

Spatial Data Integration

Hossein Mohammadi¹, Andrew Binns², Abbas Rajabifard³, Ian P. Williamson⁴
Centre for Spatial Data Infrastructures and Land Administration
Department of Geomatics
University of Melbourne

- 1: hosseinm@pgrad.unimelb.edu.au
- 2: a.binns@unimelb.edu.au
- 3: abbas.r@unimelb.edu.au
- 4: ianpw@unimelb.edu.au

Keywords: Multi-sourced Spatial Data, Integration, Interoperability, SDI

Introduction

Today's world relies heavily on spatial data to manage the natural and built environments. Monitoring and controlling the impact of human activities on the environment and the impact of natural environmental changes (including natural hazards) on the built environment is a major policy driver for many jurisdictions. The mitigation of natural hazards including tsunamis, earthquakes and landslides and securing citizens against them has become another priority of nations, especially after the tsunami within the Asia/Pacific region. September 11 was also a major point in history which attracts attentions towards providing more efficient counter-terrorism initiatives.

In many countries meeting sustainable development objectives including social cohesion and environment protection together with economical growth is the most overwhelming policy driver. In these countries most of legislations and decisions are made to meet sustainable development objectives.

All of the above mentioned activities try to control the natural and built environments and monitor the impacts of one environment on the other one. To control and monitor built and natural environments, the components of these environments need to be integrated to provide a holistic factual model of the real world. Effective access and use of spatial data has been addressed by developing SDIs (Spatial Data Infrastructures) which one of its objectives is to address and provide requirements for effective data integration.

The effective integration of built and natural environmental datasets is an ultimate goal of many spatial decision making systems which has not been fully achieved, however technical integration and interoperability of multi-sourced spatial data have received much attention. The integration of multi-sourced spatial data due to the diversity of data providers needs more than technical tools and considerations. Institutional, social, legal and policy requirements must also be taken into consideration in order to achieve effective integration.

This paper aims to highlight and discuss technical issues together with legal, social, institutional and policy issues associated with the technical integration of datasets. This paper also will address the findings and observations of some case studies which have been conducted in order to investigate the problems of effective data integration.

A case study investigation has been done on datasets from two Australia's states (Victoria and New South Wales) together with data from Australia's national mapping agency (Geoscience Australia). Datasets from these jurisdictions have been investigated against each other and problems and findings have been highlighted in this paper. These jurisdictions' mapping agencies have been visited and together with the investigation results of six other international countries within Asia/Pacific region which are currently being undertaken, will be incorporated into designing technical tools for more effective integration including a

framework, an integration data model and a web-service. In designing these tools, marginal non-technical issues will also be taken into consideration and non-technical issues will also be discussed. However the main focus of the paper is on technical issues.

Spatial Data Integration Issues

Spatial data integration is being done in most of 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 behaviours, legal considerations, institutional arrangements and policy issues which are attached to datasets differ across different providers and makes integration problematic both from technical and non-technical perspectives.

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-sourced datasets is not only the match of datasets geometrically, topologically, and having a correspondence of attribute (Usery *et al.* 2005), but also providing all social, legal, institutional and policy mechanisms together with technical tools to facilitate the integration of multi-sourced datasets. Figure 1 shows technical data integration and surrounding inconsistency issues.

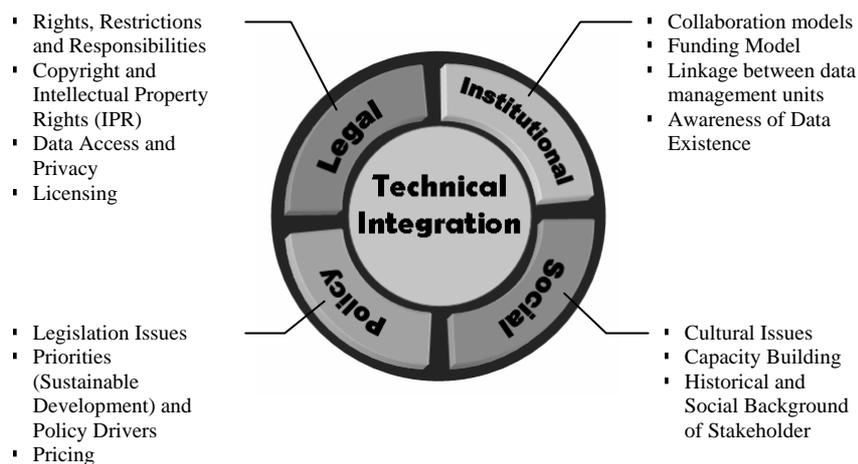


Figure 1. Technical integration in the context of non-technical considerations

These inconsistencies have been categorized in different classes due to the source of their problems. These issues are caused by technical heterogeneity, institutional structure, policy issues, legal concerns and social effects of the integration.

The most prominent technical issues recognised so far include: standards, interoperability, vertical topology, semantic, reference system, data model, metadata, format, and data quality. Technical issues have also non-technical considerations. Some important ones have been outlined below.

The non-technical phases provide the basis and requirement for the technical integration and technical integration is assessed in the context of non-technical environments. Non-technical stage also consists of the provision of technical requirements to facilitate integration such as arrangements of a single standard of data model amongst a jurisdiction, development of intermediate standards and solutions to facilitate the interaction between different systems, raising awareness of existing

spatial data through comprehensive data catalogues and spatial data search engines; development of collaboration models and so forth. This stage is mostly handled by managers, strategy- and policy-makers in interaction with other layers of spatial data stakeholders. On the other hand, the technical integration stage mostly deals with practical conflation of datasets and technical people and end users are the most significant players of this scene (Figure 2).

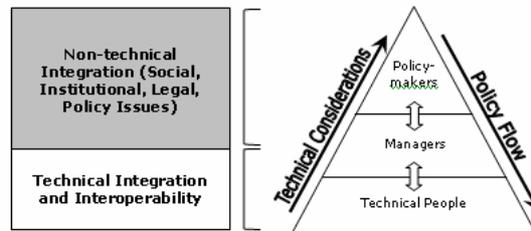


Figure 2. Different levels of data integration

The collaboration between stakeholders, business models together with funding models, making users aware of data and, finally, data management approaches are key issues within the institutional arrangement of SDI.

On the policy side policy drivers, priorities of the nations, pricing and legislation have been found the most major issues.

From a legal perspective, the following issues are prominent:

- Rights, Restrictions and Responsibilities (RRR)
- Copyright and intellectual property rights (IPR),
- Data access and privacy, and
- Licensing

These issues may differ in jurisdictions and even in different levels of a single jurisdiction and are also very challenging.

Table 1. Integration issues

Technical Issues	Institutional Issues	Policy Issues	Legal Issues	Social Issues
<ul style="list-style-type: none"> ▪ Computational Heterogeneity (Standards and Interoperability) ▪ Vertical Topology ▪ Semantic ▪ Reference System and Scale ▪ Data Quality ▪ Data model ▪ Metadata ▪ Format 	<ul style="list-style-type: none"> ▪ Collaboration models ▪ Funding Model ▪ Linkage between data management units ▪ Awareness of Data Existence 	<ul style="list-style-type: none"> ▪ Legislation Issues ▪ Priorities (Sustainable Development) and Policy Drivers ▪ Pricing 	<ul style="list-style-type: none"> ▪ Rights, Restrictions and Responsibilities ▪ Copyright and Intellectual Property Rights (IPR) ▪ Data Access and Privacy ▪ Licensing 	<ul style="list-style-type: none"> ▪ Cultural Issues ▪ Capacity Building ▪ Background of Stakeholder

Cultural differences, capacity building and equal access to spatial data are also paramount in the social category. Table 1 summarizes all issues surrounding data integration. These categories show distinct items, however it must be remembered

that they also have effects on each other and on SDI components including data, policy, network, standards and people as has been illustrated in Figure 3.



Figure 3. Components of SDI (Rajabifard 2001)

An SDI is an enabling platform to provide the users of spatial data with on-demand, correct, current, and usable spatial data. Fitness-to-use refers to the capability of the data to be linked properly to other datasets in order to align with the requirements of the applications and decision support systems. As a consequence, one of the key objectives of SDI is to facilitate the proper integration of multi-source datasets.

Mentioned non-technical issues hinder effective data integration; however in the following section the technical side is focused and articulated.

The Technical Issues of Data Integration

There is a great body of literature highlighting different aspects of multi-source data integration. They have highlighted the heterogeneity and inconsistency of the spatial information (SI) and SI systems and most of them have attempted to address these impediments by documenting the technical inconsistencies. Nevertheless, in many cases the technical inconsistency arises from non-technical problems. Technical integration mostly occurs as a result of other marginal issues; belonging to social, institutional, legal and political inconsistencies of different custodians and relevant organizations.

Over the last decade these needs have attempted to be overcome by establishing SDI where one of the key objectives is the integration of multi-source datasets, specifically cadastral (built) and topographic (natural) spatial data (Rajabifard and Williamson 2004). This need has been reported in many publications and international declarations such as the Bogor declaration (FIG 1996) which urges the establishment of National SDIs to ensure uniform data integration. Despite the new development in SDI, there are still difficulties in designing appropriate spatial data infrastructures to integrate built and natural spatial data (Parker and Enemark 2005) and many jurisdictions face many difficulties in integrating cadastral and topographic spatial data (UNRCC-AP 2003). SDI establishment should take a further step by investigating issues in the integration of cadastral and topographic data sets (UNRCC-AP 1997) .

From a technical perspective there are plenty of publications which have investigated the technical issues connected to data integration. Fonseca *et al.* (2002) maintain two aspects for technical heterogeneity of spatial data and services which are semantic and computational heterogeneities.

Hakimpour (2003) believes that there are four categories of heterogeneity between interoperating systems including conceptual, spatial schema and semantic heterogeneities. These categories have been clarified by the following example. Integrating two road datasets may cause any of the following potential heterogeneity problems:

1. Heterogeneity in the *conceptual modelling*: One system represents a street as an object class and the other as a relation.
2. Heterogeneity in the *spatial modelling*: Streets can be represented by polygons (or a segment of pixels) in one system, while being represented by lines in the second system.
3. *Structure or schema heterogeneity*: Both systems hold the name of a street whilst one keeps information about the sidewalk and the other one the width. Or, lines representing the street in one system use the DXF format while in the second system they use the IGES format.
4. *Semantic heterogeneity*: One system may consider only the paved part as a street, while the other considers the paved part and the sidewalk as a street. One may define a street as a paved way used by automobiles with sidewalk and the other defines street as any type of way used by automobile inside a residential area.

Simpson (2002) highlights some technical disparities including scale, resolution, compilation standards, source accuracy, registration, sensor characteristics, currency, temporality, or errors. Donaubauer (2005) addresses one of the major international challenges of building SDIs which is linking distributed heterogeneous spatial information resources from different data providers in an application-oriented and user-oriented way. Finn *et al.* (2004) conclude the most significant problem in data integration including several components including differences in datum, projections, coordinate systems, data models, spatial and temporal resolution, precision, and accuracy. Typical problems in integration are also introduced comprising of naming conflicts, scale conflicts, precision and resolution conflicts (Peedell *et al.* 2005).

These publications together with Australian integration projects highlight and address different aspects of technical spatial data integration. Major issues are: computational heterogeneity (standards and interoperability), vertical topology, attributes, reference system, data quality, data model, metadata and format (Figure 4).

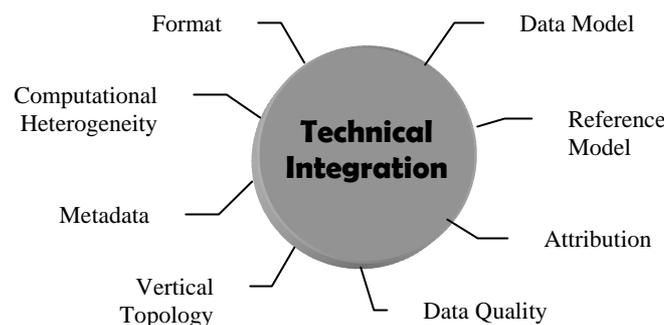


Figure 4. Components of SDI (Rajabifard 2001)

To investigate technical and non-technical issues which will be incorporated into development of required tools and framework, a research has been designed which will be discussed below.

Research Methodology

The proposed research design and stages for the study are incorporated into five major steps as illustrated in Figure 5.

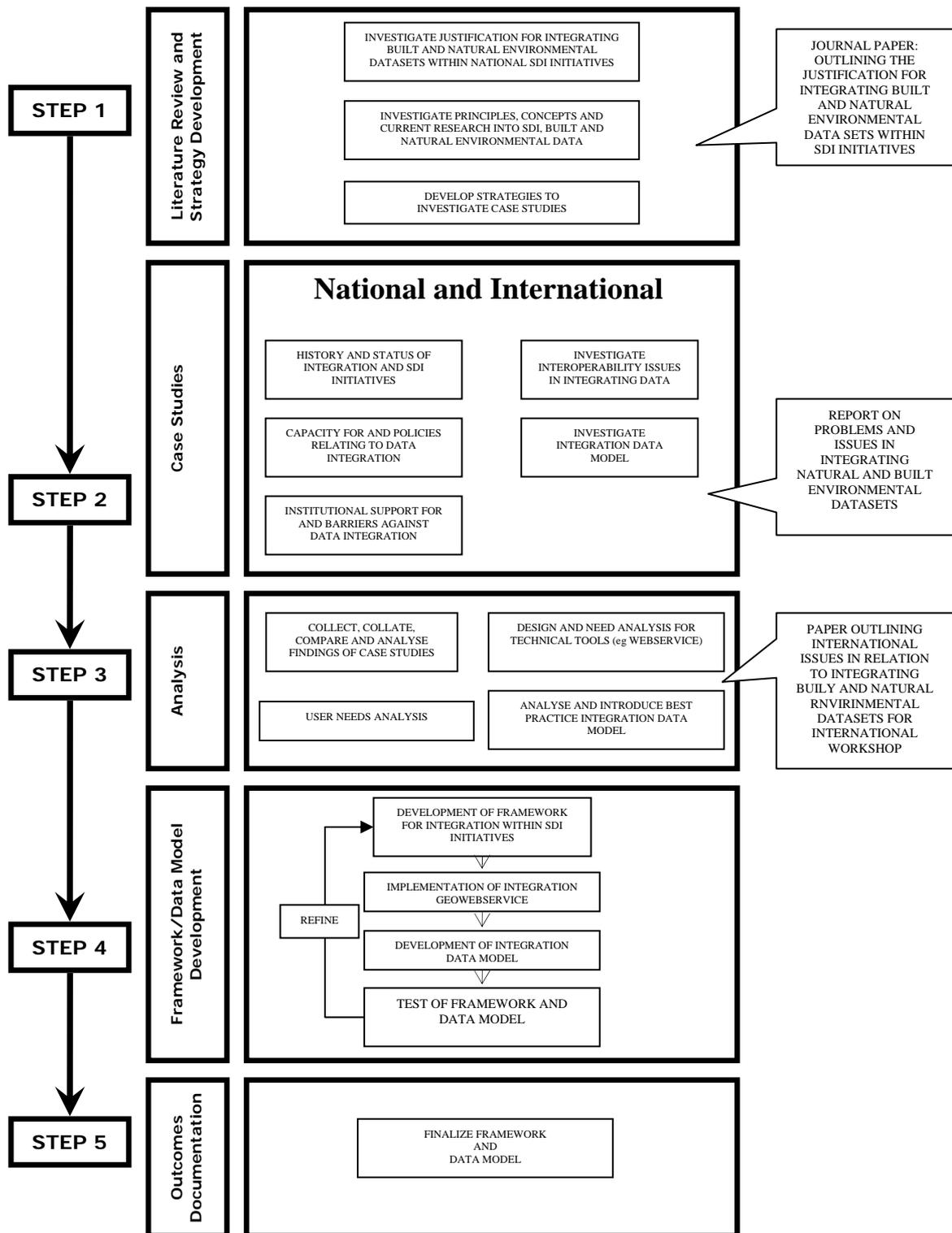


Figure 5. Research design and stages

This research can be broadly grouped into five major steps including a literature review together with strategy development, case study investigation, analysis, framework and data model development and outcomes documentation. To establish the theoretical background of the hypothesis, a wide literature review has been undertaken. The research also adopts both national and international case studies. The literature review provides the basis for the development of a research strategy and

highlights the significant issues which must be taken into consideration through case study visits. The development of the framework and data model will be done based on the outcomes of case study analysis.

Literature Review and Strategy Development

A literature review including the assessment of the existing materials has been undertaken in the areas of SDI concepts, natural and built environmental datasets characteristics, integration and interoperability and data modelling. The concept of SDI, its components and its mission will provide an understanding of the role SDI plays in facilitating the proper use and integration of multi-source datasets. Investigating significance, usage and characteristics of data integration and interoperability concepts will assist to provide a better perception of the topic. A good understanding of data modelling and applying its principles to investigate and assess existing integrated data models is also important. As a consequence, this will lead to development of a data model for integration.

The literature review will also assist in establishing an initial strategy framework for the analysis of the case studies. It is expected that the framework will change progressively as the research progresses.

Case Studies

To achieve a better understanding of actual data coordination and SDI initiatives across a broader spectrum, five countries in Asia-pacific region have been chosen as case study countries. Each country is different and hence there will be diversity in size, population, organizational framework and maturity in spatial data coordination and this forms a great combination of different case studies (Figure 6). The communication channels available through PCGIAP and Melbourne University help to investigate the case studies. Also, international case study countries in the Asia/Pacific region are very keen to be involved in this research and have expressed their voluntarily participation through provision of reports, required information and data.

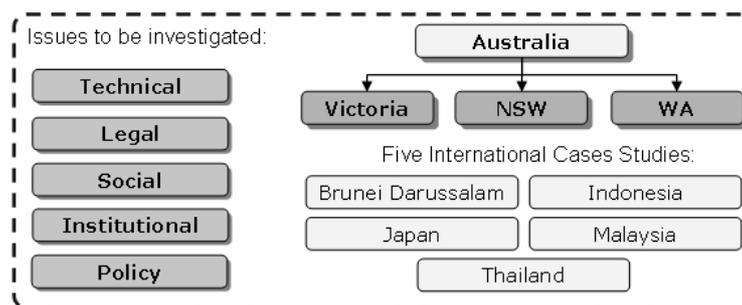


Figure 6. Case Study Countries and issues to be investigated

Australia, Brunei Darussalam, Indonesia, Japan, Malaysia and Thailand will be investigated through research progress and the outcomes of the project (data model, framework and associated tools) will be applied to these case studies.

Within Australia, through the activities of Geoscience Australia (GA), ANZLIC-the Spatial Information Council and the Public Sector Mapping Agency (PSMA), the

institutional and technological arrangements to facilitate integration of cadastral and topographic data are partially developed and the country has begun to implement a National SDI. However there are still several technical and non-technical issues that have not been solved. Therefore, considerable research still needs to be undertaken to identify an appropriate national framework to resolve interoperability issues of national topographic datasets and state/territory cadastral (and other relevant) datasets and to deliver the level of integration required to assist development of policies aimed at sustainable development.

The states of Victoria and NSW together with PSMA and GA have been visited and reviewed as part of the case study. Investigation of data coordination and integration issues has been done through a series of visits to Victoria, NSW, PSMA and GA's major spatial data providers and Western Australia will be done in the near future.

Some of the early findings and outcomes of these visits are as follows:

- Different agencies are at different levels of maturity in database management and collaborations.
- Cadastral data model is time-variant (through time-stamp) in contrast to topographic data model
- Different data models are used by different jurisdictions
- Content of metadata is not adequate for proper integration
- Feature level metadata is more efficient rather than theme-based metadata
- Attribution of same data is different across jurisdictions
- Diversity in format, precision, currency, coordinate systems and content in datasets
- Inconsistencies in specifications, feature categorizing, standards and topology
- Social characteristics do play a key role in sharing and integrating data
- Virtual data repository (an alternative for federated database) is a more appropriate solution for integration of different systems
- Integration of different data with different IP, pricing and privacy issues

As a part of Australian case studies and in order to facilitate the research steps, three short term projects have been defined and conducted on assessment of spatial data within states of Australia and providing geo-webservice requirements for data integration.

As a part of research the international case studies are also designed to gain access to both developed and transitional countries in order to broaden the focus of research and enable a comparison with Australia. Countries in the Asia-Pacific region are still developing their institutional arrangements and technological strategies to achieve collection and interaction of data, though developed countries are testing strategies on a project basis. The development of National SDIs varies from country to country, with policy creation and technical development generally being higher priorities for research (especially in transitional countries) than institutional arrangements and data

interoperability. A considerable amount of research into developing a national framework for automation of updating and integrating datasets is needed. International case studies are reporting on integration issues within their jurisdictions through the PCGIAP channel.

The analysis stage will be undertaken to provide the basis for development of a data model, framework and integration geo-webservice. These steps are firstly the analysis, collation and comparison of the outcomes of case studies to provide input for development of the framework. Analysis of different existing data models used for integration of data sets all over the world will be undertaken in order to introduce a best practice integration data model through suggesting new revisions to existing data models to meet the needs of the broader range of users. The investigation of webservices is another step of the analysis phase which provides the requirement to understand webservices concepts and to develop an improved webservice to demonstrate the principles of data integration in alignment with standards and current frameworks, especially those which have been introduced by OGC.

There are different data models used for integration of different data sets. A comparison of existing data models and their efficiency in providing a proper data model for integration of different datasets, including their characteristics will be undertaken. An improved data model will be prepared.

The framework and data model will assist to develop a prototype Geo-WebService to integrate data from different sources. This also will address data, network, standards and other technical issues connected to the technical implementation of the web-service. Touching cutting edge technologies in implementing such web-service and interoperability issues is expected, of which WFS, WMS, XML and GML are some. The framework will gather different considerations and issues connected to data integration and address both technical and non-technical impediments and inconsistencies of data integration based on the experiences of different jurisdictions.

Testing and validation of the framework, data model and geo-webservice will be done through case studies. Feedback and refinements will be undertaken to the framework, data model and geo-webservice.

The last step of the research will be spent documenting supporting literature and the outcomes of case study investigations together with analysis and development steps.

To communicate the results of the research some papers have been published and for each visit, a report has been developed. Also, a web site has been designated to communicate the results and publications to the general public as well as the partners and jurisdictions which have been involved in research.

A research timetable has been outlined below, showing the future steps of research.

Future Activities

Based on findings through the literature review and case study together with short term projects, a body of literature review has been undertaken on data modelling, interoperability and geo-webservices. This will lead to development of an overall framework, improved data model and integration geo-webservice.

Activity	2005				2006				2007				2008			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Literature Review		■	■	■	■	■	■	■								
Case Study Visits and Analysis					■	■										
Data Model Analysis							■	■								
Geo-WebService Analysis							■	■								
Data Model Development								■	■	■	■	■				
Integration Geo-WebService Development								■	■	■	■	■				
Framework Development							■	■	■	■	■	■				
International Workshop							■									

Conclusion

This paper presents a research framework for investigating and better understanding issues of multi-source data integration within SDI initiatives. The primary goal is to design a methodological framework to facilitate the integration of built and natural datasets and advance the knowledge of the National SDI capacity in meeting sustainable development objectives through facilitating data integration. This will be achieved through the development of new framework and tools to integrate built and natural environmental datasets in the context of SDI. The value of data is determined by the interests and willingness of the users to spend for and use the data sets. One of the key objectives in developing the framework and tools is to provide an environment to facilitate the integration of data to use data to its maximum potential.

The diversity in data providers creates a great deal of inconsistency in the integration of datasets, including institutional, technical, social, legal and policy heterogeneity. These heterogeneities effect different aspects and components of a spatial society to facilitate the data flow, access and integration. In order to address the issues and inconsistencies, relation between these issues and SDI components has been established. This model will assist to understand and highlight integration problems in the context of SDI and will be utilized to develop a framework together with associated tools, to facilitate the integration of multi-sourced spatial data.

Acknowledgement

This paper is part of an ongoing research project titled “The Integration of Built and Natural Environmental Datasets in National Spatial Data Infrastructure Initiatives”, and is supported by ARC-linkage programme. The research is also supported by research partners Department of Sustainability and Environment, Victoria, Department of Lands, New South Wales and Geoscience Australia. The Authors acknowledge the support of the members of the *Centre for Spatial Data Infrastructures and Land Administration*, at the Department of Geomatics, University of Melbourne in the preparation of this paper and associated research; however, the views expressed in this paper are the authors’ views and not the views of the people of the centre.

- Donaubauer, Andreas (2005). A multi-vendor Spatial Data Infrastructure for Local Government Based on OGC Web Services. From Pharaohs to Geoinformatics, Cairo, Egypt.**
- FIG (1996). The Bogor Declaration, United Nations Integral Meeting of Experts on the Cadastre.**
- Finn, Michael P., E. Lynn Usery, Michael Starbuck, Bryan Weaver, et al. (2004). Integration of the National Maps. XXth ISPRS Congress. Istanbul, Turkey.**
- Fonseca, Fredrico, Max Egenhofer, Clodoveu Davis and Gilberto Camara (2002). "Semantic Granularity in Ontology-Driven Geographic Information Systems." AMAI Annals of Mathematics and Artificial Intelligence.**
- Hakimpour, Farshad (2003). PhD Thesis-Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata. Zurich, Zurich University. PhD.**
- Parker, John and Stig Enemark (2005). Land administration and spatial data infrastructure: Special forum on the development of land information policies in the Americas. Eighth United Nations Regional Cartographic Conference for the Americas. New York.**
- Peedell, Stephen, Andres Friis-Christensen and Sven Schade (2005). Approaches to Solve Schema Heterogeneity at the European Level. 11th EC GI and GIS 2005.**
- Rajabifard, A. and Williamson, I.P (2001). Spatial Data Infrastructures: Concept, SDI Hierarchy and Future directions. GEOMATICS'80 Conference. Tehran, Iran.**
- Rajabifard, Abbas and Ian Williamson (2004). The Integration of Built and Natural Environmental Datasets in National Spatial Data Infrastructure Initiatives. Eighth United Nations Regional Cartographic Conference for the Americas. New York: 7.**
- Simpson, Dan Edward and Justin (2002). Integration and access of multi-source vector data. Symposium of Geospatial Theory, Processing and application, Ottawa, Canada.**
- UNRCC-AP (1997). Resolution of the 14th UNRCC-AP. United Nations.**
- UNRCC-AP (2003). Resolution of 16th UNRCC-AP. Okinawa.**
- Usery, E. Lynn, Michael P. Finn and Michael Starbuck (2005). Integrating Data Layers to Support the National Map of the United States. International Cartographic Conference. A Corua, Spain.**



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

MOHAMMADI, HOSSEIN;BINNS, ANDREW;RAJABIFARD, ABBAS;Williamson, Ian P.

Title:

Spatial Data Integration

Date:

2006

Citation:

Mohammadi, H. Binns, A. Rajabifard, A. & Williamson, I. P. (2006). Spatial Data Integration. In Seventeenth UN Regional Cartographic Conference for Asia and the Pacific, Bangkok.

Publication Status:

Published

Persistent Link:

<http://hdl.handle.net/11343/26704>