Enabling Spatial Data Sharing through Multi-source Spatial Data Integration

Hossein Mohammadi, Abbas Rajabifard, Ian P. Williamson
Centre for Spatial Data Infrastructure and Land Administration
Department of Geomatics
University of Melbourne

Abstract

The dynamic environment of SDIs and the involvement of diverse spatial data providers present uncertainty for involving organizations. This pushes organizations to focus on cooperative data sharing relationships to deliver their objectives. Spatial data sharing provides transactions in which individuals, governments and businesses obtain access to spatial data and services from other stakeholders. However, spatial data sharing goes beyond simple data exchange and requires the provision of usable datasets. It is specifically important at multi-national level and Global SDI (GSDI).

One of the most significant and demanding characteristics of usable datasets is the readiness of spatial datasets for integration with other datasets. However it is often difficult or even impossible for users to sensibly integrate datasets from different sources. This is because of the diversity of data standards, specifications and arrangements which have been utilized by organizations. Data providers adopt spatial data standards and specifications and establish data sharing arrangement based on their requirements which may differ form other organizations. Therefore, multi-source spatial datasets are associated with technical and non-technical inconsistency and heterogeneity.

In order to facilitate the integration of multi-source spatial datasets, the investigation of the data integration process, potential barriers and challenges of spatial data integration and possible enablers and solutions is necessary.

This paper aims to provide an investigation on the spatial data integration as a compelling reason for spatial data sharing. The investigation approach is based on a number of case studies. The case study investigation has also highlighted and identified a number of technical and non-technical barriers and issues of multi-source spatial data integration. The paper also capitalizes on the analysis of the case study investigation to identify the possible tools, solutions and enablers which can be utilized to facilitate the integration of multi-source datasets. In this regard, the paper presents a spatial data integration toolbox. The toolbox consists of a number of components including spatial data validation and integration tool, associated guidelines; and data integration metadata and data specification documents. The design and development of a spatial data validation and integration tool and associated guidelines have also been presented in the paper.

Keywords: Spatial Data Infrastructure, Data Sharing, Multi-source Spatial Data Integration

1 Corresponding author, Tel: ++61 3 8344 0234, Fax: ++61 3 9347 2916, email: abbas.r@unimelb.edu.au
1. **INTRODUCTION**

Spatial data plays a significant role in delivering effective government services, informed decision makings and creating business opportunities. At the same time, the capacity to meet such user needs and to deliver services and tools within the spatial information community has gone far beyond the ability of single organizations (Rajabifard et al., 2005), especially when more value-added and integrated spatial data is required for more complex analysis. Hence, organizations move toward the sharing of spatial datasets and collecting and integrating spatial data from different sources. But the diversity of data providers, data consumers and their needs and characteristics leads to diversity in many aspects including different standards, policies and institutional arrangements (Mohammadi et al, 2008).

The diversity of approaches taken by different stakeholders and providers make multi-sourced spatial data inconsistent. The inconsistency emerges in many ways. Some spatial data sets do not comply with common standards and technical specifications. Different institutional arrangements of data providers can be also an obstacle. In many cases, this leads to weak collaboration and liaison among different stakeholders which hinders effective spatial data integration. This is obviously crucial when some spatial data applications such as emergency management require fast and on-time access to and integration of datasets. Therefore, without a comprehensive investigation of technical and non-technical issues within a single framework, effective spatial data integration can not be achieved.

SDIs promise to provide an enabling platform to facilitate the sharing and integration of spatial datasets. In order to achieve this aim spatial data integration requires a comprehensive investigation within the framework of SDI Initiatives.

2. **SPATIAL DATA SHARING**

SDIs is thought of having dynamic structure. This is addressed by both change in the nature of SDIs and the external environment including the advancement in technology (Rajabifard et al. 2002). The dynamic environment of SDI presents uncertainty for involving organizations (Omran, 2007b) which leads organizations to focus on cooperative relationships. One of these relationships is data sharing. Spatial data sharing has been defined as transactions in which individuals, organizations or parts of organizations obtain access from other individuals, organizations or parts of organizations to spatial data (Omran, 2007a).

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

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

The value of a spatial dataset rests on its 'coverage, the strengths of its representation of diversity, its truth within a constrained definition of that word, and on its availability' (Longley et al., 2001) and the integrability with other datasets (Rajabifard and Williamson, 2004). In many cases, it is often difficult or even impossible for users to sensibly combine data from different sources (Ryttersgaard, 2001). Muggenhuber (2003) identifies the continuum of spatial data management within the context of SDIs. He explains the progress made and highlights the current challenges. In this progress, GIS attracted the focuses on implementation and tuning...
of standalone tools. The next wave began with the availability of large amounts of spatial data in form of digital data. In last decade, the stakeholders realized the urgent need for sharing spatial data with other stakeholders. This aim was highly dependent on institutional arrangements, which requires cooperation and partnership. The current demand is to provide integrable and harmonized spatial data for broad and maximum use (Figure 1).

Figure 1. Spatial Data Management Continuum (after Muggenhuber, 2003)

One of the major international challenges of building SDIs is linking distributed heterogeneous spatial information resources from different data providers in an application-oriented and user-oriented way (Donaubauer, 2005). It is also inline with the goal of SDI to reduce duplication of effort among agencies, improve quality and reduce costs related to geographic information, to make geographic data more accessible to the public, to increase the benefits of using available data, and to establish key partnership with states, counties, cities, tribal nations, academia and the private sector to increase data availability (NRC, 2003).

3. SPATIAL DATA INTEGRATION

Due to the ever-increasing complexity in spatial analysis and high cost of data production, sharing and integrating of spatial data is necessary. Therefore, the data integration and interchange between organizations became an essential issue for SDI Initiatives (Pinto et al., 2003). Organizations with interests in spatial information are no more able to produce and maintain necessary spatial data. Therefore, multi-source spatial data integration has become the fundamental motivator throughout the organizations.

Different researchers have different perception of data integration and expect different results from spatial data integration. Uitermark (2001) defines spatial data integration as the establishment process of relationships among corresponding object instances in different, autonomously produced, geographic data sets of the same geographic space. Usery et al. (2005) have investigated spatial data integration and have defined it as the process of matching different datasets geometrically, topologically and establishing correspondence of attributes.

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

In some literature spatial data integration has been defined as a process. Samadzadeghan (2004) describes data integration as a number of processes including the acquisition, processing and synergistic combination of multi-source datasets. Ronsdorf (2005) believes that in all applications the vital part is to fit multi-source data spatially in order to facilitate the joint use and analysis.
Some researchers believe that institutional component is a vital component to be considered in order to fulfill effective data integration. Ryttersgaard (2001) highlights institutional component of data integration alongside with technological developments as the most important issues for efficient use of spatial datasets. Masser (2006) and Williamson et al. (2003) address institutional structure as a facilitator for spatial data integration. Van Leonen and De Jong (2007) identify institutional setting which can highly affect and facilitate data usage. Capacity building and raising awareness has been also highlighted in some literatures as a social activity to assist spatial data integration (Alexiadou and Rajabifard (2006), Williamson et al. (2003)). Some studies (Burrough and Masser, 1998 and Van Leonen, 2003) address legal component as important as technical component for effective spatial data integration. Van Leonen (2003) has identified some policy components including pricing, access and which should be considered for comprehensive spatial data integration framework.

All the definitions, outlined above, represent a particular aspect of spatial data integration. This paper capitalizes on the researches and studies in the area of spatial data integration and introduces a more comprehensive definition which overarches complex components of spatial data integration.

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

4. SPATIAL DATA INTEGRATION ISSUES AND CHALLENGES

Different jurisdictions encounter variety of issues and challenges when attempting to integrate multi-source spatial data. In order to investigate the effective factors within different jurisdictions, a number of case studies have been conducted. In this regard, seven countries (Japan, Singapore, Australia, Brunei Darussalam, Indonesia, Malaysia and Philippines) in the Asia and the Pacific region were selected through the channel and support of UN-sponsored PCGIAP (Permanent Committee on GIS Infrastructure for Asia and the Pacific). The case studies have aimed to investigate the issues and barriers of spatial data integration together with possible solutions and enabling tools within the countries with different social, governmental and geographical characteristics. Mainly, practitioners who are involved in SDI development have been involved in the case studies. The case studies also aimed at investigating the framework data which are common in SDIs.

An integration template has been designed as a standardized generic proforma to enable the discovery of information within case study jurisdictions, including matters concerned with member countries’ spatial information policies, laws and regulations, infrastructure implementation, institutional arrangements, technology, and integration issues as well as human resource and capacity building for spatial data integration.

Secondly, a questionnaire aimed to identify the main issues of integration in five major categories of technical, institutional, policy, legal and social collected information on individual items. The reports and questionnaire show the importance of solving non-technical issues along with technical issues for effective data integration (Figure 2).
Through the case studies, it was highlighted that the priorities and major challenges of each case study country drive spatial data integration initiatives. The applications and services that are developed revolve around these drivers and in respond to the national and societal priorities.

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

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

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

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

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

Figure 3 summarizes the spatial data integration drivers and also challenges within the case study countries.

The use of a SDI as the catalyst for data integration would enable users to overcome the issues and challenges of spatial data integration and then reduce duplication of effort and expense in integrating data. For this to occur effectively however, socio technical issues such as immature institutional arrangements, inconsistencies and incomplete knowledge about the availability and quality of data along with technical issues need to be resolved. The re-engineering of SDI initiatives must take these issues into account if the integration of multi-source data is to be
achieved. In this regard, an integration toolbox can assist SDIs to facilitate the integration of multi-source spatial datasets.

The spatial data integration toolbox can provide a gateway between data providers and data users. This gateway can reduce the time, cost and effort of data harmonisation. This also helps SDI developers to deliver consistent datasets to users. The toolbox includes the following components (Figure 4):

- Data Validation and Integration Tool,
- Data Integration Guidelines,
- Integrated Data Model,
- Consistent and conceptual spatial (geographical and attributes) data specifications,
- Reliable, machine readable and consistent Metadata standard and content; and
- etc.

Figure 4. Spatial Data Integration Toolbox within the Context of SDI Initiatives

5. SPATIAL DATA INTEGRATION TOOLBOX

The spatial data integration toolbox is a suite of technical products and non-technical enablers and mechanisms that aim to facilitate the integration and sharing of the multi-source spatial datasets in the context of SDI. Effective spatial data sharing is one of the major aims of SDIs. It ensures use and access of spatial data by a broad range of users. In this regard, SDI aims to facilitate the sharing of spatial data. Spatial data sharing aims to provide usable multi-source datasets to a broader range of users. It includes the interaction between useable spatial data and the stakeholders through a number of technical tool and non-technical enablers. In this regard, spatial data integration should be facilitated for utilisation and use of multi-source spatial datasets to their maximum potential.
The guideline provides a methodology for multi-source spatial data integration together with necessary tools to overcome potential technical and non-technical barriers. The guideline also elaborates the technical tools together with non-technical mechanisms and approaches which can be utilized to overcome the barriers.

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

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

The design and development of spatial data validation and integration tool is a piece of software as a data sharing gateway. The tool evaluates spatial data against a number of integration measures. This includes a number of customizable rules such as compliancy with specific projection system, currency, accuracy, geographical extent and aspatial restrictions. The compliancy of the datasets to the measure represents the fitness of datasets for data sharing and integration purposes. The compliancy to the measures, also allows the data to be a part of the data collection component of the SDI platform. Next section presents the design and development of a prototype for spatial data validation and integration tool.

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

Usable data has a number of characteristics which are defined within the context of respective jurisdictional SDI (Backx, 2005). Usable data should comply with the rules and measures which have been defined within the SDI initiatives. It includes different technical and non-technical characteristics such as certain format(s), datum(s), metadata content, restrictions on data use, quality (spatial and aspatial accuracies, currency and coverage); and pricing policy etc. In order to examine and assess the compliancy of datasets against these measures, data validation module evaluates datasets and data integration module provides necessary functions to amend datasets based on the measures and integrates datasets.

Spatial data validation and integration tool provides a number of functionalities in order to facilitate the assessment and amendment of multi-source spatial data for integration. This includes:

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

5.1. Workflow of the Prototype

The most difficult part of any design is the understanding of tasks and requirements. In most cases the problem statement for system requirements is vague (Stevens and Pooley, 2006). In order to make a solid design of the system and before agreeing whether to tackle the design, a detailed analysis of the requirements of the system is needed. The workflow of the system has been illustrated in Figure 5.

Figure 5. Spatial Data Validation and Integration Prototype Scenario
Based on the requirement of the prototype, there are a number of associations between actors of the prototype system and use cases can be identified. Three actors of the system are system administrator (SystemAdministrator), who makes configurations based on the respective SDI and spatial community. For example in case of Australia, metadata should comply with the ANZLIC metadata profile, datasets should sit within the boundaries of Australia, certain pricing and privacy policies are applicable to datasets. The administrator also defines evaluation measures. It includes certain data accuracy, datum and restrictions on data etc. Data providers (DataProvider) evaluate spatial data. It also entails the provision of spatial data and metadata. Then, a report is created which presents the result of evaluation. Based on the evaluation results and in case of inconsistency with measures, available guidelines are assigned to the inconsistency items. It may include technical solutions or non-technical consideration to overcome the inconsistency. DataProvider then can manipulate the data to meet the requirements of the prototype and if no consistency is recognized, data can be registered as a compliant data. Another actor in the system is general system user (SystemUser), who can get integrable data from the data registry. The above-mentioned use cases are summarized as follows:

In order to identify the classes and relationships between them, the entities which interact within the prototype have been singled out. It includes following classes:

- User with three subclasses of SystemAdministrator, SystemUser and DataProvider
- Measures: defined by SystemAdministrator
- Rules: defined by DataProvider
- DataRegSession: data registration session started by DataProvider
- SpatialDataLayer: spatial data layer which is evaluated
- Metadata: metadata for spatial data layer
- MetadataStandard: metadata standard which defines the structure and content of metadata
- DataSpec: data specifications
- Report: to present the results of the evaluation
- Inconsistency: items of inconsistency between data characteristics with measures and rules
- Guideline: assigned available guideline for items of inconsistency
- ManipulationTool: assigned applicable manipulation tool for technical inconsistency

Capitalizing the strength of ArcGIS and VBA programming for ArcGIS, the prototype has been developed. This included:

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

The major task in implementation phase had been the conversion of classes which have been designed as conceptual model to physical classes in the VBA
environment. This included the conversion of UML modeling language to VBA programming syntaxes.

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

- Passing configurations and parameters between interfaces,
- Data and metadata content investigation,
- Assessment of measures with data characteristics through queries and comparisons,
- Establishment of an effective link with the database to keep the record of objects and also to retrieve existing records; and
- etc.

5.2 Prototype User Interfaces

Alongside with the design and development of the objects and messages, user interfaces are also developed. The user interfaces provide communication channel with users of the system, whenever there is a need to get some input from users (such as data layer and metadata location) or whenever a result should be communicated with users (such as report on the result of validation).

In this regard a number of interfaces have been developed to interact with different type of users. There is a first general user interface which is used to determine the access level of the user including system administrator, data provider and system user. This interface (UserAccessInterface) obtains username and password from users. Based on the existing users' information in the database, the tool allows the user to follow the next steps. Other interfaces are developed to facilitate the use and interaction with system for three above-mentioned user categories (Figure 6).

**Figure 6. Prototype user interfaces and interaction between them**
Major part of the prototype has been designed and implemented for users to evaluate and register compliant data in the system. For this purpose, six interfaces have been developed as follows:

- ConfigurationInterface,
- AgreementInterface,
- InformationInterface,
- ReportGuidelineInterface,
- Re-evaluationInterface; and
- RegistrationInterface.

Through ConfigurationInterface (Figure 7), data providers select corresponding contextual configuration which has been already defined by system administrators. It includes selection of jurisdiction, metadata standard, policy and restrictions.

![Figure 7. Configuration User Interface](image)

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

- Data custodian
- Pricing
- Currency
- Scale
- Completeness
- Attribute accuracy
- Spatial accuracy
- Logical consistency

If data providers agree with the conditions, next interface provides data providers with an interface (InformationInterface) which allows them to enter data and
metadata location as illustrated in Figure 8. Based on the configuration which has been selected in previous step, corresponding information on data characteristics including format, geographical extent and attribution rules etc. are extracted from actual data and metadata. The same items which represent counterpart values and quantities to the evaluation measures can be extracted from data and metadata. Actual data contains information on some of these values. It includes geographical extent, format; attribution content and spatial reference system. This interface also allows data providers to define more rules to restrict attribute content. It is a very useful functionality as check the quality of attributes.

Figure 8. Information User Interface

![Image of Information User Interface]

ReportGuidelineInterface (Figure 9) presents the results of data validation which have been produces through comparing the measures with the information which has been extracted form data and metadata. The report also highlights the items of data inconsistency with measures. The report also assigns available and applicable guidelines to the inconsistency items. Availability of the guideline refers to a link to guideline in guideline table of the prototype database.

Figure 9. Report on the Evaluation Results

![Image of Report on the Evaluation Results]

If data layer has been amended and provided for re-evaluation, the same evaluation process is executed on data and if there is no more inconsistency (Figure 10) with measures, data layer is displayed in the ArcGIS environment (Figure 11) and a data record is also added to the database.

Figure 10. Re-evaluation after Data Amendment
5.3 Prototype Use Test

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

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

Figure 11. Data Layer Representation in the ArcGIS Environment
The dataset was tested and the outcome showed that the geographical extent is the same as the tools configuration, but the datum was WGS84, the restriction was “subject to license” and a number of features had the attribute with null value. For the datum the specification GDA’s technical manual was proposed and the null values were reported. For the restriction as it was in a lower level of restriction, it was accepted. Table 1 shows the rules which were set as default rules and also the characteristics of the dataset. It also highlights the items of incompliancy with the defined rules.

Table 1. The Result of Data Validation and Integration Prototype Use Test

<table>
<thead>
<tr>
<th>Measures</th>
<th>Jurisdiction-defined</th>
<th>Dataset</th>
<th>Incompliancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accepted Formats</td>
<td>ESRI shapefile, ESRI coverage, WFS/WMS, ERMapper ECW</td>
<td>MapInfo TAB</td>
<td>X</td>
</tr>
<tr>
<td>Metadata Standard and DTD/Schema</td>
<td>ANZLIC’s profile, anzmeta1.3.dtd</td>
<td>ANZLIC’s profile, anzmeta1.3.dtd</td>
<td>✓</td>
</tr>
<tr>
<td>Policy agreement</td>
<td>Privacy policy Restrictions</td>
<td>Subject to license</td>
<td>✓</td>
</tr>
<tr>
<td>Geographical extent</td>
<td>Australia’s boundaries</td>
<td>Australia, Victoria</td>
<td>✓</td>
</tr>
<tr>
<td>Datum</td>
<td>GDA_1994</td>
<td>WGS86</td>
<td>X</td>
</tr>
<tr>
<td>Currency</td>
<td>Not older than one month</td>
<td>Two weeks old</td>
<td>✓</td>
</tr>
<tr>
<td>Minimum and Maximum scale</td>
<td>1:1,000 to 1:1,000,000</td>
<td>1:25000</td>
<td>✓</td>
</tr>
<tr>
<td>Completeness</td>
<td>complete</td>
<td>Complete</td>
<td>✓</td>
</tr>
<tr>
<td>And user-defined rules:</td>
<td>No “null” is accepted for attributes</td>
<td>Some “null” attributes</td>
<td>X</td>
</tr>
</tbody>
</table>

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

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

5.4 Discussion on Prototype Development

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

With utilizing a tool which contains a set of rules and guidelines, the process of data validation for integration can be much easier automated. The standardized and
routine process proposed by the tool also provides a consistent approach to evaluate different data sets.

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

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

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

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

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

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

6. CONCLUSION AND FINAL REMARKS

Effective spatial data integration ensures effective sharing of usable spatial data among stakeholders. It also facilitates the use of shared datasets with less time and effort. But the integration of multi-source spatial datasets is problematic form technical and non-technical perspectives. In order to investigate the potential barriers and challenges of spatial data integration, the paper presented the conduction of a
number of international case studies. The case studies have identified technical and non-technical barriers of data integration. It includes:

- Data models,
- Metadata concerns,
- Different semantics and representations,
- Inter- and cross-organizational access, retrieval and display arrangements,
- Sharing data among organizations,
- Lack of central access gateway,
- Access policies,
- IP (Intellectual Property) and licensing; and
- Silo mentality.

In order to overcome the barriers of effective data integration, a data integration toolbox has been proposed. The toolbox comprises the spatial data validation and integration tool, spatial data integration guideline, integration data model, data integration-specific metadata and data specifications.

Spatial data validation and integration tool is an integral component of SDI which facilitates data sharing and evaluates the readiness of spatial datasets for integration. The tool evaluates datasets against a set of rules and measures which define usable and integrable datasets. The rules and measures are defined based on the characteristics of the datasets for SDI framework. SDIs define some characteristics for spatial datasets in terms of technical and non-technical rules. The compliancy of spatial datasets to the rules shows the readiness of those datasets for integration purpose.

The developed tool automates the validation of datasets against some measures and rules and facilitates the sharing and integration of spatial datasets. The proposed prototype tool has been designed and developed utilizing software design tools including UML. UML diagrams have been utilized to illustrate the architecture and components of the prototype system. The UML use-case and class diagram have been devised to identify and design the necessary components and classes of the prototype.

The prototype tool has comprised a number of components which enables administrators to set and define configurations and rules based on the requirements of their jurisdiction. It includes metadata standard, geographical extent, data accuracy, scale, restrictions on data and aspatial restrictions etc. The tool also allows the data providers to choose any combination of the rules or define some other rules on aspatial content of data and evaluate the data against the rules. In order to perform this, the tool extracts data characteristics form actual data and metadata. For this purpose, ESRI’s ArcMap GIS engine and an XML processor have been utilized to collect required information. Then the tool evaluates the characteristics of the data with rules and provides the data provider with the results which include any item of inconsistency with the rules. Some data manipulation tools also have been provided to amend and re-evaluate the data. If there is no inconsistency exists, the data can be recorded in a database and provided to the users. The use case on the prototype proved that the tool can facilitate the integration and validation of spatial datasets. The tool saves time and effort which is put for data validation comparing to manual validation. The tool also follows a consistent and structured approach which prevents unnecessary attempts and assists users to find the solution with minimum effort in minimum time. The tool also can be used as an individual data validation tool. The
users can use the tool based on their requirement. As a result, the users can define rules and measures which meets their needs better and asses datasets based on their specific rules.

REFERENCES


SAMADZADEGAN, F. (2004) DATA INTEGRATION RELATED TO SENSORS, DATA AND MODELS. XXth ISPRS Congress. Istanbul, Turkey, ISPRS.


VSC (2005) Outline for a Policy on Data Sharing Frameworks. IN VICTORIAN SPATIAL COUNCIL (Ed.).

Author/s: MOHAMMADI, HOSSEIN; RAJABIFARD, ABBAS; Williamson, Ian P.

Title: Enabling spatial data sharing through multi-source spatial data integration

Date: 2009

Citation: Mohammadi, H., Rajabifard, A. & Williamson, I. P. (2009). Enabling spatial data sharing through multi-source spatial data integration. In GSDI 11 World Conference,

Publication Status: Published

Persistent Link: http://hdl.handle.net/11343/26086

File Description: Enabling spatial data sharing through mult-source spatial data integration

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.