Spatial Data Integration Challenges: Australian Case Studies

Hossein Mohammadi, Abbas Rajabifard, Andrew Binns, Ian P. Williamson
Centre for Spatial Data Infrastructures and Land Administration
Department of Geomatics
University of Melbourne
Email: hosseinm@pgrad.unimelb.edu.au

Keywords: Spatial Data Integration, SDI, Interoperability

Abstract

To manage the complex environment it is necessary to manage all aspects through the integration of its built and natural components. The integration of multi-sourced built and natural environmental datasets is essential for facilitating government policy drivers including hazard management, counter-terrorism, environment protection and land administration. Sustainable development, which overarches these activities, is the most major motive for the Governments of Australia. Hence, the integration of multi-sourced spatial data is of importance to Australia.

Multi-sourced data integration is not only the technical collation of data, but also efforts to explore, assess, communicate with data providers, access, acquire, convert, and collate data. However fragmentation of institutions leads to an inconsistency in data coordination approaches, especially in the integration of spatial data. Inconsistency of spatial datasets has turned integration into an expensive and time-consuming task.

Spatial Data Infrastructure (SDI) aims to provide a dynamic environment in which different stakeholders are able to easily access and use data in the minimum time with minimum cost. Therefore any SDI aims to facilitate data integration to minimize effort, time and cost.

This paper investigates spatial data integratability and interoperability within Australia based on a case study methodology involving the states of Victoria, NSW and the Federal Government. Integration issues and activities are described and also best practice has been identified within each case study, and suggested way forward on integration made. These suggestions will feed in to the development of an overall framework for integration at a National level utilising SDI.

BIOGRAPHY OF PRESENTER

Hossein Mohammadi is a PhD candidate, Centre for SDI and Land Administration at the University of Melbourne. Hossein has completed a Masters Degree of GIS in KNT University in Tehran.
INTRODUCTION

Spatial data integration is one of the most common tasks in most spatial services. Spatial services often utilize different datasets from different sources to meet the diverse criteria of services. Therefore, effective data integration is a significant issue for spatial services. Despite the importance of spatial data integration, the fragmentation of institutions who manage and distribute datasets lead to an inconsistency and heterogeneity in datasets and associated technical and non-technical considerations including institutional, policy, legal and social issues. Overcoming these inconsistencies needs an enabling platform which considers different issues together and provides tools to facilitate data integration.

SDI aims to facilitate the access, use and the integration of spatial data. To meet this aim SDI provides guidelines, policies and technical tools for spatial data stakeholders to interact easily with other stakeholders and spatial data. Despite sophisticated developments in SDI, there is still no comprehensive investigation which overarches both technical and non-technical aspects of data integration in the context of SDI.

To provide basis for such a comprehensive investigation, a case study approach involving Australian state and federal governmental agencies and datasets, has been undertaken. Australia is one of the leading countries in terms of the utilization and creation of spatial information and capitalizes on sophisticated technologies and a well-developed SDI. However fragmentation of spatial information institutions together with lack of a single coordination body at some jurisdictional levels lead to a heterogeneity of approaches and policies in terms of data coordination and this makes the integration of multi-sourced spatial data a costly and time-consuming process. Fragmentation of jurisdictions is more explicit in federated nations including Australia where less technical and administrative interoperability occurs.

This paper investigates the strengths and weaknesses in data integration within Australia using a case study approach. Together with case study visits to federal agencies including GA (Geoscience Australia), PSMA (Public Sector mapping Agency), and states of Victoria and NSW, datasets produced by GA, Lands and DSE (Department of Sustainability and Environment) have been also collected for a pilot study in Victoria and NSW. This paper introduces findings of this case study analysis including the best practices and based on these findings, some recommendations are outlined. Results of this study will assist to develop a generic framework in the context of Australia’s National SDI (ASDI) to facilitate data integration as part of ongoing research.

SDI TO FACILITATE DATA INTEGRATION

Many significant applications including emergency management, land administration, urban planning and resource management rely heavily on multi-sourced spatial data. To achieve their aims, these applications need to integrate spatial data from different themes, in order to monitor and control different aspects of the environment. Many of these applications need to integrate and analyse datasets in a limited timeframe, hence effective spatial data integration is a crucial issue for these applications.

To effectively integrate data, technical and non-technical issues including institutional, policy, legal and social issues should be investigated and highlighted. There is a great body of literature on the technical integration of spatial data. Kolbe (2006) highlights the international standards as key components for data coordination. Vries (2006) believes that in order to overcome information integration challenges, there should be an interoperable platform established for users. Goodchild et al (2005) describe effective integration and
interoperability as the technical abilities of any system to input, manipulate, display and analyse data from another system. Network interoperability, metadata and data interoperability have been identified as the most straightforward aspects for maintaining effective data integration in most publications.

Data integration comprises many tasks from data exploration and communication with data holders to technical data collation. In order to effectively manage these tasks a comprehensive framework is required which takes both technical and non-technical aspects of integration into account.

SDI provides a framework including technical and non-technical tools that aim to help people to access, use and integrate spatial data effectively. SDI provides access network, policies and standards of an enabling platform which facilitates the interaction of spatial data stakeholders with spatial data (Figure 1).

**Figure 1:** SDI components (Rajabifard et al, 2001)

Rajabifard et al (2001) believe that SDI is able to establish requisite arrangements and facilitate data integration. In order to achieve this aim, spatial data integration and associated technical and non-technical issues should be investigated and discussed in the context of SDI. In this regard, an effective channel for stakeholders including data providers, VARs and users to interact with technical, policy and standardization tools to coordinate and use data properly needs to be established.

**Data integration issues**

Data integration is not only a set of technical functions including the geometrical and topological match of data and having a correspondence of attributes (Usery et al, 2005), but also the establishment of all non-technical mechanisms including legal, policy, institutional and social mechanisms to facilitate data integration.

The consideration of technical and non-technical levels of integration includes ensuring the involvement of jurisdictions in the continued development of technical, communication, and legal tools together with capacity building and the development of policy mechanisms. Effective effort is required to ensure that individual institutions move forward to the benefit of the society, the environment and the economy. SDI aims to provide a platform to consider these issues in a single framework.

In this regard, Mohammadi et al (2006) has highlighted and categorized technical and non-technical integration issues. Based on their categorization, spatial data integration comprises a core technical integration which is surrounded by non-technical issues including institutional, policy, legal and social issues. The technical issues include computational heterogeneity, data interoperability and semantic interoperability. However technical interoperability is not able solely to remove all data integration barriers. Establishing institutional, policy, legal and social interoperability is also essential for effective data integration as illustrated in Figure 2.
An institutional arrangement for effective data integration refers to the amount of interaction and openness of any individual institute to interact, communicate and liaise with others in the spatial domain. An indispensable component of an effective institutional arrangement is utilizing effective collaboration models and flexible and consistent funding models. Spatial policies are also interoperable if they can be easily executed with other policies with minimum interference and duplication of effort.

The decision to disseminate and share data is not freely taken especially in a complex environment where different stakeholders with different legal arrangements take part. Mapping the legal considerations on individual source data to the final integrated product will help to overcome the legal barriers to integration. This includes defining rights and restrictions on data and the responsibilities of stakeholders, copyright and intellectual property rights and data access and privacy policies.

The most complex and intangible areas are the social behaviours and patterns within individual jurisdictions. Historical background and cultural issues can heavily impact the way jurisdictions deal with each other. For example in some institutions a silo mentality dominates and there is little wish to share and disseminate data.

**SPATIAL DATA INTEGRATION IN AUSTRALIA**

To investigate the integration issues in Australia, a case study approach has been undertaken incorporating datasets from Victoria and NSW and also case study visits have been undertaken to GA, PSMA, DSE and Department of Lands-NSW. Findings of case studies have been investigated against ASDI strategies developed by ANZLIC. Capitalizing on case study findings, this section discusses integratability of datasets within an Australian context.

One of ANZLIC’s activities is developing a National SDI (ASDI) for Australia. One of the ASDI’s priorities is facilitating the interoperability and integratability of spatial data (ANZLIC, 2001). Achieving this aim is possible if ANZLIC provides guidelines for technical and non-technical tools which are necessary for integration.

In terms of Australia, there has been considerable effort in relation to creating technically well structured and sophisticated interoperable systems, however many non-technical issues still need to be considered. GA wishes to cooperate with states to develop a single interoperable spatial database which can be used to drive datasets at any required scale (The Australian Government 2005). PSMA has launched the LYNX product for spatial data which promises consistent and interoperable data across Australia (Position 2006). In terms of
NSW, the Spatial Information eXchange Portal (SIX) has been launched to deliver technical spatial data interoperability. Victoria has also developed a comprehensive interoperable SDE database for data exchange between stakeholders.

Although these systems have been established, they do not take into account all of the necessary issues that are needed to effectively integrate spatial data from multiple sources. Non-technical interoperability including institutional, political and legal interoperability also needs to be considered for implementing successful data integration systems, as identified by Miller (2006).

As traditional boundaries between institutions and disciplines begin to blur, users increasingly require access to information from a wide range of sources, both within and outside their own subject area. Institutional arrangements and considerations within different organizations differ in order to meet their own needs. The inconsistency of institutional arrangements can be removed through adopting common solutions wherever feasible (Miller, 2006).

In order to adopt common solutions, different levels of authorities including national, state and local council levels need to collaborate effectively. There should also be a mutual interaction between different jurisdictional levels. In an Australian context, collaboration between national and state agencies has been well-established. For example DSE in Victoria and Department of Lands in NSW have good working relationships with PSMA, ANZLIC and GA. DSE and Department of Lands have representatives on the ANZLIC standing committee, where PSMA and GA are also represented.

However, due to the cooperative nature of ANZLIC, there is no obligation for agencies to follow ANZLIC guidelines. Some jurisdictions adopt ANZLIC’s guidelines fully and according to their priorities and policies, while others do not adopt them or adopt them only partly.

Collaboration can be addressed at vertical and horizontal dimensions among jurisdictions. Horizontal collaboration refers to the interaction of jurisdictions at the same jurisdictional level such as states. Vertical collaboration refers to the interaction of jurisdictions at different level eg between a local council and state. Vertical institutional collaboration seems much better-achieved than horizontal (Figure 3). Often, states do not interact unless there is an issue which all parties have interests in and on a project-basis such as Country Fire Authority (CFA) and Murray Darling basin management (MDBI, 2006), initiatives which are project-based.

![Figure 3: Inter-jurisdictional Vertical and Horizontal Collaborations](image)

Within states, local councils often work in isolation from other local councils. Local councils interact more through state-based collaboration. One of the most unique and successful collaborative frameworks in Australia is the Property Information Project (Alexandre-
Tomlinson, 2006) which has been initiated in Victoria and provides a single framework for council-state interaction.

Table 1 summarizes the interaction channels between different jurisdictions in Australia. This table highlights good interaction between states and local councils, however it is individually established. Collaboration between local councils themselves is not so fluent, as they liaise through states.

Table 1. Interoperability Channels among Australian Jurisdictions

<table>
<thead>
<tr>
<th>Interoperability</th>
<th>Local Councils</th>
<th>Victoria and NSW</th>
<th>GA</th>
<th>PSMA</th>
<th>ANZLIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Councils</td>
<td>Through State</td>
<td>Good Interoperability</td>
<td>Through State</td>
<td>Through States</td>
<td>Through States</td>
</tr>
<tr>
<td></td>
<td>COAG</td>
<td>Individually established interoperability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victoria and NSW</td>
<td>Good Interoperability</td>
<td>Through ANZLIC and PSMA and Joint Projects</td>
<td>ANZLIC Mutual Collaboration</td>
<td>Directly</td>
<td>Directly</td>
</tr>
<tr>
<td></td>
<td>Individually established interoperability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>Through State</td>
<td>ANZLIC Mutual Collaboration</td>
<td>-</td>
<td>Directly</td>
<td>Directly</td>
</tr>
<tr>
<td>PSMA</td>
<td>Through State</td>
<td>Direct to State</td>
<td>Directly</td>
<td>-</td>
<td>Directly</td>
</tr>
<tr>
<td>ANZLIC</td>
<td>Through States</td>
<td>Directly</td>
<td>Directly</td>
<td>Directly</td>
<td>-</td>
</tr>
</tbody>
</table>

The table shows a high level of interaction amongst some stakeholders, however generally collaboration is not well developed. This needs to be taken into consideration within the context of developing a framework for data integration.

**Data integration tasks**

Technical and non-technical issues associated with data integration may influence any stage of integration. Data integration comprises different tasks. In order to effectively integrate data these tasks should be identified and surrounding issues should be highlighted.

Ostensen (2006) has identified six major services which are utilized by spatial data users in most spatial services including exploration, assessment, view and download, geo-processing and orchestrate services. These major services are more or less the tasks which are done in integration; hence by adopting these services addressed by Ostensen, six major integration tasks which are carried out by users have been identified (Figure 4) as follows:

- data exploration (discovery),
- data assessment,
- communication with data providers,
- access to data and data acquisition,
- geo-processing; and
- data collation.
Issues which effect a users ability to integrate data along each of these steps have been identified within the context of the Australian case study.

**Data Exploration**

Appropriate data exploration leads to finding the data and data source which suits 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 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.

In the case of NSW, DIPNR coordinates the state’s spatial data directory (NRDD, 2006). This service gathers metadata from different spatial data custodians in NSW and provides them through Natural Resource Data Dictionary. NRDD provides a single channel for metadata exploration. In the case of Victoria there is no state spatial data search engine, however Victoria’s metadata is provided through DSE in the form of a list of datasets and custodians (VSDD 2006). Also, there is a single national data directory in Australia. The Australian Spatial Data Directory (ASDD) at a national level provides a national data directory for data exploration, coordinated by GA. The ASDD contains metadata on most datasets created and maintained by Australian governments.

Metadata stored in ASDD follows a common standard developed by ANZLIC which is based on ISO19115 standard on metadata and all states have agreed to prepare metadata based on ANZLIC's standard. The ASDD does not contain metadata for all Australian governments' datasets. For example there is no metadata for PSMA’s products (ASDD 2006). There is outdated metadata in the ASDD including outdated contact information for some of NSW’s data (for example email address <name@dtm.gov.au> no longer exists). Beside these problems, ANZLIC's core metadata standard (ANZLIC 2005) does not provide information on attributions and data models which is essential for more complex levels of integration. However, along with ANZLIC’s metadata, states are able to develop more detailed metadata.
elements for their jurisdiction. For example NSW has developed its own metadata at feature level which helps effective data integration, but this is not undertaken in all states. Another channel to explore data is through Value-Added Resellers (VARs) who manage or create data. However each VAR may provide particular sets of spatial data and as there is no single channel to access all VARs datasets and metadata, users may not be able to find the required data.

Another issue for effective data exploration is the custodianship of spatial data. It is the data custodian who maintains data and also provides metadata through potential data directories. The problem arises when a custodian does not provide efficient metadata or more problematic when there is not a clear custodian for a particular data. One of most critical examples of custodianship problems is bathymetry data which does not have any national level custodian. At the same time, bathymetry data is critical for some national level emergency initiatives including tsunami mitigation service. Each state creates and maintains its own bathymetry data, but it is hard to gather them under a national database as there is no custodian for that.

In terms of custodianship guidelines, The Office of Spatial Data Management (OSDM) has been appointed by the federal government to develop custodianship guidelines (Baker 2001). OSDM’s custodianship guidelines have been adopted by most states; although they are not obligatory. At state level, Victoria has initiated a custodianship program to facilitate custodianship and develop a framework and common approach to manage and coordinate custodianship (VSC 2006). Within NSW however there is no central authority to coordinate custodianship and they have currently taken advantage of the OSDM’s custodianship guideline.

The tools for data exploration are provided to some extent within different jurisdiction in Australia, but still they are not comprehensive and complete, with no single point for data exploration. Sometimes users need to search more than one resource to find the data they require. Non-frequent users who are not familiar with data providers especially face more difficulties in finding data.

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. However, metadata provided by ANZLIC core metadata profile (ANZLIC 2006) does not contain information required for data assessment. In order to evaluate data for fitness-for-purpose information on data quality, data attributes and some other technical information including distribution format, scale and datum are required. Also legal issues associated with data sets including privacy and IP issues highly affect the data suitability for purpose.

ANZLIC’s core metadata profile contains some of the mentioned data, but there is no comprehensive information for this purpose. ANZLIC’s metadata profile for example does not provide information on attributes and reference to privacy and legal issues (ANZLIC 2006). Some other metadata standards including Content Standard for Digital Geospatial Metadata (CSDGM) which is used in Canada and America provide attributes, legal issues surrounding data and much more information on data (FGDC 2006). NSW has maintained a
metadata profile which contains detailed data to feature level on datasets, but it still does not contain legal issues associated with datasets (NRIMS 2006).

**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 mostly 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.

In Australia ANZLIC’s metadata template contains a section on contact information. Any data which has metadata based on ANZLIC has information on the data provider, custodian and contact person/sector; however it is often outdated and incomplete. Communicating through value-added resellers is another way to communicate with data providers. Most governmental data providers in Australia distribute data through VARs, however they also sell data themselves, but most data distribution is done through VARs.

If a user has access to a comprehensive and complete list of datasets with information on how to communicate with respective data provider/VAR and access data, it is more convenient not only to find data but also to communicate with data providers and access data. The most comprehensive data directory in Australia is the ASDD, however it suffers from a lack of comprehensive metadata and also from outdated data, especially outdated contact information.

Some states provide data directories including NSW’s natural resources data directory which provides metadata directly, while in some other cases including South Australia (SA), there is no state level data directory and metadata searches are done through the ASDD.

**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 webservices and single point of access, but also providing non-technical mechanisms including legal, social, policy and institutional considerations to facilitate data access and acquisition.

There are different challenges which data users are faced with. The users of data prefer to interact with a single access point rather than different organizations, especially for data consumers including PSMA which rely on more than one data source and need to deal with many and different organizations at different jurisdictional levels to collect their data. One of PSMA’s products is Points Of Interest (POI), which contains more than twenty different features. Custodians of each feature may differ. Exploring and finding the custodian at each state level is not easy.

Applications which require accurate and detailed data need to deal with Intellectual Property (IP) and associated legal issues. Legal considerations sometimes avert users form data acquisition. Prices and privacy policies should also be taken into consideration.

**Geo-processing**

Applications which rely on more than one data source mostly encounter inconsistency of datasets. Data providers follow their own policies, standards and specification in creating and maintaining data which may differ from other data providers. Inconsistencies in format, data model, 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 original datasets’ privacy and IP issues effects final integrated products. PSMA’s POI is a good example in which datasets from different authorities with different legal arrangements get together to form a seamless product.

Given the complexity of technical and non-technical issues associated with geo-processing, this task is not possible if there is not comprehensive information about data characteristics including datum, data model and attributes and associated non-technical issues which affect data conversion. Geo-processing provides prerequisites for data collation.

**Data Collation**

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 (visualization and data model) levels for data integration (Figure 5). 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 5. Levels of Data Collation, Adopted from Kolbe (2006)](image)

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.

Effective software which implements data collation properly needs to deal with vertical topology, data model inconsistencies, datum and projection system differences, and format, accuracy, and specification heterogeneities.

The six tasks discussed above in relation to data integration and associated issues have been summarised in Figure 6.
Figure 6: Issues associated with data integration

If the issues discussed in relation to data integration above are taken into account, then frequent users and consumers of spatial data will expend less effort and time in accomplishing the six steps. This will save time, resources and money.

CONCLUSION

Within Australia, many applications heavily rely on spatial data to respond to the complex nature of decisions on different aspects of the environment. Making such decisions is not possible unless there is an ability to control the environment through the integration of its built and natural components. The integration of built and natural datasets is also essential for facilitating the Australian federal and state governments’ priorities including hazard management, environment protection and land administration. Hence, data integration is a significant issue in Australia.

In Australia each state is responsible to undertake its own mapping and data coordination needs. Within each state, local councils are also a major provider and consumer of spatial data. They liaise with states to capture and maintain spatial data for their use. As there are many players within each jurisdiction, the inconsistencies of approaches among different jurisdictions cause heterogeneity among datasets and associated policies. Inconsistency of spatial datasets has turned integration to an expensive and time-consuming task. Inconsistencies are barriers for different integration tasks including data discovery, data assessment, communication, data access and acquisition, geo-processing and data collation. The investigation of issues associated with these six tasks is necessary for effective data integration.

Different steps of integration are associated with complex range of issues including technical, institutional, policy, legal and social issues. Addressing data integration issues and providing tools and mechanisms to facilitate data integration is one the most important aims of any SDI.

In Australia, technical interoperability has been well received and developed within current SDI initiatives, while developments and investigations in the area of non-technical interoperability have not. Only when both technical and non-technical issues are considered together, will integratability of spatial data amongst jurisdictions in Australia be well-developed.
ACKNOWLEDGEMENT

This paper is part of an ongoing research project on data integration, 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 and also project partners 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 or organizations mentioned.

REFERENCES


