Department of Geomatics, University of Melbourne, Victoria, 3010, AUSTRALIA
Tel: ++61-3-8344 9696, Fax: ++61-3-9347 2916, email: hosseinm@pgrad.unimelb.edu.au

Spatial Data Integratability and Interoperability in the Context of SDI

Hossein Mohammadi
Centre for Spatial Data Infrastructures and Land Administration
Department of Geomatics, University of Melbourne

Keywords: Spatial Data, Spatial Data Infrastructure, Integratability, Interoperability, Geo Web-Service

ABSTRACT

The number of multi-sourced heterogeneous spatial datasets continues to grow and the fragmentation of organizational arrangements has caused much technical and non-technical heterogeneity. Spatial Data Infrastructures aim to facilitate spatial data use and sharing, and can be an effective platform to aid in data integration. This paper discusses the technical and non-technical heterogeneities of multi-sourced spatial data within the holistic framework of Spatial Data Infrastructure. The paper capitalizes on research and case studies undertaken within Australia. The paper also introduces Geo-WebServices as a means of facilitating spatial data integration and interoperability. Geo-WebService can provide a platform to assess the level of integratability and readiness of multi-sourced datasets. The results of this research aim to assist practitioners in developing the necessary technical tools including geo web-services and guidelines for effective data integration.

1. MULTI-SOURCE DATA INTEGRATION AND INTEROPERABILITY

Multi-source data integration and interoperability has become a significant issue as it ensures effective access and reuse of spatial data by many spatial users and applications. This has created many opportunities and possibilities for using and applying spatial datasets in a range of services.

Many spatial applications and services try to model and analyze some aspects of the environment utilizing different criteria. These applications rely highly on multi-sourced spatial data to meet the requirements of diverse criteria. For example, the Emergency Information Coordination Unit (EICU) of New South Wales (NSW)-Australia utilizes different spatial data ranging from *fundamental* datasets including cadastre, topography, roads and imagery to *locational* data including police, fire and points of interests to *socio-economic* and *infrastructure* data including demography, valuation, public transport and utilities (Colless 2005). Many of these datasets are managed by different custodians in NSW. For example cadastre and topography is managed by local councils and Department of Lands, roads by local councils, Roads and Traffic Authority (RTA) and National Parks (Baker and Young 2005), and fire data by Department of Land, National Parks and Wildlife Service and Royal Botanic Gardens.

Table 1 summarizes some spatial datasets which are necessary for emergency management purposes and also their potential sources within the state of NSW-Australia.

Spatial data	Source
cadastre and topography	Local councils and Department of Lands
roads	Local councils, Roads and Traffic Authority (RTA) and National Parks etc.
imagery	Department of Lands, RTA, and Department of Agriculture
vegetation	Department of Land and Water Conservation, National Parks and Wildlife Services, Forests NSW, Department of Defense
fire	Department of Land, National Parks and Wildlife Service and Royal Botanic Gardens
threatened species	Department of Land, National Parks and Wildlife Service and Royal Botanic Gardens
Waste	Environmental Protection Agency, Waste Service and Local Councils

Table 1 Example of spatial data with potential sources used in NSW-Australia

As shown in the above table, different organizations are responsible for different datasets. Organizations develop their own strategies and policies in regards to capturing, managing and sharing data. The diversity of approaches utilized by these organizations leads to many technical and non-technical inconsistencies and heterogeneity among datasets.

2. SPATIAL DATA INTEGRATION CHALLENGES

Despite the importance of spatial data integration for many applications and services, the fragmentation of the institutions that are responsible for the production and management of different datasets has caused heterogeneities and inconsistencies from different technical and non-technical aspects, as illustrated in Figure 1. These inconsistencies can be classified into institutional, policy, legal and social categories as suggested by Mohammadi et al. (2006a).

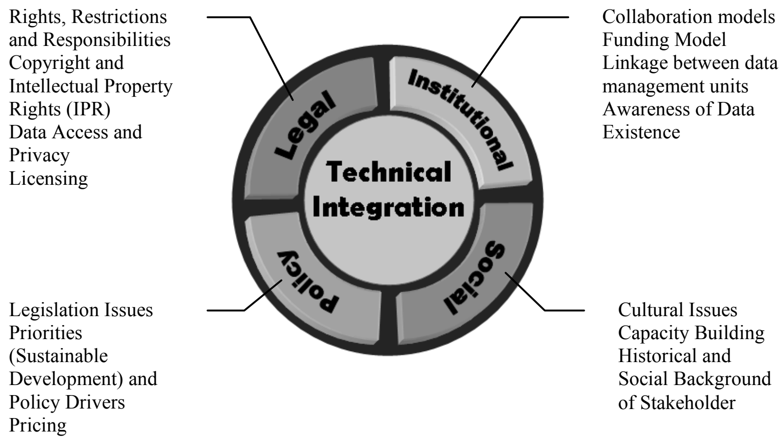


Figure 1 Technical data integration and associated non-technical issues

Technical issues including inconsistent standards, semantic heterogeneity, poor metadata/no metadata and inconsistency in data models hinder effective data integration. However the problem is not only technical in nature, with non-technical issues also hindering effective integration.

The non-technical problems including inconsistency of institutional arrangements and policies, different understanding and knowledge, capacity building, lack of regulation and lack of efficient metadata are also a concern, as has highlighted by Syafi'I (2006).

The collaboration between stakeholders, business models together with data management approaches are also key institutional issues which act as barriers against data integration. Policy drivers and priorities of nations, pricing and legislation have been found to be major issues from a policy view point. Cultural differences, capacity building and the social background of spatial data stakeholders are also paramount in the social category. 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

Table 2 has summarized technical and non-technical issues associated with spatial data integration and integratability.

Technical issues	Non-technical issues			
	Institutional issues	Policy issues	Legal issues	Social issues
▪ inconsistent standards ▪ semantic heterogeneity ▪ poor metadata/no metadata ▪ inconsistency in data models ▪ attribution heterogeneity	▪ utilizing inconsistent collaboration models ▪ funding model differences ▪ lack of awareness of data integration	▪ lack of supporting legislations ▪ inconsistency in policy drivers and priorities (sustainable development) ▪ pricing	▪ definition of rights, restrictions and responsibilities ▪ inconsistency in copyright and intellectual property rights (IPR) approaches ▪ different data access and privacy policies	▪ weakness of capacity building activities ▪ different background of stakeholder

Table 2 Technical and non-technical issues associated with spatial data integration and interoperability (adopted from Mohammadi et al 2006b)

3. SPATIAL DATA COORDINATION IN AUSTRALIA

In terms of Australia, there are several national level organizations responsible for different aspects of spatial data coordination, including Public Sector Mapping Agency (PSMA), Geoscience Australia (GA), Office of Spatial Data Management (OSDM) and ANZLIC – the Spatial Information Council. PSMA produces national coverage maps from best available data sources. GA produces and maintains national level spatial data which is mostly small scale maps. OSDM provides spatial policy and guidelines for use of data within federal governmental organizations, while ANZLIC is the peak body for development and coordination of Australian SDI (ASDI) guidelines. Collaboration between organizations at federal and state levels is mostly done through PSMA and ANZLIC. Access to data and pricing strategies are developed at different jurisdictional levels.

Access to spatial data at a national level is done on a case-by-case basis. OSDM develops federal governmental spatial data pricing policy, while PSMA has its own pricing policy for spatial data in place. Data integration is also done in-house as GA and PSMA hold almost all national level spatial datasets.

At state level, states communicate to each other through certain channels including PSMA and ANZLIC and also on a project basis. Pricing policies differ from state to state, ranging from the state of Victoria, which sells spatial data based on cost recovery policy to the state of Western Australia (WA), which provides spatial data to users at the cost of distribution.

Within states, due to large numbers of data custodians without well-established collaboration, integration and interoperability of spatial data is problematic and is a time consuming and costly process. Standards are developed by every state individually, while where applicable and if in line with state priorities, they adopt ANZLIC's guidelines. An example of a nationally consistent initiative is a metadata standard developed by ANZLIC and adopted by the states. Access channels are not singular and there are different access channels for spatial data. States produce their own spatial data with their own policies and guidelines.

Local councils produce large scale maps but follow state standards and policies on spatial data. Access to local council's data is done on a case-by-case basis. Pricing on spatial data is done by states and local councils seldom sell and distribute data. Local councils liaise with states and in some cases there is very good collaboration in place, however, there is generally little collaboration between local councils (Table 3).

	Federal	**State**	**Local council**
Data Production	PSMA, GA	State and local councils	Local council
Policy development	OSDM, ANZLIC	States and ANZLIC	State
Access	case by case basis, free to browse	for small clients through data resellers and for large clients State organizations	case by case basis
Standards	OSDM, ANZLIC, ICSM	State, ANZLIC	State
Collaboration	through PSMA and ANZLIC	With States through ANZLIC and PSMA, individually with local councils and through joint initiatives	individually with states, little collaboration with other councils through states
Pricing	OSDM, PSMA	State	State
Spatial data integration	in-house	users and third parties	users and third parties

Table 3 Spatial data management arrangements within Australian jurisdictions

The diversity of key players in the spatial data area within different jurisdictional levels with different interests and priorities leads to a diversity and complexity in data coordination arrangements and policies. The diversity of policies and coordination approaches satisfy objectives and requirements of a particular organization.

The complexity of issues associated with spatial data integration and interoperability can not be addressed and facilitated, unless there is a well-structured and holistic platform to consider all effective components and issues of spatial data integration together.

For spatial services which utilize spatial data from different sources, it is important to access data at justifiable time and cost through the fastest channel. This is not possible unless the diversity of issues is managed through an enabling platform. Such a platform establishes interoperability at technical and non-technical levels and establishes effective interaction between different technical and non-technical components including policies, standards, collaboration and access. This platform can provide a comprehensive framework which assists spatial data stakeholders to develop and design required guidelines, policies and technical tools for effective spatial data integration and interoperability.

The development of SDIs can aid in providing a basis for establishment of an enabling platform. SDIs aim to facilitate the reuse, access, integration and sharing of data. In order to achieve this aim, SDIs potentially can provide the necessary technical and non-technical tools and policies together with guidelines.

4. SDI TO FACILITATE EFFECTIVE SPATIAL DATA INTEGRATION

SDIs are being developed by many countries throughout the world as an enabling platform to assist people to access, use and integrate spatial data effectively. It includes access networks, policies and standards of an enabling platform which facilitates the interaction of spatial data stakeholders with spatial data (Figure 2).

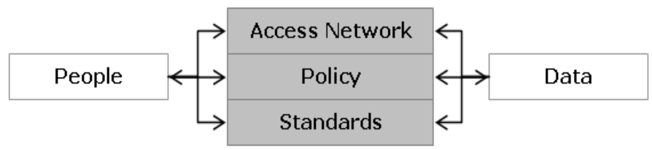


Figure 2 SDI components (Rajabifard et al 2001)

As Rajabifard et al (2001) have highlighted, SDIs are able to establish requisite arrangements and facilitate data integration. In this regard, necessary tools should be developed for stakeholders to interact with technical, policy and standardization tools to coordinate, integrate and use data more effectively. SDIs aim to address and coordinate the integration of multi-sourced data in a way which saves time and reduces costs; however for the stakeholders of spatial data, this aim has not been fully achieved. To facilitate data integration, issues and challenges need to be identified and addressed in the context of SDIs. These considerations can then assist to develop technical, policy, institutional and management tools for effective data integration.

In terms of Australia, integratability of spatial datasets is one of four main streams in the development of the Australian National SDI (ANZLIC 2003). The perception of spatial data integration and its importance among states is different which leads to a lack of knowledge of spatial data integratability issues in some states, however for other states it is a high priority. Therefore, spatial data integration has not yet been achieved fully at a national level in Australia. Special attention should be paid to data integration when developing technical mechanisms and tools taking in to account the legal, institutional, policy and social frameworks within an SDI initiative.

Data coordination and maintenance is one of the aims of SDIs. SDIs can develop guidelines for spatial data producers in order to facilitate the reuse and integration of spatial data. It consists of dictating standards, interoperability tools, semantic homogeneity among datasets, data quality and reference system, metadata guidelines, data models and attribution guidelines.

SDIs also need to take into consideration the integratability of legal, institutional, social and policy frameworks. Without interoperable collaboration, spatial datasets can not be integrated and used to their maximum potential. Consistent pricing and privacy policies together with appropriate capacity building can ensure effective data integration. SDIs can also develop guidelines and tools to act as intermediators for many inconsistent systems. These guidelines and intermediators can establish effective links between inconsistent components of multi-source spatial data, services and policies.

With the advancement of Geo-WebServices (GWS), the access and availability of datasets which are targeted by SDIs is more facilitated (Nebert 2004). Geo-webservices are significant technical tools which can also facilitate the integration of multi-sourced datasets. Geo-webservices are developed based on open standards and provide tools, services and formats which comply with interoperability concepts. Geo-webservices can access and collect spatial datasets throughout the web and local databases. Based on this, a prototype geo-webservice has been developed to assess the level of readiness of datasets for the purpose of data integration.

5. GEOWEBSERVICE TO FACILITATE SPATIAL DATA INTEGRATION

Geo-webservices accomplish spatial data interoperability at data and service levels by utilizing common standards and specifications (Gould 2001). Capitalizing on opportunities created by geo-webservices, multi-sourced spatial datasets and services can be accessed and utilized. Spatial data integration can also be facilitated with the aid of geo-webservices.

One of the major problems in an effective data integration service is the assessment of spatial data and its level of integratability. Many services gather data from distributed sources and assemble them within a single system for integration and analysis, but the lack of integratability among datasets leads to many problems. Utilizing geo-webservice capabilities, a prototype system has been designed and developed as part of an ongoing research project on the integration of multi-sourced spatial data. This system is able to access Web Map Services (WMS) and Web Feature Services (WFS) as well as other common spatial formats, throughout the web and local databases. If other complementary information including availability of metadata, any restrictions on data or pricing and privacy policy are not extractable from data or data provider (not user), this is requested from user. Data collected from different sources is then assessed against integratability requirements based on a set of predefined criteria including the compatibility of format, datum, coordinate system, bounding box, availability of metadata, any restrictions on the use, manipulation and distribution of the data, and pricing and privacy policies associated with any particular data (Figure 3).

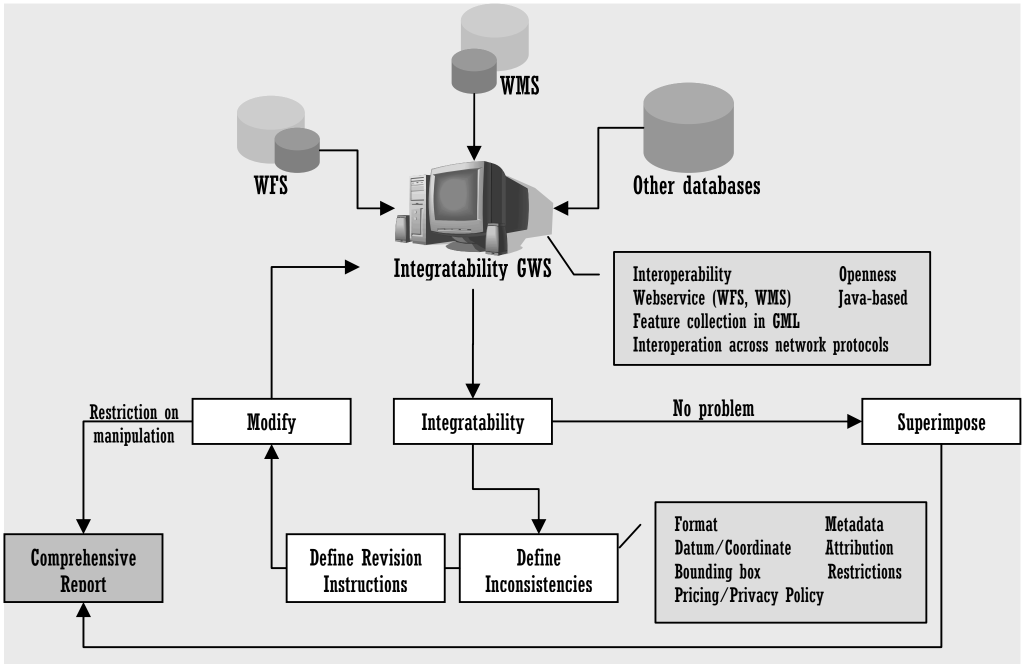


Figure 3 Prototype GWS flowchart

If there are no inconsistencies or any restrictions on access and use, which hinders integration recognized, datasets are superimposed into the system. Otherwise, the items of inconsistency are identified and a revision instruction is prepared to help the user amend data components. If there is no restriction on manipulation or use of the datasets, if the revision items have been met, the geo-webservice process is undertaken again to identify any possible problem. This process is repeated untill all criteria are met and the user is able to see the data and any information that comes with the data.

The prototype has been tested utilizing datasets from Australian organizations including those provided by Federal and the State organizations including Geoscience Australia (GA) and states of NSW and Victoria and also some other datasets from different sources distributed worldwide including datasets from Geography Network Canada (2007). The result of the test showed that there were problems in the integration of datasets including geometrical mismatch, lack or incompleteness of metadata, attribution inconsistency and restrictions on data, even among datasets from a single jurisdiction.

Figure 4 is a snapshot of the developed geo-webservice which has collected data from different sources within Australia for a region near Melbourne based on above-mentioned assessments.

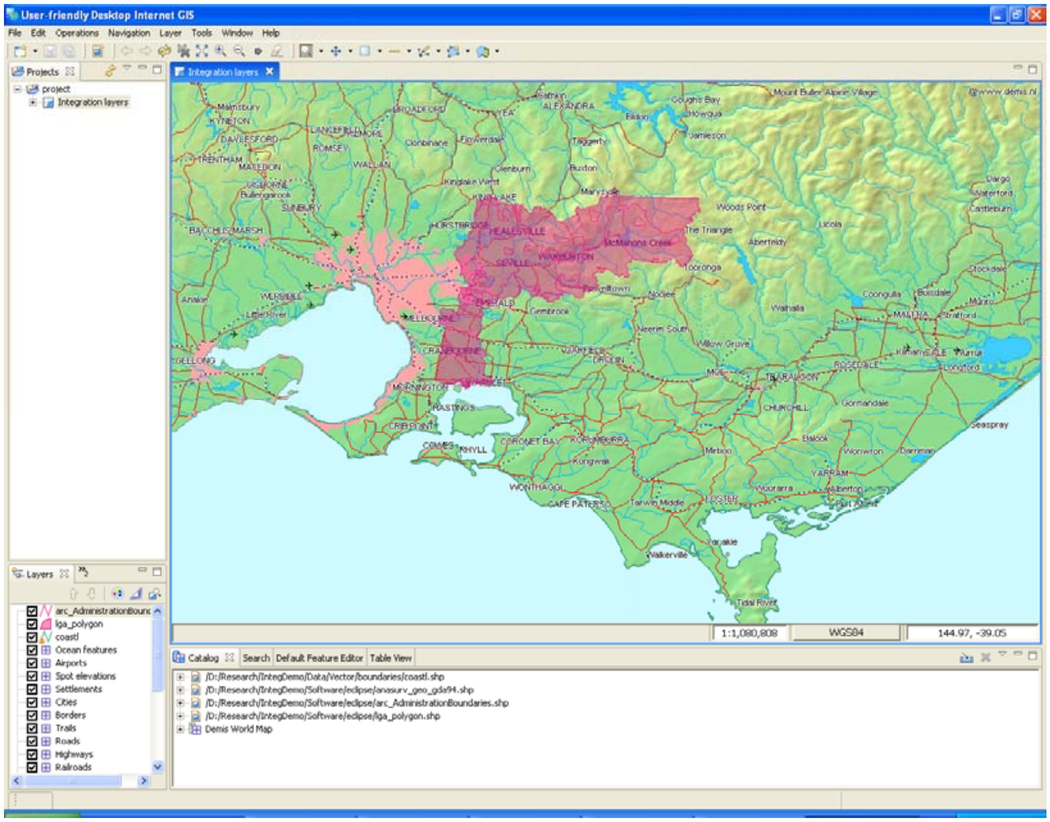


Figure 4 Prototype GWS snapshot (What does this figure tell me?????)

The design and implementation of the integratability geo-webservice and also the results of the test leads to some observations and recommendations as follows:

- Utilizing geo-webservices to asses the integratability of multi-sourced datasets can saves time and money
- The integratability test is an effective process which can decrease the workload associated with data integration
- Users and data providers need instructions for data integration
- Detailed metadata which contains data integration information including restrictions on data, privacy and pricing policies and attribution facilitates data integration
- Machine-readable metadata and complementary data highly facilitates data integration
- Data integration at attribution and data model level requires more investigation and sophisticated tools

At this stage the number of criteria is limited to the items mentioned above, but the prototype is under further development to adopt more criteria and also provide a separate data-custodian-centric instruction to help custodians of data to prepare datasets which comply appropriately with the integration criteria. The integration facilities of geo-webservice are limited to the superimposition of datasets, but further developments are needed to integrate attributes and also at more advanced levels to integrate data models.

The prototype will also be applied to the datasets from a number of countries in Asia-Pacific region through the channel of the Working Group 3 of the United Nations supported Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP) to test the issues within countries with different characteristics. This will also identify more issues and considerations which should be taken into account for future developments.

The outcomes of the research including the tools developed can not only be customized based on the requirements of different jurisdictions, but also can assist spatial data policy makers to develop the necessary guidelines on how to prepare and amend data for effective data integration.

6. CONCLUSION

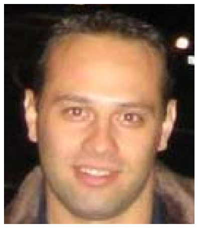
Spatial data integration is the most commonly performed task for many crucial spatial services including emergency management services. Despite the time and costs associated with data integration, due to the fragmentation of spatial custodians and inconsistency of approaches, effective spatial data integration has not been achieved in many cases. Many technical and non-technical issues result from these inconsistencies.

In terms of Australia as one of the leading countries in spatial data coordination, spatial data integration is still problematic and associated with many technical, institutional and policy, social and legal issues which hinders effective data integration at different jurisdictional levels. As a consequence, inconsistent collaboration models, pricing and access policies together with technical inconsistencies in standards, data model and metadata hinders spatial data integration and interoperability. To facilitate spatial data integration a holistic platform should be established which addresses technical and non-technical.

To fulfill this task SDIs can be utilized which aim to provide the necessary tools to facilitate spatial integration. SDI provides a platform of people with interaction to spatial data through technological components including access networks, standards and policies which can facilitate data integration and address associated technical and non-technical issues. SDIs can also provide guidelines for linking inconsistent data, services and data coordination policies. This framework can then be utilized by spatial data stakeholders to develop institutional arrangements, legal and policy tools and also social capacities to facilitate the integration of multi-sourced spatial data so that it is used to its maximum potential.

Geo-webservices provide effective technical tools to facilitate the access and integration of multi-sourced datasets. In this regard geo-webservices can by utilized not only to integrate multi-sourced spatial datasets but also to asses the integratability of these datasets. The prototype GWS evaluates integratability of multi-sourced spatial data against some criteria and provides instructions and guidelines to amend data based on criteria. This geo-webservice can be customized based on the requirements of different organizations and jurisdictions to meet specific criteria of that particular organization or jurisdiction. This tool can also help users to prepare datasets before integration and also can assist practitioners to develop required guidelines and specifications to be used by data users and providers to prepare data before use.

7. ABOUT THE AUTHOR



Hossein Mohammadi is PhD candidate, Centre for SDI and Land Administration at the University of Melbourne. Hossein has completed Masters degree of GIS in KNT University in Tehran. Hossein's interested areas are SDI, Spatial data integration and interoperability, and Geo-WebServices.

hosseinm@pgrad.unimelb.edu.au
Tel: ++61-3-8344 9696
Fax: ++61-3-9347 2916

8. REFERENCES

ANZLIC (2003). Implementing the Australian Spatial Data Infrastructure, ANZLIC-the Spatial Information Council.

Baker, A. J. and F. R. Young (2005). Digital Mapping Data Currency Through Sharing: A Practical Study. SSC2005 Spatial Intelligence, Innovation and Praxis, Melbourne, Australia.

Colless, R. (2005). Interoperability & Security in the Emergency Services Arena. http://www.anzlic.org.au/pubinfo/2413335134.html. 27th March

Geography Network Canada (2007), Geogaphy Network Canada data services. http://www.geographynetwork.ca/data/freedata.html. Access date: 20th March 2007

Gould, M. (2001). OGC: A Framework for Geospatial and Statistical Information Integration. Joint UNECE/Eurostat Work Session on Methodological Issues. Tallinn, Estonia.

Mohammadi, H., A. Rajabifard, A. Binns and I. P. Williamson (2006a). The Development of a Framework and Associated Tools for the Integration of Multi-Sourced Spatial Datasets. 17th UNRCC-AP. Bangkok, Thailand.

Mohammadi, H., A. Rajabifard, A. Binns and I. P. Willaimson (2006b). "Bridgind SDI Design Gaps to Facilitate Multi-source Data Integration." Coordinates 2(5): 26-29.

Nebert, D. D. (2004). The SDI Cookbook, Global Spatial Data Infrastructure (GSDI).

Rajabifard, A. and I.P.Williamson (2001). Spatial Data Infrastructures: Concept, SDI Hierarchy and Future directions. GEOMATICS'80 Conference. Tehran, Iran.

Syafi'i, M. A. (2006). The Integration of Land and Marine Spatial Dataset as Part of Indonesian SDI Development. 17th UNRCC-AP. Bangkok, Thailand, UN.