Seventeenth United Nations Regional Cartographic
Conference for Asia and the Pacific
Bangkok, 18-22 September 2006
Item 7 of the provisional agenda*

INVITED PAPERS

THE DEVELOPMENT OF A FRAMEWORK
AND ASSOCIATED TOOLS FOR THE INTEGRATION OF
MULTI-SOURCED SPATIAL DATASETS

Submitted by The University of Melbourne **

* E/CONF.97/1
** Prepared by Mr. Hossein Mohammadi, Mr. Andrew Binns, Mr. Abbas Rajabifard and Mr. Ian P. Williamson, Centre for SDIs and Land Administration, Department of Geomatics, The University of Melbourne.
Introduction

The world of today heavily relies on spatial data to manage the natural and built environments. Monitoring and controlling the impact of human activities on environment and the impact of natural environment changes (including natural hazards) on built environment is a major policy driver for many jurisdictions. The mitigation of natural hazards including tsunami, earthquake and landslide and securing citizens against them has become another priority of nations, especially after Indonesia’s tsunami. September 11 was also a remarkable point in history which attracts attentions towards providing more efficient counter-terrorism initiatives to secure citizens.

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, if they meet sustainable development objectives.

All 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 the factual model of 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 address 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.

Case study investigation has been done on two regions datasets from two Australia’s states (Victoria and New South Wales) and also same regions’ 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 findings of six other international countries within Asia-Pacific region will incorporate in designing technical tools for proper 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 differs across different providers and makes integration problematic both from technical and non-technical perspectives.

From 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 mechanism together with technical tools to facilitate the integration of multi-sourced datasets.

Figure 1 shows technical data integration and surrounding inconsistency issues. This should be mentioned that these issues do not include all non-technical issues which are related to datasets, and contain those non-technical issues which are related to the integration of datasets.

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 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).
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 collaboration between stakeholders, business model 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,
- Privacy, and
- Licensing

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational Heterogeneity (Standards and Interoperability)</td>
<td>Collaboration models</td>
<td>Legislation Issues</td>
<td>Rights, Restrictions and Responsibilities</td>
<td>Cultural Issues</td>
</tr>
<tr>
<td>Vertical Topology</td>
<td>Funding Model</td>
<td>Priorities (Sustainable Development) and Policy Drivers</td>
<td>Copyright and Intellectual Property Rights (IPR)</td>
<td>Capacity Building</td>
</tr>
<tr>
<td>Semantic</td>
<td>Linkage between data management units</td>
<td>Pricing</td>
<td>Data Access</td>
<td>Background of Stakeholder</td>
</tr>
<tr>
<td>Reference System and Scale</td>
<td>Awareness of Data Existence</td>
<td></td>
<td>Privacy</td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td></td>
<td></td>
<td>Licensing</td>
<td></td>
</tr>
<tr>
<td>Data model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cultural differences, capacity building and equal access to spatial data are also paramount in the social category. Table 1 summarizes all mentioned issues. 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)

Mentioned non-technical issues hinder effective data integration; however in following section 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, still there are difficulties in designing appropriate spatial data infrastructures to integrate built and natural spatial data (Parker and Enemark 2005) and many jurisdiction face many difficulties in integrating cadastral and topographic spatial data (UNRCC-AP 2003). SDI establishment should take a further action by investigating issues in the integration of cadastral and topographic data sets (UNRCC-AP 1997).

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.

From technical perspective there is plenty of publication which has 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. These categories have been clarified by following example. Two systems sharing data representing streets may confront 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, semantic, reference system, scale, data quality, data model, metadata and format.
To investigate these problems together with non-technical issues which will incorporate to develop required tools and framework, a research design has been proposed in following section.

**Research Methodology**

The proposed research design and stages for the study are incorporated into five major steps as illustrated in Figure 4.

**Figure 4.** 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.

The literature review has been undertaken in the areas of SDI, natural and built environmental datasets, integration and interoperability and data modelling. The concepts, missions and components of SDI together with the significance and applications of integration, interoperability and data modelling will provide an understanding of the integration of multi-source datasets and its requirements. Through the literature review a research strategy will be developed and will be utilized in case study investigations.

To achieve a better understanding of the issues and impediments of spatial data integration, and as part of case study, a series of visits and meeting have been conducted at different jurisdictional levels. The aim of these visits is to understand the role of the key players of the spatial data scene and the way in which they coordinate the spatial data, including the inconsistencies they or their users encounter in using data. Victoria, NSW and WA’s spatial data organizations as well as two national level agencies (GA and PSMA) are to be visited and investigated. Victoria’s major spatial data stakeholders (in public sector comprising of SII, Land Victoria, land Stewardship and Biodiversity), NSW’s Department of Lands, GA and PSMA have been visited to date, with WA to be visited in the near future.

The analysis stage will be undertaken to provide the basis for development of a data model, framework and integration geo-web service. 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 the 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 incorporate to design and 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-web service will be done through case studies. Feedback and refinements will be undertaken to the framework, data model and geo-web service.

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

*Literature Review and Strategy Development*

A literature review 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 5). 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 5. Case Study Countries and issues to be investigated](image_url)

These countries 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. The case studies are as follows:

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 (as discussed in section 1.6). 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 within the next 5 months.
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 contrary 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 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 proper solution for integration of different systems.
- Integration of different data with different IP, pricing and privacy issues.

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 will report on integration issues within their jurisdictions through a PCGIAP workshop on integration and also will provide required information and data.

As a part of Australian case studies and in order to facilitate the research steps, three short term projects have been defined and conducted with the assistance and input from final year students. The projects conducted and coordinated so far are as follows:

2. Technical assessment of built and natural datasets integration using NSW and GA’s datasets.

The first two projects have already been completed including a detailed technical report identifying the issues and problems in integrating data from three different jurisdictions.

The third project aims to investigate different issues and requirements for developing a prototype Geo-webservice to integrate data from different sources. This ongoing project will also address data, network, standards and other technical issues connected to technical implementation of the webservice.

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 result and publications to the general public as well as the partners and jurisdictions which have been involved in 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. Mentioned steps and activities will be also documented in a thesis and will be submitted. A timetable showing the future steps of research has been outlined below.

<table>
<thead>
<tr>
<th>Activity</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature Review</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Case Study Visits and Analysis</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Confirmation report</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Model Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geo-WebService Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Model Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration Geo-WebService</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Framework Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thesis Writing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Thesis Submission</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>International Workshop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 and associated tools together with associated tools to facilitate the integration of 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.
References


Simpson, Dan Edward and Justin (2002). Integration and access of multi-source vector data. Symposium of Geospatial Theory, Processing and application, Ottawa, Canada.


BIOGRAPHICAL NOTES

Hossein Mohammadi is currently a PhD Candidate at the Centre for SDIs and Land Administration, Department of Geomatics, the University of Melbourne. He has completed his Masters degree in GIS at KNT University in Tehran.

Andrew Binns is a Research Fellow and member of the Centre for Spatial Data Infrastructures and Land Administration at the University of Melbourne. He previously worked with the Cooperative Research Centre for Spatial Information (CRC-SI) investigating the development of Virtual Australia. He has also worked as part of a project team who looked into the development of a marine cadastre for Australia. His research areas include marine cadastre and administration, SDI development, Land Administration and Remote Sensing.

Abbas Rajabifard BSurv (KNT), MSc (ITC), PhD (Melb), is Deputy Director of the Centre for Spatial Data Infrastructures and Land Administration, and a Senior Research Fellow in the Department of Geomatics at the University of Melbourne. He has been an Executive Board member and National representative to the United Nations sponsored Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP) and is currently WG3-Research Coordinator. He is a Chief Investigator of an Australian Research Council project looking into the development of marine cadastre and in particular
investigation into the development of a seamless SDI model covering both land and marine environments. His current research and interest are spatial data management, SDI design and development (both land and marine environments), spatially enabled platforms and SDI capacity building.

**Ian Williamson, AM, FTSE**, is Head, Department of Geomatics, University of Melbourne, Australia, where he is Professor in Surveying and Land Information, and Director of the Centre for Spatial Data Infrastructures and Land Administration. He is Chair, Working Group 3 (Cadastre) of the United Nations sponsored Permanent Committee for GIS Infrastructure for Asia and Pacific (PCGIAP). He was Chairman of Commission 7 (Cadastre and Land Management) of the International Federation of Surveyors (FIG) 1994-98 and Director, United Nations Liaison 1998-2002. His teaching and research interests are concerned with designing, building and managing land and marine administration systems, cadastral, and land and geographic information systems in both developed and developing countries. He has consulted and published widely within these areas.

**CONTACTS**

**Mr Hossein Mohammadi**  
PhD Candidate, Centre for SDIs and LA  
Department of Geomatics  
The University of Melbourne  
Parkville, 3010  
AUSTRALIA  
Tel. + 61 3 8344 9696  
Fax + 61 3 9347 2916  
Email: hosseinm@gpgrad.unimelb.edu.au

**Mr Andrew Binns**  
Research Fellow, Centre for SDIs and LA  
Department of Geomatics  
The University of Melbourne  
Parkville, 3010  
AUSTRALIA  
Tel. + 61 3 8344 9692  
Fax + 61 3 9347 2916  
Email: a.binns@unimelb.edu.au

**Dr Abbas Rajabifard**  
Department of Geomatics  
The University of Melbourne  
Parkville, 3010  
AUSTRALIA  
Tel. + 61 3 8344 0234  
Fax + 61 3 9347 2916  
Email: abbas.r@unimelb.edu.au

**Professor Ian Williamson**  
Department of Geomatics  
The University of Melbourne  
Parkville, 3010  
AUSTRALIA  
Tel. + 61 3 8344 4431  
Fax + 61 3 9347 4128  
Email: ianpw@unimelb.edu.au
Author/s:
Mohammadi, H.; Binns, A.; Rajabifard, A.; Williamson, I. P.

Title:
The development of a framework and associated tools for the integration of multi-sourced spatial datasets

Date:
2006

Citation:

Publication Status:
Published

Persistent Link:
http://hdl.handle.net/11343/34923

File Description:
The development of a framework and associated tools for the integration of multi-sourced spatial datasets