



MODULE 2: Data Management Principles

Building capacity to implement natural resources information management systems.

www.nlwra.gov.au

MODULE 2

Table of Contents

	Guide for managers	ii
	Context	ii
	Actions	ii
	Acknowledgements	iii
	Guide to symbols	iv
2.1	Introduction	1
	2.1.1 What is data management?.....	1
	2.1.2 Guiding principles	1
2.2	Why do we need to manage data?	2
	2.1.3 File naming conventions and directory structure	2
	2.1.4 Key drivers for improved data management.....	4
	2.1.5 Statutory requirements for data and information management	4
	2.1.6 Benefits of good data management	5
2.3	Principles of good data management	6
	2.1.7 Data policy	6
	2.1.8 Data ownership.....	6
	2.1.9 Data documentation and metadata compilation	7
	2.1.10 Data quality, standardisation, harmonisation and audit	7
	2.1.11 Data life-cycle control	8
	2.1.12 Data custodian.....	8
	2.1.13 Data access and dissemination.....	9
2.4	Establishing a data policy	9
	2.1.14 Data acquisition	9
	2.1.15 Fitness for purpose and point of truth.....	10
	2.1.16 Data care (custodial duties)	10
	2.1.17 Data use and exchange.....	11
2.5	Implementing data management – key roles and responsibilities	11
	2.1.18 Data management ‘champion’.....	11
	2.1.19 Data policy manager	12
	2.1.20 Data custodianship.....	12
2.6	Additional support	13
2.7	Glossary	14
	Attachment 2–1 Data cleansing approach of Glenorchy City Council	15
	What is data cleansing?	16

Product Number: PN21205

ISBN: 978-0-642-37155-3

Guide for managers

Context

All agencies, departments and resource centres collect, create, store and utilise data in some form—most of which has been obtained at considerable cost. It has been estimated that 90 per cent of the costs of establishing a GIS can be attributed to the development of the thematic datasets, and that in excess of 80 per cent of all data have some geographical components and can therefore be referenced to geographical locations such as points, lines or areas.

Experience reveals that treating data as long-term assets and managing them within a coordinated framework produces considerable savings and ongoing value.

Most jurisdictions throughout Australia have established policies governing access and use of data and information, along with custodianship guidelines and information management strategies. Proper processes and procedures for data and information management are the foundation of an efficient management system.

This guide provides background information on 'best practice' for managing data as assets and valued resources. The primary audience is those responsible for managing spatial data and information, although the principles described are equally applicable to other types of data.

Policies and procedures are required to guide the transition from tactical, project-based data collection and management to a strategic information infrastructure that will inform decision making on a wide range of issues. In many situations data are incomplete, not easily accessible, not up-to-date, and often lack any documentation on accuracy and reliability. As such, utilising the data outside of the immediate discipline or agency that collected them can be difficult.

Module 2: Data management principles provides a framework within which the activities of data collectors, managers and information providers can be integrated.

If organisational management is convinced that the tangible benefits from investing in data management outweigh the costs, then it will be given the priority it requires.

Actions

Managers should focus on the need to achieve an integrated information management solution. Such a solution would successfully combine leadership, people, computer hardware and software applications and data into a framework ensuring the appropriate tools and rules are in place to maintain the data and turn them into useful information products in support of operations and decision making.

The goal will be achieved through the formalisation of an infrastructure, production of guidelines, and the development of standards and procedures to support data management and processing. Managers should facilitate the development and implementation of a data policy which addresses the following key elements:







- creation of an easily accessible data system which can efficiently disseminate data collected as part of project activities to the widest range of users
- development of core datasets as baseline products
- provision of 'best practice' quality assurance mechanisms and procedures to produce validated, well-documented datasets that meet priority information requirements
- archiving of all data collected to ensure their availability for multiple use and to safeguard the investment for future use
- improvement of the effectiveness and efficiencies of policy and program development through the coordination of data and information activities
- provision of timely and up-to-date data and information products to support a wide range of activities.

Acknowledgements

This module draws heavily on material produced by the UK Intra-governmental Group on Geographic Information working group on Principles and Practice of Geographic Information Data Management. The source of this material is duly acknowledged.

Guide to symbols

The following symbols are used throughout the Toolkit as a guide to users, and draw attention to important issues and information.

	Information which readers should take particular note of
	Best practice information
	Tips for readers—based on experience and aimed at saving time and resources
	Caution—readers are advised that particular care should be taken or that the subject issue may be complex
	Additional information
	Capability raising—used to show a signpost to a higher capability level
Bold Text	Used to highlight a particular issue
Boxed Text	Highlighting of issues specifically related to ANZLIC or the Audit

2.1 Introduction

2.1.1 What is data management?



This document addresses the key aspects of data management giving consideration to the following 'guiding principles':

- Don't re-invent the information management wheel.
- Where possible capture data once for multiple/generic use.
- Avoid duplication in data acquisition—share and co-operate wherever possible.
- Use existing systems/facilities wherever possible.
- Manage data to maximise their value both during and after the project that they were collected for.
- Give priority to the broadest value data—of benefit to multiple processes.
- Develop and implement metadata standards and documentation procedures.
- Where possible develop and implement (or adopt existing) data publishing standards.
- Contribute to long-term strategic goals for data and information management
- Select the most robust organisation with the broadest span of interest as the most appropriate custodian of high-value general use information.
- Reinforce and support data custodians and where possible negotiate access.

2.1.2 Guiding principles

The term 'data management' embraces the full spectrum of activities involved in handling data, including:

- data policy
- data ownership
- data documentation and metadata compilation
- data quality, standardisation, harmonisation and audit
- data life-cycle control
- data custodianship
- data security and access constraints
- data access, data sharing and dissemination/licensing arrangements
- data publishing.



1 ►► 2

Individual people or departments have documented data management processes, policies and procedures. Use the sections in Module 2 as a guide to link the documentation of processes, policies and procedures to actual business processes within an NRM region.

2.2 Why do we need to manage data?

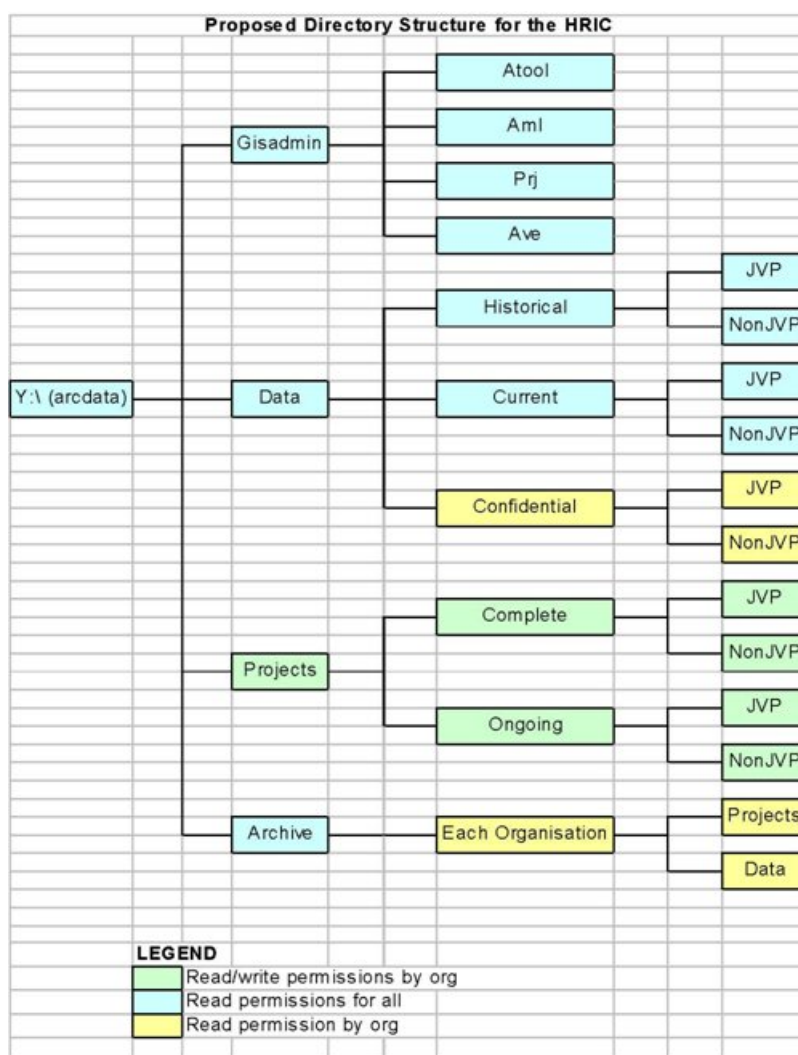
Most government projects and agencies own large amounts of data and information which have potential benefit for a wide range of user groups. All groups and agencies suffer from limited resources and as a result are under increasing pressure to maximise efficiency and effectiveness wherever possible. As such, expenditure on data should be used effectively and data sharing arrangements established for data capture (e.g. field survey/monitoring) or acquisition (e.g. satellite imagery). In practice NRM groups often hold data covering numerous themes from a variety of sources. This can lead to data management issues and inefficiencies if not dealt with in a structured fashion. The following information provides guidelines to assist in managing data in an efficient and effective manner.

2.1.3 File naming conventions and directory structure



To be efficient, data managers need a file structure and naming convention that is practical and reflects the nature of the agency's or group's core business. If the majority of the organisation's data are tiled (or map-sheet based) it is likely that a file and naming structure based on the tiles is used to store data. If the core business involves temporal (time-based) monitoring with regular updates, (e.g. every month a new dataset is produced) a file and naming structure that reflects the time series is more appropriate.

An example of a file naming and directory structure within an NRM setting from the Herbert Resource Information Centre's Data Principles and Guidelines publication is given below.



The Herbert Centre is a joint initiative with a number of partners including local government and industry. The above document can be found at the following web address:

http://www.hric.org.au/hric_site/hric_info/Policies/Policies.asp.



Another important reason for effective and efficient management of data is to ensure the continuity of the corporate knowledge base. Management practices in the form of standards, protocols and procedures are required so others can find and utilise data. **Lack of adherence to simple data management practices can cause major problems if staff changes occur and new operators are not familiar with the existing system.** In many cases this can lead to data becoming 'lost', inaccessible or 'sterile' and therefore not usable, resulting in severely reduced value and return on investment. The costs of agencies maintaining out-dated or redundant datasets can be very significant.



Agencies that need to maintain data for historical purposes and regulatory requirements may require special archiving practices to reduce maintenance and storage overheads.

2.1.4 Key drivers for improved data management

There are many factors 'driving' the need for improved data management including:

- the strong direction from the Australian, state and territory governments to improve services and make more effective use of resources
- increasing recognition that data collected at public expense must be properly managed to make them accessible to the public, to realise their potential and justify their considerable production and maintenance costs
- increasing pressure from customers for easier and quicker access to the right data and information to be made available at little or no cost
- increased focus within organisations and governments on the need to rationalise and combine data in order to improve efficiency and add value
- data publishing standards to assist in the development of seamless, consistent datasets between jurisdictions
- controls required by data owners governing the use of their data to safeguard intellectual property rights and to maintain confidentiality of sensitive data.

Regardless of the specific business context, most agencies and NRM regional bodies producing data will be judged on the ease with which the data and information are made available, as well as the quality of the information. Those able to publish, share, access, integrate and use information will benefit most.

2.1.5 Statutory requirements for data and information management

NRM regional groups, state/territory and local governments may be required by law to collect, manage and archive datasets in particular ways. Statutory obligations may be placed on the type, frequency and extent of archiving required. Given that these requirements may vary across jurisdictions (e.g. database on noxious weeds or invasive animals) it is recommended that NRM groups contact the relevant state or territory agency to clarify any statutory requirements for data and information management.

For example, in Western Australia the State Records Act (2000) requires that a records management plan—outlining archiving practices—is produced by all government agencies. To facilitate the process the Western Australian State Records Commission produced a self-evaluation guide to assist agencies in developing such a plan. Likewise the Australian Capital Territory 'Territory Records Office' has a dedicated website providing extensive guidance on developing record keeping policy and programs as part of the Territory Records Act 2002. Refer: <http://www.territoryrecords.act.gov.au/>.

2.1.6 Benefits of good data management



Data management policies and procedures ensure that data are treated as valuable assets. Implementing such policies and procedures yields many benefits. In general the benefits of good data management are reflected through:

- **Better decision making:** Ready access to existing spatial data is essential for many decision-making tasks such as protecting the environment, development planning, managing assets, improving living conditions, and national security. This leads to improved decisions being made at local, regional, national, and global levels on issues of environmental, economic, and social importance.
- **Maximising use:** Ready access to NRM regional groups' and other agencies' data will encourage more extensive use of a valuable public resource for the benefit of the community.
- **Avoiding duplication:** By sharing data the need for separate bodies to collect the same data will be avoided resulting in significant cost savings in data collection and maintenance.
- **Maximised integration:** By adopting common standards for the collection and transfer of data more integration of individual and often disparate databases is possible.
- **Custodianship:** The identification of custodians for the principal datasets enables users to identify those responsible for implementing prioritised data collection programs and developing data standards.
- **Equity of access:** A more open data transfer policy ensures better access by the whole community.
- **Security and access constraints:** Data is accessible only by those authorised to do so ensuring integrity and confidentiality where required.
- **Communication:** Communication on many levels (even program goals, objectives and results) is enhanced.
- **Improved data publishing efficiencies:** Adherence to standards and procedures streamlines data publishing activities making it more efficient for others to utilise existing datasets.
- **Partnership and new business process opportunities:** Value-added benefits created by providing new (and often unexpected) services to the user community.

2.3 Principles of good data management

2.1.7 Data policy

The first step for any organisation or NRM group wishing to implement good data management procedures is to define a data policy. This is a set of broad, high-level principles which form the guiding framework in which data management can operate. In most cases these have been identified at national and state or territory levels and can be readily adopted by NRM groups. See Section 2.4 for further information on establishing a data policy.



3 ▶▶ 4

Business processes are reviewed to ensure compliance with an organisation's or NRM regional body's documented data management processes, policies and procedures.

2.1.8 Data ownership

A key aspect of good data management involves the clear identification of the owner of the data. In most cases this is the NRM regional body or organisation which originally commissioned the data and has managerial and financial control of them. The data owner generally has legal rights over the data, along with copyright and intellectual property rights. This applies even where the data are collected, collated or disseminated by another party by way of contractual agreements, etc.

Data ownership implies the right to exploit the data, and in situations where the continued maintenance becomes unnecessary or uneconomical, the right to destroy them. Ownership can relate to a data item, a merged dataset or a value-added dataset. Intellectual property rights can be owned at different levels, e.g. a merged dataset can be owned by one organisation, even though other organisations own the constituent data. If the legal ownership is unclear, the risk exists for the data to be wrongly used, used without payment of royalty to the owner, neglected or lost.



All data, information and knowledge must have an 'owner' so, in this sense, whatever is produced by an agency or group must have an 'owner'.

As such, it is important for data owners to establish and document the following:



- the ownership, intellectual property rights and copyright of their data so they are safeguarded
- the statutory and non-statutory obligations relevant to their business to ensure the data are compliant
- the policies for data security, disclosure control, release, pricing and dissemination

- the agreement reached with users and customers on the conditions of use, set out in a signed memorandum of agreement or licence agreement, before data are released.

It is important to ensure data ownership information is included within the metadata description and related data documentation.

2.1.9 Data documentation and metadata compilation

All datasets should be identified and documented to facilitate their subsequent identification, proper management and effective use, and to avoid collecting or purchasing the same data more than once.

To provide an accurate list of datasets held by the NRM regional body or organisation, a catalogue of data should be compiled. This is a collection of discovery level metadata for each dataset, in a form suitable for users to reference. These metadata should provide information about the content, geographic extent, currency and accessibility of the data, together with contact details for further information.

All business-related datasets, once catalogued, should also be documented in a detailed form suitable for users to reference when using the data. These detailed metadata should describe the content, characteristics and use of the dataset, using a standard detailed metadata template.

2.1.10 Data quality, standardisation, harmonisation and audit

Good data management ensures datasets are able to successfully meet current needs and are suitable for further exploitation. The ability to integrate data with other datasets is likely to add value, encourage ongoing use of the data and contribute to recovering the costs of collecting the data.

To maximise the potential and use of datasets, NRM regional bodies and organisations should:



- use standard data definitions and formats
- define quality standards and apply the appropriate validation processes to each dataset
- ensure that data are quality assured and approved as fit for purpose before use or release
- encourage the use of appropriate state, national and international standards, in particular those which are relevant to geographic information.



ANZLIC has prepared a suite of inter-related policies and guidelines aiming to assist organisations achieve 'best practice' in spatial data management. This includes:

- [Guidelines for Custodianship of spatial data](#)

- [*Policy Statement on Spatial Data Management*](#)
- [*Metadata protocol and standard metadata profile*](#)
- [*Guiding Principles for Spatial Data Access and Pricing Policy*](#)
- [*Privacy Guidelines for Spatial Data*](#)
- [*Access to Sensitive Spatial Data*](#)

ANZLIC is currently developing or updating guidelines about 'liability'.

ANZLIC's guidelines and policies are available at:

<http://www.anzlic.org.au/policies.html>

An example of using processes for improving data quality from the Glenorchy City Council is provided at Attachment 2–1. This council is using a procedural approach to 'cleanse' its spatial data which ensures both accuracy and consistency of use.

2.1.11 Data life-cycle control



Good data management requires the whole life cycle of data to be managed carefully. This includes:

- business justification, to ensure consideration has been given to why new data are required, as opposed to existing data being amended, how data can be specified for maximum use including the potential to meet other possible requirements, and why the costs of handling, storing and maintaining the data are acceptable and recoverable
- data specification and modelling, processing, database maintenance and security, to ensure that data will be fit for purpose and held securely in their own databases
- ongoing data audit, to monitor the use and continued effectiveness of existing data
- archiving to ensure data are maintained effectively until they are no longer needed or are uneconomical to retain.

2.1.12 Data custodian

A number of focal points often exist within government agencies and NRM groups where data are gathered, compiled and analysed as part of the normal course of business. Each jurisdiction has a role to undertake in managing the data over and above the responsibilities of data owners.



An appointed agency, group or position (not a person, as the responsibility should remain with the position if an individual transfers to another job) is given formal responsibility to act as custodian of each major dataset by the owner. This agency, group or position should be made responsible and

accountable for the management and care of the data holdings under their control, in line with the defined data policy.

2.1.13 Data access and dissemination

This aspect will depend on the business and financial policy of the organisation or NRM group, however as a guide the following information is provided:

- Public access to data should be provided where possible
- Access to data should be granted to customers and commercial organisations when the request complies with the organisation's, group's or resource centre's business strategy, and does not infringe on any copyright or intellectual property rights, or any statutory or non-statutory obligations
- The right to use or provide access to data can be passed to a third party subject to agreed pricing and dissemination policies.

Within the NRM sector, ANZLIC and the Audit have developed a Data Access and Management Agreement signed by all jurisdictions in September 2001. The agreement provides for consistent access arrangements to datasets held by the Audit and its jurisdictional partners—refer:

<http://www.anzlic.org.au/get/2375374673>.

2.4 Establishing a data policy

The Australian and state/territory governments have overarching policies which govern data management practices. The following information provides a guide to assist regional groups in establishing a data policy statement.

2.1.14 Data acquisition

- All projects and activities, which give rise to substantial datasets should establish at the outset whether suitable data already exists in a potentially usable form, or whether new data need to be acquired.
- Prior to data collection activities being approved, the project must establish how the data will be acquired, who will be responsible for full exploitation of them and, how the benefits will be maximised and shared.
- Subsequent data handling and storage needs must be considered and plans put in place ensuring databases are maintained in such a way that maximum use can be made of them.
- Consultation should be carried out with respective state, territory and Australian government representatives to determine the correct protocol, methodology or

classification procedures to use. In many cases specific guidelines for data collection and management are available.

Examples include:

Salinity Mapping Methods in the Australian Context: User Guide available from the internet: <http://www.nrm.gov.au/publications/salinity-mapping/index.html>

Guidelines for land use mapping in Australia:
http://adl.brs.gov.au/mapserv/landuse/nat_scale_tec.html

Surveying and Mapping Nationally Significant Weeds:
<http://www.affashop.gov.au/PdfFiles/pc13456.pdf>

2.1.15 Fitness for purpose and point of truth

- Prior to using any dataset it is recommended the user undertake an assessment to determine the appropriateness of the dataset, or fitness for purpose, for the intended use. This involves assessment of such criteria as scale, resolution, accuracy, classification system and integrity of the dataset.
- Where possible (especially where more than one version of a dataset exists) users should obtain clarification from relevant state or territory jurisdictions to determine the authoritative, or point of truth, dataset. Using outdated or unofficial datasets for certain themes can lead to major problems in analysis and a lack of interoperability and integration with other datasets. For example, the IBRA Region dataset for Australia has undergone a number of revisions. Similarly the NRM regional boundaries have been changed over time. Users need to make sure they have the most recent dataset for their analysis.

2.1.16 Data care (custodial duties)

- Databases should be managed closely, with clear responsibility for custodianship established and individuals made accountable for ensuring data custodian procedures are followed.
- Data should be held securely in their own database, and adequate provision made for their long-term care. Disaster recovery and back-up procedures should also be in place.
- All data should be validated and quality assured prior to being used or archived.
- Easy access should be given to data holdings, both for staff and bona fide external customers and users.
- Data which are not required to be retained (for legal reasons or otherwise), should not be destroyed or put at risk without first exploring all possibilities.

Note: It is rare for natural resource thematic data to be destroyed.

2.1.17 Data use and exchange

- Memoranda of agreements or licence agreements (with respect to the subsequent use of the data) should be established with users and customers who receive data. These should include confidentiality declarations and conditions of use.
- Intellectual property rights should be protected in relation to any development of information, by specifying any restrictions on the use of the data in formal licensing arrangements.
- Adequate provision needs to be made for the widest possible public access to data and associated metadata.
- Pricing agreements should consider the cost of recovering the handling of data and information, in line with any policies or overarching obligations that may apply.



ANZLIC and the Audit have developed a **model agreement** for use in NRM programs, which incorporates guidelines for custodianship, metadata, archiving, accessing, data licensing and pricing within an operational context. For information refer to the Model Data Access and Management Agreement: <http://www.anzlic.org.au/get/2375374755>.

2.5 Implementing data management – key roles and responsibilities



To be successful, data management procedures must be implemented across the whole organisation under the guidance of a member of the executive board or committee. **It is often good practice to identify a data management ‘champion’ at this level who is prepared to take this role and see it through.** Other key roles are the data policy manager and the data custodians assigned to key datasets.

The following information is provided to help organisations and groups establish these key roles and implement good data management policies and procedures.

2.1.18 Data management ‘champion’

The data management ‘champion’ is responsible for:

- Ensuring that policies on data management are in line with overarching policies and obligations at a higher level (e.g. state/territory)
- Directing the development, implementation and maintenance of the detailed data policies, standards, procedures and guidelines across the whole organisation or group
- Reporting progress to the executive board on the performance achieved against targets set for improvement in data quality, and the value gained from effective data management.

In some situations, especially those involving a number of sites, a data management steering group may also be required.

2.1.19 Data policy manager

The data policy manager may require the help of local data managers to undertake the following tasks:

- developing and maintaining the data policy statement and other corporate guidance
- appointing and monitoring the performance of data custodians
- issuing guidance and training activities for staff
- ensuring local practice in individual business areas meets the standards set for the whole organisation
- ensuring the organisation maintains a central metadata resource.

2.1.20 Data custodianship



Data custodians are responsible for ensuring the following minimum standards are applied for each dataset:

- The dataset must be documented in the catalogue following the standards for discovery metadata, to enable the ownership, intellectual property rights, custodianship and accessibility factors to be determined.
- The policy for exploiting the dataset and making it available to other parties must be agreed and documented.
- The dataset and its conditions of use must comply with all statutory and non-statutory obligations of the organisation and any overarching policies at a state, territory and Australian government level.
- The data must follow standard classifications and definitions where appropriate, and must comply with all relevant standards, codes of practice and other protocols.
- The data must be fully validated and quality assured with sufficient detailed metadata to enable their use by third parties without reference to the originator of the data.
- The data must be stored, managed and accessed in line with agreed data management and security/confidentiality policies.
- The release or use of data by internal and external users must be authorised with agreement to the conditions of use documented.

- The costs and benefits of continuing to maintain the dataset must be reviewed periodically.

2.6 Additional support



Considerable information on data management policies is available from respective state, territory and Australian government jurisdictions. Though not exhaustive the following web resources can be used as a starting point.

ANZLIC – a suite of Guidelines and Policies available at:

<http://www.anzlic.org.au/policies.html>

State and territory material

Queensland: <http://www.qsiis.qld.gov.au>

Western Australia: <http://www.walis.wa.gov.au/>

New South Wales: <http://www.canri.nsw.gov.au/policies/> with additional information on the NSW Information Management Framework available at:

<http://nrims.nsw.gov.au/policies/imf/index.shtml>

Tasmania: <http://www.dpiwe.tas.gov.au/>

ACT: http://www.actpla.act.gov.au/tools_resources/maps_land_survey

Victoria:

<http://www.land.vic.gov.au/Land/lcnic2.nsf/childdocs/F31E2DE1F7D75F504A256A4F0017DA3E?open>

Northern Territory: <http://www.ntlis.nt.gov.au/>

South Australia: <http://www.environment.sa.gov.au/mapland/sicom/sicom/index.html>

Herbert Resource Information Centre – Guidelines and Policies:

http://www.hric.org.au/hric_site/hric_info/Policies/Policies.asp



Information on data standards

For additional information on key data criteria refer to Toolkit *Module 4: Spatial data priorities, standards and compliance*.

2.7 Glossary

Data: A collection of facts, concepts or instructions in a formalised manner, suitable for communication or processing by humans or computer.

Data custodian: Organisation or agency with a specific guardianship responsibility for datasets. Their concern specifically relates to security and confidentiality requirements embodied in legislation, and they should be familiar with the access limitations and the quality of datasets under their control. Given that the custodian organisation does not necessarily have managerial or financial responsibility for the datasets under their control, they can be different agencies or groups from the data owner.

Data owners: Individuals or groups, who are responsible for the management and financial accountability of a dataset. Data owners also have legal ownership rights to the dataset even though it may have been collected, collated or disseminated by another party.

Data policy: A broad set of high-level principles forming a guiding framework in which data management can take place.

Information: Data that have been value-added, processed and interpreted. People work with and act upon information. *Geo-information* is a specific type of information that involves the interpretation of spatial data. It has been argued that it is the transformation of data to decision making which brings a level of understanding and knowledge, and the ability to make informed interventions to improve the management of natural resources.

Metadata: The summary information or characteristics of a set of data. In the area of geographic information or information with a geographic reference this generally means the what, who, where, when and how of the data.

Spatial Data: Data that have some positional (location) values related to them.

Attachment 2-1 Data cleansing approach of Glenorchy City Council

Note: the following example is taken from the local government spatial information toolkit. Its source is duly acknowledged.

The following discussion is from Graham Hammond—a GIS professional working in Glenorchy City Council in Tasmania—based on his previous experience in Maroochy Shire Council. Data cleansing is an interesting subject that provides an example of what local governments (and NRM groups) are working towards. The procedure is still in the planning phase and Graham has kindly provided the following example to demonstrate the ideas behind this increasingly important process.

What is data cleansing?

Data cleansing relates to the basic cleaning of spatial data and attribute data (things like removing gaps, overlaps, checking field names and data spelling, etc.).

At Glenorchy we are in the very early days of developing a data cleansing policy.

This process is long and costly. When a council starts looking at what data they have it becomes a very large job. We will probably only extend our cleansing to datasets regularly used, and then cleanse specific datasets as needed. The risk of litigation of not having clean data, and/or at least associated metadata describing their use and accuracy coupled with a licence agreement saying how they will be used, is getting greater, so most people need to bite the bullet, or accept the limitations of not doing it. Below you will find a draft on how we might proceed.

Data cleansing procedure (Draft Version 0.1)

1. **Cleansing spatial data:** (a little platform specific)—basic cleaning of spatial data and attribute data, removing gaps, overlaps, checking field names and data spelling, etc.
2. **Directory structure and naming convention**
(The important thing here is a 3-tiered process):
 - a. Working directory—where you work on copies of data that will be cleansed to become the new data
 - b. Validation and areas where the data are elevated to so the data steward can sign off on them being correct to the best of their knowledge.
 - c. Repository—an area where completed data are located—this is where data should be used from (mapping or intranet/internet mapping). These data should be 'read only' to all but a small number of people.
3. **Constructing metadata**
4. **Assigning a data steward and custodian department**—a department or person who is responsible for the data. As most GIS people are just constructing or editing the data, the department who owns the data should be the steward. This means visually appraising the data and signing off that they think them correct.
5. **Data maintenance plan**—no point cleaning them if no plan for maintenance. I envisage a procedure to allow for timely review of datasets, i.e. cadastre may need to be maintained weekly when other data can be archived and not maintained.
6. **Data licensing**—each cleaned dataset must be covered by a licence agreement if used offsite.