【GIS Industrialization】

Metadata Strategy, Data Directory System and Emerging National Spatial Data Infrastructure in Australia

Wei Pei

GIS Branch, State Forests of NSW
423 Pennant Hills Road, Pennant Hills, NSW 2120 Australia

I. INTRODUCTION

Increasing public and political concern for proper environmental assessment and management, land-use conflicts, the issues surrounding ecologically sustainable development and biodiversity are resulting in a demand for high quality data on Australia's natural resources.

Due to the diversity of natural resources information, the metadata required to describe these resources are equally diverse. Over time an array of metadata formats have evolved which enable various organisations, agencies, academic, community or private industry to tailor metadata to specific needs. In most cases, the capture of metadata is easily and more readily achieved at the highest level, simply because summaries of datasets require less description and more general, while at lower levels, additional information is recorded. The development of any metadata standards must recognise this complex and multi-layer relationship and provide a facility to permit the capture of metadata at different levels and at different times.

In order to achieve a better understanding towards the complexity and multi-layer relationship associated with the capture of metadata, it's necessary to have a brief description of the spatial data directory systems in Australia. Using this as a starting point, metadata strategy for national, jurisdictional, theme-based and custodian agency data directory can be explored.

II. TERMINOLOGIES OFTEN USED IN AUSTRALIA

Metadata: refers generally to dataset descriptions and is equivalent to 'meta-information' (which is possibly a better term to use because 'metadata' is commonly used in data management to refer to data dictionaries).

> The popularly used metadata is normally *feature-based, cover-based, project-based* and *dataset-based*. There are also metadata about hardware and software.

Directory: refers to a collection of descriptions of datasets together with subject and spatial indexing. There are three types of directories most commonly seen in Australia:

> *National directory*: directory managed by Commonwealth government of Australia and functions as the repository of the master keyword thesaurus and gazetteer of spatial objects. An example of national data directory developed at an early stage is: National Directory of Australian Resources (**NDAR**), and **ASDD** at a later stage by National Resource Information Centre (**NRIC**)(1).
>
> *Jurisdictional directory*: directory managed at State/Territory level. Considerable efforts have been undertaken in Queensland, Western Australia, New South Wales, South Australia,

Declaration: The views expressed in the article are those of the author's. They don't necessarily reflect the views of any Australian government organisation or agency.

1082-4006/97/0301~2-60$3.00

Victoria to develop their spatial data directory. For example, in August 1995 Department of Land and Water Conservation(**DLWC**) in the State of New South Wales published the first **NSW Natural Resources Data Directory**. This directory was published in CD-ROM/Disk form for use on IBM PC computers and included an easy to use graphical search capability.

Theme-based directory: directory developed by theme-based organisations. For example, Australian Coastal and Marine Directory (**Blue Pages**), or Australian National Geoscience Information System (**NGIS**)(2).

Custodian agency: an organization which has the custodianships to take responsibility for collecting and maintaining of the dataset.

III. NATIONAL CONTEXT

It is estimated by the Economic Studies and Strategies Unit of Price Waterhouse[3] that for the period 1989-94 approximately $1 billion has been spent in Australia on investment in geospatial data. This investment produced benefits within the economy in the order of $4.5 billion. The study also found that this investment has saved users approximately $5 billion.

The US Government have long recognized the management of geospatial data as an important national issue and as early as April 1994 an Executive Order was made by President Clinton on "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure".

In Australia there is a widespread support within Government and industry for the development of an Australian Spatial Data Infrastructure (**ASDI**)(3) in order to support decision making, facilitate timely access to and share reliable consistent geospatial data by government agencies and community, and maximize integration of datasets . Development of the ASDI is being managed by the Australia New Zealand Land Information Council (ANZLIC).

This emerging ADSI will comprise a distributed network of databases, linked by common policies, standards and protocols to ensure compatibility. It will eventually become an information window to Australia's physical, natural and administrative environment.

IV. EVOLVING '*directory*' CONCEPT

An important component in this emerging ASDI is the implementation of a "national directory system" with a distributed network structure to help access Commonwealth and State/Territory spatial data.

The concept of national directory is evolved from **'central databank'**, a facility first initiated by the National Resource Information Centre in 1988 (NRIC, a branch of Bureau of Resource Sciences, federal Department of Primary Industries and Energy (**DPIE**)(4)). By 1990, it is widely accepted that **'national directories/indexes'** for land information management are necessary to facilitate decision making. The original design for a national directory was for a single centralised system with NRIC carrying out all the metadata collection and maintenance, which later turned out to be a large task far beyond the resources of NRIC.

Under the recognition that a dataset description is best maintained by that dataset's custodian, this led to a concept of **'node model'**. Each State/Territory directory would form a node with State/Territory being responsible for the maintenance of the directory entries at that node. Ideally, these node directories would then be accessed from one or more user interfaces. The node provides periodic updates to the national directory, creating an accumulation of all relevant information held at the node level — thus forming a **'distributed directory system'**.

The optimum fully distributed system is currently being evaluated by the ANZLIC Metadata Working Group and is expected to be progressively implemented from February 1998.

Prior to full implementation of a fully distributed system, a process common software has been developed by NRIC to facilitate the transfer of dataset descriptions between State/Territory installation and the national centre. This is **FINDAR directory software the prototype Australian Spatial Data Directory (ASDD) system** (Facility for Interrogating the National Directory of Australia Resources)(5). FINDAR prototype ASDD allows metadata records from subordinate nodes, which have been flagged as suitable for transfer to the national node, to be uploaded to that node. The software comprises three main components: tables of attribute, keyword and spatial data for the direc-

tory entries; a gazetteer of geographic entities for spatial indexing and searching; and a thesaurus of standard terms for subject indexing and searching.

In addition to the distributed directory system model(NDAR ASDD), a number of high-level spatial data directories have been developed at national, jurisdictional level independent of each other, with some of them theme-based. An example is the Australia Land Information Group(**AUSLIG**) metadata directory(6).

V. ANZLIC GUIDELINE ON CORE METADATA ELEMENTS

A critical step towards the implementation of a national directory system for spatial data is the development of "**ANZLIC Metadata Guidelines on Core Metadata Elements**"(7). The Australia and New Zealand Land Information Council (ANZLIC), is the peak intergovernmental body responsible for spatial data management for Australia and New Zealand. It currently manages a National Strategy for land and geographic information. To support the adoption of the Guidelines ANZLIC also produced and freely distributes a Microsoft Access Metadata Entry Tool.

This is a guideline consistent with the Content Standards for Digital Geospatial Metadata developed by US Federal Geographic Data Committee(**FGDC**), and with the Australia New Zealand Standard on Spatial Data Transfer **AS/NZS 4270**. However, Australia's approach for the mandatory metadata items is deliberately less ambitious than what has been attempted in the US. The US counterpart consists of some 220 items (elements) which are intended to describe digital geospatial datasets adequately for all purposes, while in ANZLIC Guidelines, 31 core metadata items (grouped into 9 categories) are identified as generally common for all types of data. (Table 1).

The International Standards Organization (ISO) has a draft Metadata Standard developed through its Geomatics Technical Committee ISO/TC 211 which has been developed with input from FGDC, ANZLIC, Dublin Core and other metadata guidelines. This will have two levels, a "Cataloguing Metadata" level similar to ANZLIC and Dublin Core and a complete level which includes some 300 elements. It is expected that ANZLIC will adjust the ANZLIC core elements to the ISO Level 1 when finalized.

At the national level, one of the natural resource data directory which has adopted ANZLIC Metadata Guidelines is the Prototype Australian Spatial Data Directory (**ASDD**)(8) managed by National Resource Information Centre (NRIC) on behalf of ANZLIC, a prototype of the new version of national directory system which consists of updated description of Key Commonwealth and State/Territory datasets. It supersedes the former National Directory of Australian Resources (NDAR).

The optimum fully distributed system is currently being evaluated by the ANZLIC Metadata Working Group and is expected to be progressively implemented from February 1998.

The national programs provide the framework for any corresponding activities at State/Territory level. In State of New South Wales, NSW Metadata working group was formed in March 1997 to take advantage of national activities. The NSW Natural Resources Data Directory data collection and conversion to the **ANZLIC Metadata Entry Tool** is now close to completion and the ANZLIC Tool has been supplied for use in several key NSW natural resource agencies(9).

VI. ANZLIC 'PAGES' CONCEPT ADOPTED BY DIFFERENT LEVELS OF NATURAL RESOURCE AGENCIES AND INDUSTRY

In order to include additional meta-information which is not included in the National Metadata Directory System, ANZLIC metadata has adopted a '**Pages**' concept as the basis for a national metadata framework, where more general information is recorded at the highest level (Page 0) and additional information is recorded at lower levels (Page 1, Page 2). A conceptual indication of how the Pages Concept is the foundation of the national directory system is shown in Figure 1.

This 'Pages' concept has been adopted by the **RACAC** (Resource and Conservation Assessment Council) in the Regional Forest Agreements Project (**RFA**) of NSW for documenting additional metadata on datasets, software and hardware used by stakeholders of the project.

Subsequent pages (i.e. Page 1, Page 2, etc.) provide the opportunity for data custodian agencies at the national, state, local government, academic, community or private industry levels to include addi-

Table 1. Core Elements: A Summary

Category	Element	Comment
Dataset	Title	The ordinary name of the dataset.
	Custodian	The organisation responsible for the dataset.
	Jurisdiction	The state or country of the Custodian.
Description	Abstract	A short description of the contents of the dataset.
	Search Word(s)	Words likely to be used by a non expert to look for the dataset.
	Geographic Extent Name(s) OR	A picklist of pre defined geographic extents such as map sheets, local government areas, catchments, that reasonably indicate the spatial coverage of the dataset.
	Geographic Extent Polygon(s)	An alternate way of describing geographic extent if no pre-defined area is satisfactory.
Data Currency	Beginning date	Earliest date of data in the dataset.
	Ending date	Last date of information in the dataset.
Dataset Status	Progress	The status of the process of creation of the dataset.
	Maintenance and Update Frequency	Frequency of changes or additions made to the dataset.
Access	Stored Data Format	The format or formats in which the dataset is stored by the custodian.
	Available Format Type	The formats in which the dataset is available, showing at least, whether the dataset is available in digital or nondigital form.
	Access Constraint	Any restrictions or legal prerequisites applying to the use of the dataset, eg. licence.
Data Quality	Lineage	A brief history of the source and processing steps used to produce the dataset.
	Positional Accuracy	A brief assessment of the closeness of the location of spatial objects in the dataset in relation to their true position on the Earth.
	Attribute Accuracy	A brief assessment of the reliability assigned to features in the dataset in relation to their real world values.
	Logical Consistency	A brief assessment of the logical relationships between items in the dataset.
	Completeness	A brief assessment of the completeness of coverage, classification and verifi cation.
Contact Information	Contact Organisation	Ordinary name of the organisation from which the dataset may be obtained.
	Contact Position	The relevant position in the Contact Organisation.
	Mail Address 1	Postal address of the Contact Position.
	Mail Address 2	Aust and NZ: Optional extension of Mail Address 1.
	Suburb or Place or Locality	Suburb of the Mail Address.
	State or Locality 2	Aust: State of Mail Address. NZ: Optional extension for Locality.
	Country	Country of the Mail Address.
	Postcode	Aust:Postcode of the Mail Address. NZ: Optional postcode for mail sorting.
	Telephone	Telephone of the Contact Position.
	Facsimile	Facsimile of the Contact Position.
	Electronic Mail Address	Electronic Mail Address of the Contact Position.
Metadata Date	Metadata Date	Date that the metadata record for the dataset was created.
Additional Metadata	Additional Metadata	Reference to other directories or systems containing further information about the dataset.

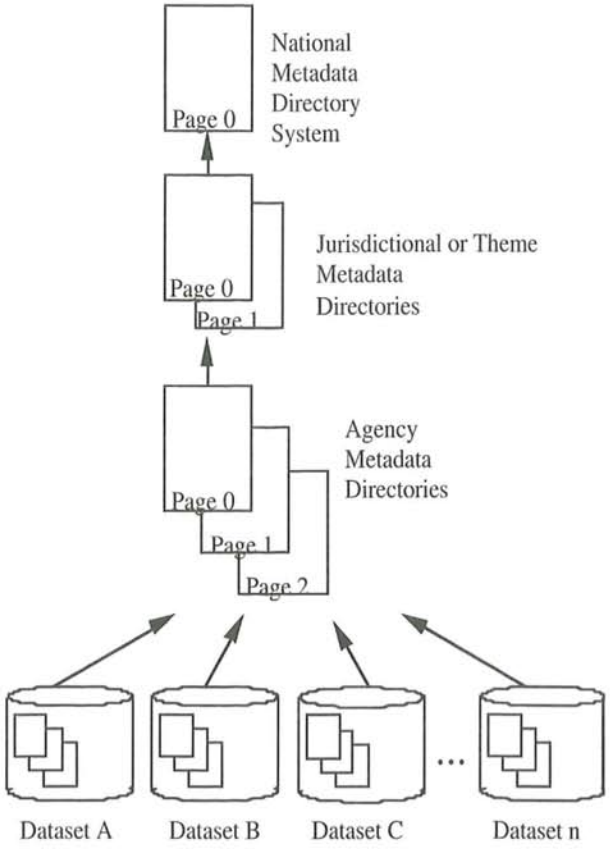

Figure 1. The Pages Concept

tional information not required in Page 0. This additional information may be in the form of sub-elements of specific Page 0 core metadata elements or entirely new and unrelated metadata elements. However, in order to ensure uniformity, it is suggested that any new metadata elements should be consistent where possible with corresponding metadata elements in the FGDC Content Standards.

VII. FEATURE-BASED METADATA SYSTEM AS BASIC META-INFORMATION

In order to retain corporate knowledge of the characteristics and processing history of datasets that an agency uses, the most detailed level of metadata is needed to ensure the efficient management and effective utilization of data within a custodian agency.

A **feature-based metadata system** is developed by the GIS Branch of the **State Forests of NSW**, with ESRI Australia Pty Ltd providing consulting services. The system is built on the top of Arc/Info GIS software and maintained by State Forests of NSW. The design of the metadata system is to tag every arc in an Arc/Info coverage with a metadata-id, which is stored in metadata table(.MDA). This .MDA table is linked to arc attribute table(.AAT) with metadata-id as common item. Therefore this allows the attributes in the .MDA to appear to be part of the .AAT. The tagging of arc with metadata-id is carried out during the process of data capture and literally built in the process of data capture, making this a compulsory step in data collection.

With this metadata system to provide the detailed, comprehensive information on data characteristics and data processing history, metadata required for higher level directory systems can be summarized or rolled up from the feature-based system.

VIII. CONCLUSION

There is a close link between metadata standards and the structure of spatial data directory system. Metadata standards and directory management software should be designed to provide flexibility in describing datasets which exist in a wide range of formats and in meeting the needs of different users. It should also facilitate the capture of metadata at different levels and different times.

It is widely recognized that metadata is best collected and maintained by custodian agency. Data custodian is also at the best position to summarize low level metadata (such as the feature-based metadata) into higher levels, for example, cover-based, project-based and dataset-based. These higher level metadata can then be submitted to nodal data directory. Here the items that are suitable for parent directory be tagged and uploaded into data directory above. Through the transfer of information towards higher levels, a summarisation of information has been achieved. At the peak, national data directory holds an accumulation of metadata stored at its subordinate directories and duplication of data storage is avoided. It also leaves the metadata maintenance to node/custodian, thus facilitate timely update of national data directory.

The distributed data directory system endorsed by ANZLIC is the directory model that could facilitate the flow and accumulation of metadata information. Combining with directory management software as interface, data summarisation, keyword indexing and spatial searching can be automated.

From ANZLIC experience, it is advisable to keep compulsory metadata items to the generally common, basic items that are common for all types of datasets. With the "Page" structure, additional metadata information can be accommodated by adding "pages" when more detailed information is required. The 'Pages' approach endorsed by ANZLIC is an important step towards the forming of metadata standards with flexibility as well as uniformity and consistency.

In order to ensure the speedy implementation of metadata standards, metadata entry tools should be produced by peak government organisations and freely distributed to all the major government agencies.

As stated in draft ASDI, the future trend is to develop a distributed network of databases, linked by common policies, standards and protocols to ensure compatibility and timely update of spatial information.

IX. HOW TO GET MORE INFORMATION ABOUT METADATA ACTIVITIES IN AUSTRALIA ?

Here is a List of Useful Web Sites and Metadata On-Line Directories (by the sequence they appear in the text). The Environmental Resources Information Network (**ERIN**) internet site supports a collection of pointers to "standards" servers and is a good starting point for surfing the internet for data standards. Please check it up in reference (10).

(1) National Directory of Australian Resources (NDAR) developed by National Resource Information Centre(**NRIC**):
http://www.nric.gov.au/nric/data/ndar_overview.html

(2) ANZLIC www sites of Related Interest supports a collection of pointers to various jurisdictional, and theme-based natural resources data directories:
http://www.anzlic.org.au/anz_site.htm

(3) Commonwealth Spatial Data Committee(**CSDC**), Draft Commonwealth Position Paper on the Australia Spatial Data Infrastructure(**ASDI**), Draft 4, 17 November 1997.
http://www.auslig.gov.au/pipc/csdc/sdi4.htm

(4) Bureau of Resource Science (**BRS**):
http://www.brs.gov.au

(5) Johnson, B.D., Shelley, E.P., Taylor, M.M., & Callahan, S., 1991 — The FINDAR directory system: a meta-model for metadata. In Medyckyj-Scott, D, Newman, I, Ruggles, C, & Walker, D, 1991 (Eds) — **Metadata in the Geosciences**. Group D Publications Ltd, Loughborough, UK.

(6) Australia Land Information Group(**AUSLIG**) Metadata: http://www.auslig.gov.au/meta/meta.htm

(7) **ANZLIC** Metadata Guidelines on Core Metadata Elements from ANZLIC (Australia New Zealand Land Information Council):
http://www.anzlic.org.au/metaelem.htm
http://www.anzlic.org.au/anz_sdts.htm
http://www.anzlic.org.au/anzstrdd.htm

(8) Prototype Australian Spatial Data Directory (**ASDD**), supersedes NDAR, currently being populated with records which follow the ANZLIC Core Metadata Guidelines, managed by National Resource Information Centre(NRIC), supercedes NDAR in reference (1):
http://www.nric.gov.au:80/nric/data/data.html
http://purl.nla.gov.au/net/asdd

(9) Draft New South Wales Natural Resources Metadata Management Strategy, 1997, NSW Natural Resources Information Management Steering Group (**NRIMS**) NSW Metadata Working Group.

(10) A collection of pointers to "standards" servers from Environmental Resources Information Network(**ERIN**):
http://www.environment.gov.au/gis/gis_standards.html

(11) A discussion forum on ANZLIC Guideline: Core Metadata Elements supported by ERIN:
Subscribe by sending an email containing the message, "subscribe ozmeta-l", to **majordomo@erin.gov.au**