are three sub-classes in the wooded-land class. DAFF (2008) defines this class as follows:

> an area, incorporating all living and non-living components, that is dominated by trees having usually a single stem and a mature or potentially mature stand height exceeding 2 metres and with existing or potential crown cover of overstorey strata about equal to or greater than 20 per cent. This definition includes Australia's diverse native forests and plantations, regardless of age. It is also sufficiently broad to encompass areas of trees that are sometimes described as woodlands … and consists of three categories as follows:
>
> - woodland: 20–50 per cent crown cover, height less than 20 metres
> - open forest: 51–80 per cent crown cover, height more than 20 metres
> - closed forest: 81–100 per cent crown cover, height more than 20 metres

Figure 7.25 illustrates the ontology defined by DAFF for wooded land. In this classification, forest, woodland and others are three subclasses of wooded land. Forest has two subclasses of open forest and closed forest. Each of these subclasses is identified by quantitative attributes of height and density, which are kept in an Excel file.

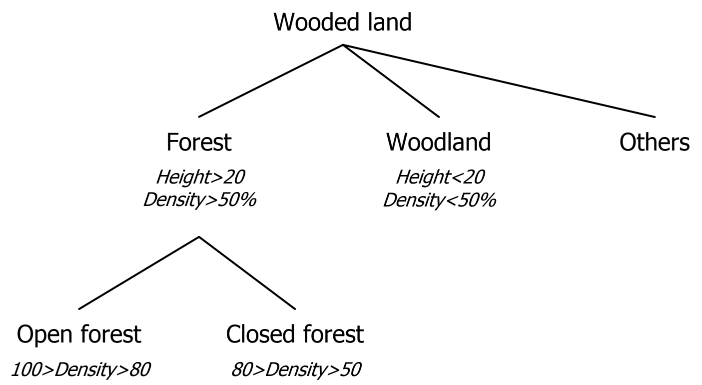


Figure 7.25. DAFF wooded-land ontology

A geographical representation of the wooded lands also exists, which contains polygons of wooded lands. However, the main dataset of forests, which is kept as part of SII's VicMap suite, is the TREE-DENSITY layer which contains forest data with their density attribute in the form of quantitative attributes including dense, medium density and low density (Figure 7.26).

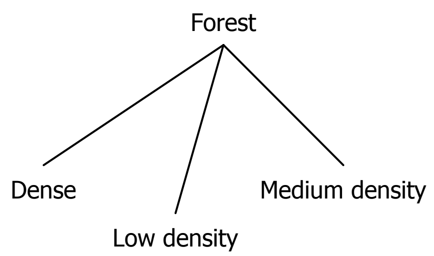


Figure 7.26. SII's forest ontology

But with visual integration of these two layers it is identified that the TREE-DENSITY layer does not represent the forests correctly (Figure 7.27).

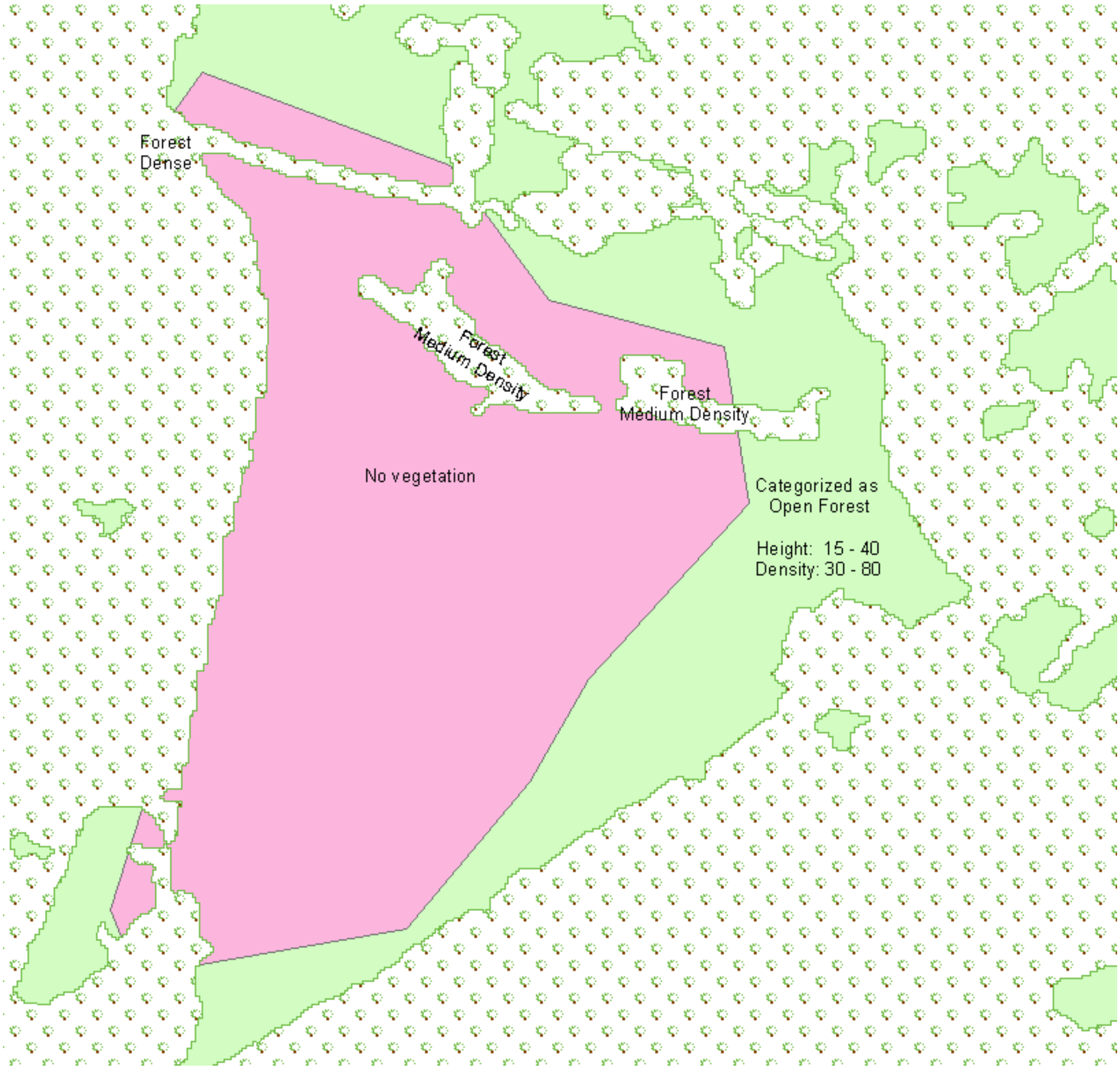


Figure 7.27. The visual integration of the TREE-DENSITY layer with the wooded lands layer

The plain green and pink areas represent wooded lands and areas with tree symbols are areas covered by TREE-DENSITY. The figure shows that there are areas with no vegetation, which have been categorised as forests.

In order to develop a new classification of forests, the two ontologies presented above have been integrated into one ontology (Figure 7.28).

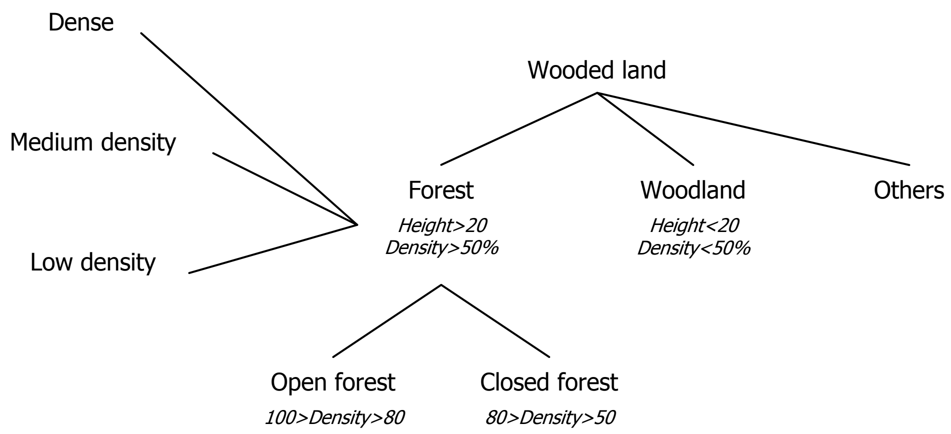


Figure 7.28. Integration of DAFF's and SII's ontologies

Now based on the new ontology presented in Figure 7.28, those features from forest layer (TREE-DENSITY) are reclassified in forest class, which sits within the borders of forest features from the wooded-land layer. In order to reclassify the features in a new class, a new feature selection operation has been executed. In this operation, all features from the TREE-DENSITY layer, which sits within the boundaries of forest feature layers from wooded-land layer, are extracted in a new layer. This removes the features from SII's ontology (and consequently from the TREE-DENSITY layer), which are not forests in DAFF's ontology.

This example is a demonstration of ontology-based reclassification of spatial features. More rules can be applied for more accurate results. There are also some relationships between features that can be used as ontology integration rules. In some cases there are also some restrictions that can be incorporated into the ontology to gain a better reclassification. The new classes can form a new data model based on the integration of ontologies of source communities.

There are also some barriers to the effective integration of ontologies and these require serious attention in order to implement this method. These barriers are listed below:

- lack of common vocabulary among participating ontologies
- lack of detailed ontology for spatial features
- spatial and aspatial content of features do not reflect the source ontology.

7.6. Integration-oriented Metadata

Metadata, commonly defined as "data about data", is a structured summary of information that describes data (SEDAC, 2006). The term, however, is not restricted to descriptions of data. More broadly defined, metadata is descriptive information about any object or resource as diverse as geospatial and non-geospatial datasets, data analysis tools, computer models, websites, graphics and textual information. It may be more up to date to define metadata as "supplementary information at a higher level of abstraction of information on a lower level of abstraction". At a minimum, metadata consists of the standard bibliographic information that supports resource discovery (discovery-level metadata). However, it generally contains information that supports a wider range of operations, such as management, evaluation, access and use. Thus, a comprehensive metadata standard would contain sufficient descriptors to allow for automatic processing of the data by ensuring machine-readability of a self-explanatory format. Data that is not documented in accordance with a standard cannot be found by queries to search engines. Such data does not exist or at least is known only to a comparatively small community of insiders.

There are a number of standards including ISO TC211, CSDGM, SERF, DCMI and ANZLIC's metadata standard utilised by different jurisdictions. These standards differ greatly in the level of information they support. Essentially, one can look at both the uses of metadata, and the various standards along a continuum of complexity. The most basic record enables data and resource discovery, much like records in a library catalogue, whereas the most complex provides essential information for processing and interpreting data, much like a user manual. Metadata facilitates comparisons between datasets from different sources. Whenever it is placed in a searchable index, it enables searching of domain-specific information, such as geographic location, title or data type. Metadata may also serve as a tool for organising and maintaining an organisation's investment in its data, by providing a systematic way of recording information about the data it produces. Metadata may even provide protection for the producing organisation if a conflict arises over the misuse of data. In essence, metadata is documentation that can answer who, what, when, where, why and how questions, describing every facet of the data or resource being documented – its content, quality, accessibility, collection methods, processing and availability (SEDAC, 2006).

Metadata is used often, especially by end-users to explore, assess and integrate datasets. The focus of this document is the integration side of metadata use.

The use of metadata elements is vital to data integration and validation alongside other data coordination operations such as data manipulation, access and exploration. Different metadata that is used internationally provides different elements. Some elements are essential for integration purposes, while not all metadata standards provide them.

This section tries to identify the metadata requirements for effective data integration. This report also assesses different international metadata standards against integration requirements.

7.6.1. Effective Metadata for Effective Data Integration

Metadata generally contains information that supports a wide range of operations. Integration relies heavily on information provided in metadata. Integration consists of different steps including data exploration, data assessment, communication with data providers, access and acquisition, geo-processing and data collation. Metadata content including data quality, contact details, rights and restrictions on data, attributes description, spatial reference systems, scales and so on, is used for effective integration.

Different metadata standards are used by different jurisdictions. Each metadata standard provides a particular set of information on data. However, much information data including titles, reference systems, abstracts and so on is common among all metadata standards.

The information content of metadata that is used to identify data is essential for data integration. This includes the description of data, data categories and applications. This will help users to find the most suitable data for their use. Data descriptions and data categories help users to find out the subject of data. Information on data applications is helpful information that assists users to assess the fitness for purpose of data and gives them ideas about what the data is good for.

Metadata on attributes is needed for users who need to integrate spatial data through attributes or those who need more non-spatial data than spatial data and need to analyse integrated datasets using attributes. Information on attributes including attributes description and type helps users to choose the data that contains the most appropriate non-spatial data for that particular application.

One of the most significant problems of integration is the currency of datasets. Sometimes users need the most current data or need data at a certain time. A date/time stamp of not only metadata but also data itself is a helpful tool for this purpose.

Quality information assists users to assess the fitness-for-purpose of data and find out if any particular data fits their application or not. Since utilising spatial data with poor quality beside good-quality data could lead to poor results, the quality of data is an essential part of metadata for data integration. Quality information includes lineage, spatial and aspatial accuracy.

Restrictions on data can restrict the effective integration of data and can greatly affect the effective use of data. Restrictions on data including access and use restrictions avoid full use of data and users tend to use data with fewer obligations.

Contact point information is a necessity for communication with data providers to access and acquire data.

Distribution information is also provided in some metadata standards and helps the users find out the way data is distributed and accessible. It includes format and data transfer methods.

Technical information including spatial reference systems, boundary coordinates and formats can assist users to overcome technical difficulties of data integration. Acquiring and using spatial data of unknown spatial reference systems or formats lead to a time-consuming and costly process. The data model is used as a tool for feature integration at model level. Datasets can be linked and integrated to others through the data model.

7.6.2. Metadata Standards and their Content

There are many metadata standards adopted internationally. Each metadata standard has been developed to meet embedding jurisdictions' requirements. Different metadata standards provide a set of elements. From the integration perspective, each of these standards has strengths and weaknesses to some extent.

In this section the most popular and famous standards at international level are highlighted and described. These include ISO199115/TC211 (International Standards Organization/Technical Committee 211), DIF (Directory Interchange Format), SERF (Service Entry Resource Format) and DCMI (Dublin Metadata Core Initiative) and some national metadata standards including standards adopted by ANZLIC (Australia and New Zealand), CSDGM (Content Standard for Digital Geospatial Metadata adopted by The United States and Canada).

ISO 19115/TC211 Metadata Standard

ISO TC 211 Geographic Information/Geomatics is responsible for the ISO geographic information series of standards. The international community, through the International Organization of Standards (ISO), has developed and approved an international metadata standard, ISO 19115. As a member of ISO, the US is required to revise the CSDGM in accord with ISO 19115. Each nation can craft their own profile of ISO 19115 with the requirement that it include the 13 core elements.

The objective of this international standard is to provide a clear procedure for the description of digital geographic datasets so that users will be able to determine whether the data in a holding will be of use to them and how to access the data. By establishing a common set of metadata terminology, definitions and extension procedures, this standard will promote the proper use and effective retrieval of geographic data. Supplementary benefits of this standard for metadata are to facilitate the organisation and management of geographic data and to provide information about an organisation's database to others. This standard for the implementation and documentation of metadata furnishes those unfamiliar with geographic data with the appropriate information to characterise their geographic data and it makes possible dataset cataloguing enabling data discovery, retrieval and reuse (ISO Standards, 2006).

DIF Metadata Standard

The Directory Interchange Format (DIF) was the product of an Earth Science and Applications Data Systems Workshop on catalog interoperability (CI). The workshop recommended that a "... first step towards data system interoperability, Catalog Interoperability (CI), the ability to find information about data held at other sites ..." (Major & Olsen, 2006) be made.

The Catalog Interoperability Working Group (consisting of several US federal and international agencies) defined the type of information and level of detail that would be contained by the DIF. After several revisions from the scientific community, the DIF was formally approved and adopted by a CI science advisory group at a CI workshop in 1988.

The DIF structure has been flexible enough to evolve with growing metadata requirements, especially for the geospatial disciplines. Then the geospatial community began work towards the development of an international standard for geospatial metadata.

The DIF is used to create directory entries that describe a group of data. A DIF consists of a collection of fields that detail specific information about the data. Eight fields are required in the DIF; the others expand upon and clarify the information.

The DIF allows users of data to understand the contents of a data set and contains those fields that are necessary for users to decide whether a particular data set would be useful for their needs (Major & Olsen, 2006).

SERF Metadata Standard

The Service Entry Resource Format (SERF) is another defacto standard established by NASA's Global Change Master Directory (GCMD). It is used for directory entries describing a service, rather than data. The SERF was established for describing services directly related to the “processing, viewing, analysis, archival, retrieval, production, interpretation, acquisition, formatting, or indexing of Earth science data” – the tools a GCMD user may need for manipulating data. Like the DIF, it too is intended to serve the user community in discovery. It is a very simple standard, only requiring seven elements.

DCMI Metadata Standard

The Dublin Core Metadata Initiative (DCMI) is an organisation dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialised metadata vocabularies for describing resources that enable more intelligent information discovery systems. The DCMI provides simple standards to facilitate the finding, sharing and management of information. DCMI does this by:

- developing and maintaining international standards for describing resources
- supporting a worldwide community of users and developers
- promoting widespread use of Dublin Core solutions.

ANZLIC Metadata Standard

ANZLIC’s mission is to provide leadership for effective management and use of land and geographic information to support economic growth, and the social and environmental interests of Australia and New Zealand. Key objectives under the headings: data, infrastructure, standards, access, industry development and organisational framework, are the focus of efforts to provide this leadership.

A working group was formed by the ANZLIC Advisory Committee to work on the following tasks to improve community access to data:

- produce a Metadata Framework paper that identifies and defines the mandatory metadata elements of a national land and geographic data directory system, discussing creation, maintenance and directory custodianship issues
- using the Metadata Framework paper, promote the concept of a national data directory system to help determine the priorities and issues for implementing a national directory
- develop an implementation plan for a national data directory system, including procedures for transfer of metadata between jurisdictions and the national directory system
- develop and circulate for comment, a discussion paper on national guidelines for developing land and geographic data quality, in a form suitable for developing into an Australia/New Zealand standard (ANZLIC, 2006).

The ANZLIC guidelines have been developed to promote a consistent standard of description for this small number of core metadata elements that are generally common for all types of data and designed to indicate what data exists, its content, geographic extent, how useful it might be for other purposes and where more information about the data can be obtained. The purpose is to make information about all available data freely available so that existing data can be reused for other purposes if it is suitable, reducing the duplication of effort.ANZLIC can actually be considered as a standard on its own because it can be compared to the FGDC work

(2006) and mainly the ISO19115/TC211 which has had extensive Australian input, particularly from interests associated with ANZLIC.

The US approach, developed by the Federal Geographic Data Committee (FGDC), specifies the structure and expected content of some 220 items (elements) which are intended to describe digital geospatial datasets adequately for all purposes. The ANZLIC approach is deliberately less ambitious than what has been attempted in the US. Arguments advanced in support of the more modest objective rely on experience to date with the creation of high-level directories in Australia.

While ANZLIC has not adopted the US approach, the Australia–New Zealand framework is, as far as possible, consistent with the guidelines on Digital Geospatial Metadata produced by the US FGDC and with the Australia–New Zealand Standard on Spatial Data Transfer AS/NZS 4270. The reasons for this are:

- Many organisations are already using these standards for their data management activities.
- Some vendors of software are providing templates and other support to the implementation of the standards.
- These standards are being implemented in some discipline or theme areas where there is international exchange of metadata.

However, until now, there has been no unifying set of metadata elements that could be used as the basis for the development of national metadata standards. ISO 19115/TC211 will provide this unifying set of metadata elements.

ANZLIC has based it on the FGDC work and then has made its own core elements and categories without taking them from another developed standard. They did not proceed to any mapping and comparison with another namespace.

To assist with the implementation, ANZLIC has developed a run-time software tool to support the collection of metadata and to ensure consistent description of core metadata elements. This software tool, based on Microsoft Access, is available for use by dataset custodians throughout Australia and New Zealand. The Data Entry tool may be used within organisations to manage the metadata database (SCHEMAS, 2006).

CSDGM

The Content Standard for Digital Geospatial Metadata (CSDGM), more commonly referred to as the FGDC standard, was developed specifically to provide a common set of terminology and definitions for documenting digital geospatial data. It has been the federally endorsed metadata standard for geospatial data in the US since 1994. It is currently used in more than 200 national and international catalogs and clearinghouses.

The standard is large and fairly complex and specifies the content of some 330 metadata elements, though only a subset of the 330 elements is mandatory. Elements are specified as “mandatory”, “mandatory if applicable”, or “optional”. Using only the minimum number of mandatory elements may be sufficient for data discovery. It may also be an appropriate set for initial documentation. However, to assist in data transfer and processing, the minimum required elements may not be sufficient.

Despite the size and level of complexity of this standard, it is fairly flexible. It can be modified for specific data types by employing an endorsed "profile" or "extension", which is simply an addition or a simplification to the standard.

To date, two profiles have been endorsed by the FGDC – the Biological Data Profile and the Metadata Profile for Shoreline Data. Several more are under development, such as the soon to be released Thematic Supplement for Geospatially Referenced Cultural and Demographic Data Metadata. Although this standard was developed for geospatial data, the Biological Data Profile, developed for the National Biological Information Infrastructure (NBII) of the US Geological Survey (USGS), is a prime example of how the CSDGM can be used for data that is not geospatial in nature (such as results from laboratory-based research). The CSDGM is used in the NBII Clearinghouse initiative to provide metadata-based descriptions of non-data resources, such as software tools, data applications, reports, websites and other information products (SEDAC, 2006).

7.6.3. Effectiveness of Metadata Standards for Spatial Data Integration

Based on characteristics required for integration as defined above, metadata standards are compared in this section. Each metadata that has the most characteristics is more suitable for data integration purposes. Table 7.2 summarises these characteristics for the above-mentioned metadata standards.

Table 7.2. The comparison between metadata standards according to the integration requirements

		ISO/TC211	DIF	SERF	DCMI	EEA-MSGI	ANZLIC	CSDGM
Data Identification	Title	√	√	√	√	√	√	√
	Abstract/Description	√	√	√	√	√	√	√
	Category/Theme	√	√	√	√	√	√	√
	Application	x	√	x	√	x	√	√
Attributes	Attributes List	x	x	x	x	x	x	√
	Attribute Description	x	x	x	x	x	x	√
	Attribute Type	x	x	x	x	x	x	√
Date/Time Stamp	Metadata Date/Time Stamp	√	√	√	√	√	√	√
	Data Time Stamp	x	√	√	x	√	√	√
Quality Information	Lineage	√	√	√	x	x	√	√
	Spatial Accuracy	√	√	√	x	√	√	√
	Aspatial Accuracy	x	√	√	x	x	√	√
Restrictions (Access/Use)		x	√	√	√	√	x	x

Contact Point Information		√	√	√	√	√	√	√
Distribution Information		√	√	√	√	√	√	√
Technical Information	Spatial Reference System	√	√	x	x	√	√	√
	Boundary Coordinates	√	√	x	√	√	√	√
	Scale	√	√	√	√	√	√	√
	Format	√	√	√	√	√	√	√
Data Model		√	x	x	√	x	x	√

Based on outcomes of Table 7.2, it is shown that CSDGM is the best available metadata that contains the necessary components to support multi-source spatial data integration. It includes attributes lists, descriptions and types and data models. However, because of the complexity and detail degree of CSDGM, which hinders regular update and maintenance, most organisations that have already utilised this metadata standard, opt to adopt metadata standards with less complexity and details. Among others ISO/TC211 is widely utilised by many organisations (Moellering, 2005).

The items that have been mentioned as facilitators of data integration can be encapsulated in spatial data specification documents that can accommodate more detailed information.

7.7. Spatial Data Specification Document

A data specification document can provide an invaluable source of information on different aspects of spatial datasets. Data specification is not bound to any standard and it can contain any kind of information about spatial datasets. Spatial data custodians utilise data specifications to introduce the datasets. General (or detailed) descriptions of data, its origin, potential audiences and users, associated datasets, spatial and aspatial content, database rules, access, pricing and privacy policies, detailed quality information and data models can be described in data specification documents.

Many of these items are quite critical for spatial data integration. For example, detailed description of spatial features that describes the spatial and aspatial characteristics of the features, the spatial and aspatial restrictions and relations can be utilised for reclassification and building integrated spatial data models. Attribute details can be utilised to join and integrate spatial layers at the attribute level.

Despite the significance of the information content of spatial data specification documents, data custodians provide very general information on spatial datasets. Table 7.3 summarises the information that has been included in spatial data specification documents for the cases of Europe (INSPIRE), State of Victoria and Geoscience Australia.

Table 7.3. Features of the spatial data specification document for INSPIRE, the state of Victoria, and Geoscience Australia

INSPIRE	Victoria	GA
• Scope (of the Document) • Overview • Specification scopes • Data product identification • Data content and structure • Reference systems • Data quality • Metadata • Delivery • Data capture (optional) • Portrayal • Additional information (optional)	1. Custodian 2. Jurisdiction 3. Contact information 4. Description 5. Content 6. Production and/or acquisition methods 7. Source 8. Generalisations within the data 9. Currency & status 10. Data creation dates 11. Maintenance and update frequency 12. Standards and specifications 13. Legislative requirements 14. Future plans 15. Data schema 16. Data model 17. Database design 18. Data dictionary 19. Areas of application	1. Brief description 2. Size criterion 3. Attributes 4. Database rules 5. Relationship rules to associated datasets 6. Map rules 7. Related features 8. Related products

INSPIRE (Myrind, 2008) has developed and proposed a profile for data specification which contains useful information on data product identification including title, abstract, topic category, geographic description, purpose (use cases and user requirements) and spatial representation type (e.g. vector). INSPIRE's profile also contains a narrative description part that provides informal descriptions of datasets, consistency between spatial datasets, identifier management, an optional modelling of object references and profiles of spatial and temporal schema. The narrative description also presents application schema and feature catalogue that are derived from the UML model in shared repository. Horizontal and vertical reference systems are also outlined in the document. Data quality is also included in this document, in accordance with ISO 19113. Some restrictions and rules are also defined in the quality section. These include topological and logical consistency rules (e.g. curves do not overlap or touch interior, i.e. curves only touch at their ends and do not intersect or overlap).

Victoria also maintains very comprehensive spatial data specification documents. This document consists of a broad range of information including description, attributes, data models, access constraints and quality information (SII, 2005). The document is distributed alongside datasets and also provides a wide range of information that can be utilised for data assessment and integration.

Geoscience Australia (2007) also maintains a brief data specification. GA's data specification contains a brief description on data, a number of rules and criteria and associated layers and features.

These three cases show the information abstraction level and details that are provided by custodians. In the case of Europe, the data specification document is standardised, while in the case of Australia each data provider develops its own data specification. There are a number of essential pieces of information that are highly critical of some data integration activities. Conceptual definition of data features that consists of spatial and aspatial characteristics of features, the restrictions and relations to other features are very important for integrated data model. Detailed attribute content that details the attributes, types and restrictions is also necessary for database integration. Data models and relations to other features also can help users to integrate datasets at model level.

In order to maintain and update an effective spatial data specification document a well-established arrangement among the custodians is highly important. A problem also raised during the data specification study, which is crucial to the access and use of these documents, is that these documents are not discoverable and accessible through defined gateways. Therefore, a data specification registry service can overcome this issue.

7.8. Chapter Summary

Effective spatial data integration ensures effective sharing of usable spatial data among stakeholders. It also facilitates the use of shared datasets with less time and effort. In order to achieve this aim, this chapter has presented the design and development of a number of the key components of the spatial data integration toolbox. It comprises the spatial data validation and integration tools, spatial data integration guidelines, integration data models, data integration-specific metadata and data specifications.

The spatial data validation and integration tools are an integral component of SDI that facilitates data sharing and evaluates the readiness of spatial datasets for integration. The tool evaluates datasets against a set of rules and measures that define usable and integrable datasets. The rules and measures are defined based on the characteristics of the datasets for SDI framework. SDIs define some characteristics for spatial datasets in terms of technical and non-technical rules. The compliancy of spatial datasets to the rules shows the readiness of those datasets for integration purposes.

The developed tool automates the validation of datasets against some measures and rules and facilitates the sharing and integration of spatial datasets. The proposed prototype tool has been designed and developed utilising software design tools including UML. UML diagrams have been used to illustrate the architecture and components of the prototype system. The UML use-case and class diagram have been devised to identify and design the necessary components and classes of the prototype.

The prototype tool comprises a number of components that enable administrators to set and define configurations and rules based on the requirements of their jurisdiction. It includes metadata standards, geographical extent, data accuracy, scales, restrictions on data and aspatial restrictions and so on. The tool also allows the data providers to choose any combination of the rules or define some other rules on aspatial content of data and evaluate the data against the rules. In order to perform this, the tool extracts data characteristics from actual data and metadata. For this purpose, ESRI's ArcMap GIS engine and an XML processor have been utilised to collect required information. Then the tool evaluates the characteristics of the data with rules and provides the data provider with the results that include any item of inconsistency with the rules. Some data manipulation tools also have been provided to amend and re-evaluate the data. If no inconsistency exists, the data can be recorded in a database and provided to users. The use case on the prototype proved that the tool can facilitate the integration and validation of spatial datasets. The tool saves time and effort of data validation compared to manual validation. The tool also follows a consistent and structured approach that prevents unnecessary attempts and assists users to find the solution with minimum effort in minimum time. The tool also can be used as an individual data validation tool. Users can use the tool based on their requirements. As a result, users can define rules and measures that meet their needs better and assess datasets based on their specific rules.

This chapter has also presented the structure and content of effective spatial data integration guidelines and proposed a methodology for spatial data integration. It covers the potential technical and non-technical barriers to spatial data integration and also available enablers for the issues. The development of the guidelines is highly dependent on the requirements of the respective jurisdictions and may contain any combination of the issues and solutions that have been proposed in this research.

The integration data model has also been discussed in this chapter. An ontology-based approach has been introduced for integration data modelling. This approach proposes a reclassification of the features based on their conceptual descriptions. The components of the specification form a number of spatial and aspatial characteristics and also a number of relations to other features. These characteristics can be converted to spatial and aspatial contents and also restrictions and relations to other entities. A comprehensive definition of the feature that covers spatial and aspatial restrictions and relations to other features can promise a holistic conceptual model. This model is independent of any model that is utilised by different stakeholders. The chapter then applied this approach to combine ontologies and reclassify the forest feature for the wooded land layer for the State of Victoria, Australia.

The chapter then has presented the appropriate content of spatial metadata and data specification documents that contain necessary information for effective data integration. The chapter has proposed that reliable, measurable and machine-readable content of metadata can greatly facilitate the integration of multi-source datasets. This information can be utilised in the data validation tool to extract information on data and also for the integration data model. In this regard, a number of metadata and data specifications have been investigated and best practices with further recommendations have been discussed. The best practice of the CSDGM metadata standard contains some significant information on aspatial content and the data model of datasets.

The chapter then discusses the appropriate content of the data specification document for data integration. The chapter exemplifies a number of data specification documents that lack necessary information for data integration. Some of the key contents include ontology definition of data features, feature-level metadata and data models.

It is the responsibility of the data custodian to maintain appropriate data specification and metadata documents. These documents can facilitate not only the multi-source spatial data integration but also other activities on spatial data including data modelling, data assessment and data manipulation.

The chapter presents the components of the multi-source spatial data integration toolbox with reference to SDIs as a data-sharing platform. The chapter contains the deliverables and outcomes of the research, which will be concluded in Chapter Eight. Chapter Eight concludes the research project and discusses whether the initial objectives of the research project have been achieved. It also provides some concluding remarks and implications and outlines a number of areas that require further research.

Chapter 8

Conclusions and Future Research

8.1. Introduction

The effective integration of multi-source spatial datasets is critical for delivery of their government services, informed decision making and the creation of business opportunities. As a result, many application and services rely heavily on multi-source spatial datasets. However, the diversity of standards, arrangements and specifications that data providers utilise, cause many technical and non-technical problems.

This research investigated the potential issues of, challenges and barriers to effective spatial data integration and also addressed the possible solutions and enablers to overcome these issues. The result of this study reaffirms the significance of the development of necessary tools and requirements for effective spatial data integration within the context of SDI initiatives. In this regard, a data integration toolbox has also been proposed and key components of the toolbox have been demonstrated and developed.

This chapter examines the outcomes achieved during this research, highlights the significance of the research project to theory and practice, reflects on the original research problem and suggests directions for future research efforts.

8.2. Research Aim and Objectives

As highlighted in first chapter, the major aim of this research has been:

"*to identify the potential technical and non-technical barriers to effective spatial data integration and also to address possible and available solutions and enablers to overcome the issues within the context of holistic framework of SDI initiatives*".

In order to achieve the aims of the research, Chapters Four and Five have conducted a number of international and Australian case studies to investigate and identify the technical and non-technical challenges and barriers to effective spatial data integration. Chapter Four has discussed the spatial data integration activities and possible issues and challenges within Australia and also seven countries in the Asia-Pacific region. The results and findings of these chapters have identified a number of tools and facilitators (including the data validation and integration tools, associated guidelines, spatial data integration data models, metadata and data specification documents) that facilitate the integration of multi-source spatial datasets and also overcome the challenges and barriers. Chapter Seven has discussed and presented the development of the toolbox components.

The research has also fulfilled its objectives. The objectives of the research included:

8.2.1. Objective 1: Investigate, identify and understand the potential technical and non-technical barriers to integrate multi-source datasets

This research has investigated and identified the potential barriers to spatial data integration through the literature review and the conducting of a number of case studies. The literature review provided a basis for understanding spatial data integration and its possible technical and non-technical barriers. The case studies have also been conducted for a detailed investigation of the challenges in different jurisdictions.

The diversity of the participating case study countries allowed the identification of diverse issues and challenges that each country may encounter according to its specific conditions. The research has also discussed the investigation of a number of technical and data-assessment case studies in Australia. This included the investigation and assessment of actual spatial datasets for three Australian major data providers at federal and state levels. Technical visits and assessments also have been conducted to study the data integration activities and respective barriers and challenges within a number of organisations.

The outcomes of the case studies provided a detailed and comprehensive study on the issues and challenges of data integration. They also identified a number of potential solutions that assist in overcoming the barriers.

8.2.2. Objective 2: Identify available enablers for effective spatial data integration

The research has capitalised on the detailed case study investigation and technical visits to identify the key components and tools that can assist practitioners and SDI coordinators to tackle the issues and challenges of effective spatial data integration. Many tools have been developed so far to overcome the data integration issues, but the research has confirmed that there are areas that need more investigation and development.

The identified components have formed a suite of data integration toolbox components. The components of the toolbox include a data validation and integration tools, associated guidelines, integration data models, integration metadata and data specification content. The need for the toolbox components together with their benefits has also been highlighted.

8.2.3. Objective 3: Design and develop the key components of a spatial data integration toolbox

In order to facilitate data integration, the data validation and integration tools have been developed to evaluate and assess the readiness of spatial datasets for data integration. This was fulfilled through the evaluation of spatial data characteristics with a set of customisable and predefined measures and rules. The use test of the tool affirmed that the tool can prevent the time-consuming and inconsistent data validation process.

Associated guidelines have also formed a part of the toolbox. The guidelines discuss the data integration methodology and major challenges which should be considered and also the possible solution for each challenge. An ontology-based approach for reclassification of features has been also presented. This approach has capitalised on the conceptual definition and characteristics of spatial features and has developed a new classification model that overarches the requirements of different data providers. The appropriate content of metadata and data specification documents to encapsulate necessary information and structure for spatial data integration has been also presented.

8.3. Conclusion

The research problem, which has been defined in Chapter One (section 1.2.1), identified that "*multi-source spatial datasets comply with diverse standards, specifications and arrangements; therefore the integration of multi-source spatial* *datasets is associated with technical and non-technical issues that hinder the use of spatial datasets to their maximum potential and usability".*

The research project has confirmed the problem exists in many spatial communities worldwide. The problem inhibits the use of multi-source spatial datasets to their maximum potential.

The research has also outlined that the technical and non-technical challenges of spatial data integration cannot be addressed by a purely technical approach and requires the holistic framework of SDI to overcome the problems.

8.4. Recommendation for Further Research

The outcomes of this research have highlighted a number of areas that require further research. Hence, future research efforts could be directed in the following areas.

> *First,* the development of the data validation and integration tools showed that the achievement of its objectives is highly dependent on the comparable and measurable rules and criteria. These rules and criteria are used as a basis to measure data specification and check the readiness of datasets for integration. They also provide a comparative method for different users to assess and evaluate their datasets based on customisable and predefined rules. The achievement of this aim requires a comprehensive investigation of the data characteristics and specifications and the sources to obtain this information. In this regard, detailed investigation of the specifics that can be collected from actual data, metadata, data specification documents or other sources is required. Also for future study is how the descriptive specifications of the datasets can be converted to measurable rules or how measurable components of the descriptive information can be extracted. An example is the data quality information in the metadata. The scope of this research did not enable further investigation into all measures and rules of integration. Therefore, it is recommended that the details of the rules and criteria for data integration, the appropriate metrics for measuring the readiness of datasets for integration with other datasets, and the approach to measure and compare all components of the data specification are investigated within the SDI initiatives.
>
> *Second,* data models are of high significance for spatial database management, effective querying and data analysis, as they provide the structure, relationships to others and architecture of the features within a data layer. It is more critical in the integration of datasets, as in most cases the integration product is utilised for data analysis and joint querying. Therefore, through a well-developed integration data model, effective data analysis can be delivered. The research highlighted that inconsistency in the diverse data models is difficult to manage. This is because of the data model structure, involving features, feature specifications and relations. Therefore, ontology-based data reclassification and extraction have been identified as an effective approach. The ontology-based data reclassification indicated that by utilising ontology vocabulary and concepts, spatial data features can be more effectively classified within integrated data models. This research has introduced the ontology as a tool that can facilitate the integration of data models that consequently facilitates queries and analysis of integrated datasets. According to its scope, this research did not allow more development on the ontology data reclassification. As a result, it is suggested that the utilisation of

ontology for comprehensive data model integration and its impact on the ease of queries and analysis is investigated.

Finally, many different success factors have been proposed for SDIs. The success of SDIs rests in the achievement of their objectives. One of the major objectives of SDIs is to facilitate the sharing of multi-source spatial resources especially spatial datasets. The integration of multi-source datasets has been discussed as the compelling reason and as an enabler to deliver spatial data sharing. As a result, multi-source data integration can indicate the success of SDIs in achieving their data-sharing aims. The measurement of the readiness of spatial data to be integrated with other SDI datasets may provide a useful research test-bed for examining the impact of SDIs on business, government service delivery, decision making and spatial society and as a result the assessment of the SDIs' success.

8.5. Final Remarks

The development of SDIs is contributing to the delivery of government services, the improvement of decision making and the creation of business opportunities in many countries. The integration of multi-source spatial datasets has been hindered by a number of technical and non-technical issues and challenges. Therefore, facilitating data integration and addressing associated issues are essential for SDIs to achieve their aims. Effective spatial data integration is the main focus and key objective of the developments within the SDI initiatives.

The outcomes of this research and the development of the spatial data integration toolbox components (especially the data validation and integration tools) have the potential to improve the success of SDIs in providing usable and integrable spatial datasets to governments, businesses and decision-makers.

References

1Spatial. (2008). 1Spatial Spatial GIS Solution. Accessed: 10 June. 2008. http://www.1spatial.com/

Alexiadou, S., and Rajabifard, A. (2006). Developing National SDI platform for Greece. *Coordinates,* March 2006,

ANZLIC. (2003a). Implementing the Australian Spatial Data Infrastructure. Accessed: 21 September. 2007. http://www.anzlic.org.au/get/2381384577.doc

ANZLIC. (2003b). Implementing the Australian Spatial Data Infrastructure: Action Plan 2003-2004. Accessed: 10 February. 2008. http://www.anzlic.org.au/get/2381384577.doc

ANZLIC. (2004). *ANZLIC Strategic Plan 2005-2010 - Milestone 5: National Framework Data Themes*. ANZLIC, www.anzlic.org.au/get/2442847451.pdf.

ANZLIC. (2006). *ANZLIC Metadata Profile*. ANZLIC-the Spatial Information Council.

ANZLIC. (2008). spatial information related terms. Accessed: 10 May. 2008. http://www.anzlic.org.au/glossary_terms.html

Arctur, D., and Zeiler, M. (2004). *Designing Geodatabases*. ESRI Press, Redland.

Arnow, D., Dexter, S., and Weiss, G. (2003). *Introduction to Programming Using Java: An Object-Oriented Approach (2nd Edition)*. Addison Wesley.

Artiso. (2008). Visual Case Tool - UML Tutorial. Accessed: 4 November. 2008. http://www.visualcase.com/tutorials/uml-tutorial.htm

ASDD. (2007). Australian Spatial Data Dictionary. Accessed: 4 May. 2007. http://asdd.ga.gov.au/

Australian Government. (2002). Reforming mitigation, relief and recovery arrangements, High Level Group Recommendations. Accessed: 3 May 2006. 2006. http://www.dotars.gov.au/localgovt/ndr/nat_disaster_report/recommendations.aspx

Axmann. (2008). Feature Manipulation Engine. Accessed: 12 May. 2008. http://www.axmann.at/downloads/fme_documents/fme_intro.pdf

Backe, K., and Edwards, D. (2005). U.S Army's S&T on Spatial Data Integration. Accessed: 12 September. 2006. http://www.ncgia.ucsb.edu/projects/nga/docs/BackeEdwards_Position.pdf

Backx, M. (2003). Gebouwen redden levens. Toegankelijkheidseisen van gebouwgegevens in het kader van de openbare orde en veiligheid, Faculty of Civil Engineering & Geosciences, Deptartment of Geodesy, Geo-Information & Land Development Section, Delft University of Technology.

Baker, A. J., and Young, F. R. (2005). "Digital Mapping Data Currency Through Sharing: A Practical Study." Paper presented at the SSC2005 Spatial Intelligence, Innovation and Praxis, Melbourne, Australia, 12-16 September, 2005.

Baker, H. (2005). State of the data: Victoria's roadmap to spatial integration. *Position* August-September 2005, 59-60.

Bell, D. (2003). UML basics: An introduction to the Unified Modeling Language. Accessed: 20 June. 2008. http://www.ibm.com/developerworks/rational/library/769.html

Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS Interoperability. *International Journal of Geographical Information Science* vol. 12, 299-314.

Blakemore, M. (2004). Reflection on the Usefulness of Spatial Information: Globalization, Infrastructure and Agenda. *GIM International* 18, 11-13.

BOSSI. (2006). To Develop a NSW Spatial Information Strategy. Accessed: 9 July. 2008. http://www.bossi.nsw.gov.au/_media/bossi/pdf/notices_and_initiatives/BOSSI_NSW_Develop_SI_Strategy.pdf

Brand, M. (1998). *Theme Paper: Global Spatial Data Infrastructure: Policy & Organisational Issues*. Accessed: 10 July. 2008. http://gsdidocs.org/docs1998/canberra/theme.html

Brassel, K. E., and Weibel, R. (1988). A Review and Conceptual Framework of Automated Map Generalization. *International Journal of Geographical Information Science* 2 (3), 229-244.

Braun, D., Sivils, J., Shapiro, A., and Versteegh, J. (2000). What is UML? Accessed: 28 September. 2008. http://atlas.kennesaw.edu/~dbraun/csis4650/A&D/UML_tutorial/what_is_uml.htm

Bray, T., Paoli, J., and Sperberg-McQueen, C. M. (1998). Extensible Markup Language (XML). Accessed: 10 December. 2007. http://www.w3.org/TR/1998/REC-xml-19980210

Brodeur, J., Bedard, Y., Edwards, G., and Moulin, B. (2003). Revisiting the Concept of Geospatial Data Interoperability within the Scope of Human Communication Processes, pp. 243-265. Blackwell Publishing Ltd.

Brydon, M. (1997). Access Tutorial 14: Data Access Objects. Accessed: 20 December. 2006. http://fisher.osu.edu/~muhanna_1/837/MSAccess/tutorials/dao_intr.pdf

Buch, V. (2002). Database Architecture: Federated vs. Clustered. Accessed: 2 May. 2008. http://www.oracle.com/technology/tech/windows/rdbms/ClusterComp.pdf

Buehler, K. (2003). *OpenGIS Reference Model*. OGC Inc.

Burgess, W. S. (1999). *Vertical Integration of Spatial Data*. Resources, M. D. o. N., The Maryland State Government Geographic Information Coordinating Committee (MSGIC) and Maryland Local Government GIS Committee (MLOGIC),

Burke, R. (2003). *Getting to Know ArcObjects: Programming ArcGIS with VBA*. ESRI Press.

Burrough, P., and Masser, I. (1998). *European Geospatial Information Infrastructure Opportunities and Pitfalls*.

CANRI. (2001). CANRI System & Data Architecture. Accessed: 9 December. 2007. http://www.canri.nsw.gov.au/activities/projects/2001/eoi/086_architecture.html

CANRI. (2006). Community Access to Natural Resource Information (CANRI) Program. Accessed: 19 May. 2008. http://www.anzlic.org.au/get/2389441055.ppt

Cecconi, A. (2003). Integration of Cartographic Generalization and Multi-Scale Databases for Enhanced Web Mapping, Mathematic Science Faculty, Zurich University, Zurich.

CGDI. (2001). *Canadian Geospatial Data Infrastructure-Architecture Description*. CGDI Architecture Working Group, http://www.geoconnections.org/architecture/architecture_description.pdf.

Chrisman, N., and B. J. Niemann. 1985. Alternate routes to a multipurpose cadastre: Merging institutional and technical reasoning. Auto Carto 7: Proceedings: Digital Representations of Spatial Knowledge, March 11-14, 1985, Washington, D.C., and Falls Church, Va.: American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, pp. 84-89.

Clark, D. (2004). Performance, Learning, Leadership and Knowledge - Understanding. Accessed: 11 June. 2008. http://www.nwlink.com/~Donclark/performance/understanding.html

Clarke, K. C., Parks, B. O., and Crane, M. P. (2002). *Geographic Information System and Environmental Modeling*. Prentice Hall.

Clarke, R. (2004). Information Infrastructure. Accessed: 1 June. 2008. http://www.anu.edu.au/people/Roger.Clarke/II/

Clausen, C., Rajabifard, A., Enemark, S. and Williamson, I. (2006). *Awareness as a Foundation for Developing Effective Spatial Data Infrastructures*. FIG, Munich, Germany.

Cleveland, H. (1982). Information as Resource. The Futurist December 1982, p. 34-39.

Coleman, D., and McLaughlin, J. (1998). Defining Global Geospatial Data Infrastructure (GGDI): Components, Stakeholders and Interfaces. *Geomatica* 52 (2), 129-143.

Colless, R. (2005). Interoperability & Security in the Emergency Services Arena. Accessed: 27th March 2006. 2007. http://www.anzlic.org.au/pubinfo/2413335134.html

Commission of the European Communities. (2004). Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL: Establishing an infrastructure for spatial information in the Community (INSPIRE). Accessed: 5 December. 2006. http://www.ec-gis.org/inspire/proposal/EN.pdf

Conybeare, C. (2003). *Report on the ANZLIC Counter Terrorism Project*. ANZLIC, Australia.

Cowen, David J., and Craig, W. J. (2003). A Retrospective Look at the Need for Multipurpose Cadastre, Surveying and Land Information Science, Vol. 63, No. 4, 2003, pp. 205-214.

CRC.SI. (2004). *Emergency Management Demonstrator Project Plan*. CRC.SI.

Crisostomo, B. A. (2007). The Philippine Experience in the Development of a National Spatial Data Infrastructure. Accessed: 3 November. 2008. http://www.geoinfo.ait.ac.th/download/SCOSA2007/4_SDI_Activities_by_countries/6_Philippines.pdf

CSCI. (2007). A Quick Guide to The Unified Modeling Language (UML). Accessed: 1 November. 2008. http://www.csci.csusb.edu/dick/samples/uml0.html

DAFF. (2008). What is a forest? Accessed: 14 September. 2008. http://www.daff.gov.au/brs/forest-veg/nfi/forest-info/what-is

Dalrymple, K., Williamson, I.P. and Wallace, J. (2003). Cadastral Systems within Australia. *The Australian Surveyor* 48(1), 37-49.

Dangermond, J. (2006). ESRI User Conference 2006 Update. Accessed: 10 July. 2008. http://www10.giscafe.com/nbc/articles/view_weekly.php?articleid=296573&page_no=2

De Santos, J. L. C., deBy, R. A., Apers, P. M. G., and Magalhaes, C. (2002). *Facilitating Interdisciplinary Sciences by the Integration of a CLOSi-based Database with Bio-metadata*, ISPRS Commission IV, Symposium 2002. Ottawa, Canada, July 9-12, 2002.

De Vries, M. (2005). Recycling Geospatial Information in Emergency Management Situation: OGC Standards Play an Important Role, but More Work is Needed. Accessed: 22 May. 2006. http://www.directionsmag.com/article.php?article_id=2019

Donaubauer, A. (2005). "A multi-vendor Spatial Data Infrastructure for Local Government Based on OGC Web Services." Paper presented at the From Pharaohs to Geoinformatics, Cairo, Egypt, 2005.

Donker, F. W., and Van Loenen, B. (2006). "Transparency of accessibility to government-owned geo-information." Paper presented at the 12th EC-GI&GIS Workshop, Innsbruck, Austria, 21-23 June 2006, 2006.

Edwards, D., and Simpson, J. (2002). "Integration and access of multi-source vector data." Paper presented at the Symposium of Geospatial Theory, Processing and application, Ottawa, Canada, 2002.

Egenhofer, M. J. (1993). What's Special about Spatial?--Database Requirements for Vehicle Navigation in Geographic Space. P. Buneman and S. Jajodia (eds.). *SIGMOD Record 22(4): 398-402.*

EUROGI. (1997). *Priority Plan for Legal and Economical GI Aspects*. EUROGI, R.,

European Commission. (2006). *Report of International Workshop on SDIs' Cost-Benefit/Return on Investment: Assessing the Impact of Spatial Data Infrastructure*. European Commission and Institute for Environment and Sustainability, Ispra, Italy.

Executive Office of the President. (1993). The National Information Infrastructure: Agenda for Action Executive Summary. Accessed: 12 June. 2008. http://www.ibiblio.org/nii/NII-Executive-Summary.html

Executive Office of the President. (1994). *Coordinating Geographic Data Acquisition and Access, the National Spatial Data Infrastructure*. Executive Order 12906, F. R., 1767117674, Executive Office of the President, USA,

Executive Office of the President of the US. (2002). Circular No. A-16. Accessed: 7 April. 2008. http://www.whitehouse.gov/omb/circulars/a016/a016_rev.html

FGDC. (2006). Geospatial Metadata Standards. Accessed: 3rd October. 2006. http://www.fgdc.gov/metadata/geospatial-metadata-standards

FIG. (1996). *The Bogor Declaration, United Nations Integral Meeting of Experts on the Cadastre*.

FIG. (1999). *The Bathurst Declaration*. The United Nations.

Finn, M. P., Usery, E. L., Starbuck, M., Weaver, B., and Jaromack, G. (2004). *Integration of the National Maps*, Istanbul, Turkey.

Fonseca, F. (2001). Ontolody-driven Geographic Information Systems, The University of Maine.

Fonseca, F. (2005). System Heterogeneieties Analyses of Interoperable Geospatial Information Systems.

Fonseca, F., Egenhofer, M., Davis, C., and Camara, G. (2002). Semantic Granularity in Ontology-Driven Geographic Information Systems. *AMAI Annals of Mathematics and Artificial Intelligence-Special Issues on Spatial and Temporal Granularity* 36, 121-151.

Geiger, J. G. (1998). *Information Connection: Recognizing a Good Data Warehouse Data Model*.

GeoConnections. (2008). The GeoConnections Initiative. Accessed: 2 May. 2008. http://www.geoconnections.org/publications/Technical_Manual/html_e/toc.html

Geoscience Australia. (2007). NTMS Specifications. Accessed: 25 October. 2008. http://www.ga.gov.au/mapspecs/topographic/v5/appendixA_files/RoadTransport.jsp

Gerasimtchouk, R., and Moyaert, L. (2007). *Integrating CAD and GIS Data at Mineta San Jose International Airport. AIMS*

Goodchild, M., and Gopal, S. (1989). *Accuracy of Spatial Databases*. Taylor and Francis, London.

Goodchild, M., Egenhofer, M. J., and Fegas, R. (1998). *Interoperating GISs: Report of the Specialists Meeting.*, Santa Barbara, CA.

Goodchild, M. F., Kyriakidis, P. C., Schneider, P., and Sifuentes, J. (2005). "Uncertainty and interoperability: the areal interpolation problem." Paper presented at the Fourth International Symposium on Spatial Data Quality (ISSDQ 05), Beijing, August 25–26, 2005, 2005.

Gordon, M. T. (2007). "Report for BOSSI." Paper presented at the Sptially-enabled Government 2007, Hyatt Hotel, Canberra, 14-15 August 2007, 2007.

Gore, A. (1998). *The Digital Earth: Understanding our planet in the 21th century*. California Science Center, Los Angeles.

Groot, R., and McLaughlin, J. (2000). *Geospatial Data Infrastructure: Concepts, Cases and Good Practices*. Oxford University Press, New York.

Gruber, T. (1992). What is an Ontology? Accessed: 21 October. 2008. http://ksl-web.stanford.edu/kst/what-is-an-ontology.html

Grus, L., Crompvoets, J., and Bregt, A. (2006). *Multi-view SDI assessment framework*, Porto, Portugal - p. 120 - 122.

Grus, L., Crompvoets, J., and Bregt, A. K. (2007). Multi-view SDI Assessment Framework. *International Journal of SDI Research* 2, 33-53.

GSDI. (2005). Definition of the GSDI. Accessed: 8 July. 2008. http://www.ipgh.org/GSDI/default_eng.htm

GSDI. (2008). Global Spatial Data Infrastructure Association. Accessed: 10 April. 2008. http://www.gsdi.org/Default.asp

Haas, L., and Lin, E. (2002). IBM Federated Database Technology. Accessed: 23 February. 2008. http://www-128.ibm.com/developerworks/db2/library/techarticle/0203haas/0203haas.html

Hakimpour, F. (2003). Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata, Zurich University, Zurich.

Hammer, J., and McLeod, D. (1993). AN APPROACH TO RESOLVING SEMANTIC HETEROGENEITY IN A FEDERATION OF AUTONOMOUS, HETEROGENEOUS DATABASE SYSTEMS. *International Journal of Intelligent & Cooperative Information Systems, World Scientific* 2, 51-83.

Hanseth, O., and Monteiro, E. (1998). Understanding Information Infrastructure. Accessed: 5 june. 2008. http://heim.ifi.uio.no/~oleha/Publications/bok.html

Harvey, F. J. (2002). *Semantic Interoperability and Citizen/Government Interaction*, Ottawa.

Hodgins, B. W. (1989). *Federalism in Canada and Australia: historical perspectives, 1920-1988*. Frost Centre for Canadian Heritage and Development Studies, Trent University, Peterborough, Canada.

ICSM. (2002). Geocentric Datum of Australia Technical Manual. Accessed: 23 August. 2006. http://www.icsm.gov.au/gda/gdatm/gdav2.3.pdf

INSPIRE. (2002). *INSPIRE Architecture and Standards Position Paper*.

INSPIRE. (2006). INSPIRE: INfrastructure for SPatial InfoRmation in Europe. Accessed: 7 April 2006. 2006. http://inspire.jrc.it/home.html

INSPIRE. (2007). INSPIRE Directive. Accessed: 10 May. 2007. http://www.ec-gis.org/inspire/directive.cfm

Iqbal, A. (2007). Integrating Spatial Datasets Using Road Networks from Heterogeneous and Autonomous Datasets, Geomatics Department, The University of Melbourne, Melbourne.

ISO Standards. (2006). ISO/TC211. Accessed: 5 November. 2006. http://www.isotc211.org/

Jones, M., and Taylor, G. (2004). Data Integration Issues for a Farm Decision Support System. *Transactions in GIS* 8, 459-477.

Kamis, R. (2007). "NSDI Development in Brunei Darussalam." Paper presented at the Workshop on Spatial Data Infrastructure Jointly Organised by SCOSA & Geoinformatics Center, Asian Institute of Technology Sponsored by Japan Aerospace Exploration Agency (JAXA), Thailand, 22-24 February 2007, 2007.

Kasturirangan, K., and Ramamurthy, V. S. (2001). "NSDI: Strategy and action plan." Paper presented at the The NSDI Vision, India, 2001.

Kazuhiko, A. (2007). "Further Advancement of Utilizing Geospatial Information in Japan." Paper presented at the International Workshop on Spatial Enablement of Government and NSDI – Policy Implications, Seoul, Korea, June 2007, 2007.

Kazushige, K. (2007). "Standardization of Data and ISO Standards." Paper presented at the Spatial Data Infrastructure Workshop, Geoinformatics Centre, Asian Institute of Technology, 23 February 2007, 2007.

Keretho, S. (1999). Software Engineering with UML. Accessed: 14 May. 2008. http://www.cpe.ku.ac.th/~sk/Research/OO_Software_Development_with_U/Software_Engineering_with_UML/software_engineering_with_uml.html

Khoo, V. (2007). "e-Initiatives in Singapore s Cadastral Survey System." Paper presented at the International Workshop on Spatial Enablement of Government and NSDI – Policy Implications, Seoul, Korea, June 2007, 2077.

Klinkenberg, B. (2003). The True Cost of Spatial Data in Canada. *The Canadian Geographer* 47,

Kolbe, T. H. (2006). Interoperable Integration of 3D Models over the Internet for Emergency Preparedness and Response. Accessed: 22 May. 2006. http://www.directionsmag.com/article.php?article_id=2038

Krek, A. (2002). An Agent-Based Model for Quantifying the Economic Value of Geographic Information. Translated by. (Frank, A., ed.), Vol. 26, GeoInfo Series, Vienna, Technical University Vienna.

Lanter, D. (1992). Intelligent Assistants for Filling Critical Gaps in GIS (A Research Program). National Centre for Geographic Information and Analysis (Department of Geography). Report 92-4. Accessed: 10 April 2009. http://www.ncgia.ucsb.edu/Publications/tech-reports/92/92-4.PDF

Lin, S., Miller, L. L., Tsai, H.-J., and Xu, J. (2001). *Integrating Heterogeneous Distributed Data Environments with a Database Specific Ontology*.

Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W., eds. (2001). *Geographic Information Systems and Science*. John Wiley & Sons Ltd, Chichester, England.

Major, G., and Olsen, L. (2006). What is a DIF? Accessed: http://gcmd.gsfc.nasa.gov/User/difguide/whatisadif.html

Martin, R. C. (2008). UML Tutorial: Class Diagrams. Accessed: 6 July. 2008. http://www.objectmentor.com/resources/articles/umlClassDiagrams.pdf

Masser, I. (1999). All Shapes and Sizes: The First Generation of National Spatial Data Infrastructure. *International Journal of Geographical Information Science* 13 (1), 67-84.

Masser, I. (2005). *GIS World - Creating Spatial data Infrastructure*. ESRI Press, Redlands, CA.

Masser, I. (2006a). Emerging Trends and Key Strategic Issues: Multi-level Implementation of SDIs. *GIM International* February 2006, 31-33.

Masser, I. (2006b). Multi-level Implementation of SDIs. *GIM International* 20,

Masser, I., Rajabifard, A., and Williamson, I. P. (2007). Spatially Enabling Governments through SDI Implementation. *International Journal of Geographical Information Science* Vol. 21(July), 1-16.

Matindas, R. W., and Purnawan, B. (2004). "Development of National Spatial Data Infrastructure in Indonesia." Paper presented at the FIG Working Week 2004, Athens, Greece, May 22-27, 2004.

McDougall, K. (2006). A Local-State Government Spatial Data Sharing Partnership Model to Facilitate SDI Development, Geomatics Department, The University of Melbourne, Melbourne.

Meixner, H., and Frank, A. U. (1997). GI Policy - Study on Policy Issues Relating to Geographic Information in Europe. European Commission DG XIII: Brussels

Microsoft. (2007). Microsoft XML Parser (MSXML). Accessed: 2008. 2 November. http://support.microsoft.com/kb/269238

Miller, L. L., and Nussar, S. (2003). "An Infrastructure for Supporting Spatial Data Integration." Paper presented at the *Federal Conference on Statistical Methodology*, Washington, DC, 2003.

Miller, P. (2006). Interoperability. Accessed: 18th May 2006. 2006. http://www.ariadne.ac.uk/issue24/interoperability/intro.html

MIT. (2006). Programming in ArcGIS using Visual Basic for Applications. Accessed: http://web.mit.edu/gis/www/programming_iap2006/

Moellering, H. (2005). *World Spatial Metadata Standards*. Elsevier.

Mohammadi, H., Rajabifard, A., Binns, A., and Williamson, I. P. (2006). *The Development of a Framework and Associated Tools for the Integration of Multi-source Spatial Datasets*, 17th UNRCC-AP, Bangkok, Thailand.

Mohammadi, H. (2007). Spatial Data Integration: a Necessity for Spatially Enabling Government. In *Towards a Spatially Enabled Society* (Rajabifard, A., ed), pp. 333-341. Melbourne University Press, Melbourne.

Mohammadi, H., Rajabifard, A., Binns, A., and Williamson, I. P. (2007). *Spatial Data Integration Challenges: Australian Case Studies*, SSC Conference, 14-18 May, Hobart, Australia.

Mohammadi, H., Rajabifard, A., and Willaimson, I. (2008). SDI as Holistic Framework. *GIM International* 22 (1),

Montalvo, U. W. d. (2003). *Mapping the determinants of spatial data sharing*. Published by Ashgate Publishing, Ltd.

Montoya, L., and Masser, I. (2004). Management of Natural Hazard Risk in Cartago, Costa Rica, Habitat International, Volume 29, Issue 3, September 2005, Pages 493-509

Morales, J. (2006). A Framework for GI-Services: Towards Service-oriented Spatial Data Infrastructure. *GIM International,* March 2006, 30-33.

Muggenhuber, G. (2003). Spatial Information for Sustainable Resource Management. *International Federation of Surveyors* September 2003, 18.

Muilu, J., Peltonen, L., and Litton, J. (2007). The federated database - a basis for biobank-based post-genome studies. *European Journal of Human Genetics* 15 (7), 718-723.

Muller, J.-C. (1991). Generalization of Spatial Databases. In *Geographic Information Systems* (D. J. Maguirre, M., and Rhind, G. a. D., eds.). Longman, London.

Myrind, G. (2008). INSPIRE Data Specification. Accessed!0 November. 2008. http://www.ec-gis.org/Workshops/inspire_2008/presentations/04_4_INSPIRE%20Data%20Specification%20Example.pdf

Najar, C. R. (2006). Model-driven approach or integrated metadata-Spatial data management, University of Rostock, Zurich.

National Research Council. 1980. *The need for a multipurpose cadastre*. National Academy Press. Washington, D.C..

National Research Council. (2003). *Weaving a National Map*. The National Academic Press, Washington, D.C.

National Technology Alliance. (2005). Project Profile: Conflation of Geospatial Feature Data. Accessed: 13/10/05. http://www.nta.org/GI/Conflation.pdf

Nebert, D. D. (2004). *The SDI Cookbook*. Global Spatial Data Infrastructure (GSDI).

Nordin, A. F. (2007). "Spatially Enabling Government: The Malaysian Case." Paper presented at the International Workshop on Spatial Enablement of Government and NSDI – Policy Implications, Seoul, Korea, June 2007, 2007.

Omran, E.-S. E. (2007a). Spatial Data Sharing: From Theory to Practice, Wageningen Universiteit, Wageningen.

Omran, E.-S. E. (2007b). Spatial Data Sharing: Bridging the Gap. Accessed: 10 August. 2008. http://www.urisa.org/omran

Onsrud, H. (1995). The Role of Law in Impeding and Facilitating the Sharing of Geographic Information. In *Sharing Geographic Information* (Onsrud, H. J., and Rushton, G., eds.), pp. 292-306. Rutgers: CUPR Press.

Onsrud, H. (2004). Geographic Information Legal Issues. *Encyclopedia of Life Support Systems(EOLSS), Developed under the auspices of the UNESCO*

Onsrud, H. and Rushton, G. (1995). *Sharing Geographic Information*, Centre for Urban Policy Research, New Brunswick, New Jersey, xiii-xviii.

Onsrud, H. Barbara Poore, Robert Rugg, Richard Taupier, and Lyna Wiggins. (2004). The Future of the Spatial Information Infrastructure. In *A Research Agenda for Geographic Information Science* (McMaster, R. B., and Usery, E. L., eds.). Boca Raton: CRC Press.

Ordnance Survey. (2003). German Experience with PAI: DEW example. Accessed: http://www.ordnancesurvey.co.uk/oswebsite/pai/pdfs/german_experience_DEW.pdf

OSDM. (2007). OSDM's Role. Accessed: 20 December. 2007. http://www.osdm.gov.au/OSDM/About+OSDM/default.aspx

Ostensen, O. (2006). *The contribution of international standards to spatial data infrastructure and disaster management*, Bangkok.

Ouksel, A. M., and Sheth, A. (1999). Semantic Interoperability in Global Information Systems. *ACM SIGMOD Record* 28 (1), 5-12.

Parker, J., and Enemark, S. (2005). *Land administration and spatial data infrastructure: Special forum on the development of land information policies in the Americas*, New York.

Paull, D. (2006). *An industry perspective from PSMA Australia*. Victorian Spatial Council.

PC Magazine. (2008). Federated database. Accessed: 10 May. 2008. http://www.pcmag.com/encyclopedia_term/0,2542,t=federated+database&i=43083,00.asp

PCGIAP. (2006). International Workshop on the Integration of Built and Natural Environmental Datasets within National SDI Initiatives, 21st September

2006, Bangkok, Thailand. Accessed: 7th November. 2007. http://www.geom.unimelb.edu.au/research/SDI_research/Integrated/workshop%20index.html

Peedell, S., Friis-Christensen, A., and Schade, S. (2005). *Approaches to Solve Schema Heterogeneity at the European Level*.

Pinto, G. D. R. B., Medeiros, S. P. J., Souza, J. M. D., Strauch, J. C. M., and Marques, C. R. F. (2003). Spatial Data Integration in a collaborative Design Framework. *Communications of the ACM* 46,

Piwowar, J. M., and LeDrew, E. F. (1990). Integrating Spatial Data: A User's Perspective. *Photogrametric Engineering and Remote Sensing* 56, 1497-1502.

Position. (2006). Lynx Launched. *Position* December 2006-January 2007,

Rajabifard, A. (2007). Preface. In *Towards a Spatially Enabled Society* (Rajabifard, A., ed. Centre for SDIs and Land Administration, Geomatics Department, The University of Melbourne, Melbourne.

Rajabifard, A., Binns, A., and Williamson, I. (2005a). "Creating an Enabling Platform for the Delivery of Spatial Information." Paper presented at the Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis: The national biennial Conference of the Spatial Sciences Institute, Melbourne, Australia, September, 2005.

Rajabifard, A., Binns, A., and Williamson, I. P. (2005b). "Development of a Virtual Australia Utilising an SDI Enabled Platform." Paper presented at the FIG Working Week/GSDI-8, 16-21 April 2005, Cairo, Egypt, 2005.

Rajabifard, A., Binns, A., and Williamson, I. (2005c). Developing A Platform to Facilitate Sharing Spatial Data. *Coordinates* Vol. 1(7), 30-32,

Rajabifard, A., Feeny, M.-E., and Williamson, I. (2002). Future Directions for SDI Development. *International Journal of Applied Earth Observation and Geoinformation* 4 (1), 11-22.

Rajabifard, A., and Williamson. I. (2001). *Spatial Data Infrastructures: Concept, SDI Hierarchy and Future directions*, Tehran, Iran.

Rajabifard, A., and Williamson, I. (2003). *Anticipating the Cultural Aspects of Sharing for SDI Development*. SSC, Canberra.

Rajabifard, A., and Willamson, I. (2004a). Regional SDI Development. *Journal of Geospatial Today* 2,

Rajabifard, A., and Williamson, I. (2004b). *The Integration of Built and Natural Environmental Datasets in National Spatial Data Infrastructure Initiatives*, New York.

Robertson, A. (2004). *Department of Land Information Project Participnat Response - An internal report compiled for Project 3.1 CRC-SI*. CRC-SI, Melbourne University, Melbourne.

Ronsdorf, C. (2005). Positional Accuracy and Integration of Geograhic Data. *Coordinates* September 2005,

Ryttersgaard, J. (2001). *Spatial Data Infrastructure-Developing Trends and Challenges*, Nairobi, Kenya.

Samadzadegan, F. (2004). *DATA INTEGRATION RELATED TO SENSORS, DATA AND MODELS*. ISPRS, Istanbul, Turkey.

SCHEMAS. (2006). European Forum for Metadata Schema Implementers. Accessed: 9th November. 2006. http://www.schemas-forum.org/metadata-watch/third/section5.7.html

Scott, K. (2007). PSMA Australia: URISA 2007 ESIG Award Application For LYNX. Accessed: 12 August. 2008. http://www.urisa.org/files/PSMA_ESIG.pdf

SEDAC. (2006). Metadata Review. Accessed: 6th October 2006. 2006. http://sedac.ciesin.org/metadata/overview.html

Sen, S. (2005). Semantic Interoperability of Geographic Information. *GIS Development* 9, 18-21.

Sharwood, S. (2005). *Emergencies put mapping on the agenda*, Sydney.

Shedroff, N. (2001). *An Overview of Understanding*. Hayden/Que, Indianapolis.

Sheth, A. P. (1998). Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. In *Interoperating Geographic Information Systems* (Goodchild, M. F., Egenhofer, M. J., Fegeas, T., and Kottman, C. A., eds.). Kluwer.

SII. (2005). *Product Description - VICMAP TRANSPORT*. Spatial Information Infrastructure, Strategic Policy and Projects, Department of Sustainability and Environment.

SII. (2008). Government information and services about land and property. Accessed: 1 November. 2008. http://www.land.vic.gov.au/land/lcnlc2.nsf/Home+Page/Land+Channel~Home+Page?open

Spyns, P., Meersman, R., and Jarrar, M. (2002). Data Modeling versus Ontology Engineering. *SIGMOD Record 31(4):12-17. ISSN: 01635808.*

Stevens, P., and Pooley, R. (2006). *Using UML, Software Engineering with Objects and Components* (Second ed.). Addison Wesley, Essex.

Sundaresan, S., and Hu, G. (2005). "Schema Integration of Distributed Databases Using Hyper-Graph Data Model." Paper presented at the IRI -2005 IEEE International Conference on Information Reuse and Integration, 15-17 Aug. 2005, 2005.

Teorey, T. J. (2006). *Database Modeling and Design: Logical Design* (4th ed.). Morgan Kaufmann.

Uitermark, H. T. (2001). ONTOLOGY-BASED GEOGRAPHIC DATA SET INTEGRATION, Twente University, Deventer.

Ulubay, A., and Altan, M. O. (2002). "A Different Approach to the Spatial Data Integration." Paper presented at the Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, 2002.

UNRCC-AP. (1997). *Resolution of the 14th UNRCC-AP*. Nations, U., http://www.gsi.go.jp/PCGIAP/unrcc/feb97.htm.

UNRCC-AP. (2003). *Resolution of 16th UNRCC-AP*, Okinawa.

UNRCC-AP. (2006). Seventeenth UN Regional Cartographic Conference for Asia and the Pacific, Bangkok, Thailand, 18 - 22 September, 2006. Accessed: 7th November. 2007. http://unstats.un.org/unsd/geoinfo/17thunrccapdocuments.htm

Uschold, M., and Gruninger, M. (1996). Ontologies: Principles, Methods and Applications. *Knowledge Enginnering Review* 11, 93-156.

Usery, E. L., Finn, M. P., and Starbuck, M. (2005). *Integrating Data Layers to Support the National Map of the United States*, A Corua, Spain.

Van Loenen, B. (2003). *The impact of access policies on the development of a national GDI*, Munster.

Van Loenen, B. (2005). The role of information policies in the development of geographical information infrastructure, Delft University of Technology, Delft University of Technology, Delft.

Van Loenen, B. (2006). Developing geographic information infrastructures; the role of information policies, Delft University of Technology. Delft: DUP Science.

Van Loenen, B., and De Jong, J. (2007). Institutions Matter: The Impact of Institutional Choices Relative to Access Policy and Data Quality on the Development of Geographic Information Infrastructures. In *Research and Theory in Advancing Spatial Data Infrastructure Concepts* (Onsrud, H., ed, pp. 215-229. ESRI press, Redlands.

Vector1media. (2008). PSMA Australia purchases 1Spatial s Radius Studio. Accessed: 20 August. 2008. http://www.vector1media.com/top-stories/corporate-news/psma-australia-purchases-1spatial%92s-radius-studio/

Visser, U., Stuckenschmidt, H., Schuster, G., and Vogele, T. (2002). Ontologies for Geographic Information Processing. *Computers and Geosciences* 28, pp: 103-117.

VSC. (2004). *VSC's Strategic Plan*. Victorian Spatial Council,

VSC. (2007). *Developing the Victorian Spatial Information Strategy 2008-2010*. Victorian Spatial Council.

VSIS. (2005). Spatial Information Business Information Guidelines for Victoria. Accessed: 10 May. 2007. http://www.land.vic.gov.au/CA256F310024B628/0/16C76A8453D11382CA2572D80018D4B8/$File/VSIS+2004-2007+Business+Information+Guidelines+May+2007.pdf

WALIS. (2008). Spatial is special! Accessed: 10 October. 2008. http://www.walis.wa.gov.au/education/stis/project/spatial_is_special

Wallace, J., Williamson, I., Rajabifard, A., and Binns, A. (2006). Spatial Information Opportunities for Government. *Spatial Science Journal* 51 (1),

Walsh, N. (2008). A Technical Introduction to XML. Accessed: 4 November. 2008. http://xml.com/pub/a/98/10/guide0.html?page=2

Warnest, M. (2005). A collaboration model for national spatial data infrastructure in federated countries, Department of Geomatics, Melbourne, Melbourne.

Weaver, B. (2004). *Implementing of the National Map Road Database*, Nashville, Tennessee, The United States.

Webopedia. (2008). What is Interoperability? Accessed: 12 February. 2008. http://www.webopedia.com/TERM/I/interoperability.html

Wikipedia. (2000). Visual Basic for Applications. Accessed: 6 September. 2008. http://en.wikipedia.org/wiki/Visual_Basic_for_Applications

Wikipedia. (2003). Extensible Markup Language (XML). Accessed: 1 Novembver. 2008. http://en.wikipedia.org/wiki/Xml_parser#Processing_XML_files

Wikipedia. (2007). Unified Modeling Language. Accessed: 4 November. 2008. http://en.wikipedia.org/wiki/Unified_Modeling_Language

Wikipedia. (2008). Spatial Data Infrastructure. Accessed: 10 June. 2008. http://en.wikipedia.org/wiki/Spatial_Data_Infrastructure

Williamson, I. P. (2004). "Building SDIs - The Challenges Ahead." Paper presented at the Proceedings of the 7th GSDI Conference, 2-6 February, Bangalore, India., 2004.

Williamson, I. P. (2006). Is Spatial Special? *Position* 17.

Williamson, I. P. (2007). *Spatially Enabled Government*. University of Melbourne, Melbourne, Australia.

Williamson, I. P., Chan, T.O. and Effenberg, W.W. (1998). Development of Spatial Data Infrastructrures - Lessons Learned from the Australian Digital Cadastral Databases. *GEOMATICA* 52, 177-187.

Williamson, I. P., Fauzi, A., and Rajabifard, A. (2007). *Report of Working Group 3 Meeting to the PCGIAP*, Seol, Korea.

Williamson, I. P., Rajabifard, A., and Feeney, M.-E. F. (2003). *Developing Spatial Data Infrastructures: From Concept to Reality*. Taylor and Francis, London.

Williamson, I. P., Rajabifard, A., and Binns, A. (2006). Challenges and issues for SDI Development. *International Journal of SDI Research* 1, 24-35.

Zaslavsky, I., Marciano, R., Gupta, A. and Baru, C. (2000). XML-based Spatial Data Mediation Infrastructure for Global Interoperability, Proceedings of the 4th Global Spatial Data Infrastructure Conference, Cape Town, South Africa, March 2000.

Zaslavsky, I., He, H., Tran, J., Martone, M. E., and Gupta, A. (2004). "Integrating Brain Data Spatially: Spatial Data Infrastructure and Atlas Environment for Online Federation and Analysis of Brain Images." Paper presented at the 15th International Workshop on Database and Expert Systems Applications, 2004.

Zaslavsky, I., Memon, A., and Memon, G. (2005). Integration across heterogeneous spatial data and applications Cyberinfrastructure Project. *GIS Development* 9, 28-31.

Appendix 1

Publications

ISI Journals

Mohammadi, H., Rajabifard, A., and Williamson, I. (2009). Development of an Interoperable Tool to Facilitate Spatial Data Integration in the Context of SDI, International Journal of GIS (In Press)

Mohammadi, H., Rajabifard, A., and Williamson, I. (2009). Spatial Data Integration Geo-Web Service, World Applied Science Journal (In Press)

Book Chapters:

Mohammadi, H. (2007). Spatial Data Integration: a Necessity for Spatially Enabling Government. In *Towards a Spatially Enabled Society* (Rajabifard, A., ed,) pp. 333-341. Melbourne University Press, Melbourne.

Mohammadi, H., Rajabifard, A., Binns, A., and Williamson, I. (2008). Spatial Data Integrability and Interoperability in the Context of SDI. In *The European Information Society – Taking Geoinformaiton Science One Step Further* (Bernard, L., Friis-Christensen, A. and Pundt H., ed,) pp. 401-413. Springer, Berlin.

Refereed Papers:

Mohammadi, H., Rajabifard, A., and Williamson, I. (2009). Enabling Spatial Data Sharing through Multi-source Spatial Data Integration, GSDI Conference, Rotterdam, The Netherlands, 15-19 June 2009.

Mohammadi, H., Rajabifard, A., Binns, A., and Williamson, I. (2007). *Spatial Data Integration Challenges: Australian Case Studies*, Spatial Science Conference May 2007, Hobart, Australia.

Other Publications:

Mohammadi, H., Rajabifard, A., Binns, A., and Williamson, I. (2006). *The Development of a Framework and Associated Tools for the Integration of Multi-Sourced Spatial Datasets*, 17^{th} UNRCC-AP, Bangkok, Thailand.

Mohammadi, H., Rajabifard, A., Binns, A., and Willaimson, I. (2006). Bridging SDI Design Gaps to Facilitate Multi-source Data Integration. *Coordinates* 2, 26-29.

Mohammadi, H., Rajabifard, A., and Willaimson, I. (2008). SDI as Holistic Framework. *GIM International* 22 (1).Mohammadi, H., Rajabifard, A., and Willaimson, I. (2008). *The Integration of Multi-source Spatial Data.* International Seminar on Land Administration Trends and Issues in the Asia and Pacific Region, 19-20 August 2008, Kuala Lumpur, Malaysia.

Appendix 2

Permission from PCGIAP-WG3 for Inclusion the International Case Study Country Reports in Research

From: Ian Philip Williamson <ianpw@unimelb.edu.au>
Subject: RE: Ask for permission to use PCGIAP-WG3's country reports for inclusion in research project
To: Hossein Mohammadi <h.mohammadi@pgrad.unimelb.edu.au>
Cc: PCGIAP Secretariat <sec@pcgiap.org>

Dear Hossein,

Approval granted. I acknowledge that this was related to the work of PCGIAP and agreed by the WG.

Ian Williamson
Chair, WG3 (Spatially Enabled Government), PCGIAP

From: Hossein Mohammadi [mailto:h.mohammadi@pgrad.unimelb.edu.au]
Sent: Wednesday, 3 December 2008 11:11 AM
To: Ian Philip Williamson
Subject: Ask for permission to use PCGIAP-WG3's country reports for inclusion in research project

Dear Professor Williamson,

I am writing to you to ask for permission to use the PCGIAP's Working Group3's member countries' technical reports and material for inclusion in the research project on the integration of multi-source spatial data integration.

Thank you for your time and consideration.
I look forward to your reply.

Yours Faithfully,

Hossein Mohammadi

Hossein Mohammadi
PhD Candidate
Center for SDI and Land Administration
Geomatics Department, The University of Melbourne
Parkville, 3010, Victoria, AUSTRALIA
Tel: ++61 3 8344 9696, Fax: ++61 3 9347 2916
Mobile: ++61 425 80 7993
email: hosseinm@pgrad.unimelb.edu.au

Appendix 3

International Case Study Integration Template

PCGIAP-Working Group 3

and

The University of Melbourne

International Workshop
On

Integration of Built and Natural Environmental Datasets within a National SDI

in conjunction with the 17th UNRCC-AP and PCGIAP Meeting

20th September 2006
Bangkok, Thailand

An objective of Working Group 3 (WG3) of the UN sponsored "Permanent Committee on GIS Infrastructure for Asia and the Pacific" (PCGIAP) for 2005-2007 is the development of a framework and associated tools to facilitate the integration of built (cadastral) and natural (topographic) environmental datasets within a National SDI initiative.

PCGIAP, through WG3, with the support of the Centre for Spatial Data Infrastructures and Land Administration, the University of Melbourne, is to hold a dedicated workshop **(20th September 2006)** for the discussion of the integration of built and natural environmental datasets within a National SDI during the 17th UNRCC-AP Conference from **18-22 September 2006** in Bangkok Thailand. The Workshop will review the national administration of SDI and data integration within countries in Asia and the Pacific region based on a common template to identify problems, issues, similarities and differences in spatial data infrastructures; institutional arrangements; current data integration methods; technology and human resource and capacity building in data integration.

The attached template is a guide to assist member nations to prepare a country report of their National SDI and data integration issues and activities (used for presentation at the Workshop). Could you please return the completed template by the **15th August 2006**. Completed templates will aid in creating an integration framework and associated tools.

This template will aid the research team to better understand and describe:

- *History of integration of built and natural environmental datasets and related National SDI initiatives.*
- *Capacity for and policies relating to data integration of cadastral and topographic datasets.*
- *Institutional support for and barriers against data integration of cadastral and topographic datasets.*
- *The technical, jurisdictional, institutional, legal and land policy perspective surrounding cadastral and topographic datasets in a National SDI.*

- *Other countries experiences and initiatives in integrating data in order to identify best practice.*

This will lead to the development of a model and framework for integration of these two forms of data capable of being used in diverse jurisdictions in support of sustainable development.

Country/state:	…………………………………………
Name of contact person:	…………………………………………
Affiliation, Organization:	…………………………………………
Function, Position:	…………………………………………
Address:	…………………………………………
Email address: **Tel, Fax**	………………………………………….

Integration of Built and Natural Environmental Datasets within National SDIs

Project Overview:

Sustainable development and meeting "the triple bottom line" (economic, social and environmental objectives) requires an understanding of the natural and built landscape in order to observe and monitor change and to create realistic simulations of the evolving environment. This requires access to both built and natural environmental datasets. Over the last decade these needs are being addressed by establishing spatial data infrastructures (SDI) where one of the key objectives is the integration of these datasets, and specifically cadastral (built) and topographic (natural) spatial data. The drive to establish SDIs is also driven by a need for governments and businesses to improve their decision-making and increase efficiency (Gore, 1998), as well as the advent of accessible, powerful information and communications technologies.

Amongst spatial data, cadastral and topographic datasets are the most important for describing the built and natural environment. These datasets are the ‘foundation data’ (Groot and MacLaughlin, 2000) in modern market economies. Cadastral datasets are the accumulation of individual property boundary surveys undertaken by land surveyors. By nature, cadastral data is very different to topographic data that is produced at medium to small scales over large regions using various techniques.

In all countries, these foundation datasets were developed to serve different purposes and are usually managed separately. This separation is recognised as a barrier to implementation of sustainable development. Duplication imposes unjustifiable costs on data collection and maintenance. The datasets should adopt the same overarching philosophy and data model to achieve multi-purpose data integration, both vertically and horizontally (Ryttersgaard, 2001). Merging of these datasets at a local level has been achieved to some degree, however, attempts to integrate the datasets at a national level, even where SDIs are well developed, has been difficult and problematic internationally.

Data Integration:

Spatial data integration is being done in most spatial services to some extent at different levels. Users of spatial data gather data from data providers and integrate them to meet their needs. Social behaviors, legal considerations, institutional arrangements and policy issues which are attached to datasets differs across different providers and makes integration problematic both from a technical and non-technical perspective.

From a technical point of view, some applications superimpose data layers geometrically in order to analyse and monitor them against each other without the establishment of any interrelationship amongst features and layers, while some other services integrate datasets based on topology relations between features or based on relationship between feature classes or attributes through data models.

The integration of multi-source datasets is not only the match of datasets geometrically, topologically, and having a correspondence of attribute, but also providing all social, legal, institutional and policy mechanism together with technical tools to facilitate the integration of multi-source datasets.

I. Country Report

Briefly describe the national administration of spatial data and data integration using the following structure. We would like you to fill out each of the four topics A-D. Section A is generic and answers may be similar to those provided for the cadastral template. Sections B-D however are focused on National SDI and Data Integration.

As we are trying to collect comparable information, we ask you to leave the headings as they are and fill out the empty space provided below. If you feel that you need further headings, feel free to add them for your own purpose. We have provided some further information in italics to help fill in each section. Please complete to the best of your knowledge – something is better than nothing.

A. Country Context

Geographical and Historical Context

Description of the basic geographic context, i.e. population, size of country, etc. as well as other outstanding geographic features. Description of the country's history in terms of relevant periods, e.g. colonization, and political development.

→

Current Political and Administrative Structures

Description of the current political and administrative structures, such as the political system, number of states or provinces, etc. and how this may affect efforts to integrate spatial data.

→

B. National SDI Context

History and Status of National SDI Initiative

Description of the origins and the development of National SDI initiatives.

For example, is the National SDI initiative based on collaboration or legislation? Are the major SDI activities occurring at a National government level or Sub-national government level? Include information on the development of SDIs at all levels within your country, eg. State SDI, Local SDI, involvement in Regional SDI etc.

→

Have Core Datasets been defined within the SDI structure?

Core datasets, sometimes called reference data, are the basic data that everyone involved with spatial information uses. For example, what are the core datasets? Have custodianship guidelines concerning data maintenance and control been created for the core datasets?

→

Describe the data acquisition and access mechanisms within the SDI.

For example, is there an effective clearinghouse or portal for sharing of spatial data? Are there effective partnership arrangements in place to share data? Are pricing, licensing, reproduction principles etc. defined to help govern data access? Do these principles govern all datasets or just framework datasets? At what level do they apply eg. national/state/local? Are there standardized, frequent and documented update cycles for spatial data?

→

If your country is not developing a National SDI, are there any plans on developing an SDI in the future?

→

Historical Outline of Built and Natural Environmental Data Development

Please describe the origins and development of both cadastral and topographic data, along with other built and natural environmental datasets..

Include names of agencies that have been involved in cadastral and topographic data development and management. Have they been developed separately from different backgrounds eg. Cadastral from a property perspective, topographic from a GIS environmental perspective?

→

Current Administration of Built and Natural Environmental Data

Please describe current institutional and management arrangements that govern cadastral and topographic data.

Include names of organisations, agencies and government departments that are involved in cadastral and topographic data at various political levels (eg. local, state, national, regional) and within various administrative areas.

→

Please describe current institutional and management arrangements that other built and natural environmental data.

→

Please describe the metadata arrangements for built and natural environmental datasets

For example, are there effective and consistent Metadata management tools in place for built and natural environmental datasets? Are these well documented and accessible? Are the arrangements standardized across all datasets at all jurisdictional levels? Is metadata searchable?

→

Please describe the data format or conceptual model for built and environmental datasets, especially topographic and cadastral datasets.

For example, is there a common data model for cadastral data at all jurisdictional levels? Is there a common topographic data model? Are data models interoperable? Is there a conceptual model developed in order to better understand and define the relationship and hierarchical structure of topographic and cadastral data and encourage data integration?

→

C. Institutional Framework for Integration – Data Provider

Please provide information on how spatial information is managed from a data provider perspective

For example, is the private sector involved in the management and administration of built and natural environmental datasets? If yes, please explain how. Description of the role that legislation and other instruments of governance play in the administration of built and natural environmental datasets. How do agencies interact? Flow of spatial data and relationships within and between agencies.

→

Please describe the tools that are used to manage spatial information.

What are the tools, for example,. modeling tool, software, etc. Are tools used accessible, open source, useable, documented, user-friendly? Are generally acknowledged interoperability standards used eg. ISO 9001, OGC?

→

At what scale do you produce built and natural environmental datasets such as cadastre and topo?

→

D. Institutional Framework for Integration – Data User

Please describe the major data uses.

List and describe the major data users (including private sector, academia, public sector etc) and their most commonly used data layers.

→

Describe current services and products that are available to data users and customers.

For example, are these services and data integratable? Are there any services that utilize integrated data for applications? How do data providers support customers (single dataset or integrated product delivery)? Do issues such as pricing, intellectual property and privacy detract from your ability to create integrated products and services?

→

E. Issues in the Integration of Built and Natural Environmental Datasets

Need for Integration

For example, is data integration a priority for your jurisdiction? What benefits do you gain from the integration of built and natural environmental datasets? What are the drivers for integration in your jurisdiction eg. environmental protection, hazard management, sustainability, counter terrorism etc?

→

Major Issues in attempting to integrate built and natural environmental datasets

What are the major issues hindering the ability to integrate multi-source datasets on a national level?

Include any perceived issues and barriers which your country may face in attempting to integrate these types of data, especially within the context of a National SDI. Do cross-jurisdictional issues play a role within a National context? Are issues of a technical, social, institutional or policy perspective?

→

Please list and describe the outcomes of any attempts to integrate built and natural environmental datasets at a national or state type level.

Where these one off projects or an attempt to create a long term, integration solution? Who was the driving force behind such integration initiatives (eg. land administration, environmental *management etc). What was the end result (eg. common data model, integration framework, tools etc). Where some of the issues mentioned in the precious section solved? If yes, how?*

→

II Questionnaire

Please answer to the best of your ability.

1.0 Policy Principles:

What are the current issues that policy needs to address within your jurisdiction in regards to spatial information?

Has your jurisdiction developed any policies on data integration? If so, what is the capacity for and policies relating to data integration within your jurisdiction?

How much of the spatial information policy and initiatives is user driven? Please give examples of user driven projects or initiatives.

Are there policies or guidelines in place for the following aspects of spatial data? If yes, at what level do these policies apply (please tick National, State/Provincial or Local)?

Policy	National Level	State/Provincial Level	Local Level
1. Management	[ ]	[ ]	[ ]
2. Data Model	[ ]	[ ]	[ ]
3. Metadata	[ ]	[ ]	[ ]
4. Custodianship	[ ]	[ ]	[ ]
5. Pricing	[ ]	[ ]	[ ]
6. Access	[ ]	[ ]	[ ]
7. Distribution	[ ]	[ ]	[ ]
8. Privacy	[ ]	[ ]	[ ]
9. Security	[ ]	[ ]	[ ]
10. Purchasing, Procurement	[ ]	[ ]	[ ]

2.0 Institutional Principles

What are the current institutional issues hindering the integration of data within your jurisdiction? Please tick level of importance.

Issue	Very Important (5)	Important (4)	Neither (3)	Not Very Important (2)	Not Important At All (1)
1. Funding	[ ]	[ ]	[ ]	[ ]	[ ]
2. Collaboration	[ ]	[ ]	[ ]	[ ]	[ ]
3. Awareness of data existence	[ ]	[ ]	[ ]	[ ]	[ ]
4. Licensing	[ ]	[ ]	[ ]	[ ]	[ ]
5. Data Access	[ ]	[ ]	[ ]	[ ]	[ ]
6. Other………………………….	[ ]	[ ]	[ ]	[ ]	[ ]
7. Other………………………….	[ ]	[ ]	[ ]	[ ]	[ ]

How is spatial information accessed within your jurisdiction? Please tick if access mechanism is the primary method of accessing data (used constantly), secondary method (used occasionally) or not used at all.

Access Mechanism	Primary	Secondary	Not Used
1. Paper maps (tourist maps, detail maps, charts)	[ ]	[ ]	[ ]
2. Directory (ie. Street Directory), set of indexed maps	[ ]	[ ]	[ ]
3. CDROM or other portable (digital) medium	[ ]	[ ]	[ ]
4. Email (attached file)	[ ]	[ ]	[ ]
5. Online, Internet (ie. Data Directory / Data Atlas / Map Viewer)	[ ]	[ ]	[ ]
6. Local Area Network (LAN), Wide Area Network (WAN), other communication network	[ ]	[ ]	[ ]
7. Other………………………………..	………..▶	………..▶	………..▶

How is spatial information managed within your jurisdiction?

SI Management	Please Tick
1. Centralised (National Government)	[ ]
2. Decentralised (State/Provincial/Local)	[ ]
3. Other……………………………….	[ ]

How are spatial data initiatives funded within your jurisdiction?

Funded By	Please Tick
1. Government (public sector)	[ ]
2. Cost recovery of data	[ ]
3. Private sector	
4. Public/private partnership	[ ]
5. Other………………………………	[ ]

3.0 Technical Principles:

What are the current technical issues hindering the integration of data within your jurisdiction? Please tick level of importance.

Issue	Very Important (5)	Important (4)	Neither (3)	Not Very Important (2)	Not Important At All (1)
1. Computational heterogeneity	[ ]	[ ]	[ ]	[ ]	[ ]
2. Vertical topology	[ ]	[ ]	[ ]	[ ]	[ ]
3. Reference system	[ ]	[ ]	[ ]	[ ]	[ ]
4. Data quality	[ ]	[ ]	[ ]	[ ]	[ ]
5. Metadata	[ ]	[ ]	[ ]	[ ]	[ ]
6. Data format	[ ]	[ ]	[ ]	[ ]	[ ]
7. Other………………………………..	[ ]	[ ]	[ ]	[ ]	[ ]
8. Other………………………………..	[ ]	[ ]	[ ]	[ ]	[ ]

What standards organisations or bodies for spatial data does your jurisdiction subscribe, member or adhere to?

Standard	Please Tick
1. International Standards Organisation ISO, Technical Committee for Geographic Information / Geomatics - TC 211	[ ]
2. National Standards Committee or Body	[ ]
3. Open GIS Consortium OGC	[ ]
4. World Wide Web Consortium W3C	[ ]
5. Other	…………▶

4.0 Legal Principles:

What are the current legal issues hindering the integration of data within your jurisdiction? Please rank.

Issue	Very Important (5)	Important (4)	Neither (3)	Not Very Important (2)	Not Important At All (1)
1. Copyright	[ ]	[ ]	[ ]	[ ]	[ ]
2. Intellectual property	[ ]	[ ]	[ ]	[ ]	[ ]
3. Data access	[ ]	[ ]	[ ]	[ ]	[ ]
4. Privacy	[ ]	[ ]	[ ]	[ ]	[ ]
5. Data licensing	[ ]	[ ]	[ ]	[ ]	[ ]
6. Other………………………………..	[ ]	[ ]	[ ]	[ ]	[ ]
7. Other………………………………..	[ ]	[ ]	[ ]	[ ]	[ ]

5.0 Social Principles:

Who are the major users of cadastral and topographic data within your jurisdiction?

Data User (Cadastral)	Major User	User	Sporadic User	Not a User
1. Government – technician	[ ]	[ ]	[ ]	[ ]
2. Government – manager	[ ]	[ ]	[ ]	[ ]
3. Private sector	[ ]	[ ]	[ ]	[ ]
4. Academia	[ ]	[ ]	[ ]	[ ]
5. Military	[ ]	[ ]	[ ]	[ ]
6. Other………………………………..	[ ]	[ ]	[ ]	[ ]

Data User (Topographic)	Major User	User	Sporadic User	Not a User
1. Government - technician	[ ]	[ ]	[ ]	[ ]
2. Government – manager	[ ]	[ ]	[ ]	[ ]
3. Private sector	[ ]	[ ]	[ ]	[ ]
4. Academia	[ ]	[ ]	[ ]	[ ]
5. Military	[ ]	[ ]	[ ]	[ ]
6. Other………………………………..	[ ]	[ ]	[ ]	[ ]

What capacity building initiatives are currently underway within your jurisdiction in regards to spatial information development?

6.0 Spatial Data Infrastructure

What are the main data sets available within your National SDI? Please list.

Data Set	Please Tick
1. Geodetic reference, survey network	[ ]
2. Cadastral, ownership, property boundary information	[ ]
3. Topography	[ ]
4. Land use, zoning, planning	[ ]
5. Native title	[ ]
6. Road networks, road centre-line data	[ ]
7. Utilities and essential services infrastructure	[ ]
8. Transportation	[ ]
9. Geographic names, localities and administrative boundaries	[ ]
10. Street Address	[ ]
11. Aerial or Satellite Imagery	[ ]
12. Elevation and Bathymetry	[ ]
13. Hydrology	[ ]
14. Vegetation	[ ]
15. Forestry	[ ]
16. Mineral resources	[ ]
17. Agriculture	[ ]
18. Environment	[ ]
19. Other	…………▶

What is the cost for data available through the SDI?

Cost	Please Tick
1. Free (open access)	[ ]
2. Cost of transferring the data	[ ]
3. Full cost recovery (cost of transfer plus cost of creation, updating etc)	[ ]
4. Other……………………………………	[ ]