supervising scientist report **124**

Data management systems for environmental research in northern Australia

Proceedings of a workshop held in Jabiru, Northern Territory, 22 July 1995

Edited by Max Finlayson and Ben Bayliss

supervising scientist

supervising scientist report **124**

Data management systems for environmental research in northern Australia

Proceedings of a workshop held in Jabiru, Northern Territory, 22 July 1995

Edited by Max Finlayson and Ben Bayliss

supervising scientist

This report should be cited as follows:

Finlayson Max & Bayliss Ben (eds) 1997. *Data management systems for environmental research in northern Australia: Proceedings of a workshop held in Jabiru, Northern Territory, 22 July 1995.* Scientist Report 124, Supervising Scientist, Canberra.

The Supervising Scientist is part of Environment Australia, the environmental program of the Commonwealth Department of Environment, Sport and Territories.

Supervising Scientist
Tourism House, 40 Blackall Street, Barton ACT 2600 Australia

ISSN 1325-1554

ISBN 0 642 24324 7

Printed in Darwin by NTUniprint.

Contents

Preface

The development of information technology over the past decade has heralded new opportunities and challenges for environmental researchers. At the Environmental Research Institute of the Supervising Scientist (***eriss***) this has seen the implementation of a network of personal computers within the reach of all staff. This change alone has provided immense opportunities for handling large amounts of data and rapid communication and exchange of information. The advent of technology, however, does not, by itself, guarantee that we are able to take advantage of the opportunities so presented. To do this we need to alter, or even develop from scratch, our every day work practices to maximise our efficient usage of the technology. Especially, we need to eliminate sloppy work practices in relation to handling data (including the collection, storage and interpretation steps). Our ability to embrace the technology is still only as good as our individual and corporate attitudes and skills; sadly, day by day experience demonstrates that we still have some way to go before we have a uniformly high level of commitment to corporate standards.

The challenges we face are to ensure that we do achieve uniformly high standards and to make effective use of the technology that is available, or likely to be available. To do this we need to accept the individual challenges associated with handling data and the corporate challenges of maintaining accessible and valid records of all data and their whereabouts. It is unarguable that ***eriss*** and other institutes have not successfully done this in the past and possibly vast amounts of data have been under-utilised due to staff turnover and programmatic changes. The first step in turning around this unsatisfactory situation is to develop protocols that ensure the basic information for all datasets is recorded and available. In other words, all project management protocols require a framework for establishing a core set of information that can be readily understood by others. The days when one person alone knew the location of the sampling sites, the sampling methods and the whereabouts of the data should be long behind us. It is acknowledged that this ideal can not be achieved without training and the development of awareness and corporate responsibility, but these are part of the process and not excuses to delay or even derail the process.

In an effort to develop superior data handling approaches and to take advantage of the opportunities presented by new technology ***eriss*** purposefully chose to seek external advice on the development of a metadatabase and the use of other information systems for environmental research purposes. For this reason, a small group of external environmental information experts was assembled and combined with a likewise interested group of ***eriss*** staff. This combined group was asked to provide initial guidance to ***eriss*** in its endeavours to ensure it was at the forefront of information technology and data management responsibility. A workshop format with invited short presentations was the chosen vehicle, but it was realised that, to a large extent, this was akin to preaching to the converted. However, even the converted need positive reinforcement and support. The workshop format was designed to make us aware of the further opportunities to use data management systems. We fully appreciated the need to tackle the real challenge of aligning all staff members with the basic corporate-level attitudes that would ensure we did progress from the bad old days when data management was, at the best, an individual concern. The workshop was not asked to address this attitudinal problem, but we found ourselves returning to it with regular monotony. We could not escape the 'people' factor.

Thus, we have taken steps to develop our capacity to take advantage of the opportunities for environmental research, with a special emphasis on wetland research in this instance. With this material and information resource we can more confidently face the future challenge of

making it happen across the board. No more should we be plagued by inadequate data storage processes—the single paper copy that was put in this drawer, but does not seem to be there now syndrome—they are very much a thing of the past. We have the technology and assumedly the knowledge and interest and courtesy to make effective use of it and thereby demonstrate our realisation of the value of our data. This workshop should be seen as just one of many steps in this process. We realise fully that the process for developing a corporate-wide interest in these issues is very much an internal one; there was no expectation that expert advice would solve our problems, but it could strengthen our resolve.

In the latter we were successful and for this we warmly thank the external experts. Now it is over to our internal self-interest and altruism.

Workshop summary: Data management systems for environmental research in northern Australia

C Max Finlayson[†]

Abstract

A review of information services for environmental research at ***eriss*** has identified the need to develop a metadatabase and to identify the usefulness of data management systems specifically for wetland research purposes. These systems will form part of a corporate information service that will underpin the goal of providing advice on selected environmental issues through research. A metadatabase was seen as the first step in preserving the data resource of the Institute and providing an information base for expanded research endeavours. The potential value of relational databases, decision support systems, Geographic Information Systems and the Environmental Resource Information Network (ERIN) were also outlined.

1 Introduction

The Environmental Research Institute of the Supervising Scientist (***eriss***) is reviewing its information management systems to take advantage of opportunities provided by advances in technology and restructuring of the research program (see SSARR 1995, Johnston in press). As a component of the restructuring of the Institute during 1994, a wetland protection and management research program was established to complement the existing environmental impact of mining research program. Along with this move it was recognised that research projects conducted under the auspices of the wetland program would benefit both by access to past research information and by the utilisation of new information management systems. Thus, it was anticipated that judicious use of 'new' technology would enhance the utility and hence value of much of the data collected over the previous 10–15 years.

In developing a strategy to maximise benefits from the 'new' and the 'old' the needs of the Darwin based operations of the allied Office of the Supervising Scientist (***oss***) (see SSARR 1995) were also considered. As the ***oss*** makes use of the information management services at ***eriss*** it was included in the planned review of these services, although it was noted that its needs did not totally correspond to those of ***eriss*** (Devonport 1996).

As part of the information system review process a workshop was held in Jabiru on 22 July 1995 to

- comment specifically on a proposal to develop a metadatabase for ***eriss***
- identify the usefulness of other data information systems for wetland research purposes at ***eriss***.

The first objective represents an institute-wide development (Devonport 1996) and it is envisaged that it will eventually involve all current information services and procedures at ***eriss*** (eg registrar, library, computing, editorial and publishing, and research). The second objective represents a further step in developing the wetland protection and management

---

[†] Environmental Research Institute of the Supervising Scientist

research program. The development of further data management systems was one of a number of recommendations that came from a wetland research workshop held in Jabiru 17–20 March 1995 (see Finlayson 1995 for the workshop proceedings).

This paper presents a summary of the discussions held during the July 1995 information systems (specifically for data management) workshop held in Jabiru. A list of workshop participants is given in Appendix 1 and a copy of the program in Appendix 2. Papers submitted by participants were used as resource documents and are contained within this report. They describe the key features of metadatabases, relational databases, decision support systems, Geographic Information Systems and the Environment Resource Information Network (ERIN). Accompanying the papers is a report on information systems needs at ***eriss/oss*** that initially stimulated and then strengthened our resolve to investigate the development of a metadatabase (see Appendix 3).

2 Background

eriss has embarked upon a project to develop a metadatabase. We plan to i) establish protocols that will enable us to produce a catalogue of the datasets that we have accumulated since the Institute was established in 1978 and ii) capture metadata for all existing and future data sets. This may seem an ambitious and even elusive target, but we see nothing wrong with being ambitious. Data are both expensive and valuable and we see this exercise as having inordinate value for our future research operations. We do not currently have an integrated ***eriss/oss***-wide data management system. Rather, data are stored in a fragmented array of databases that have been maintained by individuals or small groups associated with specific programs or projects. Given major changes to the role and structure of ***eriss/oss***, a regular turnover of staff and a filing system that was formerly split between Jabiru, Darwin and Sydney (before being relocated to Canberra in 1992), we are in constant danger of losing the collective corporate memory that is, as with many organisations, the mainstay of our information management system. This is not seen as satisfactory given current technology and the recognised value of our data information.

To ensure that our fragmented approach to information management did not become a liability we heeded the advice of an external consultant to undertake the task of constructing a metadatabase (Devonport 1996) whilst also pursuing other data management tasks. Devonport advised us to pursue a number of tasks to achieve the ultimate goal of having an integrated decision support system for our scientific endeavours. These tasks were:

- formulation of a policy for the management of information resources
- establishment of a metadatabase
- establishment of a corporate database
- integration of GIS and remote sensing into the information system
- use of the information system as a decision support system

It is perhaps instructive to point out that the report from Devonport was initially an assessment of GIS needs and was a follow-up to a specific assessment of GIS for a component of the Institute a year before (Riley et al 1994). Before locking ourselves into irreversible decisions about GIS, Devonport advised us to first take stock of our current data and information management system. If a GIS platform is required by ***eriss/oss*** it will be based on the wider platform provided by a corporate information management policy backed

by a comprehensive catalogue (or dictionary) of previously collected datasets (ie the metadatabase).

Thus, we have taken initial steps to develop a policy for managing our information systems and to simultaneously investigate the establishment of a metadatabase. The development of a metadatabase was addressed at this stage as it would also provide the means of responding to a recommendation made by Barrow et al (1994) to review and document the wealth of baseline research information that has been amassed by ***eriss/oss*** over the past 15 years. Specific talks on the nature and use of metadatabases were presented as a prelude to a draft protocol for a metadatabase being presented to stimulate discussion (see other papers in this volume).

3 Discussion on metadatabases

The following points outline the major comments on the development and use of metadatabases made by the participants. The comments should be read in conjunction with the resource papers provided by selected participants and included in this volume. The points given below convey the general thrust of the discussion and purposefully do not present details or technical information which will be addressed when the protocol for establishing a metadatabase at ***eriss/oss*** is agreed.

Table 1 Major issues to consider when developing a data management system for wetland research. These are key issues raised by participants in a summary session and are not in priority order. However, *all* issues were considered important, if not critical, by the participants.

Know the origin of the information/data being described
Provide geographic coordinates and/or describe the extent/expanse of the site
Document details of the survey design and the type of data collected
Avoid the trap of publishing meaningless pretty pictures
Maintain the corporate commitment to developing the data management system
Adopt a flexible structure for recording data
Provide training for all users on a regular basis
Develop a prototype to test all assumptions and processes
Identify the users and define their needs in an interactive manner
Adopt a 'whole of life' approach for data management
Incorporate data from other relevant sources with attention to security and access
Keep in mind the questions that need to be answered
Develop a structured framework for developing a data management system
Maintain the impetus and respond to user needs through simple processes

The key points made in a summing-up session are presented in table 1. The points in this table reflect the final comments from the participants and do not correspond directly with the numbered comments given, with more explanation, below. At the conclusion of the discussion participants were specifically asked to simply list the major point(s) that they wished to see recorded as a result of the discussions; they were not asked to explain the reasons for their choices nor to prioritise them. In this manner we obtained an indication of key issues that we would need to keep in mind when proceeding further with our goal of developing a metadatabase.

3.1 A metadatabase is a tool that is used to access 'information about information'. At one level it can be described as a catalogue, but it can be much more than a simple listing of data sets. At a more complex level it can provide the means to easily share and access data between users and sites, to identify gaps and weaknesses in the data, to assess the value and need for new databases, and to monitor the currency and usefulness of data. Thus, it is an integral component of the research planning and management process and it does not stand alone. This last point is critical. A metadatabase that is not underpinned by a corporate policy on information management is unlikely to achieve the goals of its creators. Without this support we may develop a very nice metadatabase, but not actually make effective use of it!

3.2 An important function of a metadatabase is to feed information about data back to the user. Thus, it is an interactive tool for the benefit of the user. The effectiveness of the database is, therefore, not independent of its users. As the users will provide the final judgement on the usefulness of the database they should be intimately involved in the developmental process. If the potential users are not keen to interact and develop and use the database it is not likely to be a success. The need for the database and its usefulness must be accepted by the users before its development. If these criteria are not met it could be a fatal waste of resources to proceed without a concomitant effort to bring the users onside. User friendliness is necessary, but is not very useful if the database is not accepted by the users. Therefore, corporate support and active involvement of all users are essential. It is noted that users need to consider the needs of the Institute and not be locked into sectoral and/or selfish individual pursuits.

3.3 An efficient data management system should bridge the gap between collecting the data and making the datasets available for analysis or for use by other people. This includes the ability to query the database to identify data relevant to specific tasks or needs, some of which would not have been thought of when the data were collected. It also requires a mechanism to feed the information back to the users. The ease with which users can access the database is a critical factor. Training may be required to ensure that this is done with ease.

3.4 There is a temptation to view the establishment of a metadatabase as a goal in itself. This should not be the case. As pointed out by Devonport when introducing the concept of a metadatabase, it is but one component of a corporate information system. Linking the metadatabase to other data management systems could produce problems with systems compatibility, but it is expected that such problems will be less of a concern in the future and should, in any case, be examined before any major and irreversible (ie without incurring great costs) decisions are taken. The metadatabase is neither the start nor the end of the process; it is but one part of the data management process that enables users to achieve their own individual and corporate workplans now and, critically, in the future.

3.5 A metadatabase is a necessary tool for effective data management, especially when an individual or an institution is faced with a rapid proliferation of data and types of data. It is essential that its development and use are based on realistic assessments of needs, benefits and resources for now and for the future. Devonport (1996) and House (1996) both presented a number of fields of information that can all be seen as useful. But what is a realistic level of information to include in such a database? What is the minimum or core information that is required to describe the datasets and make them useful to users? For ***eriss/oss*** this will require an assessment of the likely users both within ***eriss/oss*** and elsewhere at this moment and in the future. Thus, it will require an active and imaginative input from staff at ***eriss/oss*** and equally active and imaginative management decisions about the further development or direction of the research program.

3.6 A key purpose of a metadatabase is the preservation of data sources. The information gathered should not only provide a user with the knowledge to access a particular set of information, but also comment on the condition of the information (ie is it safely stored or is it likely to deteriorate unless certain actions are taken). Thus, it also provides a mechanism for assessing the quality of the medium upon which the data are stored. The latter is particularly important when data have not been kept in ideal conditions (eg climatic variations) and, especially, when no other record exists.

3.7 Given different needs for data and immense differences between the data that may exist in multi-disciplinary organisations such as ***eriss/oss***, it may be necessary to establish tiers of information and detail for users. The tiers should be constructed to avoid burdening users with information that they do not require. Such tiers may also be useful for the exchange of information between metadatabases maintained by other organisations. The structure of the database should firstly serve the needs of the current users, but decisions that restrict future (and even unknown) uses should be very carefully considered. Security of data is necessary, but this should not be an excuse for limiting the usefulness of the data. It must not be forgotten that ***eriss/oss*** is part of a much larger governmental system that collates and provides data to a broad range of users. Individuals within ***eriss/oss*** may be custodians of the data, but they do not own the data.

3.8 The development of a system for specific institutional needs may be restricted by soft/hardware constraints. Thus, the availability and suitability of both software and hardware needs to be addressed. When considering the choices available it could be instructive to review industry or governmental standards for the exchange of information between data sets. The Australian and New Zealand Land Information Council (ANZLIC) has initiated discussion on standard approaches to the exchange of data between users and database systems (see ANZLIC Working Group on Metadata 1995). ***eriss/oss*** should ascertain what advantages and constraints such standards place on its plans for a metadatabase.

3.9 It was specifically emphasised several times that ***eriss/oss*** should firstly identify its critical or core need for a metadatabase and then involve the users in the developmental process. With the users behind the process and with strong corporate support the reward should be a metadatabase that meets the needs of both the individual users and the corporate body. Once this basic planning process has been established those responsible for developing and implementing the protocols need to concentrate on the following:

- What data are available?
- Where are the data located?
- How are the data described?
- Who has access/possession of the data?
- What condition are the data in?
- Are the data secure?

It was also recognised that some staff may need to be made aware of the value of a metadatabase and/or aligned with institutional needs for data management and security. In the latter instance it was also recognised that not all staff would necessarily naturally accept the need for such institutional requirements ahead of their own personal research tasks. Alignment of such staff with the institutional processes is critical.

3.10 Following the establishment of a metadatabase it is essential to not only describe the data, but to also assess their quality and potential usefulness. Simply, are the data useable? Are they reliable? If there are question marks over the accuracy or usefulness of the data these must be attached to the entry in the metadatabase. If, for any reason, this assessment has not been done, or cannot be done, this should also be recorded. The level of documentation available with the data will prove critical in this process. Poorly documented data are extremely difficult to deal with. Any doubts about the data must be expressed. Where possible, contact names should be given, but it is recognised that if such people have left the organisation these may not be very useful.

3.11 The establishment of a metadatabase provides the impetus to develop a basic documentation process for all newly collected data. At the commencement of each data collecting exercise a basic set of information and documentation is required. This ensures that the metadatabase is not simply seen as a listing of 'old data sets' and therefore effectively irrelevant for current data collection.

3.12 A metadatabase should be seen as an asset to both the corporate user and the individual researcher. Both levels of usage are essential ingredients of the management process within a research institute, but concern was expressed that, generically, too many researchers do not see such corporate goals as commensurate with their own goals. This situation may be lamentable, but it should not be ignored, as it does exist and is a potential combat zone for the blind or unwary corporate-oriented manager (campaigner).

3.13 ***eriss*** has an immediate requirement for a metadatabase—to provide the means of assessing the value of previously collected data and using this as the basis of a description of the knowledge base of the Institute. We have 15 years of data and we have been asked to summarise the state of our knowledge of the environment of a wonderful, diverse and contentious region. First, we need to catalogue the data. However, once we have completed this large and vital task we need to ensure that a high level of data storage and documentation is maintained.

4 Discussion on data management systems for wetland research

The potential value of Relational Databases (RD), Decision Support Systems (DSS), Geographic Information Systems (GIS) and the Environment Resource Information Network (ERIN) for wetland research at ***eriss*** was discussed. The detail and technicalities of each system were not discussed unless as background for a particular point. The papers in this volume should be consulted for further information on these information systems.

The ***eriss*** wetland research program was singled out for particular attention for three main reasons. Firstly, whilst it has only recently been established ***eriss*** has previously collected a large body of wetland/aquatic data. It was seen as crucial to assess the value and usefulness of this past data for current research endeavours. In other words, we need to assess the extent of the baseline data resource. Secondly, we expect the wetland program to contain a strong spatial component that could expand at a rate previously unseen at ***eriss***. Thirdly, the research will be particularly directed towards providing information in a form directly useable by wetland managers and planners. ***eriss*** has a goal to undertake research to provide information and advice for management purposes. We do not undertake research and simply hope that it will be used. Our institutional goal is to provide information for management purposes. Thus, individuals need to present information in a manner that is compatible with the Institute goal.

Specific recommendations to ***eriss*** were not requested in this session. Rather, we treated this session as one of information exchange that would assist our future planning and decision making. We recognise that not all ***eriss*** staff are equally familiar with the logic and value of various data management tools. Thus, the discussion was very much seen as an educational and awareness opportunity. However, it was not all one-way as a second objective was to explore ideas for cooperation and collaboration between ***eriss*** staff and invited participants. Human networking was not forgotten as we considered the intricacies and even traps of computer networks and their offshoots.

A summary of key points is presented below. In these discussions it was again clear that any information system should be contained within an overall information management policy that linked data collectors to users and contained an adequate training component. Many of the points made about developing a metadatabase are also relevant to other information systems; these are not repeated, but it is emphasised that they should be read and assessed for their relevance to all data management systems. Similarly, the specific features of particular data management systems are not presented; these are available in more technical documents and workshop proceedings (eg Riley et al 1993, Supervising Scientist 1995).

4.1 RD and DSS are used to enhance access to information contained within databases. As such, they should not be seen as separate from other databases. They are linked to the information system, in part, to reduce the necessity for re-entry of data and to easily provide the sources of the information.

4.2 Prior to developing a RD/DSS the needs and attitudes of the users towards such systems need assessing. User antagonism or disinterest require careful attention. It is also essential that care is given to the actual identification of the users. Who will benefit most from the RD/DSS? The users should be seen as the justification for the system.

4.3 Standardised designs backed by valid hypotheses should underpin the data that are contained within a RD/DSS. Any limits on the data should be vividly and succinctly displayed if the data are to be actually included within the data system. The purpose for which the data were originally collected should also be identified and decisions on the data's applicability made and rated along with the reasons for a particular decision. Thus, the quality of the data and rationale for the collection of the data, or both, are critical: quality assessment must be included and displayed to the users. The applicability of the data for different purposes will be dependent on the rationale and means of collecting.

4.4 Simple descriptive and more rigorously collected statistical data both have value in decision making processes, but not usually equally. The value of the data in a DSS is also linked to the nature of the data; why were the data collected? Similarly, the temporal nature of the data should be considered. Long-term data should be used wherever possible.

4.5 It is critically important to develop a prototype system and/or undertake a pilot study of any data management system. By critically evaluating a prototype, great savings in effort and expense can be made. In effect, will the system perform as is expected? This can only be effectively assessed using specific and real examples even if they are only a subset of the total data available.

4.6 A GIS provides an invaluable means of transferring spatial information from a variety of datasets to users who may not otherwise have easy access to them. This is particularly so where there are a multitude of users with different needs and knowledge of the data and information available.

4.7 There is no limit to the type of data that can be included in a GIS with, for example bibliographic and library information being included along with numeric datasets. The emphasis is on the spatial aspect of the data and in some instances it may be preferable to use non-spatial databases instead of GIS. A GIS is not the only database format available.

4.8 A GIS should not be seen as a static data source. As further information becomes available it will grow and even develop new directions. Making the users aware of the nature of the data is crucial. There is little point in adding data to a GIS and keeping the users in ignorance of its existence. Thus, interaction with the users is not only important at the design and development phase, but also in the maintenance and extension (if appropriate) phases as well.

4.9 Interaction with the users is crucial to ascertain what decisions need to be made. What questions will the users require to be answered by using the GIS? Do users understand what a GIS can provide?

4.10 As with other monitoring programs a pilot study is essential. Critical datasets should be identified and their successful incorporation into the GIS or DSS etc tested. As long-term datasets form the very basis of many monitoring programs the procedures required to obtain and continually update the GIS or DSS should also be tested during the pilot study. This testing could effectively be undertaken using specifically chosen reference sites.

4.11 The means of transferring information to users is constantly changing. The utilisation of CD-Rom and the Internet will be partly dependent on user attitudes, access and familiarity with the medium. It can not be assumed that technological advances will be readily accepted or available to all users. The further development and acceptance of CD-Rom, for example, should be beneficial in this respect.

4.12 The use of established information networks is encouraged in order to reduce individual operating costs and to take advantage of other people's developments. There is no need to reinvent the motor vehicle (the wheel is now well known), but there is every reason to combine talent and develop a slicker and more efficient vehicle. Compatibility between handling systems and components is very useful. But, even here it is worth assessing the usefulness of a generic system (a Holden ute) versus a 'fancy' custom-made system (a Porsche coupé)—it all depends on your specific needs.

4.13 Existing data and information systems should be accessed and explored thoroughly before embarking on new developments. For example, how can ERIN contribute to the data and information needs of ***eriss/oss***? Being a government service within the same environment portfolio as ***eriss/oss*** there may well be ideal opportunities for advice and collaboration. Similarly, other agencies with similar interests and data needs may be useful collaborators. It is also possible that ***eriss/oss*** could feed into the ERIN data systems.

5 References

ANZLIC Working Group on Metadata 1995. A metadata framework for land and geographic data directories in Australia and New Zealand: Discussion paper. The Australian New Zealand Land Information Council, Canberra. Unpublished paper.

Barrow J, Bowmer K & Davy D (compilers) 1994. Report of the consultancy on the Alligator Rivers Region Research Institute. Supervising Scientist for the Alligator Rivers Region, Canberra, Australia. Unpublished paper.

Devonport C 1996. ***eriss*** information system opportunities for improvement. In *Data management systems for environmental research in northern Australia: Proceedings of a workshop held in Jabiru, Northern Territory, 22 July 1995*, eds Max Finlayson & Ben Bayliss, Supervising Scientist Report 124, Supervising Scientist for the Alligator Rivers Region, Canberra, 54–62.

Finlayson CM (ed) 1995. *Wetland research in the wet-dry tropics of Australia.* Workshop, Jabiru NT 22–24 March 1995, Supervising Scientist Report 101, Supervising Scientist, Canberra.

House T 1996. ***eriss*** metadatabase development: A starting point. In *Data management systems for environmental research in northern Australia: Proceedings of a workshop held in Jabiru, Northern Territory, 22 July 1995*, eds Max Finlayson & Ben Bayliss, Supervising Scientist Report 124, Supervising Scientist for the Alligator Rivers Region, Canberra, 23–29.

Johnston A (in press). Introduction. In *Development of a stream biological monitoring program in the Alligator Rivers Region, northern Australia*, Proceedings of a Workshop, 24 September 1993, University of Canberra, eds CM Finlayson, CL Humphrey, & RWJ Pidgeon, Supervising Scientist Report, Supervising Scientist for the Alligator Rivers Region, Canberra.

Riley SJ, Devonport C, Waggitt PW, Burrough PA, Milne AK & Skidmore AK (eds) 1994. Proceedings of the ARRGIS 93 workshop. Jabiru NT 12–13 August 1993, Internal report 139, Supervising Scientist for the Alligator Rivers Region, Canberra. Unpublished paper.

Riley SJ, Devonport C, Waggitt PW & Fitzpatrick B (eds) 1993. *NARGIS 93: Proceedings of the North Australian Remote Sensing and Geographic Information Systems Forum.* Darwin 9–11 August 1993, Supervising Scientist for the Alligator Rivers Region and the Australasian Urban and Regional Information Systems Association Inc [Monograph no 8], AGPS, Canberra.

SSARR (Supervising Scientist for the Alligator Rivers Region) 1995. *Supervising Scientist for the Alligator Rivers Region: Annual report 1994–95.* AGPS, Canberra.

Supervising Scientist 1995. *NARGIS 95: Proceedings of the 2nd North Australian Remote Sensing and Geographic Information Systems Forum.* Darwin 18–20 July 1995, Supervising Scientist and the Australasian Urban and Regional Information Systems Association Inc [Monograph no 11], AGPS, Canberra.

Information navigation architecture: The metadata network

John Blackburn[†]

Abstract

The components and development of a metadata network are described. The need for metadata is explained in terms of a 'table of contents' of the corporate data resource. As this resource can change rapidly the database needs to be maintained and available for query by the users. Effective data management and productivity in accessing information can be achieved via a well structured metadatabase.

Introduction

Data sources and stores in the form of datasets and databases can reside on one or more computer systems within a local or wide-area network (LAN/WAN). This network supports access to the organisation's corporate information systems and access to a wider information/ data community. The distribution and configuration of the network can vary depending on the nature of business, geographic positioning, departmental separation and functional grouping of the organisation. For example a typical network configuration is Head Office and Regional Office connected 'nodes'. Each office node may have databases that are specific to their business/region and will also need access to head office databases for general, common or centralised purposes. The reverse situation also operates.

For the Manager of Information Services the objective is to organise and configure this collection of 'true' data stores as a seamless or 'virtual' master data store, ie the distributed data network. The 'table of contents' of the master data store needs to be arranged in the same way.

The metadata network

The metadata network is like an intelligent 'table of contents' of the corporate data resource, which is changing and growing constantly. Hence the term *metadata*—'data about the data'. It is information in its own right and more than a catalogue or directory, because it includes descriptive information elements designed to be queried and provide information about content, lineage, quality, coverage, source, access, etc.

Each 'true' database or data collection will have a discrete set of metadata fields that describe its contents in full or in part, depending on the metadata requirements. These requirements emanate from what the organisation decides are necessary query elements; what may be required or recommended by standards; and what security limitations may be imposed. This discrete set of metadata fields form the particular metadatabase containing the relevant information about the contents of one or more datasets or databases, on that node in the computer network.

---

† Genasys II Pty Ltd, North Sydney

Creating and maintaining the metadata network

A metadatabase (or metabase) is created by interrogating the source databases and datasets and 'loading' the metadatabase fields with the relevant metadata. The same 'loader' maintains the metadatabase on a frequency depending on changes in the source data.

As metadata requirements change by theme or form, a new metadatabase will be created. More than one metadatabase is needed when a set of metadata (therefore the source databases) is logically and thematically different, eg the content of customer databases would be greatly different from plant and equipment databases. This separation in metadata may be dictated by functional or departmental separations in the organisation(s). Therefore, there can be more than one metadatabase on a node in the network, as well as other metadatabases on other nodes in the network.

The metadata server

The network of metadatabases is managed by a metadata server (figure 1). The server registers and connects/disconnects each metadatabase so that queries can be passed to all registered and connected metadatabases.

Results of queries made from the client graphical interface (GUI) are returned back to the client for display to the enquirer by the server. The client software can run anyway on the network (LAN or WAN).

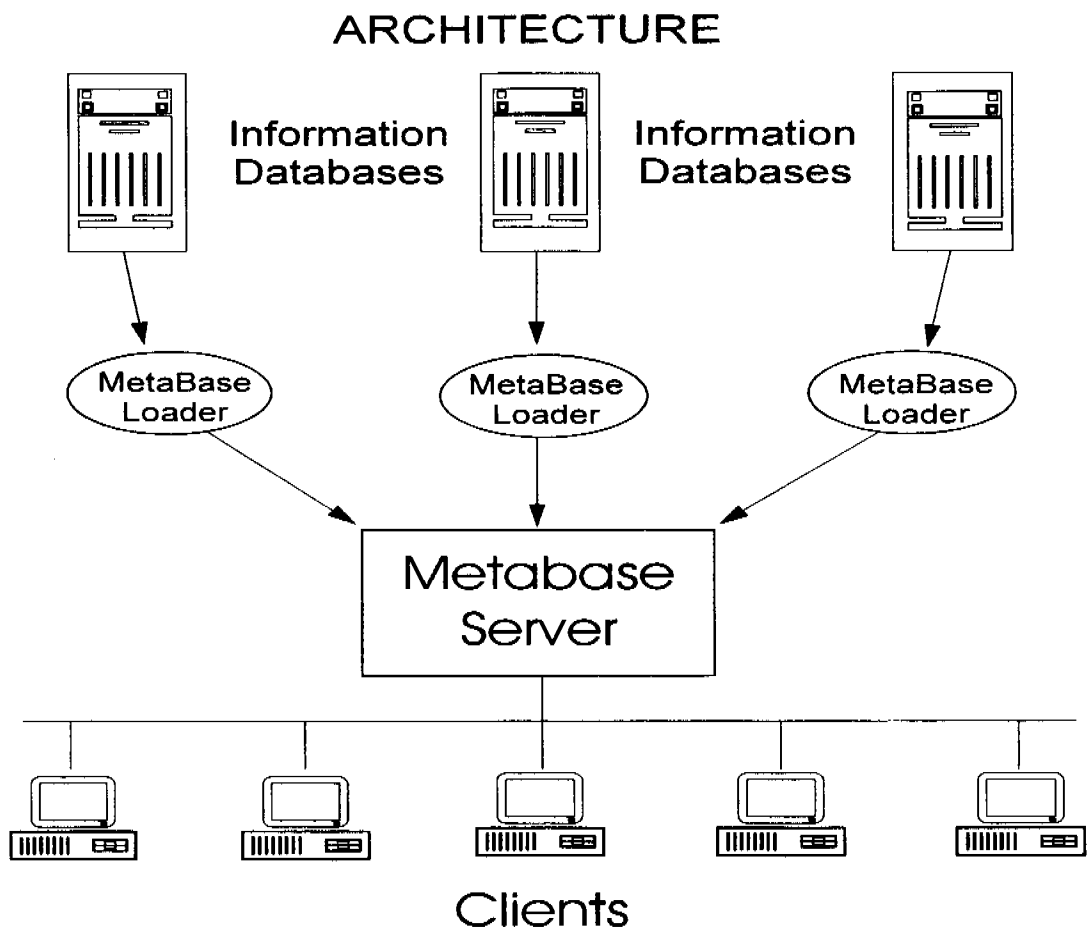


Figure 1 The metadata network architecture as presented in the Genasys program 'Henri – The Navigator'

Specialised metadata utilities

There are some specialised metadata utilities as part of the server configuration designed to aid and support metadata searching and query formulation. For example, a gazetteer aids spatial location searching on place or region name. Another example is a thesaurus which aids keyword, subject term or concept matching. Concept matching is an important utility that will find relevant datasets even if the keyword doesn't appear in the metadata. A calendar helps define date ranges or time periods as 'eras'. There can be more than one of these specialised metadata utilities which can be for public or private purposes, ie defined by industry, discipline, or by the organisation.

Summary

The metadata network is a modern information navigation architecture for effective data management and for productivity in accessing information about corporate data resources. The next logical step is to access and use the actual data that have been found. The above concepts are summarised in attachment 1.

Attachment 1 Summary of metadata concepts

a Metadata objectives

- to inform about metadata trends
- understanding of method and technology
- introduce a modern information navigation architecture
- case studies

b The requirement for metadata

The need is to:

- document data we have collected and created
- make these data known to others
- find someone else's data

c What is metadata?

Information describing data, ie 'data about data'

- content
- accuracy
- reliability
- when created
- by whom etc

All together it is a 'catalogue'

d What is metadata for?

- effective data management, including a method of understanding data resources
- efficiency in planning and conducting projects, ie cutting the time to find necessary datasets
- identifying holes in data coverage
- bridging the gap between 'knowledge' of the data and using the data, ie launching applications

e The use for metadata

Metadata = data management

Managers need metadata technology to:

- control the proliferation of databases
- know where are the gaps in data collection
- assess the value of new database projects
- monitor the currency and usefulness of their data

f The technology response

- dimensions of typical queries of metadata:
  - What, where and when?
- a data network approach—the information architecture

g The technology response

Dimensions of metadata—the scalability required:

- level 1—commonly used/required fields
- level 2—fields specified by standards
- level 3—field specific to an organisation
- consistency with proposed standards viz: ANZLIC (published)

h Technology components

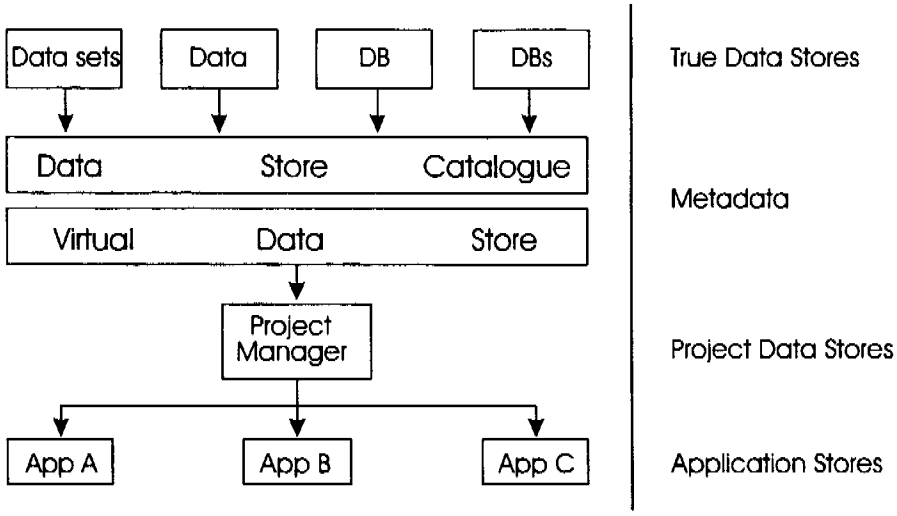
Three key components:

- graphical user interface (the client)
- server that manages one or more metadatabases:
  - thematic and functional separation
  - gazetteers, thesauri
- loaders that create and maintain the metadatabases

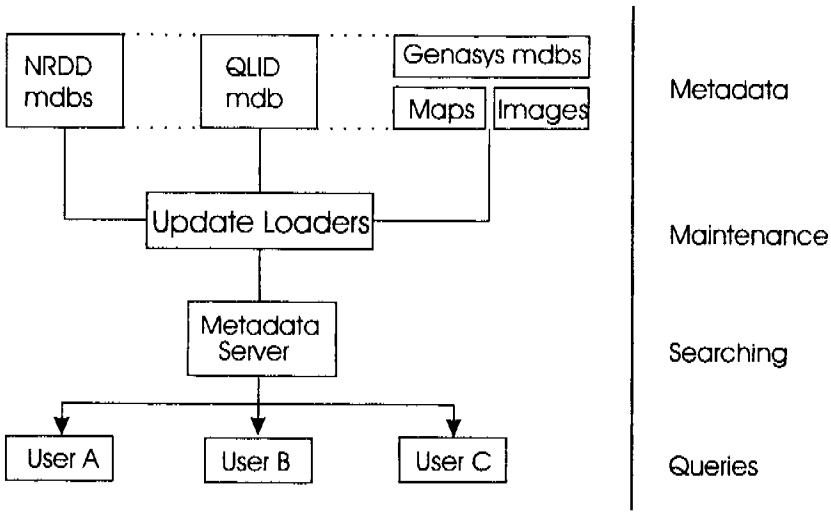
i Technology development

- a generic directory approach
- Importance of spatial query
- Customer and industry focus
- Industry testing

j Metadata = the virtual data store



k Virtual data store



l Metadata standards

- metadata in data transfer formats, eg SDTS, DIGEST
- metadata standards FGDC, ANZLIC (published)

Metadata: Some national and international perspectives

Paul Shelley[†]

Abstract

The concept and use of metadata has expanded with the use of computer technology and the recognition of data/information as corporate assets. The National Directory of Australian Resources (NDAR) has been established to assist with the provision of information to support decisions about ecologically sustainable development. It contains descriptions of some 6000 datasets. ANZLIC has encouraged the use of NDAR and the development of compatible systems.

NDAR is likely to become a cooperating node to the International Directory Network. The USA government has adopted content standards for metadata and this is being utilised in some GIS software. UNEP has also proposed and is testing standards for the Global Resources Information Database.

Introduction

Metadata—data about data—is a word which appears to have been established in the literature since about 1988 (NASA 1988). However, it describes a concept which has been around for a very long time in the form of catalogues and lists of publications.

The advent of computers, the growth of data collections and the recognition of these data collections and associated information as corporate assets, however, has resulted in the need to record or register metadata. The development of this field over the past six or seven years has produced a new vocabulary of terms and concepts by which to describe data and information. Terms such as directory, catalogue, inventory, attribute and metadata itself are taking on new meanings as specialists endeavour to put some structure into this new activity (NASA 1993; Shelley & Johnson 1995).

In addition, computer software is being developed to cater for the needs of both data managers and data users and there has been considerable activity in the development of metadatabases at agency, national and international levels. This paper presents a summary of some of the activities going on in this area.

National developments in metadata

The National Directory of Australian Resources (NDAR) was initiated in 1989 following the establishment of the National Resource Information Centre (NRIC) in 1988. NRICs primary mission is to facilitate the flow of information to support decisions about ecologically sustainable development. NRIC achieves this through the development of research and demonstration projects in close collaboration with clients. See attachment 1 for contact information.

To properly achieve its role, NRIC saw an immediate need for a directory of natural resources and related datasets which provided details on what data were available, what their format was, where they were located, and how they could be accessed.

---

† National Resource Information Centre, Canberra

To adequately describe data, a set of attributes was developed in close consultation with federal and state/territory agencies. This covered such things as name, abstract, ownership and custodianship, content, lineage and quality, spatial referencing, format, region covered and availability. This set of attributes is summarised at attachment 2.

In the absence of any appropriate commercial software, a directory system—FINDAR—was developed to provide a flexible, spatially-searchable metadatabase. The flexibility of FINDAR allows datasets and other types of information to be described in ways appropriate to the type of information. Further details are available in Johnson et al (1991).

NDAR has been operating now for six years and currently contains descriptions of some 6000 datasets and related information on Australia's natural resources. It is a distributed system with five active directory nodes in federal and state (Victoria and WA) agencies providing metadata to the central searching node. A report of the current status of NDAR and its future directions can be found in Shelley (1995). See attachment 1 for information on access to NDAR.

Computer-based directory developments are also being actively pursued in Queensland (Queensland Land Information Directory), New South Wales (NSW Natural-Resources Data Directory), and South Australia (SA Directory of Spatial Information). Each of these has required a set of descriptive attributes to be developed. While there was no deliberate move to develop compatible systems, there is sufficient commonality in the different sets of attributes used to facilitate metadata exchange among agencies. See attachment 1 for contact information for these state directory systems.

NDAR has been endorsed by the Australia New Zealand Land Information Council (ANZLIC) as the national directory and all jurisdictions have been encouraged to develop systems that allow exchange of metadata. An ANZLIC working group recently developed a core set of metadata attributes to facilitate metadata capture and metadata exchange among different directory systems. This 'standard' has been endorsed by ANZLIC and all jurisdictions have undertaken to adopt and promote it. Its use and acceptance will be reviewed in one to two years. A summary of the core attribute list is at attachment 3.

In early 1995, Standards Australia published the Spatial Data Transfer Standard as AS/NZS 4270:1995. This includes a specification for the metadata which must accompany a dataset; its primary focus is on data lineage and quality. See attachment 1 for information on availability.

International developments in metadata

In the late 1980s, the US National Aeronautics and Space Administration (NASA) established the NASA Master Directory and the Global Change Master Directory (GCMD). A Directory Interchange Format (DIF) was developed to provide a structured way of capturing and recording metadata, initially of remotely-sensed and space-related datasets, but more recently of a much wider range of data.

The GCMD project forms the basis of the International Directory Network (IDN) which comprises identical coordinating nodes in the USA, Japan and Italy, together with a range of cooperating nodes in many other countries. NRIC has agreed that NDAR will become a cooperating node and we are currently discussing arrangements with NASA.

An event in the USA which will have an international effect was the publication in 1994 of a metadata standard, *Content standards for digital geospatial metadata*, by the US Federal Geographic Data Commission. This is a very comprehensive specification and is a mandatory

requirement for all relevant datasets produced by USA federal government agencies. Already GIS software companies are writing modules that produce data documentation to the new standard. An example of this is ESRI's Arc/Info *Document* module which extracts data documentation information from a GIS project and presents it in the FGDC standard format. See attachment 1 for information on availability of this standard.

Another international activity was the initiation in late 1994 of a metadatabase specification for the Global Resources Information Database (GRID) of the UN Environment Program. This will be applied to all GRID centres which to date have managed their data and metadata in a variety of ways. A data and metadata 'housekeeping' software tool has been developed and is currently being tested by GRID centres and several other agencies.

In the UK, the Global Environmental Network for Information Exchange project (GENIE), based at the University of Technology at Loughborough, has developed a metadatabase system which has been installed in a number of agencies to provide a distributed directory facility. The future of the GENIE project is not certain owing to funding and management problems.

References

Johnson BD, Shelley EP, Taylor M & Callahan S 1991. The FINDAR directory system: A metamodel for metadata. in *Metadata in the Geosciences*, eds DJ Medyckyj-Scott, I Newman, C Ruggles & D Walker, Group D Publications, Loughborough, UK, 123–37.

NASA 1988. *Directory Interchange Format Manual*, Version 1.0, July 13 1988. NSSDC/WDC-A-R&S.

NASA 1993. *Directory Interchange Format Manual*, Version 4.1, April 1993. NSSDC/WDC-A-R&S.

Shelley EP 1995. The National Directory: Current status and future directions. *Proceedings of the Third National Conference on the Management of Geoscience Information and Data*, Australian Mineral Foundation, Adelaide, Australia, 18–20 July 1995 24, 1–7.

Shelley EP & Johnson BD 1995. Metadata: Concepts and models. *Proceedings of the Third National Conference on the Management of Geoscience Information and Data*, Australian Mineral Foundation, Adelaide, Australia, 18–20 July 1995 4, 1–5.

Attachment 1 Supplementary information

A Contacting NRIC

Addresses

National Resource Information Centre
PO Box E11
Kingston ACT 2604

1st Floor, John Curtin House
22 Brisbane Avenue, Barton ACT 2600

Tel: 06 272 4688
Fax: 06 272 4687

National Directory of Australian Resources

External users may access NDAR via a user interface supplied by NRIC. In addition, NRIC can carry out searches of the directory on behalf of infrequent users.

NRIC also has a Web site: http://www.nric.gov.au

B State systems

West Australian Land Information Directory

This FINDAR-based directory is available as an online service through the WALIS Web Home Page:

http://www.walis.wa.gov.au

For further information contact:

Director
WALIS Office
PO Box 2222
Midland WA 6056
Tel: 09 2737046
Fax: 09 2737691

Queensland Land Information Directory (QLID)

Directory is available on diskette for an annual subscription which includes update service.

Contact: Department of Lands
Locked Bag 40
Coorparoo Delivery Centre
Qld 4151

NSW Natural Resource Data Directory (NRDD)

The current edition of this directory can be purchased on diskette and CD-ROM. An update is proposed for mid 1996.

Contact: Office of Land Information Policy & Coordination
Department of Land & Water Conservation
PO Box A2134
Sydney South NSW 2001

Tel: 02 228 6052
Fax: 02 223 8650

Internet: olipac@slim.slnsw.gov.au

South Australian Spatial Information Directory

The current edition of this directory is available on diskette. Regular updates are planned.

Contact: Data Management Branch
Department of the Environment & National Resources
282 Richmond Road
Netley SA 5037
Tel: 08 226 4946

Victoria Directory of Geographic Data Sets

This directory is available in booklet form.

Contact: Office of Geographic Data Coordination
Level 5
436 Lonsdale Street
Melbourne Vic 3000
Tel: 03 9603 9100
Fax: 03 9603 9199

C. Spatial Data Transfer Standard

This is an Australia/New Zealand Standard and has been published in three volumes:

AS/NZS 4270.1:1995	Geographic Information Systems: Spatial data transfer standard. Part 1: Logical specifications (216 pages)
AS/NZS 4270.2:1995	Geographic Information Systems: Spatial data transfer standard. Part 2: Spatial features (64 pages)
AS/NZS 4270.3:1995	Geographic Information Systems: Spatial data transfer standard. Part 3: ISO8211 encoding (52 pages)

These are available through Standards Australia offices in all capital cities (check your telephone book) or through the mail order sales group in Sydney (Tel: 02 746 4600; Fax: 02 746 3333).

D. Federal Geographic Data Committee (USA)

To obtain copies of the standard, contact:

FGDC Secretariat
c/o U.S. Geological Survey
590 National Center
Reston, Virginia 22092 USA
Tel: + 703 648-5514
Fax: + 703 648-5755
Internet: gdc@usgs.gov

The standard is also available by anonymous file transfer protocol (FTP) from:

FGDC.ER.USGS.GOV under /var/ftp/pub/metadata

If a download through a WWW browser is preferred, set 'Load to local disk' and press one of the following options:

- Wordperfect 5.0 Version
- PostScript

[Note: The WordPerfect file is set up for 10-point, Times Roman font on an HP LaserJet 4 printer. It can be printed using other fonts or printers, but the sections that refer to page numbers (ie the table of contents and list of data elements with page numbers) may have to be regenerated. The file contains codes that aid efforts to regenerate the page numbers.]

Attachment 2 Attributes in 'standard' dataset description in the National Directory of Australian Resources

Section	Attributes
Identification	Name, acronym, abstract, owner and other organisations associated with the dataset.
Data items	Name, description and, where applicable, spatial resolution of each item in the dataset. Items can be grouped where appropriate.
Spatial identification	Type of spatial referencing, projection, coordinate units and feature types.
Spatial coverage	General and detailed information on the area covered by the dataset.
Dataset information	Working form, working medium, size, applicable hardware and software, interchange format and supporting documentation.
Data currency	Custodian details, data collection start and end dates, dataset update frequency, future proposals, archive details.
Data lineage and quality	Data collection method, source material, data processing details, positional and attribute accuracy, consistency and completeness.
Ordering information	Access restrictions, output products and charges, supplier information and order procedure.
Keywords	Keywords describing the dataset suggested by person providing the entry.
Organisation/position information	Additional information about the custodian or supplier organisation/position.

Attachment 3 Summary of ANZLIC core metadata list

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

Attachment 3 Summary of ANZLIC core metadata list (cont)

	Geographic extent: name(s)	A picklist of pre defined geographic extents such as map sheets, local government areas, catchments, that reasonably indicate the spatial coverage of the dataset
	OR	
	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 verification
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

eriss metadatabase development: A starting point

Tony House[†]

Abstract

eriss needs a catalogue of its data resources—a metadatabase. This paper will briefly describe how ***eriss*** arrived at this conclusion through seeking external advice and internal consultation. It will describe not only why ***eriss*** needs such a database, but also what we expect of it. Initial steps to develop the database have been taken by ***eriss***, but we are keen to further explore the concepts and possibilities before going further. Thus, our ideas are presented for your comment.

Introduction

eriss intends to create a metadatabase—a database that contains data about data. The aim of this paper is to describe briefly the impetus for this development; to describe what ***eriss*** expects of such a development; to give some indication of the situation we are starting with; to put forward the results of a first step towards a metadatabase; and to give some guidelines as to where we go from here. The development of a metadatabase is seen as an integral component of the further development of ***eriss*** into a broad and multi-faceted environmental research institute (A Johnston pers comm). This initiative recognises the value of the data resources that ***eriss*** has amassed over approximately the last 15 years.

The initial plan was to incorporate a Geographic Information System (GIS) into the operations to support the Institute's research, monitoring and environment management functions (Riley et al 1992). A consultancy was set up with Genasys II to look at how this aim could be best achieved. After some discussion, the terms of reference of the consultancy were soon broadened to encompass information systems in general at ***eriss*** (see Devonport 1996).

Devonport (1996) noted that an integrated corporate information system at ***eriss*** did not exist, and reported that such a system was a prerequisite for developing a corporate GIS. Nevertheless, this report did state 'the implementation of a metadatabase was identified as the highest priority for the improvement of the present information system'. Given this basis, steps were taken to actively develop a metadatabase as a central component of a corporate information system.

The objectives

The minimum requirements for a metadatabase at ***eriss*** are that it will provide a structured mechanism to:

- identify what datasets ***eriss*** has
- identify the essential qualities of these datasets
- identify where the datasets are and how to access them.

It will, in essence, provide a means to catalogue the numerous datasets that have been generated by ***eriss*** over the past 15 years (approx) and in the future. This can be the heart of

---

[†] Environmental Research Institute of the Supervising Scientist

a corporate information system into which a GIS can be integrated. The metadatabase should be integrated into the ***eriss*** information system such that most of the elements of metadata for a dataset are collected as a part of normal activity (ie collecting and recording metadata is not seen as an adjunct or burden to collecting research data).

The user interface to the metadatabase is expected to provide enough information to the user to allow them find their way around the metadatabase. This should allow a user with very little knowledge of the contents of the metadatabase to find datasets of interest, as well as letting users who know precisely what they are looking for to locate it with a minimum of fuss. It should also provide the opportunity to 'browse' through the metadata—to thumb through the catalogue to get a feeling for what sort of research is being done and has been done by ***eriss***.

It is expected that the metadatabase will contain enough information about each dataset for a prospective user to be able to decide whether or not the dataset is of use in his/her current project. If he decides it may be then the database should also tell him who to see or what to do to get more detailed information and indeed how to gain access to the dataset.

The current state

The past and current data management situation at ***eriss*** could be considered a classic demonstration of why an organisation, particularly a scientific organisation, should have a corporate metadatabase. A very fragmented 'system' of data management has arisen from a combination of the following factors:

- the Institute has existed for 15 years (approx)
- there has been no corporate data or metadata system
- the Institute encompasses a wide variety of scientific disciplines
- staff at the ***eriss*** have been more or less grouped by scientific discipline
- research personnel have had a fairly high rate of turnover
- research directions have shifted significantly

These factors have combined to produce a very fragmented 'system' of data management. ***eriss*** has reached the point where some datasets may lose their usefulness because the knowledge about the datasets—the sampling methods, the accuracy of the figures, the analysis procedures, where the data were collected, ie the metadata—has been lost to the Institute. In effect, this represents a potential loss of information.

The lack of an integrated information system for the Institute has led the individual research groups, and sub-groups, to put their own systems together to handle the datasets they generate, and, although quite good, are nevertheless isolated stand-alone systems. Knowledge of the existence of such systems may extend beyond the small group of people who use them, but the knowledge of the details of the data they contain almost certainly does not. This information, the metadata, may be recorded in a wide variety of ways: notebooks, field notes, Institute registry files, computer file, and quite often in someone's head. This sort of approach to data systems, when combined with the age of the Institute, the rate of staff turnover and the changes in the research priorities of the Institute, has simply magnified the problem.

The Institute is currently on its third generation of computer systems (the fourth if you count the first few years without one), each new generation essentially incompatible with the

previous. This means some datasets the Institute currently has are no longer readily accessible. The change in other technologies used by various disciplines within the Institute over the years has meant a change in the 'nature' of the data that have been collected. In some cases this has affected the collection of long-term datasets. These qualitative changes which should have been recorded in a metadatabase may be recorded somewhere, or the people collecting the data simply 'know' of the changes and allow for it in their analyses.

The wide variety of scientific disciplines active within the Institute has meant that the datasets they have generated have taken a variety of forms. This not only means the obvious forms such as notebooks, survey sheets, a variety of computer based formats, published papers etc, but also means sets of physical items such as water and soil samples, samples of plants, leaves and pollen, calibration samples for assorted testing equipment. Again such datasets are usually very specific to a particular area of research. The people active in that area have usually documented the datasets and how to access them, but such information is usually only accessible to those people.

Our first step

As a first step for ***eriss*** towards a metadatabase discussions were held with representatives from the various groups within the Institute. The aim was to produce a basic set of items (metadatabase fields) that could be used to describe the variety of datasets at ***eriss***. The requirements of each group varies but there were essential components that all groups identified and these represent the core items of each metadatabase record.

The items have been generated for the most part on the basis of posing the question: if you wanted to find a particular dataset how would you describe it—what questions would you ask? The responses to this question are grouped together below.

What it is	The project the dataset was associated with
	The subject/topic of interest
	The title of the dataset
	The description of the dataset
	The purpose of the dataset
Who did it/has it	Names of people who generated the dataset
	Names and contact information of the custodians
When/where	The time period over which the dataset was created/collected
	The geographical location to which the data pertains
How to access it	Where to find the dataset
	What form is it in
	What is the procedure for accessing the dataset
Any related material	Related references, publications, datasets

A number of other items describing a dataset were mentioned as being important once the dataset had been identified:

Data quality	Some indication of the accuracy of the data
	Some indication of the completeness of the data
	Some indication of the developmental history of the data
Currentness	Is the dataset still updated
	Is it still being developed
Access or use restrictions	Are there limits on who can access the data
	Are there limits on how the data may be used

Three other items were included as a matter of housekeeping for the metadatabase:

A record ID	A unique identifier for this record in the metadatabase
Date of entry/update	When was this entry made or last updated
Updated by	Who entered/updated this record

These basic fields are expanded upon below.

What it is:

Project	The project identifier. This is currently a unique project number but has not always been so nor will it necessarily stay so. A project title in text is expected.
Subject	A free text statement of the subject of interest for example 'Calibration sources' or 'seasonal pattern of floodplain vegetation'
Title	Title of dataset in text
Description	Free form text describing the data set
Purpose	Free form text describing why the dataset was created

The first three of these should be compulsory for every entry in the metadatabase, the last two would be useful for most entries.

Who:

Authors	Names of people in the team that generated the dataset
Custodian	Name and contact information of person currently responsible for the dataset

When/where:

Time period	Period of time over which data was collected eg Jan 1993 – Mar 1994. It may be that dates down to the day might be necessary.
Where	The geographical area involved. Initially this will probably be a word description eg 'Magela Creek'. It may need a second field for a numeric format in AMG, UTM or similar co-ordinate system.

How to access it:	
Location	Current physical location of the dataset. This can be an indication of which filing cabinet in which office, the name of a publication, the name/location of a computer file.
Media	A word description of the physical form of the dataset, eg computer diskette, physical sample in jar, A5 notebook
Format	The internal structure of the dataset, eg ASCII file, WINWORD file, executable files, ring bound folder
Procedure	How to access the dataset— which programs to run etc

Related material:	
Other references	Reference to related materials, publications etc in free form text

Data quality:	
Data accuracy	An indication of the accuracy of the dataset
Data completeness	An indication of the completeness of the dataset
History	An indication of the development history of the dataset— the type of analysis, data smoothing and error correction that have been applied to this dataset

Currentness:	
Work status	The dataset still being developed/created. A single word entry from a set of choices
Updates	A single word description of the update policy for the dataset eg daily, weekly, monthly, quarterly, as-required, none

Housekeeping:	
Record ID	Used internally by the database system and generated by that system
Record date	Maintained automatically by the system
Updated by	Maintained automatically by the system generated from the login ID of the user

A diagrammatic summary of the above is presented in figure 1.

Where to from here?

eriss has developed a concept of what it wants from a corporate metadatabase. We are less sure of how to get there. We have taken the first step in assessing our users' requirements for such a system, as documented on the preceding pages. But from there we need guidance. We need the advice and assistance of people who have already done what we have only started to do. For that reason, we convened a metadatabase workshop. The task ahead of us could evolve as follows:

- feedback from this workshop being used to modify the basic field items that have been proposed;
- the development of a trial system with a structure using the modified fields;

- testing that system using as wide a range of datasets as possible to check that it does fulfil the user needs;
- expand the trial system into a full corporate metadata system;
- use the metadatabase as the framework for producing a corporate information system, including in that a GIS.

The second and third of these points will almost certainly have to be repeated several times.

That's the situation. We know where we are, we know where we want to go, but we need your assistance in getting there. We are here to listen to your advice before we proceed.

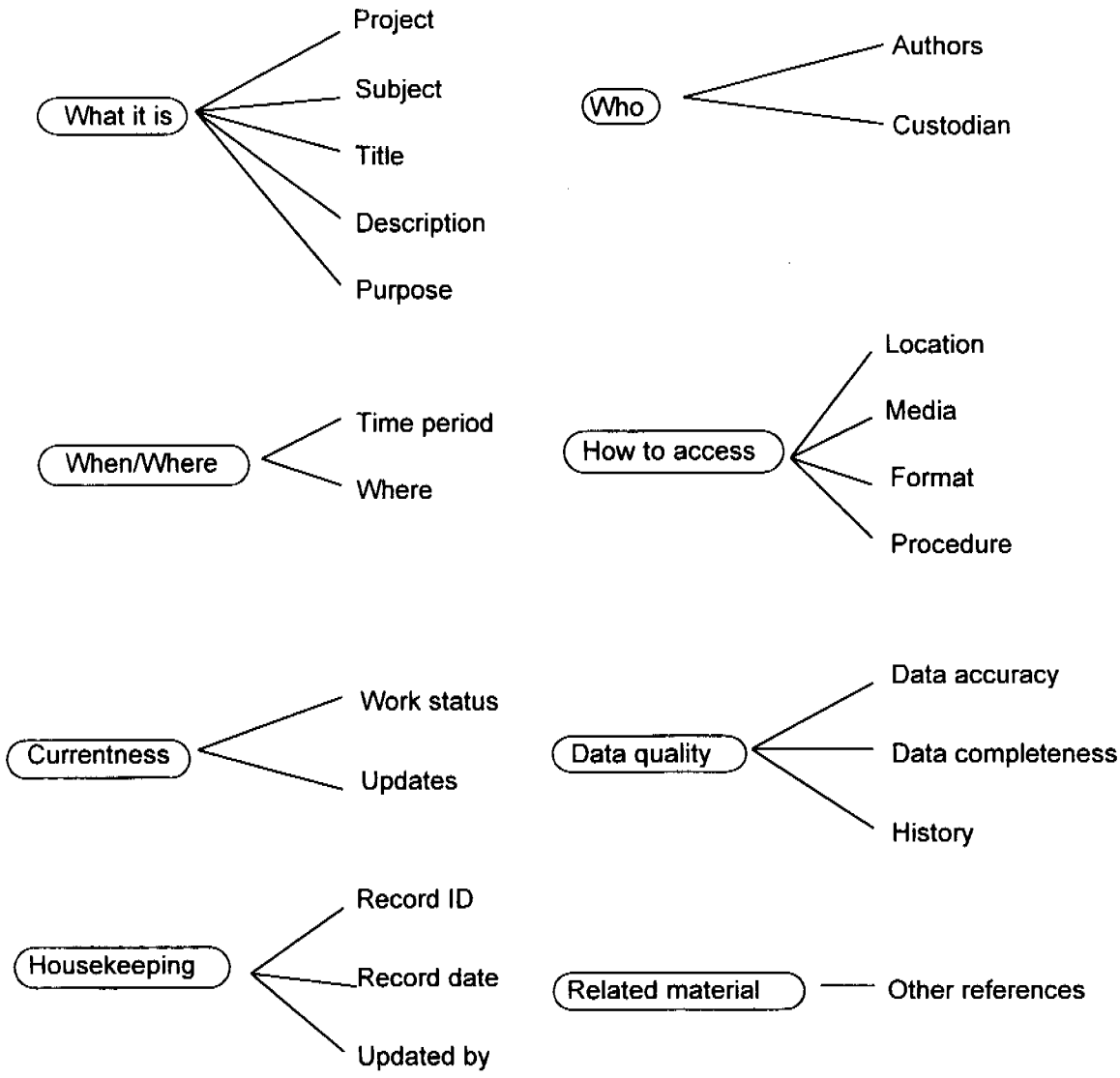


Figure 1 The main elements of a record in the metadatabase grouped by function

References

Devonport C 1996. ***eriss*** information system: Opportunities for improvement. Internal report 219, Supervising Scientist for the Alligator Rivers Region, Canberra. Unpublished paper.

Riley SJ 1992. The role of Geographic Information Systems in the Office of the Supervising Scientist. in *Proceedings of the GIS and Environmental Rehabilitation Workshop.* Darwin 4–5 September 1992, eds C Devonport, SJ Riley & SM Ringrose, Supervising Scientist for the Alligator Rivers Region, AGPS, Canberra, 92–99.

Bibliography

Crossley D 1996. *WAIS through the web—Discovering environmental information.* Environmental Resources Information Network (ERIN), Canberra.

Environmental Resources Information Network (ERIN). *ERIN Distributed Spatial Data Library: Metadata Guide.* http://kaos.erin.gov.au/general/spatial_data/

US Federal Geographic Data Committee (FGDC) 1994. *Content standards for geospatial metadata.* ftp://fgdc.er.usgs.gov/gdc/metadata/

Wadlow R 1995. *Metadata for Geographic Information Systems.* Consultant report, ESRI Australia, Sydney.

Relational databases for environmental and biological data: Combining existing datasets—Points and pitfalls

Margaret Cawsey[†]

Abstract

It is now generally accepted that great advantages, in terms of increased data accessibility, are conferred by holding data in relational database management systems. However, it is crucial that a logical process of database development be followed in order to achieve the purposes for which the data have been collected. The data model must adequately reflect the purpose of the project. The database must be placed in context within the organisation in which it is to function. The data items must be standardised and defined accurately. Where existing data, previously collected for a different purpose, are to be combined within a single database, particular attention must be given to data definition, to ensure that data are not accidentally used inappropriately. The database structure must be designed carefully, to avoid propagation of errors caused by data redundancy and other design flaws which might lead to breach of integrity. Problems encountered during the physical building of the database must feed back to the design and definition phases to ensure that mistakes are not made and that necessary data items are not left out. The data must be verified to the highest possible standard to ensure their integrity and value for use. Throughout the database development process, it is important that those directing the process understand the data, their purpose and the underlying assumptions of the intended analyses, otherwise errors can occur and data may be used inappropriately.

Introduction

The common trend in the world of data management is a move away from flat file datasets to multi-table databases, managed within computer management systems that facilitate input, output, storage, interrogation and maintenance of the integrity of the data (Belbin et al 1995). Most database management systems now recognise that relational databases provide the best and most functional of these systems, and if relational features are not automatic (eg automatic linking of data items, or columns, with the same names) then they are usually readily programmable.

The benefit conferred by a relational database lies in its capacity to represent a data model that accurately reflects the real-world relationships between data items. The more closely a data model reflects the interconnections and hierarchical nature of the real-world system it is supposed to represent, the more robust the model, ie it is more flexible and adaptable and easy to develop/maintain, and less prone to requirements for radical restructuring. The design of the database is the crucial element that allows the maximal advantage to be achieved using a relational database.

However, a database can only be as useful as the data it holds. A database is not a goal in itself, but must be designed to meet the goals of the project(s) it serves. It doesn't matter how well a database is constructed if, in the final analysis, the data are inappropriate to answer the questions that the users need to ask.

---

† CSIRO Division of Wildlife and Ecology, Lyneham, ACT

Biological and environmental systems are complex, and many things must be taken into consideration when designing integral data models to reflect the complex relationships inherent in such systems. The rigorous approach to identifying and defining required/available data items, which must be undertaken as part of the process of database design, is an excellent first step in breaking complex systems into component parts that can be realistically managed in both data management and in analysis. It also facilitates the recognition of the appropriate (and inappropriate) linkages between data items, and thus the construction of a robust data model.

Goals and purposes of relational databases

I define the goal of a database as ultimately to make the appropriate data readily accessible in the best forms to meet the requirements of the users.

The purpose of the database depends on the subject matter, objectives and goals of the users, and the organisational context in which the database is to exist (Goodell 1992). The purpose requires the utmost attention in the first instance, before the design of a database is attempted. The questions that must be answered first are: why do you want a database, who will use it and what products do you expect from it?

Definition of purpose requires the detailed definition of the data/information that researchers expect to get from the database. You must know what you want to get out of a database before you can decide what data items have to go into it, and the form those data items should optimally take. In order to know this, the researchers must have a good notion of the analyses they wish to do and the hypotheses or issues they wish to address. The data type, sampling method and accuracy of measurement must be suitable to the proposed analyses (Austin 1995).

When a database is being designed for a new research project, this ought generally to co-occur with the experimental design of the project itself. The purpose is then explicitly tied to the project's goals and objectives from the outset. This ensures that the worst possible scenario cannot occur *viz* collection of data which cannot contribute to testing the null hypotheses, or perhaps only to testing irrelevant null hypotheses. Data collection in the field is an expensive operation, and collection of useless/unnecessary data is to be avoided at all costs.

Where a single database is to be designed to make maximal use of existing data—data from field exercise(s) which have been carried out by other research groups, each of which will have had different purposes—then we face a different suite of problems. The researchers should have a clear idea of the purpose of such a database before they start. It is best at this stage to couch requirements in terms of what people want rather than limiting the exercise at the beginning by only considering what they think they can get. Also, it is wise to be open-minded and keep an eye out for other opportunities and advantages that might arise as you proceed.

There are many advantages to combining existing datasets into a single database, which cannot be conferred by using multiple databases. The major advantage lies in the potential increase in the number of degrees of freedom; often, biological datasets are too small and sparse to allow any confidence in statistical analyses. Another significant advantage lies with the ability to compare and contrast datasets and data items from different sources, which provides the potential for generation of new ideas and approaches. This also permits global assessment of gaps/biases in sampling, which would otherwise be very difficult. A collection of separate databases may place unnecessary limits on the way data are used. This paper will particularly address the points and pitfalls associated with combining existing datasets into a single database.

Context

All potential users should have a vested interest in the development, well-being and use of the database. They must own the idea, and believe that it will be useful. A beautifully designed database which no-one will use is useless. Thus, it must be placed into context within the organisation, with both managers and users involved in the ideas, design and implementation (Goodell 1992).

It is also important to place the database and database management system in context within the framework of other (computer-based) tools you anticipate using, eg GIS, data analysis packages, expert systems, decision support systems, graphics packages etc. All your interconnecting systems should be able to communicate data to and fro with as little external programming and fiddling as possible. Ideally, the database management system should be able to access GIS files and vice versa, but I haven't yet found this capability to be generally available.

Data definition

Having defined the purpose of the database and identified its context, the next step becomes one of deciding on (an) absolute minimum set(s) of data required to meet the 'purpose'. For example, site location would be required absolutely for all sample-sites. In the forest research context in which I work, the minimum dataset for predictive modelling of distributions of tree species comprises location, plot size, and the record of presence/absence for the nominated canopy tree species. All other data items required for modelling can be derived from maps, if they are not already available, and from climate estimation software. For vegetation classification of the canopy trees into 'communities', the constraints become more strict. That is, the minimum dataset requires the presence/absence records of all canopy tree species extant in the study area. For fauna modelling the situation becomes more complex, as animals respond to the environment at a different scale than do the trees (Braithwaite et al 1983, Lindenmayer et al 1991, Pausas et al 1995), and variables that measure habitat structure become necessary for the modelling.

A database design may incorporate several compatible data models which can provide data suitable for many different analyses and different scales. However, not all component datasets may meet the required criteria for each analysis, and facility will have to be made within the database and data dictionary to note which component datasets are suitable for what, and thus ensure that data are not used inappropriately.

The component datasets must be critically compared to discover the lowest common denominators amongst data items. For some datasets it can become immediately obvious that a small amount of extra data acquisition, eg recording topographic position, slope, aspect, elevation from maps, would be a cost-effective investment for the corresponding increase in value of the dataset to the whole exercise. Extra field sampling can also raise the value of the final database as a whole, adjusting biases and gaps (geographic, environmental) in sampling that are obvious even after many datasets have been combined. If, over all datasets, the lowest common denominators are too low to allow meaningful conclusions to be drawn from analyses, and the costs of augmenting the data to the appropriate standards are too high, then the purpose of the database cannot be met, and continuation with the task of combination of datasets into a single database becomes an exercise in futility.

The compilation of a metadatabase, or data dictionary, is an excellent way of examining existing data, and defining the data items from each component dataset, to look for compatibility or lack thereof. Belbin et al (1995) have examined the requirements and

limitations experienced by ERIN in dealing with many and varied existing datasets and databases in Australia. Although under normal operations a particular organisation is unlikely to experience problems on the scale that ERIN encounters, many of the points made by Belbin et al are salient at all scales.

At the end of the information-gathering stage you will have a comprehensive list of all data items and their definitions, over all component datasets, and have begun to develop a good idea of how they relate to each other.

Survey design

Although this point relates particularly to the definition of data, I think it important to emphasise the real need to understand the survey design employed for every component dataset. On the face of it, the same data items may seem to have been collected in the same way for several datasets. However, the physical and temporal designs of surveys can lead to different degrees of spatial and temporal autocorrelation for example, and thus different assumptions must be accounted for in any modelling exercises. Misunderstandings about the meanings of certain variables (eg 'abundance' data), acquired because the exact context and method of data collection have not been teased out, have led in the past to analyses having to be redone. It is imperative for the reputations of the organisation and the researchers that their work stands up to scientific scrutiny. These days there are also legal implications inherent in accidentally drawing scientifically insupportable conclusions because the data structure has been misinterpreted.

Physical database design

The purpose of this paper is not to explain the finer details of the art of database design, and I would be presumptuous to try. The manuals of most database management systems emphasise the need for good design. There are computer-based training courses in relational theory, and companies which run training courses. Texts on design tend to become rapidly outdated, however, the basics of defining the relationships between data items do not change (see Kent 1983, Smith 1985, Teorey et al 1986). There are also computer packages, called 'CASE tools', that can virtually design your database for you (see Gibson 1989, Orr et al 1989, McClure 1989a,b). I haven't used them, and so can make no recommendations either way.

The information gathering stage is the first and most important part of the design phase. After this phase, it is likely that the 'entities' (tables) and their 'attributes' (columns) will have become obvious, as will most of the relationships between them. Normalisation (Kent 1983, Smith 1985), which is the process of reducing data redundancy (duplication) to an absolute minimum, will further elucidate tables and columns. The closer you get to the 'perfect' normalisation the more robust your database is likely to be.

However, in my experience, you don't have to go overboard on normalisation to the point where it drives you crazy. Perfect normalisation may be useful (and achievable) if you are designing an airline schedule database, but the complexity of the hierarchy of relationships with biological and environmental data makes it sometimes counter-productive. If you always search/extract data using a particular primary indexing column, and too much time is taken with nested sub-selects, then ignore the finer points of normalisation and duplicate that column in other tables. Just make sure the column rules and constraints that you build into the database are sufficient to cover for this contingency.

You should never allow the 'rules' of database design to prevent the users from getting what they need. Basically, the procedure is an exercise in commonsense.

Draw 'mud-maps' of your database, tracing the linkages between tables. This will help you make sure that all tables link to another. You should have no data lost because there is an unlinked table wandering around in your database.

Building the database—likely problems

You may have agreed, in the design phase, that a particular data item requires storage. But which version? Often different projects sample the same general data item, but in slightly different ways. The differences may not necessarily lead to mutual exclusivity, but researchers need to be absolutely aware of the assumptions implicit in any set of data they analyse and draw conclusions from. They need to know the suitability or otherwise of data collected in various ways for different kinds of analytical approaches (see Belbin et al 1995, Austin 1995, Austin et al 1995). Thus, it is important to document every little difference.

There are penalties associated with storing too many data items. The storage of 'unnecessary' data items costs in terms of time taken to verify (and collect) them, space on the disk, and in space taken in the database, which equates to time taken in transaction processes. This is very important for databases with high transaction rates (eg airline schedules). I have found that databases built for biological/environmental data do not generally have high transaction rates; researchers usually wish to take a selected dataset away and analyse/display it using some other tool, rather than to constantly and intensively interrogate the database itself. Also, once data are in and verified, they rarely require revision/update.

With most database management systems, size limitations to a database (in terms of numbers of tables and columns) are now virtually non-existent. Also, the costs of storing null values (missing data) are nil. Therefore, all columns pertaining to a given primary identifier can be placed into the one table. For example, all 'unchanging' environmental variables pertaining to sites can be placed in one table, whether or not they belong to the common minimum dataset, or are only specifically related to (a) particular dataset(s). This allows the logical structure of the data model to be maintained, while allowing different versions of the same basic data item to be held for different datasets.

There are 7 general categories of data problems likely to be encountered while building a database.

Data items recorded dubiously

An example of this in forest research is a situation where plot size has been recorded as a standard size, although in many cases that *actual* plot size was half that size, and values (eg numbers of stems in different size classes) have been doubled to make up to the supposed plot size. This has particular implications for stem abundance, and for species richness, which are known to be a function of plot size.

Data items of dubious value

An example of this has been judged to be a variable called 'soil texture', which has been judged by a panel of experts to carry no significant meaning for analysis, in comparison with other variables which are easier to record in the field without taking actual soil samples.

Data items with the same meaning but different levels of categorisation

There are hazards associated with aggregation of categories when attempting to make data items compatible. For example, if one dataset holds topographic position in 6 categories, and

another in 5, and the 6 categories are easily aggregated into 5 (by combining mid- and upper slopes to a straight 'slope' category for example), it is dangerous to assume that the best course of action is to only store the 5 category variable for all datasets. You may lose information crucial to the purpose of a particular component dataset; perhaps the difference between upper and mid-slope might be significant when modelling habitat preference for some arboreal mammal species. If you aggregate at the storage stage then you lose the capacity to disaggregate at the analysis stage without going back to the raw dataset. You never lose the capacity to aggregate at any stage.

Data items with the same meaning but different methods of 'measuring' them

An example of this is 'measurement' of basal area. One component dataset in the FOREST database has it calculated simply by multiplying the mid-point of each of 8 size classes of DBH by the number of stems in each class, and accumulating the results. The other does the same for the first 7 size classes, but adds the *actual measured DBH* of all size class 8 trees. This tends to increase the values of basal area somewhat, but whether this is significant is uncertain.

Data items which should have been stored in the database and weren't

In my experience, the penalties for not storing data items which are judged initially to be 'unnecessary' and subsequently discovered to be significant, or even only interesting, are far worse, and more time-consuming to mitigate. An example for the forest research is the omission of the actual DBH/basal area of the largest trees. It has become obvious that arboreal mammals become more abundant where trees are over a particular DBH (Pausas et al 1995). It is much easier to verify and include data items from a dataset all at one time, while the details of the structure and meaning of that dataset are fresh in one's mind, than to return at some later date to retrieve a data item when one has forgotten all the details. In some cases circumstances cannot allow time to retrieve the data, and the possible benefits of using them in analyses are effectively lost.

Unavailable data items which might be worth the investment to capture

(Already discussed above in the Data Definition section.)

Data items which are defying definition

This problem arises when you cannot get two or more scientists to agree on a definition of a data item. An example is 'vegetation community'. We have two items derived in *exactly* the same way. One scientist is happy to call it 'vegetation community' and the other isn't, leading to an impasse.

Data verification and integrity

Once any dataset has been placed into a database, then all 'old' versions of that dataset should be archived and removed from easy access. Everyone should be confident that the dataset they are analysing is the latest, most up-to-date version. Therefore, it should come from one place only, and any updates should occur in that same place. The penalties for failure in this area are the same as those incurred through misinterpretation of data structure (see above).

Data verification is an iterative process. The first pass in any analysis will usually discover (possible) data errors (eg 'outliers', Austin et al 1995) which have escaped the first nets of data verification (applied as the data are loaded into the database). This leads to error checking, correction and re-output of the data for analysis. For error checking to be possible,

it is imperative that each data item can be related to the physical field data sheets – where they still exist (field data sheets are an extremely valuable resource and should always be kept!). If an error obviously exists, but no-one knows how to correct it, then the observation is usually excluded from the analysis. Biological data are often sparse enough without losing unnecessary degrees of freedom through uncorrectable errors. In most cases, if the field data sheets are available, the error can be identified and the observation successfully retrieved. For this reason, it can be important to store data items which exist only for the purposes of allowing the data to be related to field data sheets.

There are three ways to absolutely guarantee the failure of a database:

People don't want to use it

This occurs where a database hasn't been placed effectively in the organisational and research context. Ownership and commitment are necessary from both managers and potential users.

People can't use it, even when they want to

There can be several causes:

- unfriendly interface
- database not accessible
- data in 'wrong form'
- people not trained sufficiently to be able to take advantage of the system.

People don't trust it

This usually means that data integrity has been breached, which has two general causes:

- people using old versions of data and updating/analysing those (this relates to lack of commitment and feelings of ownership from users—see point 1 above)
- data redundancy and bad data model leading to propagation of errors.

Conclusion

I cannot over-emphasise the depth of understanding of the data required when designing/building a complex database, particularly when combining datasets. Those building the database will provide the first nets that catch data error and misapprehensions about the data type and accuracy and survey design. They must understand the underlying assumptions inherent in the proposed analyses, otherwise logical errors will slip through the nets, and propagate themselves within the database. Prevention of this is more important than getting an absolutely perfect database design first off. With expertise in exploratory data analysis, problems with design will become apparent very quickly. The converse is not necessarily so.

The ultimate goal of any data collection/storage exercise is the information elicited from analysis. Analysis is a subject in itself (see Austin 1995), but it cannot be separated from the data and the way in which they are stored.

References

Austin MP 1995. Data capability, Sub-Project 3, Modelling of landscape patterns and processes using biological data. CSIRO Division of Wildlife and Ecology.

Austin MP, Meyers JA, Belbin L & Doherty MD 1995. Simulated data case study, Sub-Project 5, Modelling of landscape patterns and processes using biological data. CSIRO Division of Wildlife and Ecology.

Belbin L, Austin MP, Margules CR, Cresswell ID & Thackway R 1995. Data Suitability, Sub-Project 1, Modelling of landscape patterns and processes using biological data. CSIRO Division of Wildlife and Ecology.

Braithwaite LW, Dudzinski ML & Turner J 1983. Studies of the arboreal marsupial fauna of eucalypt forests being harvested for woodpulp at Eden, New South Wales. II. Relationship between the fauna density, richness and diversity, and measured variables of the habitat. *Australian Wildlife Research* 10, 231–247.

Gibson ML 1989. The CASE philosophy. *Byte*, April 1989, 209–218.

Goodell T 1992. Designing and building a database. in *The power of R:Base*, MIS:Press, New York, Chapter 2, 17–83.

Kent W 1983. A simple guide to five normal forms in relational database theory. *Communications of the ACM* 26 (2), 120–125.

Lindenmayer DB, Cunningham RB, Tanton MT, Nix HA & Smith AP 1991. The conservation of arboreal marsupials in the montane ash forests of the central highlands of Victoria, south-east Australia: III. The habitat requirement of Leadbeater's possum *Gymnobelideus leadbeateri* and models of the diversity and abundance of arboreal marsupials. *Biological Conservation* 56, 295–315.

McClure C 1989a. The CASE experience. *Byte*, April 1989, 235–244.

McClure C 1989b. The CASE workshop. *Byte*, April 1989, 246.

Orr K, Gane C, Yourdon E, Chen PP & Constantine LL 1989. Methodology: The experts speak. *Byte* April 1989, 221–233.

Pausas JG, Austin MP & Braithwaite LW 1995. Modelling habitat quality for arboreal marsupials in the South Coastal forests of New South Wales. *Forest Ecology Management* 78, 39–49.

Smith HC 1985. Database design: Composing fully normalised tables from a rigorous dependency diagram. *Communications of the ACM* 28 (8), 826–838.

Teorey TJ, Yang D & Fry JP 1986. A logical design methodology for relational databases using the extended entity-relationship model. *Computing Surveys* 18 (2), 197–221.

Developing decision support systems: Issues and considerations

Bruce Bailey[†]

Abstract

Decision Support Systems (DSS) are an extension of the information system concept. They are commonly constructed from: expert or rule-based systems; hypertext; graphics, sound and video; access to databases; and communication between software applications. Three case studies are used to illustrate the development and use of DSSs. Cropping 21 was developed for conservation farming and demonstrated the limitations of expert systems when dealing with semi-structured and unstructured decision making. Information ownership and 'use by date' were also issues with this system. CaLM Action Planning System (CAPS) was developed to support management by objectives for soil conservation in NSW. A prototype was used to involve users in developing the system. Options to integrate CAPS with other information systems were limited by an absence of data suitable for modern system modelling. Dune Base was developed for land and water management agencies and educational institutions. It uses hypertext and features a facility to trace the user's pathway through the system. It also takes into account the likely skills of the operator and provides a trade off between functionality and user friendliness.

Introduction

Decision Support Systems (DSS) are systems which provide the end-user with decision making information when a decision making situation occurs. Some people argue that DSSs are another generation of information systems. In reality, DSSs are merely an extension of the information system concept. The information technologist commonly uses the following technologies to construct a DDS:

- **Expert systems or rule-based systems**
  An expert system is a computer program that makes decisions or solves problems in a particular field by using knowledge and analytical rules defined by experts in the field. Experts solve problems by using a combination of factual knowledge and reasoning ability. In an expert system these two essentials are contained in two separate but related components, a knowledge base and an inference engine. The knowledge base provides the specific facts and rules about the subject, and the inference engine provides the reasoning ability that enables the expert system to form conclusions.

- **Hypertext**
  Highlighted terms in a screen-based document may be 'hit' with the mouse cursor to activate further levels of text or other media. It changes a two dimensional document into a multi-dimensional structure which can be explored along a variety of pathways at the will of the user.

† NSW Department of Land & Water Conservation

- **Graphics, sound and video**
  The multimedia 'fashion-feature' war has resulted in the development of several technologies which are particularly useful for developing DSSs. Video and sound are increasingly incorporated into DSSs, and these days a system lacking graphics would be considered mundane.

- **Access to databases**
  Most DSSs utilise database access. In the past, dealing with data files in different formats (eg xBase, Btrieve, Oracle etc) was problematic. Open database Connectivity (ODBC) techniques are helping to overcome problems associated with attaching to data files in various formats.

- **Communication between software applications**
  Dynamic Data Exchange (DDE) is also proving useful in developing DSSs which require communication between various Windows-based software packages (eg databases, spreadsheets, GIS systems).

As with the development of all information systems, there are many issues and considerations associated with the development of DSSs. There are several useful texts which discuss the DSS system development life cycle and provide a heady discussion on descriptive decision theory (see References/Further reading). The approach taken in this paper is to discuss some of the DSSs which I've developed and to utilise that very 'exact science' known as 'hindsight'.

Case study—Cropping 21

Cropping 21 (Bailey 1992) is a DSS for conservation farming which was developed during the late 1980s for the Soil Conservation Service of New South Wales. Cropping 21 was developed to assist farmers and extension officers in the adoption of conservation farming techniques in the wheat belt of southern New South Wales. In short, conservation farming implies the replacement of tillage, as a weed control method, with the 'selective' use of herbicides. The concern being that excessive tillage results in increased soil loss and soil structural decline. Many farmers experience difficulties in adopting zero-till/direct drill systems and consequently have reverted to more conventional tillage systems.

Cropping 21 was to provide the answers. It was planned as an expert system which allowed users to select the 'best' crop rotation, advise on existing and potential weed problems, deal with herbicide resistance, avoid root diseases and provide general advice on the plethora of problems associated with the adoption of the conservation farming system. The knowledge bases were to be compiled using data from numerous conservation farming trials and from the experience of extension officers, researchers, agribusiness, academics and farmers.

The Cropping 21 experience highlights a number of pertinent issues about DSS development, particularly regarding the use of expert systems. Expert systems are most useful for dealing with structured decisions, ie those we can predict will happen. We can't always predict when they will happen, but we can predict that they will happen. Expert systems are not useful for dealing with semi-structured and unstructured decisions, ie decisions which cannot be predicted or where there are too many interrelated variables to be considered. Although useful in the financial and medical fields, expert systems have a limited role to play in the environmental and resource management fields where semi-structured and unstructured decisions dominate. This point was highlighted during Cropping 21 'think tanks' composed of farmers and researchers. The think tanks were conducted to fill in the gaps in the knowledge bases and to devise 'best bet' options. Debate was common and agreement rare!

Information ownership was always an issue during the development of Cropping 21. The compilation of the knowledge bases required obtaining data from researchers in several government departments. Rivalry amongst departments and the reticence of researchers to release information unless it has been published is a constraint.

The 'use by date' of information is sometimes overlooked in the development of DSSs. Many, if not most DSSs, are developed and distributed in a project mode and the commitment to support the system in the longer term, is deferred—the 'lets see how it goes' approach. This can result in the use of systems which contain 'expired' information. This was a concern with Cropping 21 given the volume of information on herbicides and insecticides contained in its databases.

Case study—The CaLM Action Planning System (CAPS)

CAPS (Bailey 1994) was developed for the New South Wales Department of Conservation and Land Management (now incorporated into the Department of Land and Water Conservation) to provide decision support for the department's 'management by objectives' planning process.

CAPS is a PC-based, distributed system (installed in 67 offices) which allows data on the department's activities to be collated at the district, regional and state levels. It is an electronic (uses E-mail) reporting system which provides management intelligence on a monthly basis. It also includes modules for activity based costing and planning.

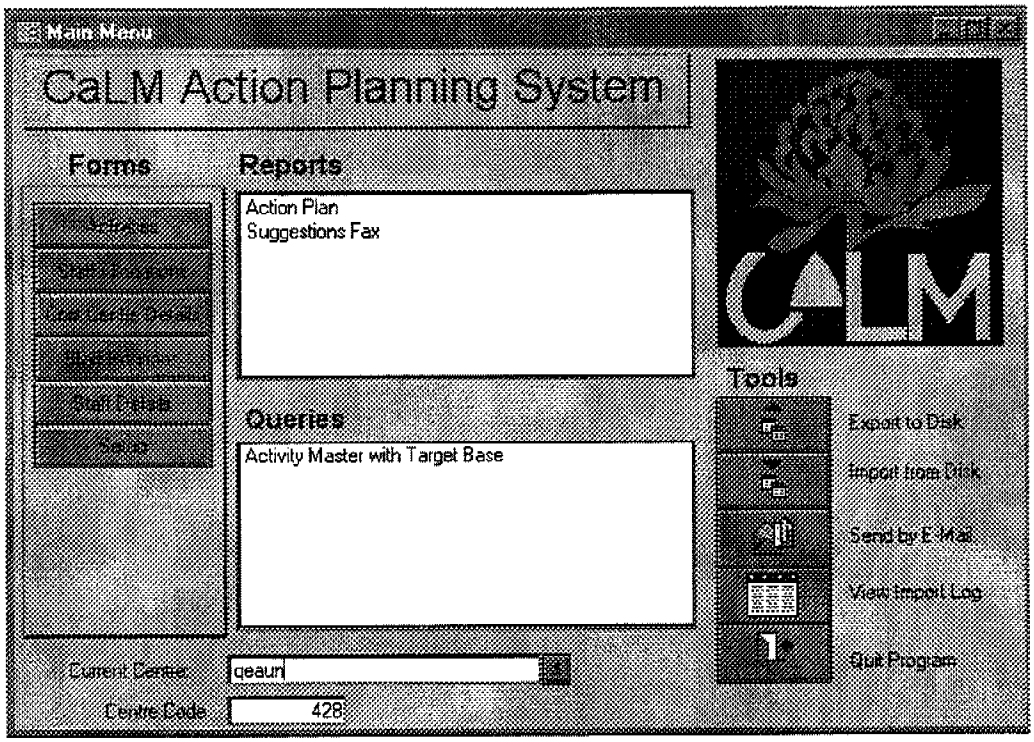
A CAPS prototype was developed early in the project and installed in selected offices. This approach was advantageous for the following reasons:

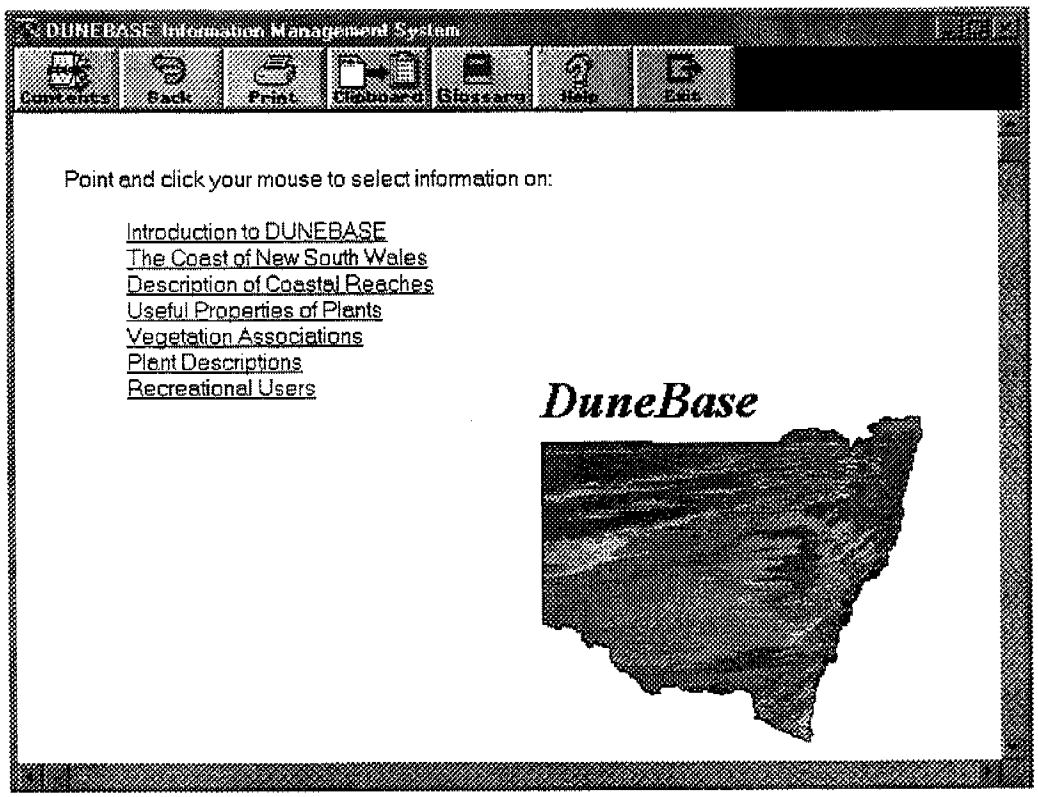
- The prototype encouraged active end-user participation. This increased end-user morale and support for the project.
- Iteration and change are a natural consequence of systems development—that is, users tend to change their minds. Phototyping better fits this natural situation since it assumes that a prototype evolves, through iteration, into the final product.
- End-users don't fully know their requirements until they see them implemented.
- Phototyping increased creativity through quicker user feedback.

Phototyping also provides an opportunity to exit a project before committing excessive funds and resources.

Often, the greatest challenge in developing DSSs is understanding, and getting users to understand what type of information is actually required to support their decisions. There are many managers, for example, who believe that 'information is power'—the more they can get, the better they can manage. This old maxim is false and could be changed to 'knowledge is power'. For information to contribute to knowledge it must be in the right form, be current, be timely in its delivery and get to the right people.

As often occurs in the development of DSSs, opportunities for extending and enhancing the system are identified. In the CAPS example, linking the system with financial and human resource management information systems would be extremely fruitful. Integrating systems in this way is not an issue in the 'ideal' organisation where a suitable information system model had been adopted and adhered to. Unfortunately, departmental information systems are seldom separated into the user interface, services layer and data layer as required in modern system modelling. The lack of separation tends to maximise interdependencies between the user, processes and the data. In this situation, expanding and integrating systems is a difficult task.


Main Menu
CaLM Action Planning System
Forms
Reports
Action Plan
Suggestions Fax
Queries
Activity Master with Target Base
Tools
Export to Disk
Import from Disk
Send by E-Mail
View Import Log
Quit Program
Current Centre
qeaun
Centre Code
428


DUNEBASE Information Management System
Contents
Back
Print
Clipboard
Glossary
Help
Exit
Point and click your mouse to select information on:
Introduction to DUNEBASE
The Coast of New South Wales
Description of Coastal Reaches
Useful Properties of Plants
Vegetation Associations
Plant Descriptions
Recreational Users
DuneBase

Case study—DuneBase

The DuneBase project (Chapman & Bailey 1995) developed from a major study of the dunes on the sand barriers of the 2000 km long coast of New South Wales. The project was concerned with three principle strands of study:

- Geomorphology, particularly the assessment of potential aeolian sand transport, dune formation and dune degradation.
- Vegetation, the species present, and their characteristic plant associations, were largely unknown, and were investigated, as were the special micro-environmental and community requirements of the highly adapted dune species, and their distribution over identifiable zones of the beach and dune.
- Land use, particularly the recreational and urban planning needs of the beach using public.

DuneBase was developed primarily for land and water management agencies and educational institutions.

When developing DuneBase it was recognised that environmental problems are usually complex and there is typically no static 'best' solution. DSSs must therefore be flexible so that users can explore and perhaps re-conceptualise a problem. To this end, DuneBase uses hypertext extensively to link information contained database files, text files, graphics and models. Hypertext, however, needs to be used wisely to prevent the creation of a pile of information 'spaghetti'. In allowing information to be connected according to the multiple associations it naturally has, by allowing the user to follow a 'stream of consciousness' process, there is also a danger that he/she will become lost. In Dunebase this has been avoided by creating a facility which retains a trace of the user's pathway through the system and (a) allows retracing that pathway step by step, or (b) returning directly to the top menu at any stage.

In Dunebase, database search functions have been simplified by developing interfaces which do not require the user to understand Structured Query Language (SQL). Many users find SQL and even Query By Example (QBE) difficult to use. Sheltering the user from SQL, however, does reduce search functionality. The inclusion or exclusion of SQL capability in the DSS represents a trade-off between user friendliness and functionality. One needs to consider the likely skills of the potential end user of the system when making this decision.

References

Bailey B 1992. Cropping 21: A decision support system for sustainable cropping. in *Rotations and Farming Systems for Southern and Central New South Wales.* eds G Murray & D Heenan, NSW Agriculture.

Bailey B 1994. *CAPS: The CaLM Action Planning System.* NSW Department of Conservation and Land Management.

Chapman DM & Bailey B 1995. *DUNEBASE: Information management system for coastal management.* University of Sydney and Department of Land & Water Conservation of NSW.

Bibliography

Hopkins LD 1984. Evaluation of methods for exploring ill-defined problems. *Environment and Planning*, B 11, 339–348.

Schiboula S & Byer PH 1991. Use of knowledge-based systems for the review of Environmental Impact Assessments. *Environmental Impact Assessment Review* 11, 11–27.

Sprague R & Carlson E 1982. *Building effective decision support systems*. Prentice-Hall, Englewood Cliffs NJ.

Waterman DA 1986. *A guide to expert systems*, Addison-Wesley, Reading MA.

Wright JR, Wiggins LL, Jain RK & Kim TJ 1993. *Expert systems in environmental planning*. Springer-Verlag, Berlin.

The use of Geographic Information Systems for wetland conservation

Richard Kingsford[†]

Abstract

Conservation of wetlands through nomination of reserves and listings as Ramsar sites may not be an adequate approach. Neither reserves nor nomination of Ramsar sites provide a representative sample of wetland types. This is partly because the information base is inadequate. As wetlands depend on their water supply, simply placing the wetland in a reserve is also unlikely to be an effective conservation measure. GIS linked with remote sensing could provide objective spatial analyses to identify key wetlands and threatening processes. The Murray-Darling Basin is an example where GIS is now being used to provide managers with information on wetlands—their location, size and unique features.

Introduction

Freshwater ecosystems have suffered considerably from human impacts (Allan & Flecker 1993). Wetlands have been drained and rivers dammed for hydroelectricity, irrigation and urban water supplies. There is increasing global recognition that wetland conservation needs to be afforded a priority status.

Wetland conservation in Australia

With this recognition there are some basic questions to answer. Firstly, what is wetland conservation? For most people it is presumably something about protection of ecosystem integrity for later generations. The next question is 'What processes do we use to achieve this goal?' although they may not actually use these words. Establishing reserves around a country is often considered the most effective way of conserving habitats, and wetlands. This has generally been the favoured approach to wetland conservation worldwide, except perhaps in Northern America (Bildstein et al 1991). Usually an important wetland area is designated as a national park or some equivalent. At an international level, the Ramsar Convention lists wetlands of international importance based on agreed criteria. Such a list tends to concentrate the minds of politicians and the community, although this Convention has no legal status in Australia and little effective power.

Further considerations

How do we know that we are reserving the right or best wetland ecosystems? We should be attempting to identify and reserve 'hot spots' (Wilson 1992), but the methodology behind identification of important wetlands (ie those worth conserving) has generally been ad hoc in Australia. Like all reservation processes, it is usually based on best available knowledge which is spatially subjective. The 1993 compendium of important wetlands in Australia illustrates this (Usback & James 1993); it was based on current knowledge which was very uneven. It was not spatially objective which means important wetlands may not have been identified. A further edition of the compendium (ANCA 1996) does contain more information, but it can not still be considered as a spatial inventory of Australian wetlands.

---

† NSW National Parks and Wildlife Service

The distribution of current Ramsar sites in Australia also illustrates the problem. New South Wales and Queensland have few compared with Victoria, Tasmania, and Western Australia. On this basis, wetland ecologists should all be working in Victoria! Clearly, there is a problem of representativeness mixed with more than a little politics.

To further illustrate this problem of bias, take one group of biota which is reasonably well known—waterbirds. There are some tremendously rich (species and abundance) wetland sites which have not been recognised until relatively recently (Kingsford 1995). This is because the state of knowledge on waterbirds, based on scientific publications, is geographically biased (Kingsford 1995). About 85% of studies on waterbirds since 1890 have been in 30% of the continent—that part outside the arid zone. Even the 15% of studies within the arid zone have been very limited in breadth. They are either species lists, band records or diet studies. The information for other aquatic fauna and flora groups is much poorer and is probably similarly biased.

As well as this problem there are generally two fundamental problems with a strategy for wetland conservation reliant on reserves: an upper limit to the reserved area and the uncertainty that reservation will effectively conserve aquatic biota. Firstly, about 5% of Australia is in conservation reserves. This may reach 10% one day, but it is unlikely that society will allow much more land area under reserves given current management and restrictive land uses. This leaves most biota out of a conservation process based on reservation. Secondly, wetlands depend on a water supply which may not be subject to conservation. There are recent examples where reserved wetlands continue to degrade because of water use upstream (eg Macquarie Marshes in NSW: Kingsford & Thomas 1995; Narau Lake in NSW: Crabb 1995). Conservation strategies need to be focused on management of threatening processes as well as reservation.

Spatial data for wetland conservation

In a survey of ranger staff in the New South Wales National Parks and Wildlife Service about wetland issues, the most frequently cited concern was the lack of information on local wetlands for making assessments of environmental impact. With little spatial data on wetlands, developments with deleterious impacts on wetlands usually proceed because counter arguments are weak.

Geographic Information Systems (GIS) combined with remote sensing is a useful tool for wetland conservation. Objective spatial analysis can be performed to identify key wetland sites worthy of reservation and threatening processes can be better managed. The GIS allows conservation biologists to work at much larger scales. This partially circumvents the tyranny of small decisions or the argument that 'there is another one just like it over the hill so we can develop this one'.

Consider the Murray-Darling Basin which is arguably the most heavily impacted river basin in Australia because of its importance for agriculture. Wetlands in the Murray-Darling Basin are perhaps among the most threatened ecosystems because there has been a considerable expansion in water diversion from major rivers, mainly to irrigation (MDBMC 1995). Expansion in the cotton industry in the last ten years has seen the use of water increase greatly as the huge public dams have been augmented by large off-river storages. These expensive structures store floodwaters, captured by pumps. There are no data on the locations of wetlands or their importance so that constructive comment may be provided for their conservation to counter such developments.

The New South Wales National Parks and Wildlife Service and the Centre of Remote Sensing and GIS at the University of New South Wales are mapping the wetlands of the Murray-Darling Basin to produce a GIS. The number of wetlands has variously been estimated to be 11 000 or 30 000 and perhaps as many as 100 000. The main aim of this project is to establish where the wetlands are and how big they are. There is no other way, besides with GIS and remote sensing, of providing such a large spatial layer of data. Ultimately, the GIS will be able to provide managers with the spatial information which will assist them to better conserve wetlands in the Murray-Darling Basin.

Conclusion

The greatest challenge for wetland conservation is to provide spatial data on the extent of the resource over a large scale. This should provide decision makers and conservationists with the least biased view of the wetland world so that the consequences of development decisions are better identified. GIS at a large scale can provide the tools for effective wetland conservation. We will know where the wetlands are and how big they are. We might even know some unique characteristics of those wetlands which can be used to identify their conservation importance.

Once the information has been obtained the next job is to provide it to the decision makers. Often these managers may be local councillors or council staff. With an ability to overlay data layers it is possible to derive information relating to conservation in a form that they can use. The information may be available on CD-Rom or through the Internet. The aim is to give wetland managers and local communities a sense of ownership through information. There should be more incentive for wetland conservation if political decision makers and the community can identify the importance of 'their' wetland and derive conservation strategies at the local level.

References

Allan JD & Flecker AS 1993. Biodiversity conservation in running waters. *Bioscience* 43, 32–43.

ANCA 1996. *A directory of important wetlands in Australia.* 2nd edn, Australian Nature Conservation Agency, Canberra.

Bildstein KL, Bancroft GT, Dugan PJ, Gordon DH, Erwin RM, Nol E, Payne LX & Senner SE 1991. Approaches to the conservation of coastal wetlands in the western hemisphere. *Wilson Bulletin* 103, 218–254.

Crabb P 1995. Managing water resources: Seeking unity in the interests of diversity. In *Conserving biodiversity: Threats and solutions*, eds RA Bradstock, TD Auld, DA Keith, RT Kingsford, D Lunney & DP Sivertsen. Surrey Beatty & Sons, Sydney, 162–173.

Kingsford RT 1995. Occurrence of high concentrations of waterbirds in arid Australia. *Journal of Arid Environment* 29, 421–425.

Kingsford RT & Thomas RF 1995. The Macquarie Marshes in arid Australia and its waterbirds: A 50 year history of decline. *Environmental Management* 19, 867–878.

MDBMC (Murray-Darling Basin Ministerial Council) 1995. An audit of water use in the Murray-Darling Basin. Murray-Darling Basin Ministerial Council.

Usback S & James R (eds) 1993. *A directory of important wetlands in Australia.* Australian Nature Conservation Agency, Canberra.

Wilson EO 1992. *The diversity of life.* Penguin Books, London.

Environmental information through the Internet: Using the Internet to disseminate and manage information

Ann Bull[†]

Abstract

The Australian Environmental Resources Information Network (ERIN) was established to provide information for environmental planning. It makes use of the Internet and databases can be stored at appropriate nodes rather than centrally. Databases can be queried through on-line service. Linkages to further agencies and databases are being developed. Directories and metadata are also available and can be searched. There is an emphasis on offering easy to use facilities for obtaining and disseminating information.

Introduction

The Australian Environmental Resources Information Network (ERIN) was established in 1989 with the goal to provide environmental information as required for planning and decision making. Using the latest in computing technology, ERIN is setting up a national facility to allow ready access by government agencies to key information on the Australian environment. This effort is based on cooperation with other agencies interested in environmental information and effective environmental decision-making. The information collected and compiled by ERIN is available through a computer network so that it is easily available at the point where it is required. Attachment 1 provides information on contacting ERIN.

On this basis ERIN has established a number of databases. These were designed to assist with answering some key questions on the Australian environment. These questions include:

- what environment resources are found in a particular region?
- where are particular environmental resources (such as management zones, rare species or specific habitats) found?
- what kinds of environments exist and where are they located?
- how are these environments being managed?
- is the environment changing, and by how much?

The Internet

ERIN is using the Internet and, in particular, the World Wide Web to make information available to the Federal environment portfolio and where possible also for the general public. The Internet is the largest and best known public information network. The number of users on the Internet is still growing, and is expected to continue to grow rapidly. Many different groups including governments, researchers, universities, commercial groups and community groups are now using the Internet to obtain and disseminate information. The potential of this network is enormous and is, seemingly, being continually extended.

[†] Environmental Resources Information Network, Department of Environment, Sport and Territories, Canberra

The Internet offers a number of services that are widely used, including:

- electronic mail (email)
- file transfer protocol (ftp)— for transferring files between computers on the network
- telnet—allowing users to log on to another computer on the network
- gopher—a text based information service
- Wide Area Information Servers (WAIS)— a means for indexing and finding information
- World Wide Web (WWW)—uses hypertext technology to retrieve and display information—has an easy to use 'point and click' interface.

Use of computing technologies such as Internet and the World Wide Web means that data need not be stored centrally. Instead data can be stored at a network node where they are under the control of individual custodians. This ensures the data will be updated and otherwise maintained by those best able to do so and not be lost within an enormous and anonymous central store. All data, regardless of type, should be accessible through an easy-to-use interface, which incorporates a comprehensive directory facility.

The World Wide Web, gopher and WAIS are bringing about an information revolution. They are opening up access to an ever burgeoning amount of information. Never before has it been so easy to access information or to make it available. However, this is not without problems. Already it is obvious that we are 'swamped' with information. We are overloaded and we need to have intuitive means to navigate our way through it.

The Environmental On-line Service (EOS) was developed to allow users to easily access environmental information. This information is in many forms, including maps, reports and documents, data files, databases, satellite imagery, images, bibliographies and video. ERIN is aiming to make this information readily accessible by organising it into themes and maintaining search and navigation facilities.

Each of the themes contains a store of information produced by the Department of the Environment, Sport and Territories and associated agencies, such as the Australian Nature Conservation Agency, the Bureau of Meteorology and the Australian Heritage Commission. There are also links to external organisations such as other government departments, universities, CSIRO and other groups around the world dealing in environmental information. Prominent examples include the World Conservation Monitoring Centre and the United Nations. Information includes policy documents, legislation relating to the environment, photographs, satellite imagery, videos and database access.

Presently, there is the facility to query the ERIN databases, which include specimen and survey records, as well as data pertaining to managed areas such as national parks and nature conservation reserves, and also to metadata about datasets held by ERIN and associated agencies. There are plans in the near future to have a link to the Register of the National Estate Database at the Australian Heritage Commission. The potential for further links is enormous if resources are made available and benefits are clearly identified.

The need for directories

Effective management of geographic information involves knowing what data resources are available, the uses to which the data can be applied, and the data's quality and physical location. It also involves providing simple means to find and access those datasets across a network. We need to be able to interrogate the database in order to perform queries such as

'list the datasets about this topic, satisfying the specified criteria, that are relevant to this geographic region'.

The following issues should be considered in order to implement successful data directories:

- Directories relevant to particular sites or agencies should be stored at that site, or where they can be easily updated by the relevant people.
- It should be easy to document datasets.
- Environmental information is a valuable resource. It is often expensive and difficult to generate. Before new work is carried out a search should be able to be conducted to find out what information already exists so that it is not duplicated.
- Directories should contain enough information to allow the user to determine quickly whether the dataset is suitable for their purpose, and then be able to go directly to more detailed information if necessary.

The WWW, WAIS and metadata

Information which describes a dataset is known as 'metadata'. It is not the data themselves, but rather 'data about the data'. It is analogous to library catalogues which describe books yet are not the books themselves. The aim is that the prospective user should be able to find out about the dataset without needing to access and investigate the actual data.

Metadata has two main functions:

- to provide a means to discover that the dataset exists and how it might be obtained or accessed
- to document the content, quality and features of a dataset and so give an indication of its fitness for use.

The problem with documenting datasets is that it can be an onerous and less than enjoyable task. This ensures that many datasets remain undocumented at the end of a project, in many cases rendering the data unuseable for any other purpose. In some cases the data are effectively lost. In order to develop an approach to metadata that is successful, the task of completing and maintaining metadata must be made easier. One approach is to keep a minimal set of core fields which would enable most datasets to be documented quickly and easily.

A combination of use of the WWW and WAIS searching can then be used to maintain a directory of available datasets which can be searched, and contain pointers to more detailed information about the datasets. WAIS (Wide Area Information Servers) implements a searchable front end to collections of information held at distributed sites by indexing each word in every readable document. In this way WAIS can be used to search text files of metadata. Each custodian site on the network can prepare descriptions of the datasets that they wish to make available and build a WAIS index to them. ERIN is implementing a directory of spatial datasets to serve the Department of the Environment and its agencies. Each agency on the network then has access to all metadata held at the other agencies and can search this metadata. The use of WWW allows further links to be embedded in the metadata where more detailed information is available for that dataset.

The metadata can be stored as text documents or in a database. If the metadata is within a database, it can be easier to manage and keep up to date. For example, the ERIN metadata is stored within an Oracle database. Scripts automatically generate a description for each

dataset as a text file of the appropriate fields which is then indexed by WAIS. This helps ensure that the description contains the latest and most up to date information. Custodians who do not have database facilities, however, can also store metadata in text files. Whichever way the metadata is stored, the text file can then link to other documents, files or images which further describe the data, or even contain a link which downloads a copy of the dataset onto the user's computer. For example, an ARC/INFO GIS export file can be obtained in this way. With respect to what is stored in terms of metadata, ERIN has endeavoured to follow the US 'Content Standards for Digital Geospatial Metadata'. Australia is currently adapting the 'Spatial Data Transfer Standard'(SDTS) for local use.

ERIN has also implemented a map based query to enable users to find datasets relating to a particular geographic region. The addition of a map interface allows the query to be restricted to a specific geographic region. The user can choose to query by a state, some other region, or a minimum bounding rectangle. Metadata can then be obtained for datasets that are contained wholly within the region, datasets which encompass the region, or both. The user can also choose a theme to query, for example 'marine', or query all themes. The query retrieves a list of datasets held at agencies with links to metadata and more detailed information.

GILS—Government Information Locator Service

GILS is a program being implemented in the United States to identify and describe information resources throughout the US Federal Government, and provide assistance in obtaining the information. GILS supplements other government and commercial information dissemination mechanisms, and uses international standards for information search and retrieval so that information can be retrieved in a variety of ways.

GILS aims to identify the public information resources, describe these resources, and assist people to obtain the information. GILS will thus operate like a metadata directory for the whole US Government. GILS also allows users to find information worldwide. There are two levels of searching. The first search allows a user to find the relevant agencies that hold the information of interest. The second level then allows the specific information to be discovered at these agencies.

Although GILS will encompass a huge range of information sources, US Government resources will all be identified in a common manner using the 'GILS Core Elements', ie core metadata elements. The set of locator records will be accessible on public networks free of charge. In addition, GILS contents may also be established in other ways such as CD-ROM, floppy disks, electronic mail, bulletin boards and printed material. There is also a G7 proposal that the scope of GILS be extended to become a Global Information Locator Service.

Conclusions

Internet and the WWW offer facilities for obtaining and disseminating information. In addition, the indexing capabilities of WAIS and other searching facilities make it easier to obtain information over the network and search for datasets on a particular theme or pertaining to a particular area.

If sites around the world that deal with environmental information can use standard names for certain metadata fields then queries on many data holdings at many sites, for example the 'NOAA Dataset Catalog', the 'USGS Spatial Data Discovery System', and the 'ERIN Distributed Spatial Data Library' could be carried out concurrently. This would greatly facilitate the user's ability to find the information required.

Attachment 1 Contacting ERIN via the World Wide Web

ERIN home page: http://www.erin.gov.au

ERIN Distributed Spatial Data Library: http://www.erin.gov.au/dsdl/dsdl.html

ERIN Spatial Interface: http://www.erin.gov.au/cgi-bin/spatial_interface

GILS home page: http://www.usgs.gov/public/gils/gilscopy.html

Appendix 1 Workshop participants†

Name	fax	phone	email	Address
Bailey, Bruce			bjbailey@ozemail.com.au	PO Box 576 Port Vila, Vanuatu
Bayliss, Ben	08 89792076	08 89799793	benb@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886
Blackburn, John	02 99262997	02 99262800	johnb@genasys.com.au	Genasys II Pty Ltd, Level 13, 33 Berry St, North Sydney, NSW 2060
Boyden, James	08 89792076	08 89799708	jamesb@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886
Bull, Ann	08 89792149	08 89711751	annb@eriss.erin.gov.au	**eriss**, Locked bag 2, Jabiru NT, 0886
Cawsey, Margaret	06 2413343	06 2421628	Margaret.Cawsey@dwe.csiro.au	CSIRO Division of Wildlife Ecology, Forests Dynamics Project, PO Box 86, Lyneham ACT 2602
Devonport, Chris	08 89466712	08 89467138	Chris.Devonport@ntu.edu.au	Faculty of Science, Northern Territory University, NT 0909
Finlayson, Max	08 89792149	08 89799756	maxf@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886
House, Tony	08 89792076	08 89799724	tonyh@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886
Johnston, Arthur	08 89792499	08 89799700	arthurj@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886
Keith, Gordon			gordon_kei@antdiv.gov.au	Australian Antarctic Division, Channel Hwy, Kingston, Tas 7050
Kingsford, Richard	02 5856595	02 5856444 (ext: 88)	rkingsfo@nsw.erin.gov.au	NSW National Parks & Wildlife Service, PO Box 1967, Hurstville, NSW 2200
Kitchin, Margaret			mkitchin@henric.nric.gov.au	University of New England, Armidale NSW
Le Gras, Chris	08 89792076	08 89799770	chrisl@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886
Mount, Joan	08 89792076	08 89799749	joanm@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886
Mount, Tony	08 89792076	08 89799723	tonym@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886
Shelley, Paul	06 2724687	06 2724688	pauls@henric.nric.gov.au	National Resource Information Centre (NRIC), PO Box E11, Kingston ACT 2604
Skidmore, Andrew	02 3137878	02 3854400	a.skidmore@unsw.edu.au	University of NSW, School of Geography, Sydney 2052
Waggitt, Peter	08 89814316	08 89814230	peterw@oss.erin.gov.au	***OSS***, GPO Box 461, Darwin NT 0801
Walden, Dave	08 89792076	08 89799711	davew@eriss.erin.gov.au	***eriss***, Locked bag 2, Jabiru NT, 0886

† addresses at time of going to print

Appendix 2 Workshop program

Framework for developing a metadatabase at *eriss*

Time	Topic	Speaker
0900–0910	***eriss*** data management	C Max Finlayson (***eriss***)
0910–0935	Information navigation architecture: The metadata network	John Blackburn (Genasys II)
0935–0950	Metadata: Some national and international perspectives	E Paul Shelley (NRIC)
0950–1005	***eriss*** metadatabase development: A starting point	Tony House & Gordon Keith (***eriss***)
1005–1030	Break	
1030–1245	Discussion and conclusions	

Data management systems for environmental protection

Time	Topic	Speaker
1400–1415	Relational databases for environmental and biological data: Combining existing datasets—Points and pitfalls	E Margaret Cawsey (CSIRO)
1415–1430	Developing decision support systems: Issues and considerations	Bruce Bailey (CALM)
1430–1445	The use of Geographic Information Systems for wetland conservation	Richard Kingsford (NPWS-NSW)
1445–1515	Environmental information and the Internet: Using the internet to disseminate and manage information	Ann Bull (ERIN)
1515–1545	Break	
1545–1745	Discussion and conclusions	

Appendix 3 *eriss* information system: Opportunities for development

A report prepared by Chris Devonport, Genasys Pty Ltd, May 1995[†]

1 Synopsis

1.1 Background

Following discussions between Arthur Johnston and Max Finlayson (***eriss***) and Chris Devonport (Genasys) in January 1995 a consultancy titled 'Assessment of a GIS capability at ***eriss*** and **oss**' was arranged. The terms of reference were wide-ranging and included investigating Geographic Information Systems (GIS), remote sensing, database management, links with external organisations as well as hardware, software and staffing issues relating to GIS within ***eriss***.

1.2 Interviews

On 22 February a number of ***eriss*** and ***oss*** staff were interviewed to ascertain the then existing information system environment and establish users' needs for expanding it to encompass GIS and remote sensing technology. The interview process and outcomes were summarised and circulated for comment prior to the workshop below.

The outcome of the interviews was that ***eriss*** staff felt that attention to the existing system would not only provide immediate benefits but also a more robust platform on which to build new technologies such as GIS and image processing. A number of desirable attributes for the ***eriss*** information system were identified at this time. These outcomes were discussed with Arthur Johnston and Max Finlayson and it was agreed that the consultancy focus on the ***eriss*** information system in more general terms than originally planned and place less emphasis on specific GIS and image processing requirements.

1.3 Workshop

A number of key concepts were identified as critical to the effectiveness of the ***eriss*** information system and a workshop was conducted on the 19 April to communicate the outcomes of the interviews and discuss these key concepts with ***eriss*** staff. The objectives of the workshop were to:

- raise awareness about the potential for improvement in the information system;
- determine which of these improvements were appropriate and achievable;
- prioritise their implementation;
- put in place a plan of action to achieve the first goal as soon as possible.

The workshop identified the establishment of a metadatabase as the top priority and a metadata working group was formed. A plan (including tasks, responsibilities, and timelines) to prototype and prepare a report on the implementation of an ***eriss*** meta database for presentation to a panel of experts in July was set out (see ***eriss*** internal minutes of the first meeting of the metadata group). Longer term goals identified included the establishment of a corporate database and subsequently the development of a decision support system.

[†] Genasys II Pty Ltd, Darwin, GPO Box 4011, Darwin NT 0801, tel (08) 89810144, fax (08) 89411533

1.4 Recommendations

This report concludes with a number of recommendations relating to the ***eriss*** information system. These take into account feedback from ***eriss*** staff at interviews and the workshop as well as discussions with individual staff members. They provide a starting point for the development of an ***eriss*** policy on information resource management and suggest more specifically what steps and resources are likely to be required in the future to achieve the goals outlined. It should be noted that this report addresses the needs of ***eriss*** at a conceptual level rather than a technical level.

2 *eriss* information system

2.1 Desirable qualities

One of the outcomes of the initial interviews conducted was a list of attributes that staff at ***eriss*** felt were important to the effectiveness of the existing information system. (The point was also made that the following list applies to data in both digital format and other information resources such as photographs or field notebooks.)

Information should be:

findable: a means of finding out what data there are, where they are, and how they can be accessed is required

available: data of common interest to different groups within ***eriss*** should be made more easily available across the organisation

accessible: the hardware and network tools required to store and access data need to be readily available to users

useable: if information resources are to be effectively exploited the data content, quality, and extent in space and time need to be catalogued

maintained: datasets should each be owned and maintained by a specific custodian who is responsible for keeping the data relevant and up to date

flexible: the information system must be able to accommodate different types of data (eg spatial, text, image, video)

easy to use: software applications and data exchange utilities should provide users with an interface appropriate to their task and level of training

protected: as data are made more available attention needs to be given to protecting the data from unauthorised users and from malicious or inadvertent destruction (including backup and archiving processes)

2.2 Challenges

Although each of the items in the above list may appear self-evident it is not a trivial task to achieve a system that can provide for each of these attributes at a desirable level overall. It was generally agreed that there was room for improvement under most, if not all of these headings, although it was acknowledged that the task at ***eriss*** is made more difficult by a number of factors including:

- differing interests of research, assessment, monitoring, administration and management in the use of information resources;
- the relatively small size of ***eriss*** which limits the number of specialists who can be resourced and gainfully employed outside of the core areas of interest;

- the dynamic nature of the focus of ***eriss*** responsibilities means that any system put in place must be flexible enough to cope with significant changes in information requirements.

2.3 Future direction

With these factors and the ultimate goal of an ***eriss*** integrated decision support system in mind, it was agreed that the path to achieve the desired outcome was through the following steps:

- the formulation of a policy for the management of information resources
- the establishment of a metadatabase
- the establishment of a corporate database
- the integration of GIS and remote sensing into the information system
- use of the information system as a decision support system

These steps are sequential in the sense that each step is a prerequisite for the success of the next and following steps. However, this does not necessarily mean that they cannot be pursued in parallel to achieve the end result more quickly. In addition, some feedback mechanisms will need to be put in place so that, for example, the metadatabase sensibly reflects contents of the corporate database which were not anticipated when the metadatabase was first established. The first three steps are expanded on below in more detail.

3 Information resource policy

3.1 Needs

The formulation and maintenance of an information resource policy within ***eriss*** is central to the on-going success of any actions which may be implemented in the short term. Information and people are the key resources which distinguish ***eriss*** from other organisations with overlapping roles in Australia (and particularly the Alligator Rivers Region) and information management should, therefore, be constantly under review by management. Policy and goals are a prerequisite to evaluation of changes made to the ***eriss*** information system.

An information system strategy should identify the core information needs of users and the types of service they require, as well as considering the resources available. In this context it is important to understand the components of the information system and how their interaction affects both costs and risks to the organisation.

3.2 Capability

The capability of any information system (leaving aside the issue of data) can be represented as a point in figure 1.

The four main components are hardware, software, people and the nature of the tasks to be accomplished. These can each be viewed as an axis representing minimum to maximum capability or complexity. The ***eriss*** information system is in turn made up of many different components each of which can be mapped on the graph (fig 1). For example, the retrieval of a word-processed report by a secretary using a PC is a simple task and would be mapped in the lower left quarter of the graph. In contrast, the modelling of sediment dispersal over the Magela Creek flood plain by a research scientist using purpose-written software running on a UNIX workstation would be mapped in the upper right quarter.

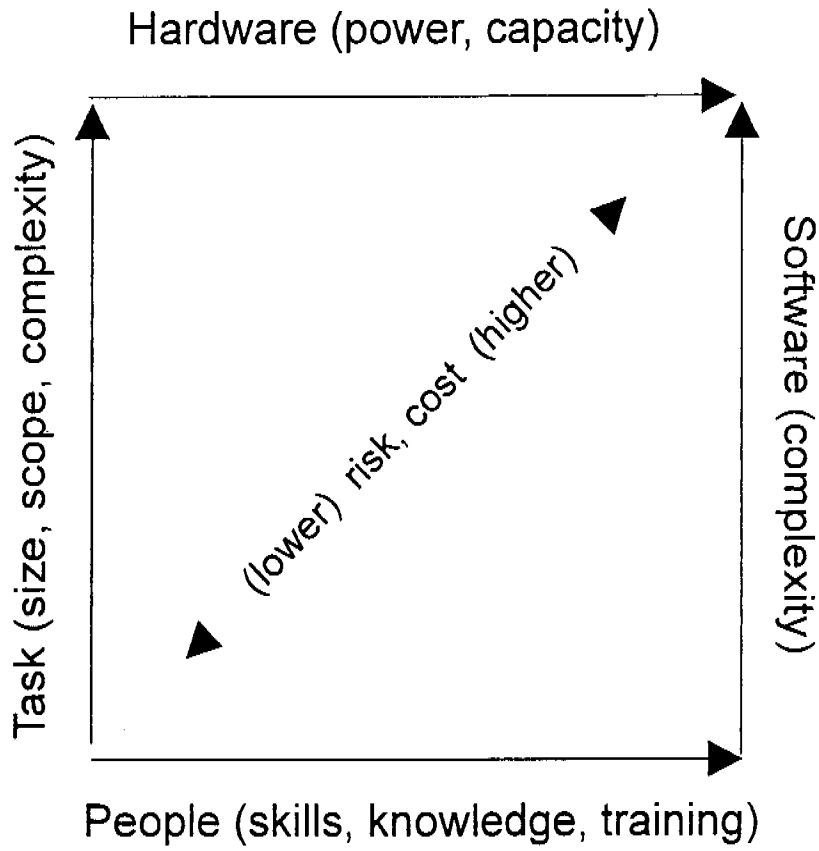


Figure 1 Information system capability model

Risk (eg losing skilled personnel, inappropriate hardware/software acquisition) and cost increase as capability on one or more axes is increased. Clearly, lower cost and associated risks are desirable and this approach has already been successfully adopted in the development of aspects of the ***eriss*** information system. It is recommended that this approach continue to be the central thrust of information system policy.

The question remains as to what to do where tasks require a capability in the upper range if benefits are to be realised. Although it may, on investigation, prove appropriate to accept the high cost/risk factors and provide the necessary resources within ***eriss***, it is recommended that strategies be developed to avoid this eventuality where possible. For example, strategies may include one or more of the following:

- the employment of external resources for the life of a specialised project;
- the employment of external using high end resources to develop a system which can then be applied using conventional resources within ***eriss***;
- collaborative arrangements with other scientific and/or academic organisations which may already have the needed resources;
- breaking down big, complex tasks into a number of smaller, simpler tasks.

3.3 Information system model

Current thinking in information technology which will support the achievement of the desirable attributes discussed above is a model which separates the information system into three layers each of which can be viewed independently of the other.

The user interface

The user interface provided for a particular purpose should be independent of the functionality and data layers to allow for the tailoring of the interface to suit the particular requirements of a task. For example, a technical assistant who is required to conduct routine analysis on specific datasets may be provided with a graphical user interface (GUI) which facilitates this task. In contrast, a research scientist may wish to make a more specific enquiry using the same data and functional tools. The GUI would not provide for his needs and adapting it for a single enquiry would not be practical so he would probably use a command line interface, ie different interfaces will often need to be provided for different purposes depending on the users' authority and skills.

Services layer

The services layer consists of the applications, development tools and utilities which are available to users of the system. They may be accessed through a GUI or at the command line as discussed above. The services should be viewed independently of the data layer as the same data may be accessed by several services. For example, digital elevation data may be used by one person using a GIS package to develop a drainage model and the same data accessed by someone using a contouring package and also by a geostatistics program to predict slope in a particular area.

Data layer

In addition to the above, separation of the data layer allows for more sophisticated data management technologies to be adapted. This is the heart of the system and its physical and logical storage and access needs to be controlled independently of the interfaces and applications available to users on the system.

This separation provides for the minimising of interdependencies between the user, the processes and the data. This approach will facilitate the adoption of client-server architecture and distributed data and processing, the implementation of modern data management strategies, and the incorporation of emerging hardware and software technologies which are addressing questions of interoperability (eg DCE, DOE, OLE, CORBA, and other standards). This is particularly relevant to the sharing of data between ***eriss***, ***oss*** sites and other organisations.

Viewing the information system this way also allows for the flexibility required to address information issues across the ***eriss*** range of interests. For example, while some applications may be adopted as standards and particular formats may be required for data of common interest, there remains room for specialist applications to be adopted within the organisation without detracting from the effectiveness of the information system as a whole.

4 A metadatabase

4.1 Attributes

Metadata provides information about data within an information system and should be easy to search on simple criteria such as key words, time ranges or spatial attributes.

It describes, *inter alia*, its:

- content (description of the data)
- spatial attributes (x, y, z coordinates in space)
- temporal attributes (time and date)
- data type (structured, unstructured)
- quality (accuracy, precision)
- physical location (place)
- media (disk, tape, CD)
- permissions (accessibility)
- data custodian (owner/maintainer)
- source (including contact details)

4.2 Types of data

eriss has information in large number of formats both digital and non-digital. The formats can be broadly categorised as follows:

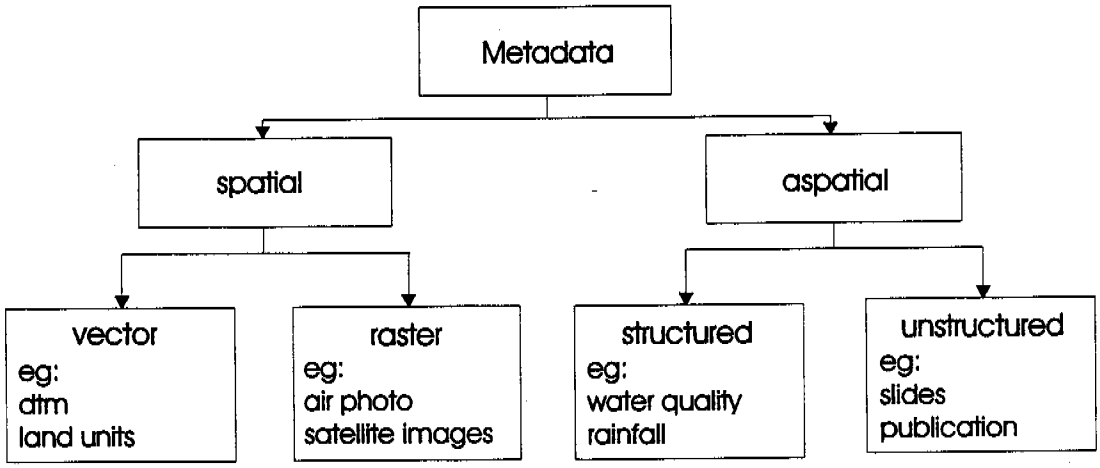


Figure 2 Types of data within ***eriss***

At present, although there are a number of separate databases and indexes held within different sections of the organisation, there is no way of making a single enquiry as to availability of data on a particular topic and/or for a particular region and/or for a particular time across the entire organisation.

4.3 Metadata working group

A comprehensive metadatabase is an essential prerequisite for the effective management and utilisation of information resources within ***eriss*** and was identified as the top priority for implementation during the workshop. A strategy to put a metadatabase in place as soon as possible was discussed and a detailed plan produced at the workshop by the metadata working group outlined tasks, allocated responsibilities and set time lines for the preparation of an implementation plan.

5.0 A corporate database

A discussion of the implementation of a corporate database is beyond the scope of this report. It will, to a certain extent, depend on the outcome of the metadatabase project which will provide the information necessary to decide whether a corporate database is needed and what form it should take.

A number of matters were raised during discussions with staff and these are listed briefly as they will become central issues when the development of a corporate database is discussed.

Corporate vs individual data

Information collected by ***eriss*** staff during the course of their employment belongs to ***eriss*** and should be stored, protected and maintained according to its importance to the organisation.

Data custodianship

The fact that data belong ultimately to ***eriss*** should not interfere with the allocation of custodial rights and duties for particular datasets to individuals—if the data are not 'owned' by an interested person they are likely to suffer from lack of maintenance and quickly become unreliable and not useful.

Access and security

A corporate database will highlight the need for control over access to data and security from intentional or unintentional damage—data held from other organisations will also need to be considered in this context.

Data management

The successful implementation of a corporate database will require the services of a data manager whose functions will include identifying data or corporate significance, allocating custodial responsibilities, controlling access to data, ensuring data integrity, determining data availability (eg online vs offline), and backup and archiving of data on and off site.

6 Recommendations

6.1 Policy on management of information resources

Planning, developing and implementing information systems with the ultimate goal of providing support for decision makers is of central importance to ***eriss***. This should be reflected by policies which recognise information as an important resource and put in place mechanisms for continual reassessment of information system objectives and accompanying needs in terms of hardware, software and staffing and provide for appropriate levels of resourcing.

It is important that senior management are involved in this process for the following reasons:

- information technology strategy is an integral part of major organisational objectives;
- implementation of information technology leads to change which must be managed;
- sectional interests within the organisation must be balanced;
- information is the lifeblood of ***eriss*** and it follows that it should be controlled by management.

An information strategy and policies on management of information resources should be given high priority and are a necessary prerequisite to the successful, on-going, development

and implementation of an information system which successfully integrates the resources listed below and leads ultimately to an effective decision support system.

6.2 Metadatabase

An ***eriss*** meta database working group was formed during the process of this consultancy. The implementation of a metadatabase was identified as the highest priority for the improvement of the present information system and a detailed plan has been formulated for the first stage of its achievement. Senior management need to continue to support this group and make available the resources that it needs to accomplish its task. The planned workshop to be held in July will provide the first major milestone for this project. The implementation phase will depend on the outcome of the workshop. An internal ***eriss*** meeting will be required immediately after the workshop to put in place the recommendations of the workshop.

6.3 Corporate database

The need for, and nature of, a corporate database in which data are managed at an organisational level will be largely dependent on the policy objectives and review of data discussed above. In this context it is important to separate the policies that relate to the information resources from those that relate to the technology that handles much of the information. The policies that relate to the information resources should be established by management in general and need to cover issues such as who has access to which classes of information, and who has the right to update, change and delete information. The technologists can then recommend the best means of implementing these policies.

The appointment of a data manager at ***eriss*** would provide expertise to assist management to put appropriate policies in place as well as skills to evaluate plans for development and supervise implementation of the information system and its various parts. As a senior manager, he or she would also carry the information resource banner in the formulation of organisational objectives and the allocation of resources.

6.4 GIS and remote sensing

Although short-term success in the implementation of GIS may be achieved for a relatively small project based on the expertise and political skills of an enterprising individual the long term integration of GIS and image processing capability into the ***eriss*** information system is unlikely to be successful unless:

- a robust general information system and good knowledge of existing data resources exists
- simple applications produce information which is fundamental to the work of potential users
- implementation is user-directed and involves their participation and commitment

Once the first of these is achieved, ***eriss*** should seek outside expertise to review the integration of GIS and remote sensing technology into the information system (the original terms of reference of this consultancy would provide a good starting point).

6.5 Decision support system

In the longer term, following implementation of a metadatabase and corporate database, the next logical step is to capitalise on the strategic value to ***eriss*** of the information available. This will involve a number of issues including the creation of a data warehouse (large database holding integrated data from operational databases) and applications which provide users with the tools they need to retrieve and analyses any data relevant to their needs without

needing to know where or how the data are stored (sometimes called expert systems). Although the baseline information and appropriate technology need to be put in place before an effective decision support system can be implemented, the information needs of ***eriss*** managers and scientists and the type of support they need should underlie the development and implementation of the system. In this context the importance of putting in place strategies and policies (outlined above) which will lead to useful outcomes in terms of decision support should not be underestimated.