1998

User response and organisational fit for information systems in Earth observation

Jeff Kingwell

Edith Cowan University

Recommended Citation

This Thesis is posted at Research Online.
https://ro.ecu.edu.au/theses/987
You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.

- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author’s moral rights contained in Part IX of the Copyright Act 1968 (Cth).

- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.
USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.
User response and organisational fit for information systems in Earth observation

by


Thesis submitted in fulfilment of the requirements for the award of

Master of Science (Information Science)

at the Faculty of Communications, Health and Science,
Edith Cowan University.

Date of submission: 20 November 1998
ABSTRACT

A group of seventy six scientists and data managers in the Australian research agency CSIRO were surveyed to establish their needs and preferences in relation to information systems for Earth observation data. After study of available alternatives, three prototype Earth observation information management systems were installed and the user response was evaluated through interview of fifteen of the group. The prototypes consisted of web-based client servers which permitted users to interrogate databases of Earth observation datasets; to search for information about sensor or satellite performance, and to retrieve data and information products. The chosen systems were CILS, the CEOS (Committee on Earth Observation Satellites) Information Location System; IDN, the CEOS International Directory Network; and IMS, NASA’s Information Management System of EOSDIS, the Earth Observing System Data and Information System. For this study, no special effort was taken to populate the system directories and inventories with local data holdings, and the prototypes were essentially mirror sites of operational data management systems used in other parts of the world.
While some of the interviewed scientists expressed enthusiasm for web-based spatial information management approaches, all indicated that improvements should be sought in the prototypes to make them more user-oriented, intuitive, and responsive. Most of the interview group were experienced remote sensing researchers who had developed their own contacts with overseas peers and data providers. Several in this category expressed the view that on-line data directories such as CILS and IDN would have limited use for them, unless the scientists changed discipline, application or geographic area of interest. On the other hand, several individual research projects or organisational units of CSIRO, as a result of these trials, were considering utilising one of more of the prototypes - particularly the IMS - to address their current unfulfilled requirements for data management. The study also found that while all fifteen of the interviewees felt they could benefit in some way from electronic information retrieval and spatial data management systems of the type assessed, it seemed unlikely that the target organisation would ever assign a sufficient priority to implement any of them in a systematic manner.
The biggest impediment to an organisation-wide approach to spatial data management for Earth observation was the low priority assigned to information management, because this activity was considered "supporting" or "non-core" in relation to the central objective of scientific research.

Results indicated that a piecemeal, decentralised or federated approach was the only means by which systems of this type could feasibly be introduced into the operating environment of CSIRO, in the absence of a major external forcing mechanism. This observation was compared to the evolution of EOSDIS, which had demonstrated a marked change from a centralised to a federated paradigm due to user preferences similar to those observed in the CSIRO case.

Key words:

CEOS; CILS; CSIRO; data management; distributed information systems; Earth observation; EOS; EOSDIS; federation; IDN; information science; IMS; organisations; remote sensing; space technology.
DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgement any material previously submitted for degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text; or

(iii) contain any defamatory material.

Signature.................................................................

Date 25 February 1999
ACKNOWLEDGEMENTS

Special thanks are due to current and former staff of the CSIRO Earth Observation Centre - Murray Wilson, Susan Campbell, and Dr Edward King - for the installation and maintenance of the NASA Information Management System, the CEOS International Directory Network, and the CEOS Information Location System.

Thanks are also due to the researchers - members of the "invisible college" of CSIRO Earth Observation - who gave their time to participate in surveys and interviews. I am also grateful to Linda Whitford and Paul Barrett (Librarian and former library technician, respectively, at CSIRO Mathematical and Information Sciences) and to Jonathan Potter, Manager, CSIRO Information Technology Services, for finding the time to be interviewed for this study.

The NASA IMS operating system was made available for this research through courtesy of Martha Maiden, Yoonsook Enloe and Lisa Shaffer (Goddard Space Flight Center, GSFC, and NASA Headquarters, respectively). The CILS operating system was made available through courtesy of Gunter Schreier and Dirk van Gulik of the German
aerospace Agency, DLR, and the Centre for Earth Observation of the European Community, respectively.

I appreciate the professional exchanges and discussions with colleagues from space agencies and research organisations of many countries, through the Committee on Earth Observation Satellites' successive Working Groups on Data; Information Network Services; and Information Systems and Services.

I am grateful for the support, during this research, of Dr David Jupp, Head of the CSIRO Office of Space Science and Applications (COSSA) and Science Leader of the CSIRO Earth Observation Centre; and Dr Brian Embleton, Executive Director of the Cooperative Research Centre for Satellite Systems and former Head of COSSA.

Finally, I am grateful for the advice and encouragement of my successive supervisors at Edith Cowan: Dr Gülten Wagner, Dr Ron Oliver, Dr Julie Johnson, and Dr Jan Ring; and to Dr Arshad Omarri, who along with Dr Wagner supplied valuable review comments on my research proposal.
TABLE OF CONTENTS

ABSTRACT .................................................................................................................. 2
DECLARATION ............................................................................................................. 5
ACKNOWLEDGEMENTS ............................................................................................... 6
TABLE OF CONTENTS ................................................................................................. 8
1. INTRODUCTION ....................................................................................................... 11
  1.1 PREFACE .............................................................................................................. 11
  1.2 ISSUES IN EARTH OBSERVATION DATA MANAGEMENT ...................................... 12
     1.2.1 Places of information in space projects ...................................................... 12
     1.2.2 A large information system for Earth observing ........................................ 17
     1.2.3 A long and winding road ......................................................................... 18
     1.2.4 Implications for other information services ............................................... 21
  1.3 RELATED ISSUES IN INFORMATION SCIENCE ................................................. 22
     1.3.1 User-centred systems ............................................................................. 22
     1.3.2 Data warehousing and distributed databases ............................................ 24
     1.3.3 Changing patterns of collaboration .......................................................... 25
  1.4 STRATEGIC INFORMATION MANAGEMENT .................................................. 26
  1.5 OPERATIONAL CONTEXT: EARTH OBSERVATION RESEARCH IN CSIRO ....... 27
  1.6 RESEARCH METHOD ......................................................................................... 29
  1.7 RECAPITULATION ............................................................................................ 30
2. LITERATURE REVIEW ............................................................................................. 32
  2.1 CHAPTER OVERVIEW ......................................................................................... 32
  2.2 DEVELOPMENT OF EARTH OBSERVATION FROM SPACE ............................ 34
  2.3 COMMERCIAL TRENDS IN EARTH OBSERVATION FROM SPACE .................. 39
     2.3.1 Policies ...................................................................................................... 39
     2.3.2 The Earth observation market ................................................................. 43
  2.4 INFORMATION MANAGEMENT IN EARTH OBSERVATION ............................ 47
     2.4.1 Overview .................................................................................................. 47
     2.4.2 Categories of Earth Observation Information Systems ............................. 55
     2.4.3 How are Earth observation data systems different? .................................. 57
  2.5 TYPICAL EARTH OBSERVATION DATA SYSTEMS .......................................... 60
     2.5.1 ESA (Europe) ............................................................................................ 60
     2.5.2 CEOS International Directory Network (IDN) .......................................... 62
        2.5.2.1 Metadata on the IDN ...................................................................... 66
     2.5.3 Australian National Spatial Data Infrastructure ......................................... 67
     2.5.4 Information Management System (USA) ................................................ 69
     2.5.5 Intelligent Satellite Information System (Germany) .................................. 71
     2.5.6 Global Earth Observation Information Network (Japan) ......................... 72
     2.5.7 GCNet (Canada) ...................................................................................... 74
     2.5.8 CEOS Information Location System (CILS) ........................................... 75
  2.6 TRENDS IN CATALOGUE SERVICES ................................................................ 77
  2.7 WIDER ISSUES IN SCIENTIFIC DATABASES .................................................... 80
     2.7.1 Database copyright ................................................................................. 80
1. INTRODUCTION

Digital information services in Earth observation

1.1 Preface

This research was undertaken to test service models and user needs relative to a class of scientific information derived from satellites observing the Earth, and the particular working environment of a group of Australian scientists employed by CSIRO and sharing a need for these data.

The main aim of the work is to understand whether efficiency in this working environment can be improved by adopting special information management systems devised to manage Earth observation data. To answer this question, I look at the nature of the technology itself, but more importantly, I study the nature of the organisation and the role information plays in helping CSIRO to meet its objectives.

In the process, I examine in detail the evolution of a large system designed to manage Earth observation data, exploring what this
evolution reveals about the changing nature of space enterprise and of information science.

1.2 Issues in Earth observation data management

1.2.1 Place of information in space projects

A verse from the field of meteorology goes:

"More data, more data
From pole to equator
Measuring everything, everywhere, all of the time"

This simple rhyme encapsulates both the power and the dilemma of satellite measurements of Earthly phenomena. The technology permits continuous observation of a wide range of parameters useful in nearly all activities which demand geographically-referenced information. However, the volume, diversity, and cost of space-based observation of Earth create unique problems for those who wish to organise or use these data and the information products derived from them.
During the “heroic” early period of space exploration (1957 to around 1987), only a small proportion of mission funds was typically devoted to data management. In consequence, Earth observation (EO) image data have been difficult to access; products derived from them have in many cases been difficult to merge with other data; and users have had to contend with a variety of idiosyncratic data formats and unpredictable data quality (Jupp, 1997; Graetz, 1996a; Office of Technology Assessment, 1993; Sarrat et al., 1995). This led, especially in the USA and Europe, to accusations by scientific bodies and budgetary authorities that the space agencies’ information management practices were wasteful and inefficient (Hasselmann, 1992; Office of Technology Assessment, 1994). In modern times, in what may be termed the “pragmatic” period of space activities (around 1987 to present), much more attention is being given to the archiving, access and application of information derived from space missions. In many cases, this new emphasis reflects a global antipathy towards high levels of public spending, in turn prompting space agencies to demonstrate tangible benefits arising from their programs (ESCAP, 1997; Austin, Macauley, Simpson & Toman, 1997; Rogers, 1998). In these cost-conscious times, voters’ interests are seen to be excited less by the idea
of travelling in space than in what we learn or otherwise gain from
doing so.

In examining the balance of resources allocated within space projects, it
is useful to distinguish between the physical assets launched (the "space
segment") and the supporting infrastructure, particularly the
information management component, back on Earth (the "ground
segment"). The emerging consensus, especially among end-users of
space-based services, is that the less exotic ground segment has in the
past been seriously neglected in comparison to the space hardware
component (Becker et al., 1993; Office of Technology Assessment,

In an important new analysis, the space policy analyst Joanne
Gabrynowicz (Gabrynowicz, 1997) examines the ground segment of
space projects in terms of a "data-centred" or "bottom-up" approach,
and a "satellite-centred" approach, in which the emphasis is the
engineering and management of the spacecraft. In the "satellite-
centred" approach, data exploitation is at best a necessary evil required
to justify further expenditure on space projects. Gabrynowicz argues
that this begrudging attitude has held sway in the past, and has limited
the extent to which new users (and potential supporters of future space
projects) have been able to experience the benefits promised when
space missions were proposed for funding.

One index of tangible return is the number and loyalty of users who
apply information returned from space missions. In the field of remote
sensing of the Earth's environment an important factor for increasing
the number and loyalty of users is improved access to data and data
products, via digital information management systems (Office of
Technology Assessment, 1994; Vetter et al., 1995).

These systems - comprising in their fullest form software, hardware,
communication infrastructure, databases, standards and procedures -
share many attributes of traditional libraries, although this connection
is seldom recognised. In particular, two elements of traditional library
management - (1) shared access by a varied group of clients to a
common pool of information, and (2) the standardisation of methods
of classification and documentation - are increasingly being
implemented within on-line information systems designed to deliver
electronic information derived from Earth observing satellites (CEOS, 1997, p.22-23).

I survey the literature on these information systems in Chapter 2. From the available field I selected three systems for further study, basing this selection on user needs within CSIRO; availability; cost; technical feasibility; and functionality. I then arranged for these three systems to be made available to CSIRO researchers via network access to a server. In this study, I refer to these test systems as "prototypes", even though all three are in operational use in other parts of the world. This term is employed for three main reasons:

(1) I wished to emphasise that the selection of information systems for this specific work environment was provisional, and subject to user feedback;

(2) The information databases contained in these systems include few data sets relating to Australia (they have yet to be "populated" with data more likely to be useful to Australian researchers); and

(3) Because the systems have been installed for assessment, only a basic support service (below the minimum that would be required for operational support) was provided during the study period.
1.2.2 A large information system for Earth observing

The most ambitious and by far the largest Earth observation data management systems is the Earth Observing System Data and Information System (EOSDIS) of the United States’ National Aeronautics and Space Administration (NASA). The objective of EOSDIS is to manage information obtained from a collection of Earth observing satellites, on behalf of a group of users ("Principal Investigators") selected by NASA. EOSDIS was conceived as an integral component of an even larger program, the ambitious “Earth Observing System”, or EOS. This program, originally costed at US$17 billion, was proposed in the late 1980s at perhaps the high water mark of international concern about the Greenhouse effect and ozone depletion. The centrepiece of EOS is a suite of complex, purpose-built satellites designed to study natural and human forces as they are manifest in changes to the Earth’s total environment (Asrar & Ramapriyan, 1995; Vetter et al., 1995; NASA, 1998a). In its early days of conceptualisation, EOS was seen as:
"...necessary to develop a comprehensive understanding of the way the Earth functions as a natural system. This includes the interactions of the atmosphere, oceans, cryosphere, biosphere, and solid Earth, particularly as they are manifested in the flow of energy through the Earth system, the cycling of water and biogeochemicals, and the recycling of the Earth's crust driven by the energy of the interior of the Earth".


1.2.3 A long and winding road

At the time of writing (October 1998), no EOS spacecraft is yet in orbit, although elements of EOSDIS are in use for handling data from earlier NASA missions (King, 1997). The decade-long germination time for EOS, combined with frequent changes in budget, policy, design philosophy, and technology, led to a dramatic transformation in EOSDIS. Its original architectural model was heavily centralised, with NASA supplying the initial data, the processing capacity, and the distribution network. Data and product distribution was initially planned to be carried out by NASA's eight data warehouses, known as
"Distributed Active Archive Centers", or DAACs. The current NASA distribution system for Earth science data is built upon the DAACs, the operations of which NASA funds. Not all the Centers are directly managed by NASA, but their operations are characterised by uniform product standards, distribution policies, priority setting and prices.

Even though the system requirements definition and the system design review procedures adopted by NASA employed a high degree of consultation with scientific users, some of the latter have, from about 1992, been highly critical of the design of ECS (the “EOSDIS Core System”). Some users, particularly from the academic community, strongly opposed the “centralised” approach initially adopted for EOSDIS. These critics regarded the system design proposed by the prime contractor, Hughes Applied Information Systems, as inefficient; unresponsive; and just too big, complex, prescriptive and expensive for their purposes. They argued that information systems of such a size would have too much inertia to be able to react in a timely way to technological, policy, market and scientific developments.
The U.S. National Research Council, reacting to these concerns about the perceived unsuitability for academic users of the centralised model of EOSDIS, critically reviewed the program in the mid 1990s. In doing so, the Academy proposed a different model for delivery of Earth observation information delivery, known as “Federation” (National Research Council, 1995a). Maiden (1996) describes Federation as a union of Earth Science Information Partners (ESIPs) with consumers of information services and preservers of information, all overlaid by protocols and standards.

In the Federation model, information providers have a large measure of autonomy, but negotiate the degree of commonality in interfaces and protocols across the Federation. The Federation model cedes development of high-order information products to a competitively-selected group of value-added service providers, both public and private, while circumscribing NASA’s role to space research, operating space assets (the Earth Observing System’s sensors and satellites), and supplying basic (or “raw”) data and information. The Federation model for EOSDIS implicitly supports the view that data users, rather than data suppliers, should hold sway in determining the market for
Earth observation data products (De Witt & Naughton, 1994). By its potential engagement of the growing private sector in Earth observation, Federation also responds to criticism that NASA ought to concentrate on its core business and leave value added services to the private sector (Space News, 1998a; Oler, 1998).

1.2.4 Implications for other information services

Although there are bound to be limitations in examining any phenomena in terms of a binary model, the Hegelian dialectic approach is often used in information science to reach a better understanding of the relationship between a “fact” or real-world observation, and the concept of which it is an example (Hirschheim, 1985).

To some extent, the dichotomy represented by the centralised and Federated models of EOSDIS is reflected in the dynamics observed within information services elsewhere. In addition, the clientele for EOSDIS closely resembles the target group in CSIRO, being geographically-dispersed, technologically sophisticated, and dependent upon access to Earth observation data. For these reasons, I have given
particular attention to analysing alternative centralised and Federated models of EOSDIS, in the context of proposing options for future Earth observation information systems within CSIRO. Further, because EOSDIS would appear to be one of the largest information systems ever planned, a detailed examination seems warranted in terms of understanding trends in the discipline of information science.

1.3 Related issues in information science

1.3.1 User-centred systems

In considering basic principles of information system design, Allen poses a similar dichotomy to that of Gabrynowicz (Allen, 1996; Gabrynowicz, 1997). He refers to “data-centric” and “user-centric” approaches in which information systems and services are implemented, respectively, either top down by engineers or experts familiar with the nuances of the information the system contains; or they are implemented consultatively, in order to empower a set of users to help solve a specific problem or achieve a specific set of outcomes.

The Federation model for EOSDIS, in Allen’s terms (ibid.), responds to the user-centric rather than the data-centric approach to
information. In the literature survey, I examine the Federation model, focussing on the advantages and disadvantages articulated from a more general perspective (Papazoglou, 1991; Thuraisingham, 1997).

In this research study, I explore whether principles established for generic information science - such as the user-centred design approach - have relevance to the policies, techniques and markets of Earth observation information management. One common technique used in information science to evaluate new systems is the user survey (Lipetz, 1980). In this study, I surveyed by questionnaire and interview a group of scientists within the Australian research organisation, CSIRO, to assess their response to the three prototype Earth observation information systems. In carrying out this survey, I investigate the appropriateness of the technology from the user perspective, and also explore which architectural model - Federated or centralised - would be most appropriate in the event that one or more of the prototypes systems was implemented for operational use in CSIRO. The results of this survey are given in Chapter 4.
1.3.2 Data warehousing and distributed databases

The information needs of companies which operate on a continental or global scale, the expansion of communication capacity, and the diversity of software and hardware systems in business use have all contributed to increasing interest in the interoperability of corporate databases distributed at a number of locations (Thuraisingham, 1997). In the literature survey (Chapter 2), I examine trends towards middleware such as CORBA (Common Object Request Broker Architecture) and the Catalogue Interoperability Protocol, designed to mediate user requests between heterogeneous and geographically separated databases. The purpose here is to better understand the impact these wider information science trends may have on data systems for Earth observation.

Similarly, the corporate finance sector has demonstrated strong interest in the concept of data warehousing (Inmon, 1992; Tanler, 1997), in which raw or lightly processed and seldom-used data are moved into background storage, leaving the “production environment” to focus more on value-added and volatile information. Through the literature survey in Chapter 2, I briefly explore the
potential relationship between the data warehouse and information systems for scientific data.

1.3.3 Changing patterns of collaboration

The process by which scientific ideas are developed and communicated has been studied by sociologists and information scientists for many years. Griffith and Mullins (1980), Rothwell (1980), and Crane (1980), for example, showed that informal communication between “invisible colleges” or self-selected peer groups (often at separate locations) was often a more enduring and effective route for sharing scientific information than formal lines of reporting based on organisational structure. More recent studies (Walsh & Bayma, 1996; National Research Council, 1993) appear to demonstrate that the efficiency of communication in the invisible college, and the potential size of the college, is being enhanced through greater use of web-based information services.

This issue is relevant for the CSIRO group I surveyed, which was both highly geographically-distributed and arranged in an administratively
complex manner. I explored the issue of informal communication though interview, the results of which appear in Chapters 4 and 5.

1.4 Strategic information management

It is clear that the choice of information system by an organisation is not influenced by technological considerations alone. A growing body of literature now explores “strategic information management”, the relationship between an organisation’s objectives and culture, on the one hand, and its information processes and priorities, on the other. I examined this issue within CSIRO, using a critical analysis of internal documentation and interviews with three information professionals, including the head of information technology services.

The analysis of this part of the research, dealing with the organisational environment, draws heavily upon recent management theory which suggests that it is futile to attempt a technological “fix” to an organisation’s information needs unless the technology is capable of servicing the user needs and conforms to the predominant culture or style of the organisation (D. Best, 1996; Webb, 1996; Orna, 1990, 1996).
1.5 Operational context: Earth observation research in CSIRO

CSIRO researchers have expressed a need (Simpson, Barton, Kingwell, Neal, & Wallace, 1995) for on-line Earth observation data, but many have also expressed reservations about “centralisation” of such services (Graetz, 1996a). Equally, system designers and operation managers are at a loss to take the first steps to establishing a coherent information system, in an environment where individual users and research laboratories seem unwilling to compromise local control of products and services, and their direct relationship with users.

Does the “Federation” model offer a way forward from this dilemma? I explored this through a questionnaire and small-scale interview of subjects from within a group of 76 researchers on the e-mail contact list of the CSIRO Earth Observation Centre. These researchers share an interest in the use of Earth observation information, but they are geographically dispersed and weakly coupled in organisational terms, being split into numerous research groups within sixteen different business groups, or Divisions.
The questionnaire and survey explored user response to the prototype information systems that were in operation over a period of approximately two years during the research program, during which time they were accessible to all the client group and to the public, via World Wide Web. Special emphasis was given in this action research to the relationship between technology and information services, on the one hand, and the culture and objectives of the host organisation, on the other.

One of the purposes of the interviews was to establish whether one or more of the test systems would be useful to the client group if implemented operationally to archive and manage data sourced from CSIRO. The significance of this aspect of the study lies in helping to determine whether further work in data management would be "profitable" for the Earth Observation Centre to pursue; and if so, to help guide the choice of specific data systems. By examining information use patterns and organisational culture, I also hope to determine whether a "centralised" or a "federated" mode of operational information system would be the most suitable in this particular work environment.
1.6 Research method

**Action research** is a method employed when a researcher emphasises a close interaction between practice and theory, becoming a protagonist in a research issue by participating with others in attempting to bring about change, often in the work place (Sandberg, 1985; Habermas, 1978). In this study I employ action research, with the cooperation of colleagues from CSIRO, in order to address the question “is it possible to improve Earth observation in CSIRO by utilising information systems already developed elsewhere?” A related question which I studied was “if users feel that one or more prototype information systems may suit their purposes, how should those systems be implemented operationally in CSIRO, given the nature of the users’ work and the role of information in the organisation?”

The research method is covered in detail in Chapter 3. Chapter 4 describes the results of the work, and conclusions are given in Chapter 5.

A Glossary of technical terms is contained in Appendix A.
1.7 Recapitulation
This study looks at a particular information environment and examines the suitability of prototype information systems introduced on a trial basis over a two-year period. Suitability is approached from a users' perspective, and is also explored in terms of organisational culture. The purpose of the research is to help improve information services for a group of CSIRO scientists whose work depends on the use of digital data obtained by satellites observing the Earth.

The information systems developed until recently for managing satellite image data have been highly "satellite-centric", in that they are custom made to suit the structure of a specific sequence of data from a specific satellite sensor. Recently, more systematic approaches have been adopted for managing multiple streams of data from different satellites and from airborne platforms. In looking at various of these systems in some detail, I also explore commonalities with the principles and methods used in better-established information service industries.

In one model for service delivery, "top down" system design has lead to a data warehouse concept, in which a single ("centralised")
management structure controls content and conditions of use. The example on which I have drawn is the original design of the NASA Earth Observing System's Data and Information System, EOSDIS.

A more recently articulated, "bottom up" and heterogeneous approach, "Federation", is predicated upon control of at least some portions of the information system being in the hands of one or more classes of users - usually intermediaries rather than end users - for example, academic researchers. The key issue for system stability and long-term client satisfaction is how standards can be maintained within such a loose management structure.

By critically reviewing changes to the design of EOSDIS over the past decade, I hope to better understand the relevance of information architecture and service concepts such Federation, in terms of the information environment of CSIRO. I also attempt to elucidate what the evolution of EOSDIS reveals about the changing nature of space enterprise and developments in information science.
2. LITERATURE REVIEW

Data systems for Earth observation and lessons from information science

2.1 Chapter overview

The focus of this research is whether an existing suite of information management systems is suitable for use by CSIRO scientists who frequently employ Earth observation data. I begin this chapter by briefly reviewing the development of Earth observing technology, and then move to a more detailed look at customised information management systems which have recently been developed for this application. In passing, I also examine the trend towards commercialisation within space industry generally, because this trend will affect developments within Earth observation and associated data services.

Next I examine trends in information science, particularly those which appear to have an influence on the specific area of Earth observation information management. In particular, I explore the concepts of data warehousing, database interoperability, and architectures for distributed data systems.
This study concerns a specific information technology within a particular work environment, the Australian research organisation CSIRO. My emphasis is not so much on "best technology" but on "fitness for purpose": that is, whether an information system is suitable for meeting the requirements of users operating within a given organisational culture.

To explore this aspect of the "fit" between information system and work place, I review literature on the role information plays in helping an organisation achieve its underlying mission. This analysis focuses on the concept of "strategic information management" in which one of the key steps in evaluating information technology is first to understand the organisation's information needs and patterns of use. I return to this issue in Chapter 4, where I explore the role of information in CSIRO in more detail.

In the present chapter, I review past analyses of CSIRO's requirements for Earth observation data.
Finally, I briefly examine literature on the research methods used later in the study.

2.2 Development of Earth observation from space

Satellites have been used to gather intelligence for national security and meteorological information for weather forecasting since around 1960 (Schnapf, Hallgren, Smith & Zbar, 1990). In many respects, these applications remain specialist areas with relatively homogeneous communities of sophisticated users. Satellite remote sensing as a generic tool for a variety of purposes in the area of environmental and resource assessment and monitoring is usually considered to have commenced in 1972 with the advent of the Landsat series of Earth observing satellites. In the 1990s, the term Earth observation came to be widely used in preference to remote sensing. The former describes a specific application while the latter is a generic technique for measuring physical attributes at a distance, for applications as diverse as industrial quality control to exploring the outer planets of the Solar System.

Pamela Mack, in her study "Viewing the Earth" (Mack, 1990), describes the evolution of the Landsat program, showing some of the difficulties
the program faced in applying technologies initially developed for
defence purposes to civilian use. Military considerations were the
primary driving force of the early Space Age and remain the rationale
for the majority of satellite missions on a global basis (Pike, 1992,
p.42).

The link between remote sensing and military surveillance from space
remains important because defence research and development is the
source of much of the technology which eventually appears in the civil
domain. In addition, the existence of military applications induces
governments to subsidise remote sensing capacity despite the low rate
of economic return on investment. Thus among other countries,
Canada, China, France, India, Israel, South Korea, Russia, Ukraine,
and United States all maintain independent remote sensing satellite
systems which produce markedly similar information products, the
market for which appears incapable of recouping the true cost of
supply. The current operating costs of the world's civilian remote
sensing satellites is estimated at over $US 1 million per day (CEOS,
1997, p.15).
Although economic difficulties in 1998 have slowed growth, in the last few years many countries in the Asian region (including Australia, Malaysia, Pakistan, Taiwan, and Thailand) have commenced or completed remote sensing satellite projects (ESCAP, 1997). In most cases, these satellites replicate data already available from European, North American, Indian and Japanese satellites.

It may be inferred from the emergence of competing regional satellite programs that other factors are playing a significant role in investment decisions by the respective governments. These factors include industry development objectives, defence needs, and national prestige (Mansell, Paltridge, & Hawkins, 1993).

The legal foundation for obtaining images of other countries from space derives from the proclamation by the United Nations of the "open sky" principle. This asserts that any country subjected to data gathering by Earth observing satellites may obtain information from the satellite operator on the same basis as any other customer - that is, in a "non-discriminatory" fashion (Perek, 1992).
There is an exception for surveillance by satellites termed "National Technical Means", a euphemism for spy satellites (Zimmerman, 1990). Data from "National Technical Means" are almost always distributed in a markedly discriminatory fashion.

The open sky concept, formally known as "Principles Relating to Remote Sensing of the Earth from Space", UN General Assembly Resolution 41/65 of 3 December, 1986, represented a compromise between the views of technically advanced Western countries and those of socialist and developing countries. The former argued that remote sensing data should be freely collected, and made available to interested parties by various means including commercial sale. Developing countries, in the political climate of the time, feared that technically advanced countries would gain unfair and potentially critical economic and strategic advantages, by obtaining more information on a country's resources than that country itself possessed.

Irrespective of the UN Principles, nations or public interest groups that do not possess satellites systems of their own may be
disadvantaged in many ways in respect of delivery of Earth observation data (Kingwell, 1988). Ways in which disadvantage may accrue to countries without Earth observing satellites include delayed access to data; unsuitability of data products; and in extreme circumstances, such as occurred to Iraq during the 1991 Gulf War, complete denial of access (McLean & Swankie, 1998).

Commercial interests are another key factor influencing the user environment. Monetarist policies emerged in several Western governments in the 1970s and spread even to former Soviet bloc countries by the late 1990s. These encouraged the "user pays" principle in relation to Earth observation data delivery. Funding models adopted to date have primarily involved public funding of the space assets and partial cost-recovery of the data processing and distribution (the "ground segment"). Cost recovery methods adopted include charges to data receiving facilities; royalty charges on users; application of copyright law to prohibit on-selling and copying; and the formation of supplier cartels (Gabrynowicz & Wood, 1991; Mansell et al., 1993).
Recently some governments have commenced outsourcing the supply of services: that is, contracting the private sector to provide the delivery of Earth observation images, or even to supply and operate the entire satellite system. This approach accounts at the moment for only a small proportion of data products (Space News, 1998a).

The issue of commercial development in Earth observation is examined more fully below.

2.3 Commercial trends in Earth observation from space

2.3.1 Policies
Mansell, Paltridge and Hawkins examined the economics of supplying Earth Observation data, questioning government subsidy of these data, especially when alternative information products were available at lower real cost (Mansell et al., 1993). They pointed to a long history of government intervention in the market: despite this decades-long support, less than one fifth of the space segment cost is recovered in revenue from data sales. They pointed out that product enhancement - value adding - was much more lucrative, but was vulnerable because these activities are acutely subject to product substitution.
"The issue of product substitution is an important factor in assessing the potential for entry in the EO industry. It is especially important in an environment where new data sources may be available at very low prices."

(Mansell et al., p.13).

Sometimes, governments adopt conflicting policies in relation to Earth observation industry development. On the one hand, policies are devised which encourage the private sector: these include preferential access to previously classified military technology; mandatory purchase of private Earth observation products by the government; subsidy of such purchases; and subsidy of satellite development or satellite operating costs. On the other hand, the same governments may stunt the growth of private sector Earth observation by over-regulating markets, or by funding competing services which supply large volumes of Earth observation data to users free, or at an artificially low price.

Some business analysts argue that despite difficulties such as the in-orbit failure of the commercial remote sensing satellite EarlyBird, the move to private-sector Earth observation has already benefited users, by encouraging innovative, entrepreneurial start-up companies.
prepared to take risks (Space News, 1998b). However, private companies will be discouraged from entering a market in which the largest customers are governments, when those same governments fund and operate competing Earth observation systems through which data products are virtually given away. This encourages costly systems in which governments design and pay inflated prices for their own remote sensing spacecraft, instead of simply purchasing data from competing providers. In this situation there is little incentive for the private sector to develop their own Earth observation spacecraft or data services (Space News, 1998a; Mansell et al., 1993; Oler, 1998).

The Reagan government privatised the operation of the pioneering Landsat satellites in 1984, arguing that this would encourage enterprise and reduce the burden on taxpayers. However critics of this move argued that the higher data unit prices under this regime led to under-use of space information, to illegal trade in data, and proved a hindrance in international global change studies (Gabrynowicz & Wood, 1991). Officials from NASA and the National Oceanic and Atmospheric Administration (NOAA) supported this view, claiming that if Earth observing spacecraft are paid for by the public on the
grounds of their utility for global change research or weather forecasting information, then the taxpayer should benefit from this investment by being able to obtain data at the incremental cost of extracting it from data archives (Williamson, 1997).

Landsat satellite management was re-absorbed into the public space program in 1992. Prices for images from Landsat 7, the first publicly-owned satellite in the series following the privatisation experiment, have been set at less than $US 600 per scene, or approximately an order of magnitude less than under the privatised regime (Williams, 1998).

Pricing space-based data, on public interest grounds, at a level which does not recover recurrent (operating) costs - let alone the capital cost of the information system - raises many interesting public policy questions. For example, could the service be supplied using other technologies (aerial photography or ground surveys) at lower cost? Would the public-good end purpose (global change research outcomes) be better served by funding the research, rather than a particular technology? In such a case, researchers may be able to competitively
choose from varied data sources, or even use the funds for types of research (modelling, for example) which did not require space-based data.

Recently NASA has been criticised for competing with the private sector in Earth observation markets, by preferring to develop Earth observation satellites under public ownership (Office of Technology Assessment, 1992). This criticism has led to small-scale contracting on the part of NASA to commercial data providers (NASA, 1998b).

The changing nature of the Earth observation market place is further examined below.

2.3.2 The Earth observation market
The market for unprocessed Earth observation data in 1994 was $US150 to 200 million per year: the market for value-added (processed and map-referenced) data was two to three times larger (Office of Technology Assessment, 1994, p.23).

The commercially-operated Spot remote sensing satellite series commenced February 1986, competing with the U.S. Landsat satellites
then operated by the private company Eosat. Data from the Spot satellite series are marketed by Spot Image of Toulouse, who are responsible for about 60% of the world’s sales of satellite imagery. Despite this market dominance, the company’s revenue barely covers operating expenses: the company estimates that it would need to double its annual revenue of about $US40 million to cover spacecraft capital costs (de Selding, 1998a).

Spot 5 is already being built, by Matra Marconi Space Systems in France, for launch in 2001. It will be slightly heavier than its predecessor (3000 kg instead of 2700), and is expected to cost about 3.6 billion Francs (about $US595 million) for the satellite, launch, and five years’ operations, about the same as Spot 4. The French Space Agency CNES spends about $US 115 to 165 million on Spot each year, while Swedish and Belgian organisations contribute a smaller amount. CNES is now reported to be unwilling to continue to subsidise loss-making and expensive remote sensing programs, and is radically altering its approach (de Selding, 1998b). CNES now plans to reduce by a factor of four the cost of post-Spot 5 Earth observation satellites (to be known as the 3S series). These new satellites are expected to be about 500 kg in
mass (reducing launch cost), and will operate at an altitude of 633 km instead of 832 km, making it slightly easier to achieve the same 2.5 m resolution. CNES officials are also reported to have adopted a new market strategy, by taking user needs as the baseline for satellite system design (ibid.).

Spot Image faces the additional difficulty that many of its major shareholders are space vehicle manufacturers. These shareholders face a conflict of interest, because they build remote sensing satellites for other companies (traditionally a profitable thing to do). By doing so, they create more image selling services which compete with Spot Image, resulting in even less profitable business. Spot Image shareholders Matra Marconi Space and Aerospatiale are both designing remote sensing satellites which they hope to sell to foreign buyers (de Selding, 1998b).

While Spot Image (like the U.S. company Eosat before it) requires ongoing government subsidies in order to pay for the space hardware, two other subsidiaries of CNES make a small profit: both Scot Conseil and GDTA supply value-added remote sensing services, an area which
other organisations have also found more profitable than data sales (ibid.).

Several new companies are developing commercial, medium to very high resolution satellite imagery services. These include Orbital Sciences Corporation with its Seastar ocean colour detector for tracing fisheries stock. WorldView Imaging Corporation plans to establish a global data service with a resolution of 3 m (Fritz, 1996). This unfortunately suffered a setback leading to dismissal of one third its staff when its first satellite failed following launch from Siberia on 24 December 1997. Space Imaging Inc. and Eyeglass International both expect to market 1 m resolution images. These companies will vertically integrate launch, satellite and service enterprises through their respective principals, Lockheed Missiles and Space Co., and Orbital Sciences Corp. The Indian Space Research Organisation, through partnership arrangements with Eosat; Antrix Corporation, and Euromap, currently holds the technological edge in the digital civil remote sensing field with their 6 m resolution and stereo capability on the IRS satellite series (Kramer, 1997).
The market focus of these services is on desk-top geospatial information, relying on the integration of highly processed and re-mapped image data with Geographic Information Systems capable of merging many layers of spatial data from a variety of sources (for example, census, cadastral, aerial, space).

We know turn from the Earth observing technologies to those used for managing the data and information products.

2.4 Information management in Earth observation

2.4.1 Overview

Earth observing satellites carry sensors which respond to light, heat, or radio signals radiated or reflected by the Earth. Data transmitted to the ground by these satellites thus contains elements of instrument performance; satellite, sensor and solar geometry in relation to the Earth frame of reference; the nature of the objects being observed; and the effects of atmosphere and other intervening material (Harrison & Jupp, 1989).

The raw or unprocessed satellite data are received directly from the satellite by ground receiving stations, or are relayed to receiving
stations via special communication satellites. These raw data are complex compound results arising from many interacting variables. A sequence of mathematical procedures is applied to separate the influence of the sensor, atmosphere and geometry, leaving behind the signal from the surface or object being observed. These mathematical treatments, which reduce sensor and satellite data into information, are known collectively as processing. The mathematical description of the processing is known as an algorithm. Understanding of what constitutes “good” processing changes over time. Users, particularly sophisticated ones such as researchers, therefore wish to preserve the option of applying improved algorithms to old data. To perform this function, it is necessary to preserve the original data in its original form, lest previous processing should irretrievably transform it (Simpson et al., 1995).

Satellite data are commonly categorised by level of processing, in order of increasing information content and decreasing volume, as follows:

- **Level 0** Satellite sensor readings
- **Level 1** Raw sensor readings, combined with auxiliary data needed to make these readings useful. Auxiliary data include information
about the satellite at the time the sensor reading was made; instrument calibration information; and quality control information. This is the minimum processing level which is useful to archive for future use.

- **Level 2** Information products showing geophysical parameters derived from the satellite measurements. This might be, for example, a set of figures showing sea surface temperatures as depicted by satellite infrared measurements at particular locations.

- **Level 3** Interpolated and analysed information products on standard global or regional map projections, showing geophysical parameters (for example, maps of sea surface temperature or ozone distribution). Often, products at this level combine satellite information with that derived from other sources (for example, temperature measurements from ships).

(Booth, 1994, p.178-9).

The major challenges in information management for Earth observation lie in acquiring raw data from the sensors on satellites (or on other platforms such as aircraft or research ships); storing these with sufficient reference information to be able to unambiguously link
the data to a specific time and geographical area; producing from these basic materials useful information products; and enabling users to locate data sources, select information products appropriate to their needs, order, and finally receive this material.

Earth observation data and enhanced products were initially exchanged as hard copy (prints or negatives). Later, they were distributed through media such as CCT (Computer Compatible Tape), tape cartridges like IBM’s 3360; small format media such as 8 mm or 4 mm helical scan tape cartridges (commonly known as Exabyte and DAT [Digital Audio Tape], respectively); and CD-ROM (Office of Technology Assessment, 1993).

With the growth of digital telecommunications infrastructure, Earth observation data are being accessed increasingly through web-based services, often custom-built. Because of bandwidth limitations, final delivery of data to the client is still often by physical media (MacDonald Dettwiler Pty Ltd., 1994a). Prior to order, however, a customer will often browse archives and select products for ordering, all via the web (De Witt & Naughton, 1994; Simpson & Harkins,
1993). In this study I concentrate on web-based services associated with managing Earth observation data and information, because this is clearly the emerging paradigm for marketing those products (Asrar & Ramapriyan, 1995; Australian Earth Observation Network, 1995; Baker & Finney, 1995; CEOS, 1997; MacDonald Dettwiler Pty Ltd, 1994b).

Large-scale Earth observation data collections are rarely thematic, but are most often sorted according to the sensor from which they were derived (Nill, 1996). To this extent these collections are primarily "data-centric" and the collection's custodians may specialise in handling data from only one sensor. This specialisation is due to the highly technical nature of the algorithms; the complexity of data processing; and the intimate understanding of the sensor behaviour required in order to develop useful information products (Simpson & Tajeldin, 1995).

Earth observing sensors on the same satellite are often built, owned and operated by separate entities, sometimes from different countries. The large Japanese satellite ADEOS (launched August 1996) is a good
example: it carried sensors from NASA and NOAA in the USA; from CNES in France; and from the National Space Development Agency, the Ministry of International Trade and Industry, and the Environmental Agency in Japan (Kramer, 1997).

This "hitchhiking" arrangement is often regulated through a formal Announcement of Opportunity issued by the satellite provider. In general, the agency or investigator supplying the sensor is also responsible for managing its data system, because in principal these groups are closest to the clientele requiring the data products. One of the consequences of ownership of sensor data by groups other than the satellite owner and operator is that generic Earth observation data collections are seldom compiled under a single centralised management (Schreier, 1996a; van Gulik & C. Best, 1996; Nill, 1996).

NASA's Earth Observing System (Enloe, 1995) and the European Space Agency's Earthnet (Fusco, 1996) are probably the most ambitious attempts so far to centrally manage large scale Earth observation data collections composed of data from multiple sensors.
The present situation of Earth observer data users is perhaps best
categorised as a set of point-to-point connections with providers, each
collection providing unique information products (Sarrat et al., 1995;
Office of Technology Assessment, 1993). A key contemporary
challenge for information systems in Earth observation is to replace
these multiple point to point connections between users and data
providers with a "market place" in which the user can simultaneously
compare various products and choose the most suitable data and
services. Requests or inquiries to this market place will be mediated
through a "middleware" layer featuring interoperable catalogues,
comparable to traditional library service union catalogues linking
special-purpose collections at numerous sites. This capability will
permit merging of Earth observation data with information from
other sources; and the capacity to manage multi-mission/sensor
requests, even at the satellite-programming stage (Sarrat et al., 1995).

The Working Group on Information Systems and Services of CEOS is
developing the “Catalogue Interoperability Protocol” to enable a
customer to interrogate multiple Earth observation databases in this
fashion (CEOS, 1998; Nill, 1996). This “global query” facility was an
integral component of the initial conception for EOSDIS (NASA, 1993).

Data price remains a strong determinant of user demand (Office of Technology Assessment, 1994; Gabrynowicz & Wood, 1991). Some advocates of network delivery of Earth observation satellite data have argued that this step would reduce data cost, alleviating the price barrier to market expansion (Australian Earth Observation Network, 1995). However, media and delivery charges are a minor component of commercial Earth observation product costs, which are driven in most cases by recovery of spacecraft operation cost, the spacecraft capital cost often being a hidden public subsidy (Mansell et al., 1993).

There is little persuasive evidence in the literature that on-line delivery would significantly reduce the price of Earth observation data.

Network access to Earth observation data and products is however regarded as desirable for other reasons:

• reduced delivery times compared to physical media;
• efficiency gains through ability to inspect ("browse") products before ordering;
• improved inventory management;
• decreased requirement to maintain local archive;
• greater uniformity of products;
• increased public support for space programs arising from greater access to information, and enhanced return on investment; and
• decreased operating costs in service delivery.

(Kingwell, Jayaraman, & Liu, 1995, p.17-18).

2.4.2 Categories of Earth Observation Information Systems

For convenience I distinguish between information locators; directories; inventories or catalogues; and information management systems, although in reality these categories often represent a continuum of technology and function.

In order of increasing capability, electronic systems for managing Earth observation data comprise:

• information locators which allow users to establish the existence and whereabouts of collections types of data. These collections are
usually described by text; information on how to contact the data custodian is supplied.

- **directories** often supply a greater level of alphanumeric detail about the type, geographical and temporal extent, access conditions and, sometimes, accuracy of the data collections. Sometimes a discrete guide is available and describes these attributes in detail.

- **inventories or catalogues** list the individual elements of a data collection (granules). A typical granule would be a scene, or digitised image of a specific dimension: for example, a scene for the SPOT 3 satellite represents an area of 60 km X 60 km with a pixel, or picture element, of either 10 m or 20 m linear size (resolution). Catalogues often contain a low-resolution copy of a scene (known as a quick-look or browse image); information required to unambiguously describe the granule (such as the co-ordinates of the corners of the image; the date on which the image was acquired; the part of the spectrum used); quality control information (such as proportion of the scene covered by cloud); and extra (ancillary) information needed to process the data (such as sun angle and spacecraft attitude).

- **information management systems** usually include all the above functions, as well as a facility for selecting processing level for the
required product; in-house management functions such as the migration of little-used data to "deeper" storage layers; usage statistics; and ordering/payment sub-systems.

Each of the systems above usually comprise several components:

• a "front end" or interface for inquiries and display of inquiry results;
• a communications protocol; and
• one or more databases containing logically-linked information.

2.4.3 How are Earth observation data systems different?

Van Gulik (1996a, 1996b) offers an insightful comparison between information systems used for Earth observation and those used in primarily text-based disciplines.

He describes the European Wide Service Exchange (EWSE), a resource location service focusing on environmental, natural resource, remote sensing and Earth observation information. Users requiring information for these applications typically require geographically-referenced data, rather than the text-based data typically required by scholars using more traditional information systems. Also, the value of
Earth observation information may critically depend upon the precise time it was acquired.

"Within these scientific community (sic) there is a clear desire to be able to locate geospatial 'environmental' information. Several large international bodies are active organizing (sic), collecting and storing such information. To the EWSE this community is of interest as its members have come to expect services which require a certain understanding of the meaning and context of the (meta) data. Unlike current services, such as Altavista or WebCrawler which are by and large (English) text and single 'word' oriented, the prime selection criteria is by it's (sic) geographic location, and to a lesser extent the time slot covered by the data. Furthermore in the above disciplines a sizable, though very dispersed and disjunct, body of information has been and still is collected; such as satellite imagery, catalogues of datasets, digital maps, etc."

(van Gulik, 1996a, p.1).

Metadata is a key concept in Earth observation data management. These are data about data: they give attributes such as the date of acquisition and the geographical area covered. Metadata play a similar
role to a library catalogue record: they make it easier for collection managers to group like articles and they enable potential users to make judgements about the utility of the information for a given purpose.

Van Gulik argued that metadata within Earth Observation information systems are often well-ordered and offer instructive examples to the broader Internet and information science worlds. Metadata standards generated for Earth observing may offer guidance for structuring generic web-based information location aids such as the proposed Universal Resource Characteristics (URC): van Gulik argues that the latter should allow for specialist "additions" such as geo-referencing, which is a characteristic search approach for users of Earth observation information. There is a risk that unless the URC approach supports searching based on geography, the information spaces of Earth observation, Geographic Information Systems, and text-base information will diverge (van Gulik, 1996b).

I will return to the important concept of metadata in Section 2.5.2.1.
Some typical examples of networked information systems in Earth Observation are now described.

### 2.5 Typical Earth observation data systems

#### 2.5.1 ESA (Europe)

The task of the European Space Agency, ESA, is defined in Article 2 of its Convention as:

> "to provide for and to promote, for exclusively peaceful purposes, cooperation among European States in space research and technology and their space applications, with a view to their being used for scientific purposes and operational space application systems"


The Agency comprises 14 Member countries and one participating Member (Canada). ESA’s Ministerial Council draws up a European space plan that spans the fields of science; Earth observation; telecommunications; space segment technologies including in-orbit
stations and platforms; ground infrastructure and space transport systems; and microgravity research.

ESA's European Space Research Institute (ESRIN) in Frascati, near Rome, operates several electronic systems for Earth observation data retrieval. The ESRIN Home Page (European Space Agency, 1998a) is the most convenient gateway into ESA's Earth observation electronic network, “Earthnet Online” (formerly known as the Guide and Directory Service, GDS). This is a comprehensive Earth observation information service, offering image browse; inventory; catalogue of products and services; mission information and numerous other features.

Earthnet Online is the European gateway to InfoSys, the CEOS Information System (IDN World Guide, 1994; European Space Agency, 1998b), an on-line service containing text and other information about the activities of the international Committee on Earth Observation Satellites, CEOS. The network of Earth observation systems supported by CEOS members is described in the next section.
2.5.2 CEOS International Directory Network (IDN)

The Committee on Earth Observation Satellites attempts to co-ordinate the operation of electronic Earth observation data networks on a global basis. Its Working Groups carry out development projects designed to ensure that national and agency networks are mutually compatible, or interoperable (CEOS, 1995).

Perhaps the most significant information search service sponsored by CEOS is the International Directory Network (IDN), a distributed database which contains directory information regarding Earth and space science data. These databases are not limited to image data, as they include atmospheric profile data and ground-based data used to validate (or check) Earth observation information. The Network system also carries instructions on how these data may be obtained.

Co-ordinating nodes in Japan (the NASDA Earth Observation Center), Europe (ESRIN); United Nations Environment Programme (UNEP, Nairobi, Kenya), and Washington DC (Goddard Space Flight
Center) serve as the regional collecting points for new information added to the network (Australian Cooperating Node, 1998).

"Cooperating" nodes in many countries operate local mirror copies of the system, and contribute new information through the nearest coordinating node (IDN World Guide, 1994). Within days of being logged at any co-ordinating node, new directory entries are copied to the other two nodes. This near continual replication cycle ensures that users accessing the network through any major node will obtain the most current information.

The IDN was established by the Working Group on Data (later the "Working Group on Information Systems and Services") and permits searches (text-controlled or free text) for:

- datasets in Earth science; Life science; solar physics; and space physics;
- field campaign and project information;
- information on some 140 data centres; and
- information on 88 different sensors used for Earth observing.
Searching can be done with Boolean free text, or through field-controlled search using combinations of terms chosen by pull-down menu.

Although life and space science data are covered, the data system is predominantly for Earth observation (with 133 data centers supplying information in this discipline as opposed to 3 in the life sciences and 9 each in solar and deep-space physics). Data Centre queries result in text information (about 200 words total) on the institution’s start date, purview, type of data held and on which media distributed; and contact details. Some contact details contain hypertext links to e-mail or URL, but most give options only for postal, facsimile or telephone contact.

The datasets referenced in the IDN database tend to be skewed to U.S. collections, because U.S. institutions have generally been more forthcoming in supplying electronic resource information, and because much of the IDN architecture originated with NASA as the "Global Change Master Directory" (Scialdone, 1992). However CEOS is active in adding details of new datasets, with ESA perhaps being one of the
more assiduous groups to update the IDN. Until about 1996, very few Australian data-gathering organisations registered metadata on the IDN, the Australian Oceanographic Data Centre being a notable exception. However there appear to have been a significant number of recent additions of details on Australian spatial data: I searched the IDN (Australian Cooperating Node, 1998) on 12 August 1998 and retrieved after about ten seconds 182 references to “Australia” in free text search. About 30 references were Earth observation images or data collections and about 40 more represented other types of geographically referenced data, from organisations such as the Bureau of Meteorology and the National Resources Information Centre.

One of the strengths of IDN is the ability to search for datasets by a set of queries (multi-parameter search) such as topic, date range, geographic coordinates, and geophysical variable (from a menu of 147 types). This utility gives users the ability to narrow down the number of possible sources for particular data. For example, using the "Dataset Information Query Form", one is able to quickly (within minutes) establish that solar physics information for the whole ionosphere, documenting sudden ionospheric disturbances, is held in two solar
physics data centers, namely the NOAA National Geographic Data Center and the NOAA Space Environment Laboratory.

2.5.2.1 Metadata on the IDN

Each IDN web site gives instructions on submitting metadata, which can be electronically lodged to any of its four primary sites, or coordinating nodes.

Metadata in IDN are written in a prescribed format. This is known as the Directory Interchange Format (DIF), adopted for NASA's pioneering Global Change Master Directory and now a de-facto international standard for creating directory entries which describe the characteristics of a dataset. A DIF has six mandatory fields:

1. directory entry identifier;
2. directory entry title;
3. parameters;
4. originating centre;
5. data centre; and
6. summary.
Up to 27 optional fields (including quality, resolution and use constraint information) expand and clarify the basic information.

Seven fields have been added to the DIF to facilitate compliance with the U.S. federally-mandated Federal Geographic Data Committee's (FGDC) Content Standard on Digital Geospatial Metadata. The new fields add information for users to make a better decision on the usefulness of a dataset. FGDC is one of several metadata format styles to which the DIF may be mapped (NASA, 1998d).

*Skinny DIF* is a shortened inventory record which consists of only mandatory fields. Skinny DIFs are put into a directory to alert users of the existence of a particular data set, and may be modified at a later time.

2.5.3 Australian National Spatial Data Infrastructure

Baker and Finney (1995) summarised recent initiatives of the Australia/New Zealand Land Information Council (ANZLIC) to identify and
model a national land and geographic infrastructure for Australia. The objectives of such an approach are to:

- produce standardised fundamental land and geographic data sets;
- avoid unnecessary duplication of effort;
- facilitate access to and applications of the data; and
- enable value-adding (integration of other data and information).

They identified several key elements of an institutional framework in this field:

- leadership ("ownership" of relevant policies and concepts; championing developments; sponsoring pilot projects; and devising technical and organisational models);
- custodianship (identified agencies/organisations which formally accept responsibility for data acquisition, storage, maintenance, quality assurance, security, access, documentation and distribution); and
- directories (containing metadata for the land and geographic data within the infrastructure and including the key characteristics of the data; information on access conditions).

(Baker & Finney, 1995).
Baker and Finney argued that a future national land (or geographic) data infrastructure must include a directory system through which potential users could determine the availability and suitability of data.

ANZLIC has developed, through consultation with "jurisdictions" of State, Commonwealth and private sector mapping and data organisations, a national standard for compiling geographic directory metadata (ANZLIC, 1996). The National Resources Information Centre is slowly implementing a national directory in which spatial datasets from Commonwealth and State agencies are cited, using the ANZLIC metadata standard.

2.5.4 Information Management System (USA)

The U.S. National Aeronautics and Space Administration (NASA), the world's largest space agency, is currently carrying out as one of its major tasks the "Mission to Planet Earth", MTPE\(^1\), which uses space-, ground- and aircraft-based quantitative measurements to increase scientific understanding of the global climate (NASA, 1995). MTPE is
said to be the largest scientific experiment in the world (Vetter et al., 1995), and has as its centrepiece the Earth Observing System (EOS), a series of complex Earth observing satellites and their attendant data systems. The first satellite in the EOS series is scheduled for launch in 1999, although related satellites taking part in the multi-billion dollar experiment were in operation before that date. NASA is developing a comprehensive infrastructure, the EOS Data and Information System (EOSDIS), to lay the basis for the archiving, retrieval and exploitation of information arising in the course of the decades-long program. EOSDIS is intended to manage all data arising in Mission to Planet Earth, whether originating from aircraft, land or space measurements. When EOS is fully deployed, the measurement systems are expected to generate more than one Terabyte ($10^{12}$) of data per day (Vetter et al., 1995).

The operating system for EOSDIS is known as IMS, the Information Management System. IMS is the gateway through which researchers may search for, select and order EOS data. The operating system is being implemented in a phased manner; the preliminary version was

---

1 MTPE was renamed "Earth Science Enterprise" in early 1998 (King, 1998)
released in 1995 (Enloe, 1995). This will be progressively upgraded as new elements of EOSDIS are completed (Colucci & Keener, 1995). Current and later versions of IMS will permit users to access pre-EOS data already held at NASA archives (Office of Technology Assessment, 1994; Maiden, 1996).

Version 0 of the IMS was installed by CSIRO in August 1996, as part of this study. The system is available for trial by CSIRO staff and the public (CSIRO-EOC Installation, 1998).

I shall return to EOSDIS in Chapter 4 where its development is critically analysed.

2.5.5 Intelligent Satellite Information System (Germany)

The Intelligent Satellite Information System (ISIS) was developed at the Remote Sensing Centre of the German aerospace research centre and national space agency, DLR. It is the central user interface with DLR's Earth observation archive, connecting users with data archives of affiliated data providers including those from the private sector and from the former Soviet Union. The ISIS server can be accessed by
modem, ISDN, X25 and the Internet, and can be operated by an ASCII interface or by a graphical interface for either PC or Sun workstations.

Through ISIS, users can select from over 30,000 digital images from 14 sensors, supported by a map browser and a geographic names lexicon. On-line browse, order and transfer are supported, and the system can mediate inquiries to other catalogues (such as NASA's IMS) through the CEOS Catalogue Interoperability Protocol (Sarrat et al., 1995).

ISIS is a highly capable archival management and user access tool, able to simultaneously interrogate multiple sensor databases (Schreier, 1996a).

2.5.6 Global Earth Observation Information Network (Japan)

Japan’s information systems for remotely sensed (Earth observation) data and information products are spearheaded by the Science and Technology Agency (STA) and by NASDA, the National Space Development Agency (NASDA, 1995a). In July 1993 the U.S. and Japanese governments began detailed discussion on the exchange and
distribution of Earth Observation information for global change studies, disaster mitigation, environmental monitoring and related fields (NASDA, 1995b). The concept of using high speed networks for such exchanges was elaborated by the G-7 Economic Summit in 1993, and a proposal to set up an international technical Working Group on Networks within the framework of CEOS was presented by the Japanese space and science agencies in Tokyo in November 1993 (NASDA/STA, 1993).

The resulting bilateral effort known as Global Observation Information Network (GOIN) was demonstrated in Tokyo in 6 June, 1995 (NASDA, 1995b). GOIN links the databases of its member organisations - U.S. universities and research bodies, Japanese government ministries and agencies, and eighteen other bodies - and enables them to be searched with a single query. The joint U.S./Japanese initiative, endorsed by President Clinton and Prime Minister Maruyama in 1995 has been described as "a model for global networking efforts being undertaken by space agencies" (NOAA, 1995).
GOIN quickly evolved into the NASDA Earth Observation Center’s Earth Observation Data and Information System, EOIS. Its objective is to serve the users of Earth observation satellite data, especially those interested in global environmental change, through provision of online services (NASDA, 1998a). The EOIS services a series of Bulletin Boards and acts as the Asian co-ordinating node for the International Directory Network of the Committee on Earth Observation Satellites (CEOS, 1995).

A test release of a web-based image catalogue service was developed within EOIS on 29 October 1997. This service (NASDA, 1998b) allows users to browse the inventory of Landsat TM, JERS-1 SAR, and ADEOS AVNIR data of limited geographic extent and obtained in the period 1 January 1996 to 10 June 1997.

2.5.7 GCNet (Canada)

The Canada Centre for Remote Sensing (CCRS) operates the Global Change Network, GCNet, comprising Canada's co-operating node of the CEOS International Directory Network. In addition to metadata, GCNet provides for geographic-based queries of image inventories;
access to Quick Look snapshots of available data; sample data inspection; and a Bulletin Board facilitating user feedback. Besides the CCRS, other Canadian agencies (including the Pacific Forestry Centre, the Institute for Space and Terrestrial Science, and the Office of Environmental Stewardship of Environment Canada) supply information through GCNet.

A more advanced system offering on-line data delivery is under study (MacDonald Dettwiler Pty Ltd., 1994b). This is virtually identical to the proposed, but later abandoned, Australian Earth Observation Network, AEON (MacDonald Dettwiler Pty Ltd., 1994a).

2.5.8 CEOS Information Location System (CILS)

CILS is a search engine for locating Earth science information and acts as an outlet for information about users, providers and products. It contains metadata on datasets, projects, people, organisations and products relating to Earth observation, remote sensing and environmental information, especially for developing countries. Information can be searched for by dataset name; geographic area
CILS was proposed by the German Space Agency DARA at the 8th CEOS Plenary in Berlin, September 1994, as part of CEOS's program for developing countries. A scoping study was then carried out by the European company GEOSCAN in 1995. DARA selected the German Aerospace Research Agency, DLR (which in 1997 absorbed DARA) to implement the project. A series of meetings with international colleagues led to installation of CILS at CSIRO Canberra; UNEP, Nairobi; DLR, Oberpfaffenhofen; and the European Union (Ispra, Italy) during August 1997; and at NASDA Earth Observation Center, Hatoyama, in February 1998 (Schreier, 1996b, 1997).

The design philosophy of CILS focuses on ease of use. The system and network requirements are non-demanding. Users can log details of their datasets, and advertise services or products, by sending an internet message to the system manager. Like the IDN, CILS constantly and automatically updates new information by replicating this to each operating site (Schreier, 1997).
2.6 Trends in catalogue services

The world's space agencies and Earth observation organisations consult on information management issues in a professional forum, the Working Group on Information Systems and Services, supervised by a non-government association, the Committee on Earth Observation Satellites (Kingwell, Wilson, Campbell, Ward & Bradshaw, 1996).

The vision of this group, in relation to catalogue service development, is to:

• implement a global, network-based marketplace for Earth Observation data;

• provide seamless access to data and services; and

• automate the process of resource discovery.

(CEOS, 1997).

The main drivers for these objectives are:

• the perceived need to increase the use of Earth Observation data;

• reduced space agency budgets;
increased interaction with private sector (diversifying source of data supply; Value Added Resellers who can build demand; sharing infrastructure costs); and

integration/convergence between in-house systems and new technologies (electronic commerce/authentication of user and transactions; high speed networks; artificial intelligence/intelligent agents; smart middleware; data archiving developments; JAVA).

(CEOS, 1997; C. Best et al., 1996).

The group observed that Earth observation (EO) data management was an example of resource discovery, and that approaches to EO metadata standardisation resembled more general attempts by the IETF (Internet Engineering Task Force) to facilitate resource location through the development of URC (Universal Resource Characteristics). This generic or systematic effort may result in better network search engines which could "capture" EO information via a facility in the generic metadata, incorporating geographic location and time range (van Gulik, 1996b).
Until these more powerful general search engines are developed, EO data searching can be improved in the medium term through interoperable protocols which would allow a single query to be brokered to a collection of heterogeneous catalogues, through a middleware layer (Retrieval Manager). This is the rationale for a major current CEOS project, the Catalogue Interoperability Protocol (CIP). CIP standardises the services needed for interaction between users and catalogues. Its middleware provides routing and translation services, helping users present searches to multiple heterogeneous catalogues (CEOS, 1998; Nill, 1996). Commercially-available tools such as CORBA and JAVA mean that this middleware can be made “smarter” to improve the speed and success of resource discovery (CEOS, 1997).

I now turn to examination of some general developments in electronic database technology, with an emphasis of those which are beginning to impact on Earth observation data systems.
2.7 Wider issues in scientific databases

2.7.1 Database copyright

A common expectation of scientists and other users of information based on the World Wide Web is that these data are "free". This belief may have arisen because of the roots of early web-use by the scientific community, for rapid exchange of nascent ideas and for everyday communication. However by the mid 1990s, the commercial implications of the web became clear to book and journal publishers as well as to entrepreneurs who wanted to introduce new electronic information services.

Publishers are accustomed to asserting Intellectual Property (IP) rights over their material. This helps protect against theft or unacknowledged use of created works, and is a concept central to the strategies and procedures of most knowledge-based organisations that operate in a commercial manner. These groups are now utilising electronic networks such as the Web to distribute products, and are attempting to assert IP rights in order to earn revenue (from royalties, access fees or similar arrangements) in doing so. Publishers will also be wary of potential loss of revenue arising from users down-loading "free"
material (for example, electronic copies of scientific journals) from the web instead of purchasing hard copies of the material.

A particular problem area is that IP rights may be asserted over the way in which information is presented or organised, as opposed to the content of the information. For example, the intellectual capital of a telephone directory publisher resides in the organisation and completeness of the directory, not in the ownership of the individual names and addresses. In a similar manner, profit-based web information services may market material that has been obtained from a variety of public and no-cost sources. Assertion of intellectual property rights over the way in which this of ensemble is presented, many researchers fear, may lead to the eventual removal of the source material from the public domain and incur additional research costs and effort.

Many librarians and users of copyright material are involved in plans for community digital libraries utilising publicly-funded electronic networks and computing facilities. These concepts are orthogonal to the vision of "knowledge entrepreneurs" represented, for example, by
the International Publishers Copyright Council (IPCC). The IPCC has embarked on a strategy of strengthening legal protection for copyright owners, in order to develop commercial opportunities in electronic publishing. By its nature, this strategy can only succeed by reducing, in the special field of databases and electronic data, "public access" provisions presently enjoyed and expected by users of written or hard copy information (Herd, 1997).

Current moves to amend international copyright law reflect in part the evolution of technology - the advent of the "electronic age". The proposed change represents a paradigm shift in favour of large-scale purveyors of knowledge, or information brokers. In the firing line are public-access institutes, including libraries and academic institutions which have traditionally defined their role, in part, as bringing information to the notice of the public. New regulations considered (but not accepted) by the World Intellectual Property Organisation (WIPO) in December 1996 could affect access to databases in widespread use, such as Dialog and Medline, as well as those being developed under various "digital library" initiatives worldwide (Woodberry, 1997).
Many scientific organisations such as CEOS and the International Council of Scientific Unions (ICSU) have warned that private ownership of electronic databases - particularly those comprising data acquired for public good purposes and at public expense - could threaten the efficient use of web-based information resources for research purposes.

Dr Angus McEwan, former Chief of CSIRO Division of Oceanography, has been working with the Australian Academy of Sciences to ensure that these concerns are recognised by those negotiating international "information trade" treaties on Australia's behalf. McEwan (McEwan, 1997) argues that the recent dependency of scientific research on electronic databases has arisen because of the insatiable need for more, and more widely based, research data. However, this trend has left these activities vulnerable to changes in access or usage conditions under which databases operate. Moves by WIPO to bring about a "pay for view" regime protected by vigorous copyright of databases could seriously limit international exchange of scientific information. Datasets that may be particularly vulnerable
include satellite observations of the Earth, and data on genome (gene) sequence. McEwan argued that the best interests of the international research community would probably require that any database copyright regime includes specific provisions for "fair use" (as in review, research and teaching); concessions for public benefit (non-commercial) use; preservation of public domain status of data later included in copyright databases; and "default conditions" under which it may be assumed that unless otherwise stated, the generators of original data will permit no-charge access to other users (ibid.).

2.7.2 Interoperability, data warehousing, and distributed architectures

2.7.2.1 Information dependency

Many organisations now realise that many of their corporate assets lie in or depend upon databases. In the past, these have been hardware-specific, creating major discontinuity and risk when the technology becomes dated and needs to be replaced. A new information skill, database migration, has evolved to help keep organisations functioning while their data are moved from one "platform" to another. A high rate of company takeovers in a volatile economy, and the growth of the "globalised" economy, mean that many organisations function at a
multitude of geographic locations, or use a variety of databases and supporting technologies. These organisations typically require information to flow through and between these different inherited or "legacy" information systems in a way that is transparent to a user.

"Data has become a critical resource in many organizations and therefore efficient access to data, sharing the data, extracting information from the data, and making use of the information has become an urgent need. As a result, there have been many efforts on integrating the various data sources scattered across several sites. These data sources may be databases managed by database management systems or they could simply be files. To provide the interoperability between the multiple data sources and systems, various tools are being developed. These tools enable users of one system to access other systems in an efficient and transparent manner."

(Thuraisingham, 1997, p.1).

**2.7.2.2 Data warehousing**

Recent analysis of information structure, especially in the corporate finance sector, appear to demonstrate the value of removing from the
workplace or "production environment" responsibility for processing and managing little-used or minimally-processed data (Inmon, 1992).

This has the advantage, it is argued, that decision makers are then more able to concentrate on the generation and evaluation of higher level (or "derived") management information. According to this view, the supporting data which helps decision making by executives should be maintained in atomic or raw form in subject-oriented databases in a logical space known as the data warehouse (ibid.).

The purpose of the Data Warehouse is to support decisions in the enterprise, by supplying raw, correlated, interpreted or other categories of data. NASA's Distributed Active Archive Centers (Asrar & Dokken, 1993) can be viewed as Data Warehouses in Earth observation.

Many corporations operate private telecommunications networks or intranets, many of which are linked to the internet. Off the shelf commercial software and web-browsing products are now emerging that make it feasible for corporations to develop "Intranet-enabled"
data warehouses which enable users to dynamically generate database queries and obtain the query result on a web browser (Tanler, 1997).

A decentralised variation to the data warehouse is the data mart, which services a single subject area, market segment or product (ibid.).

2.7.2.3 Federated and centralised database architectures

Thuraisingh (1997) proposes a generic or reference model for database systems featuring three levels of data management:

- database technology and distribution;
- interoperability and migration; and
- information extraction and sharing.

These data-managing functions sit on a "supporting layer" comprising:

- networking;
- distributed processing;
- mass storage;
- agents, and
- distributed object management.
The four data management and supporting layers supporting the highest or "application layer", which includes collaborative computing; visualisation; mobile computing; and knowledge-based systems.

There is a dichotomy between autonomy and cooperation, in relation to networks of databases. This conflict can be resolved by centralised management, in which case (some argue), the usurpation of responsibility leads to loss of identification (or "ownership") by the original database custodian, possibly with overall loss of background knowledge and efficiency. Alternatively, entities responsible for databases may negotiate on levels of cooperation, while retaining some degree of autonomy. Such collectives are called "Federations" (Thuraisingham, 1997).

Federated approaches to information management have been proposed for many years, for example by Heimbinger and McLeod (1985).

Papazoglou characterises the federated approach as follows:
"A federated database architecture allows forming a loosely coupled union of the data, by means of a collection of component databases participating in the federation. A federation consists of a number of interconnected nodes and a federal dictionary which maintains the topology of the federation and oversees entry of new nodes. In a federation there is no global schema and no central authority so the different nodes would have to negotiate as to which portion of the data they can 'see'."


In federated structures, three schemata are simultaneously necessary: namely import, export and negotiation schema.

By contrast, a "logically centralised" database architecture uses a global schema, applied through a "global manager", to address queries to all the constituent databases. In practice, this ability constitutes an advantage over federated approaches:
"The decentralised nature of a federation offers many advantages when compared to logically centralised databases. However, logically federated database architectures present some serious limitations such as the need for maintenance of a set of three complex component schema as well as the requirement for many complicated forms of communications protocols to implement a pair-wise dialogue among any two component databases in the federation. The fact that federated architectures adhere so strictly to retaining local autonomy introduces some of their most notable drawbacks and inflexibilities".


2.7.2.4 Database interoperability

Scientific research and commercial activities both rely upon rapid access to data held in a distributed collection under multiple proprietary database management systems. Users may wish to retrieve, extract and manipulate these data within a given software environment or application. Microsoft Corporation initiated the "Open Database Connectivity" (ODBC) concept, with the intention that Microsoft applications can operate upon various database management systems.
The ODBC idea expanded beyond its company-specific origin and has is becoming a de facto standard for client-server interoperability (Thuraisingham, 1997).

One can distinguish between two principal classes of databases: those that deal in relationships between parameters, and those that describe objects. The latter (object-based databases) include image data collections such as Earth observation data archives.

Recently, a specification known as CORBA (Common Object Request Broker Architecture) has been made by the Object Database Management Group. The purpose of CORBA is to allow the growth of heterogeneous, distributed, object-oriented database systems. CORBA makes it possible for heterogeneous applications and databases to interoperate through means of middleware, a level intermediate between the operating system and applications and which connects these elements. A key concept in CORBA is the Object Request Broker (ORB), which interprets communications between multiple clients and multiple servers (Thuraisingham, 1997).
Developments like CORBA and ODBC mean that it is becoming more feasible for a user to interact with a Federated decentralised database collection, and to extract information from it through a single query. A prerequisite for this capability is that each of the contributing databases must summarise and label its contents using a metadata structure addressable by the user's query.

2.7.3 Internet services and distributed high performance computing

The Internet arose from military research into robust communications, commencing in 1969 with the ARPANET, developed by the U.S. Department of Defence's Advanced Research Projects Agency. The premise was that the ability to re-route communication via surviving links should outlast the outbreak of a global nuclear war (Krol, 1992).

Use of this network for person to person communication grew slowly and apparently unintentionally. By the late 80s, a critical innovation - the World Wide Web, developed by physicists at the European Particle Physics Laboratory - enabled scientists to more easily locate and retrieve information via the Internet. In early 1993, software known as
Wanderer was created by Matthew Gray. This made it possible to "crawl" the Web, counting servers: 100 were located by this method in June of 1993. Two and a half years later, in January 1996, the count was 100,000 (McMurdo, 1997).

Internet-based information systems have been developed to make information in various native formats—such as video, text, audio—available to the community. Hyper-G is an example of such "hypermedia" information services; it was developed at the Graz University of Technology as the basis for a Europe- or world-wide "University information system" (Kappe, Maurer, & Sherbakov, 1993).

"Minitel", the French network used by over 6 million people, was one of the first demonstrations of the services which can be provided via Public Switched Telephone lines. Minitel, introduced in the late 1980s, provides households with telephone directory services; chat lines for special interest groups; access to medical databases; bibliographic networks; and 20,000 other services (Lanoue, 1994).
Use of the Internet for research purposes has grown in extraordinarily rapid fashion. Many academics use the Internet simply to keep in touch with their peers, while others use it for nearly every facet of their work.

Klobas (1995) examined use rate, and influences on use rate, for information resources provided in electronic form through a "CWIS" or Campus-Wide Information Service, in this case at the University of Western Australia. She examines this issue from three perspectives:

- organisational behaviour;
- library and information science; and
- information technology.

Klobas used the "Theory of Planned Behaviour" and the "Fitness for purpose model" to help anticipate use for the CWIS, concluding that the "Fitness for purpose" approach gives better results. In this approach, the two principal questions information resource providers should ask of potential users are:

1. how useful is this resource likely to be to you? and
2. how convenient will it be for you to use?
Here "Fitness for purpose" was defined as the extent to which an information resource is of appropriate quality for the situation in which it is to be used.

The Internet can also be used for "distributed computing", in which a computationally complex task can be sub-divided and performed at a number of separate sites in parallel. This shares the computing load, and makes it possible for particular laboratories to concentrate on number-crunching tasks closely related to their needs or expertise, while leaving other necessary, but less relevant, computing tasks to others (National Research Council, 1993).

2.8 Scientific communication

The sociologist De La Solla Price, in exploring methods of scientific collaboration, coined the term "invisible college" to describe the informal (and sometimes transient) collaboration between elite, productive scientists who apply mutual influence even at a distance. Griffith and her colleagues Garvey and Mullins elaborated the concept
(Garvey, 1980; Griffiths & Mullins, 1980), pointing out that informal communication between scientists working on similar topics, but in different formal structures, are often more robust and productive than formal reporting mechanisms. In such cases, scientists “recognise” others as peers and colleagues, while sometimes excluding from their communications those among whom they formally work. This modality of communication is common in scientific fields which are just emerging, because in this circumstance, institutional structures have not yet evolved to accommodate the new discipline and regulate the distribution of information.

There is emerging evidence (Walsh & Bayma, 1996) that increased use of “computer-mediated communication” by scientists is changing their pattern of collaboration by making it possible for researchers to share access to scarce resources such as databases. This reduces the intellectual isolation of scholars working alone or in small groups at widely separate locations, and counteracts, it has been suggested, a propensity in scientific circles to bias distribution of resources toward the eminent or fortuitously-located (ibid.).
2.9 Strategic information management

Management and information theorists have recently explored the links between an organisation's underlying purpose ("core business") and the information processes and technologies it employs, whether purposefully or inadvertently.

It has long been clear that information technology and procedures, no matter how "suitable" from a technical point of view, will not receive acceptance or resource priorities if they are associated with cultural norms markedly different from those of the "host" organisation (Horton, 1987).

Economists and information specialists now recognise a growing class of "knowledge-based" enterprise, to whom information is a commodity and their major investment (Klobas, 1997).

Strategic information resources management connotes the recognition of information as one of the key raw materials for the success of a modern organisation. Alongside staff, finance and physical assets, information is now regarded by some management theorists as the
"fourth resource" which is essential for survival in an increasingly competitive and changeable world (D. Best, 1996).

The United States General Accounting Office analysed patterns of information management and technology use in 19 organisations which it considered to be leaders in blending information policies with organisational objectives. That study (General Accounting Office, 1994) identified a number of techniques, issues and strategies for strengthening use of information management and technology in order to improve delivery of goods or services. The three key phases of this process are:

1. deciding to change;
2. directing change; and
3. supporting change.

An important component of the third aspect is the establishment of a champion, a "Chief Information Officer" who is a senior manager responsible for all aspects of information flow.
Orna (1990) pointed out that there is often a gulf between those managing information and those managing organisations. She quotes the Director General of the British Institute of Management:

"One of the biggest culture gaps in Britain is the one between those who know how to handle information and those who have the responsibility for running businesses."

(ibid., p.14)

In the strategic information management approach, the starting point is establishing the objectives of the enterprise. This is not as easy as it sounds, because there may be a discrepancy between reality as experienced by employees and clients, and the corporate image projected by brochures, reports and slogans. One must sift through the formal statements of objectives and perform a "reality check" by interviewing key people. Unless a realistic statement of enterprise objectives can be arrived at, there is limited point in designing an information system to service it.

Sangway (1989) proposed a general method for analysing an organisation's information structure and requirements in terms of its
underlying purpose. Understanding the objectives and functions of the organisation was the starting point for determining optimum information flow and improving information systems.

This concludes analysis of broad but pertinent developments in information science. I now turn to the nature of the organisation which employs the user group at the centre of this study.

2.10 Earth observation in CSIRO

2.10.1 Overview of organisation

The Commonwealth Scientific and Industrial Research Organisation, CSIRO, was founded in 1926 and is Australia's largest research body. With a budget of $689.2 million in 1996/97, CSIRO obtains about 65 per cent of its funding directly from Parliament; the remaining 35 per cent comes from "external" sources, including competitive granting schemes, research funded by industry and other users, and earned revenue (CSIRO, 1998b).
CSIRO conducts research in 22 areas ("sectors") in five principal areas: agribusiness; environment/natural resources; information technology/infrastructure/services; manufacturing; and minerals/energy. Earth observation and remote sensing supports many research activities conducted by CSIRO, especially those carried out in the climate and atmosphere; minerals exploration and mining; marine; land & water; and information technology and telecommunications industrial sectors (CSIRO, 1997a).

Earth observation is, however, not viewed by CSIRO as either a scientific discipline or an industrial sector; rather it is regarded a generic technology with multiple research and service applications. The bulk of CSIRO’s research in Earth observation takes place in about 16 of its 24 research Divisions or units, with resources and priorities being assigned in terms of the discipline-based application which the host division practices. Approximately 76 CSIRO scientists carry out Earth observation research at a total of about 29 different sites; of this group, perhaps 10 are principally concerned with data management while the remainder are principally employed on the interpretation and application of the information. Total annual
expenditure in remote sensing by CSIRO is estimated to be about $10 million (Simpson et al., 1995).

Although CSIRO possesses no Earth observation satellites, it owns and operates two satellite ground stations (in Hobart and Melbourne) and participates in the operation of two more (Perth and again in Hobart). The oldest of these ground stations has operated since 1983 (Kingwell, 1990). The rationale for CSIRO assuming a (comparatively rare) operational responsibility for these facilities is that in the past, operationally-oriented organisations (such as the Australian Centre for Remote Sensing and the Bureau of Meteorology) had insufficient technical or budgetary means, and insufficient incentive, to operate in a timely and cost effective manner the services required by CSIRO.

The CSIRO Office of Space Science and Applications (COSSA) was founded in 1984 and for many years it was assumed both inside and outside CSIRO that COSSA would somehow act a focal point for CSIRO remote sensing (see for example Aubrey, 1988). However COSSA had no research mandate and therefore had little standing, inasmuch as directing scientific research, within a culture dominated
by researchers. An indication of its lack of research independence was the oversight of COSSA by a “Steering Committee” comprising senior research management. From about 1992 to 1996 this committee comprised the four deputy chief executives responsible for the groups of research Division seen as “clients” of COSSA. From about 1996 this function was devolved to the Chiefs of the Divisions concerned. This form of supervision is rare in CSIRO and tends to be employed when several research units compete for the resources or services of a non-research (“support”) group or a joint facility.

2.10.2 Previous reviews of Earth observation in CSIRO

Several earlier studies of CSIRO’s space research programs have emphasised the need for increased focus on Earth observation data management. A 1993 report arising from meetings of research program leaders devising a new approach in CSIRO space research (Fandry, Harris, & Huntington, 1993) remarked that with remote sensing moving from simple image manipulation to a tool for modelling geophysical parameters, the ability to locate, and then merge, spatial data from a variety of geographically separate databases and archives was growing in importance. This review stated that:
"The future of remote sensing lies in the development of geophysical models to turn what we can see into what we need to know: to become solution oriented."

(loid, p.18).

This review also indicated strong support for further investigation of data access and archiving needs, and proposed NASA’s Distributed Active Archive Centers as a model for Australia.

To better coordinate remote sensing research and operations, CSIRO established in 1992 a “Multi Divisional Program in Satellite Data Acquisition and Utilisation”. The purpose of the program was to “...ensure efficient data gathering and use of Earth observation satellite data to support the research objectives of CSIRO Divisions and their client and co-operating organisations” (Simpson et al., 1995).

At the time it was established, this was the nineteenth such cross-disciplinary research effort, hence the title “MDP19”. MDP19 was coordinated by COSSA under the direction of a scientific steering committee representing 12 participating Divisions or units (Deeker & Kingwell, 1996). Expenditure by COSSA in this program
(approximately matched by combined expenditure from other Divisions) was about $AUS 400,000 per year, some 15% of COSSA's budget (Kingwell, 1995a; Deeker & Kingwell, 1996). In mid 1994, the then head of COSSA, Dr Brian Embleton, initiated a review of CSIRO (and more generally, Australian) Earth observation data needs, and of the existing research effort (MDP19) in this field. The objective was to establish world "best practice" and to help CSIRO plan, on the basis of its strategic research interests, for anticipated changes in Earth observation, such as the advent of EOS (Simpson et al., 1995).

This was probably the most thorough general survey of CSIRO Earth Observation ever conducted, and was carried out in 1994-95 by a review team chaired by an independent expert in Earth observation information systems, Dr James J Simpson of the Scripps Institution of Oceanography in San Diego. Other members of the review team were the leader of the Bureau of Meteorology's satellite group, Mr Bruce Neal; and three CSIRO staff: Dr Ian Barton, Mr Jeremy Wallace, and Mr Jeff Kingwell, of the Division of Atmospheric Research, Division of Mathematics and Statistics, and CSIRO Office of Space Science and Applications, respectively.
The Simpson review relied upon data gathered from written submissions and prior reports; and from a site-visit and interview program involving 22 separate locations and nearly 20 different organisations apart from CSIRO. Seven recommendations emerged from this study:

1. establishing an Earth Observation Centre to consolidate the bulk of CSIRO's remote sensing, reducing duplication of effort and conflict of objectives; achieve scientific critical mass for higher productivity; better cost efficiency, morale and improved high level data products;

2. support establishing of Distributed Active Archive Centres where data could be captured, archived and distributed as peer-reviewed Level 2 information products;

3. a long-term strategy for an Australian ground station network, capable of receiving data from higher frequency (X-band) transmissions from advanced Earth observing satellites;

4. a cooperative agreement between CSIRO and the Bureau of Meteorology on Earth observation data acquisition and management;
5. scientific working groups to decide upon the best procedures (algorithms) for generating Level 2 and 3 data products;
6. a continuation of the research effort in calibration and validation;

and

7. greater emphasis on the generation and distribution of value-added products (compared to the acquisition of raw data).

(Simpson et al., 1995, p.29-33).

2.10.3 Comparison between CSIRO and NASA

The recommendations of the Simpson review represent a delivery model similar to the “Federation” of collaborative interests proposed by the National Research Council (NRC) in the case of EOSDIS (NRC, 1995a). However, the Simpson review recognised that CSIRO has a research culture rather than an operational culture. In this respect, the review argued, CSIRO should avoid operating data facilities (such as the proposed Archive Centres), instead supporting their establishment by more operationally-inclined organisations, or through collective effort.

In his book *The Gods of Management*, Handy characterises research environments as being populated by existentialists, craftsmen and club
members (Handy, 1995). Operational (or role) cultures are typically populated by bureaucrats, Handy asserts. Handy argues that organisations express particular dominant cultures that made it difficult for them to fulfil certain work patterns.

NASA provides an interesting example of blended organisational cultures. Although NASA has a clear research responsibility, this is manifest as a sub-culture rather than the dominant or mainstream culture: the co-existence of normally conflicting norms of organisational behaviour was made possible by NASA's origin as separate and quasi-autonomous Centres with differing background and composition (Newell, 1980). The advent of operational responsibilities such as routine space flight led to some stresses as the organisation evolved towards a more “repetitive” or operational role model (McCurdy, 1993).

Unlike CSIRO, NASA from its outset had a strong imperative to deliver agreed results (first American satellite; first American in space; first humans on the Moon, etc); this series of fairly well-defined tasks demanded task-driven sub-cultures. As operations such as space launch
and satellite operations became routine, role-driven, repetitive and predictable (bureaucratic) subcultures became institutionalised. As an aside, I observe that the last word in the acronym ‘NASA’ is “Administration”. It would be difficult to imagine the research culture of CSIRO accepting the appearance of that word in its own title.

On the basis of organisational culture and role, it is “normal” (although not necessarily preferred) for NASA to support operational functions such as controlling Earth observing spacecraft and the archiving and distribution of their data. On the other hand, similar data management tasks tend to be conducted by CSIRO only on sufferance, to the extent they are seen as imperative for CSIRO’s “real” purpose of delivering research outcomes.

The Simpson review, in recognising this role or cultural aversion, advocated that CSIRO attempt to encourage other, more culturally-suitable organisations (such as the Bureau of Meteorology and, implicitly, the Australian Centre for Remote Sensing) to carry out the routine data collection tasks, with a correspondingly greater leadership by CSIRO of the research-oriented task of improving data products.
A problem with the Federation model for data management in Australia, as proposed in the Simpson review, is that unlike the case of the EOSDIS Federation there is no NASA or NOAA to act as the natural leader for the data collection (Level 0 and 1) effort.

On a national scale, responsibility (de facto or formal) for the collection of primary Earth observation data (Level 0 and 1) is divided between the Australian Centre for Remote Sensing; the Bureau of Meteorology; the Australian Institute for Marine Science; several universities; CSIRO; the Plague Locust Commission; and other groups. Until recently, there was little systematic effort directed towards a union catalogue or joint directory of data and product holdings, which would make a “Federation” of data providers more transparent to the user. Without such a systematic approach, it is probable that researchers and other users will continue to make sub-optimal use of existing infrastructure, or even plan to establish one-off national data networks for individual experiments (for example, see Graetz, 1996b).
The development of an Australian national data network in Earth observation may depend upon the emergence of a more complex set of organisational alliances. Possibly for this reason, progress has been slower in addressing infrastructure recommendations (data centres, acquisition networks) than on those elements of the Simpson review requiring action by only one or two organisations.

2.10.4 The CSIRO Earth Observation Centre

One of the earliest and most significant responses of CSIRO to the Simpson review was the creation in mid 1995 of the Earth Observation Centre. The appointment of Dr David Jupp as Science Leader was made in early 1996. A draft Science Plan developed by CSIRO Earth observation researchers (Graetz, Prata, Wallace, & Barton, un pub.) and the subsequent plan prepared by Dr Jupp (Jupp, 1996) demonstrated in general a stronger emphasis on application research than on data management, in comparison to the Simpson review.

These science plans argued that the end-users of Earth observation data are not concerned with the data sources or the processing methodologies but rather in the utility of the information product and
its validity. The provision of a variety of value-added information products (Level 2 and 3), and documenting their validity and quality, are therefore primary CSIRO Earth Observation Centre goals and determine the priority of resource allocation (Jupp, 1996, 1997; Graetz, 1996a; Wallace & Campbell, 1998).

One of the purposes of the research reported here is to determine if it is feasible for the Earth Observation Centre to establish in CSIRO a data management framework that would more easily permit users of Earth observation data and information to locate and access these services, even under the prevailing condition of divided responsibility for the acquisition and initial processing of the original data.

Next I review literature on the methodology adopted in this study.

\textbf{2.11 Research methods}

\textbf{2.11.1 Action Research}

This research took place in a particular context: the author was employed in an organisation which demonstrated a long-standing...
requirement for Earth observation data systems. As well as being academic research, the work was also operationally oriented, designed to bring about a change for the better in the management of Earth observation information inside the author's organisation.

In this situation, the researcher becomes a protagonist in the research activity, not a chronicler of experimental results observed in a disinterested manner. This approach is often termed "Action research".

This form of research can be especially advantageous to the sponsoring organisation, because it has the explicit objective of improving practice in an area of identified weakness. By the same argument, the research may directly benefit the researchers and their colleagues. However, there is an additional responsibility on the researcher to avoid alignment with pressure groups or partisans whose objectives may not coincide with other groupings. Research ethics are therefore an important aspect of this method (Powell, 1991).

The emancipatory and self-help aspects of this research method have been stressed by social theorists, such as Habermas, and by activists
and scholars who believe that it is a researcher's responsibility to affect change for the better in the lives of those whom one studies and works (Nissen, 1985; Sandberg, 1985; Habermas, 1978).

Bunning offers an extensive analysis of this research method (Bunning, 1994), focussing on applications in the social sciences and in social situations where it is impractical for the researcher to adopt the distant observer standpoint characteristic of other research methods.

One definition of action research emphasises its iterative method:

"Action research can be defined as a process whereby, in a given problem area; research is carried out to:

- specify the problem;
- identify a plan of action;
- monitor the effectiveness of the action; and
- identify what has been learned (sic) and how this should be communicated."

(Bennett & Oliver, 1988, p.3).
Bunning amplifies this definition, saying that the method has more aspects than can be readily contained in a brief list. Instead, he describes action research as displaying a number of characteristics, in that it is -

- *practical* (leading to improvement)

- *participative* (the researcher is a co-worker of people having the “problem”)

- *emancipatory* (lacking hierarchy and intended to make people more aware of their own ability to solve problems in workplaces)

- *interpretive* (solutions are based on views and interpretations of people involved)

- *critical* (community involved is critical/self critical)

- *representative* of the community concerned

- *re-educative* in that participants typically change their views in conformity with knowledge acquired during the research

- *multidisciplinary* and lacking disciplinary bias

- *evolving and open-ended* (leading to continuous process of inquiry and improvement).

(Bunning, 1994, p.21-22).
Trust and political validity are established by the researcher, because these (with scientific validity) are prerequisites for the remedial action intended by the research (ibid.).

The action research approach may appear illegitimately interventionist in comparison to more traditional “objective” scientific research methods, such as experimental measurement. However, in social research this method has received much recent attention, and is often preferred where the intention of the research - as is the case here - is for the researcher to join with the “subjects” as co-learners in order to collectively solve practical problems and to implement the solutions discovered. In many cases, a “hands-off” approach (one in which the researcher would neutrally ‘observe’ but not ‘participate’) may inhibit the development of understanding about the organisational context of a particular problem, undermining the ability of the researcher to resolve the difficulty (Bunning, 1994).

2.11.2 Survey techniques

An important element of this research was establishing user attitudes to particular information systems. Given that the subject group is
widely dispersed geographically, extended direct observation of work patterns was not feasible, and so survey techniques were adopted.

Surveys, especially those utilising questionnaires, are frequently used in social research or in other disciplines where human beings are key components of the issue under study. Questionnaires are convenient and relatively low-cost methods of gathering information, especially from geographically-dispersed sources. However, they require careful preparation and testing.

Characteristics of good questionnaires may include the following:

- include a return envelope;
- avoid questions which are irrelevant to the objective;
- vary the order of graded responses
- start with the general and move to the specific;
- group similar questions by topic, beginning each group with a brief explanation or description of what follows;
- use a consistent format for answers (e.g., ticks, crosses, but not both for affirmative replies); and
- contain sufficient background, often in a covering letter.
Obtaining or being able to readily devise a list of prospective respondents is almost a prerequisite for carrying out a successful survey. The most suitable potential respondents are those likely to share, with the researcher, some interest in the research topic (Wurzburger, 1987).

Moore (1983) advises use of questionnaires as a selection tool for compiling a short list of subjects for more intensive questioning, where open questions can be more readily used than in questionnaires.

Although personal interviews can be expensive, and are more time consuming than questionnaires, they tend to have a higher response rate and can be used to supplement answers given in questionnaires. Blackmore (un pub.) and Moore (1983) both give detailed guidance on questionnaire design. They stress the importance of testing the product to reduce ambiguity or bias, and of maintaining anonymity of the subject to increase the probability of frankness in the responses.
2.12 Summary

Earth observation from space is being adopted by a growing number of countries, with increasing private sector participation.

More emphasis than ever before is now being placed on successful delivery of useful data from Earth observing satellites, because governments have felt more obliged to justify their space expenditure and the private sector is obliged to meet shareholder's expectations. This is focussing attention on user (customer) needs, and in turn upon information management systems which acquire the raw data and harvest the fruits in the form of elaborately transformed information products.

A difficulty facing those who implement such information management systems is that the complex technology of Earth observation has resulted in a legacy of iconoclastic and data-centred collections (archives) operated by a wide variety of entities. Significant recent effort has been devoted to enable potential users to more easily search for information in these collections. I examined a number of these information management systems and noted that the Committee
on Earth Observation Satellites has played a key role in championing system interoperability.

I demonstrated that developments in information science, such as database interoperability, middleware, and articulation of architectural models for distributed information systems, have complemented changes occurring in the Earth observation field. The Federated architectural model has captured attention as it recognises both the geographic dispersion and the heterogeneous management of Earth observation databases. Some trends in information systems, such as a move towards a stricter copyright regime on electronic databases, warn us that better technology alone will not guarantee better service to users of Earth observation data systems.

This theme was further developed by reference to recent literature on strategic information management, which emphasised the nexus, in best practice organisations, between information use and organisational purpose. Methods of communicating scientific information were also briefly noted, with the observation that computer-mediated communication appears to be magnifying the
ability of researchers to collaborate across geographic and management boundaries.

The nature of the organisation at the centre of the study - CSIRO - was next reviewed. Unlike most organisations, CSIRO exists for research, and other activities are subsidiary to this end. The group contrasts with production, operational or service-oriented cultures. Previous studies of the organisation have shown weaknesses in its management of Earth observation research and information.

An analysis of research methods indicated that action research appeared to be a suitable approach to the research problem of choosing suitable Earth observation information systems for use in the CSIRO environment. Literature on survey and questionnaire methods to be employed were also summarised.

The research method adopted in this study is elaborated in the next chapter.
3. RESEARCH METHOD
   How the work was done

3.1 Introduction

In this chapter, I explain the components of the research and the research methods applied to each component of the work. Technical terms and acronyms are explained in Appendix A, and milestones for the study are shown in Appendix B.

Many methods have been employed in scientific research. Some have as their basis the belief that there are certainties in nature which can be uncovered through investigation or reasoning. Others are based on the concept that knowledge is conditional, and can be overturned through new discoveries or ways of thinking. Qualitative and quantitative approaches have prevailed at various times, or have co-existed as a dialectic struggle. Current thinking in the information sciences supports a plurality of research methods, depending upon the nature of the topic as well as the working conditions and social context of the researcher.
In this study I used two principal research methods - case study and action research - in order to learn about the “best” information systems for Earth observation research within a specific organisational culture. “Best” in this case is defined firstly in terms of *episteme*, what is known through observation of similar situations to be true. “Best” here is also defined in terms of the social reality of the users, including the workplace resources available to carry out their research, and the core beliefs of their peers.

The research was carried out in six phases, some of which were conducted in parallel. These six phases comprised:

1. literature survey;

2. case study;

3. examining the role of information in CSIRO;

4. selecting prototypes;

5. installing and testing prototypes; and

6. evaluating the suitability of the prototypes.

Each phase is described below.
3.2 Literature survey

A review of current practice in Earth observation information systems seemed a logical first step for this research because of the expense of developing new systems; the extent of recent international effort in this field; the specialised nature of these systems; and the prospect of gaining from the experience of those who are already expert in this area. Literature including journals; on-line resources; reports and contract studies were analysed to determine the state of the technology and to uncover trends in data management systems for Earth observation. In addition, literature relating to policies and their trends was studied, as was material dealing with more general developments in the field of information science. The generic aspects investigated included management of heterogeneous and distributed databases; data warehousing; the concept of strategic information management (the alignment between information technologies or processes and the underlying mission of the host organisation); and the differentiation between “user-centric” and “data-centric” approaches in the design of information systems.
The literature survey (Chapter 2) also included examination of relevant research methods and techniques, such as survey design and interviewing.

Since 1990, I have been an active participant in the Working Group for Information Systems and Services (WGISS), a forum for information scientists and users in the field of Earth Observation. WGISS carries out its work on behalf of the Committee on Earth Observation Satellites (CEOS), an informal body in which world space agencies, research organisations, and international scientific programs cooperate to improve technical aspects and coordination of space-based observations of the globe. WGISS can be seen as a peak council or reference group of specialists involved in research and implementation of Earth observation data systems. It comprises about 30 or 40 individuals from space agencies, research organisations, private contractors, and scientific programs using Earth observation information. CSIRO had been a member of CEOS since 1989 and was therefore eligible to participate in the Working Group. I was the CSIRO representative on it and its predecessors (the Working Group on Data and the Working Group on International Network Systems)
from 1991 to 1997. Interaction with this peer group was crucial to the
development of ideas, and access to software and techniques employed
in this research work. This interaction made it possible to access
topical technical documents not widely available otherwise.

3.3 Case study

The focus of the research is specialised data management systems for
Earth observation from space. One example in this field dwarfs all
others: the Earth Observing System Data and Information System
(EOSDIS). EOSDIS has been under development for the U.S. National
Space and Aeronautics Administration (NASA) for the past decade, at
a cost of about $US 1.6 billion. Most of the significant global trends in
information science, in technological, organisational and policy
dimensions, are reflected in the history and evolution of EOSDIS. A
detailed case study of this system was undertaken for several reasons.
The first objective was to better understand the impact upon actual
information systems of the trends determined in the literature survey.
Conversely, the EOSDIS program was so large, complex and long-
lasting as to influence the creation of new infrastructure, techniques,
philosophies and policies which may then became de facto and long-

126
lasting standards. An examination of these issues was warranted in order to better anticipate flow-on effects on other information systems. The third reason was that a component of EOSDIS - known as its Information Management System (IMS) - was an obvious candidate for adoption by CSIRO to help handling its own Earth observation data holdings. Closer analysis of the features (and possible deficiencies) of EOSDIS was a prerequisite for this prospective application.

3.4 Examining the role of information in CSIRO

In this research, I am evaluating the “fitness of purpose” of new information systems. Clearly a number of factors need to be considered. First there are the technical merits of the information system itself - its ability to manage and retrieve information. This “data-centric” aspect was examined in the literature survey and the case study. The preliminary stages of the research highlighted several other pertinent issues. The first was whether the nature of the host organisation - its culture and mission - suited the information “solution” represented by the technology in question. The second issue was whether the technology could help meet the specific day to day
information needs and work patterns of the intended user group. I explored each of these factors.

3.4.1 The information environment of CSIRO

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) is a respected and aged Australian institution which, as its title indicates, conducts scientific research. CSIRO is not a common research subject, despite its perhaps unique place in Australian society. In this component of the research, I investigated the role of information in this organisation. I used manuals, reports, web sites, personnel statistics, and internal communications from senior management, to better understand what the organisation sees as its main objectives and products. I also used these sources to consider the priority CSIRO assigns to information management and the profile of those entrusted by the organisation to perform various information tasks.

I supplemented this survey of internal literature with interviews of three information professionals - the organisation’s information technology manager, and the librarian and library technician in one of
CSIRO' research Divisions. In addition, I took part in several CSIRO internal seminars devoted to Earth observation research data or to information technology.

The theme of CSIRO's information culture was also explored through the survey of the user group, described in the next section.

3.4.2 Consultation with prospective users

The research was discussed with colleagues in the CSIRO Earth Observation Centre both before and during its course. Science planning workshops held by the Centre (in Hobart on 29-30 August 1996 and in Canberra on 29 July 1997) gave valuable opportunities for obtaining feedback on the research method and scope. The continued interaction with the survey group helped to determine attitudes of CSIRO researchers in relation to information management. In particular, the Workshops helped to establish the scientists' views on the relative importance of policy issues (such as the role of CSIRO in delivering spatial data to other agencies or to the public); on pros and cons of devoting resources to preserve either raw or processed data; and on whether information systems should be controlled by users.
rather than by information specialists (Graetz, 1996a, 1996b; Kingwell, 1996).

The present research formed a sub-project, for which I was solely responsible, within an overall data management project (Wilson, Kingwell, & Campbell, un pub.) proposed in September 1996 and subsequently accepted following internal peer review in the Earth Observation Centre.

I presented preliminary research results at a CSIRO workshop which examined mass data storage technology of the type frequently employed overseas in Earth observation data management systems. At this workshop, in early 1997, I made a presentation (Kingwell, 1997) on international developments in spatial information systems. This workshop was attended by about 30 data and information technology managers and it presented a valuable opportunity to better understand current information trends and initiatives in CSIRO.
3.4.3 Consultation with allied users

The Australian Surveying and Land Information Group (AUSLIG) has over the past decade led numerous efforts directed at developing national standards, policies and infrastructure in support of the productive use of spatial information in the national economy. Specifically, AUSLIG established and chaired the Commonwealth Spatial Data Committee (CSDC), devoted to building consensus amongst Commonwealth Departments and agencies in relation to the management and exchange of geographically-referenced data. This approach was needed to generate a consistent Commonwealth view, in relation to similar considerations involving all three levels of government, and the private sector, through the Australian and New Zealand Land Information Council (Baker & Finney, 1995).

I took part in many of these discussions and processes over the period 1991-1997, especially in the CSDC Coordinating sub-Committee and its Catalogue sub-group. Some of the issues addressed in this research, such as the desirability of adopting and promoting national standards for spatial metadata, arose in the context of considering an appropriate response by CSIRO to these national level initiatives. The practical
benefits (or otherwise) of this kind of policy are not always clear, nor is it easy, in a decentralised and heterogeneous body like CSIRO, to arrive at an organisational consensus. Feedback from CSIRO researchers on these issues, obtained through survey results and interviews during the study, may assist the development of appropriate policies in the organisation.

During 1995 and 1996, AUSLIG, the Bureau of Meteorology, CSIRO and the Australian Space Office studied the possibility of developing an indigenous national on-line data and information system for managing and supplying Earth observation products. This ultimately unsuccessful initiative was called AEON, the Australian Earth Observation Network (1995). By joining the working group carrying out this investigation, I developed a stronger appreciation of the complexity and expense of such systems, and of the desirability of more closely examining systems already developed overseas for similar purposes.
3.5 Selecting the prototypes

Two things became clear during the initial consultation stage (Section 3.4). Firstly, CSIRO scientists working in the Earth observation field perceived a need for more systematic information management. Secondly, while there was some (imperfect) awareness of the extent of international developments on this front, there was insufficient familiarity with the technical performance and availability of existing systems to enable CSIRO researchers to make an informed choice between either adopting one (or more) of them for their own use, or alternatively to develop analogous in-house systems.

For those reasons, I decided to select representative and potentially suitable pre-existing information systems of several levels of utility and complexity, and to install these as working prototypes in order to obtain feedback from CSIRO users about their suitability. There was a pragmatic reason for selecting and installing working prototype information systems prior to a full user needs analysis: this was based on the recognition that in the short term, resources were unlikely to be found which would permit the design of an information system customised to CSIRO’s precise needs. This view was reinforced by the
failure, in 1996, of the AEON initiative which attempted to develop an Australian data management system in Earth observation.

Prototypes for evaluation were selected on the following criteria:

1. technical feasibility (fit between hardware and software requirements and hardware or software already in use at CSIRO Earth Observation Centre);

2. price (had to be zero or low, as the project had limited resources for this work);

3. ready availability in the study period;

4. probability that system will be maintained or upgraded by its developer for at least several years; and

5. compatibility of system with those in use by research collaborators of CSIRO.

On these criteria, I selected for further evaluation the CEOS International Directory Network (IDN); the NASA Information Management System (IMS); and CILS, the CEOS Information Location System (Kingwell, un pub.). This selection gave a nice
symmetry, consisting of an entry-level information locator (CILS) which also provided an outreach or extension service; an intermediate level and reasonably well-populated directory (IDN), giving information about data holdings in participating agencies' data centres world-wide; and a comprehensive (and evolving) data management system capable of being used for a range of data tasks, including inventory control (IMS).

The installation and use of IMS was formalised in early 1997 through a Memorandum of Understanding between CSIRO and NASA (refer Appendix C).

3.6 Installing and testing the prototypes

Two of the three prototypes were tested during a demonstration of high-performance computing and distributed archiving, before an audience including media and about 120 senior representatives of global space agencies and research bodies at the Committee on Earth Observation Satellites Plenary in Canberra during November 1996. I conceived this demonstration (Kingwell, 1995b) as an illustration of the functions which could be achieved through a regional or global
Earth observation information network, as proposed by the United Nation's Economic and Social Commission for Asia and the Pacific (Kingwell et al., 1995). The demonstration relied upon the exchange of metadata and both real-time and archived data via the information systems of several collaborating nations and agencies in India, Japan, China and Australia. For this purpose, it was necessary to choose information management system and protocols that were interoperable with the respective national or local systems.

Resources - including the goodwill of international space agencies - were provided for the pilot demonstration. These in-kind contributions of software and related technologies much reduced the cost of the subsequent phases of the research.

Two of the three data management systems used in the research (IDN and IMS) were installed by CSIRO Earth Observation Centre staff in preparation for the CEOS demonstration, and were maintained for another 24 months for the evaluation. The software installation and data population was by automatic file transfer, coordinated with the U.S. National Aeronautics and Space Administration, Goddard Space
Flight Center. This arrangement was brokered via the CEOS Working Group on Information Systems and Services and ensured that the database contents in the versions installed in Canberra remained up to date mirrors of the U.S. sites. However, staff time resources did not permit upgrading the Canberra operating versions of the IMS software as these evolved at Goddard.

The third operating system, CILS, was installed in Canberra in August 1997 with the assistance of the German space agencies DARA and DLR.

All three information systems were accessible to CSIRO researchers (and to the public) through an Internet server isolated (firewalled) from internal programs and services.

3.7 Evaluating the suitability of prototypes

3.7.1 Target group selection

One of the acknowledged difficulties in carrying out surveys lies in locating the target group of possible respondents (Wurzburger, 1987).
As noted by Moore (1983), a questionnaire survey is useful, but this has more value when used as a preliminary step for smaller scale interviews. To obtain a detailed perspective of client’s needs, and their views in terms of the centralised and the federated models being tested, I surveyed 76 CSIRO scientists through questionnaire, obtaining 32 replies. Fourteen of the respondents (and one additional scientist who joined the group after the questionnaire had been completed) expressed interest in further discussion on this topic. I interviewed each of these fifteen scientists, either face to face or by telephone if a meeting was not possible.

The 76 subjects comprising the target group (those contacted for the initial questionnaire) represented the potential pool of users of Earth observation data systems within CSIRO. Contact details for the group were obtained courtesy of the program leader of the CSIRO Earth Observation Centre, who used the list to interact with CSIRO researchers whose work involved Earth observation. These researchers worked primarily in small groups, based in a total of twenty nine different laboratories administered by a total of sixteen different business units (Divisions or major research groups) in CSIRO.
I had known most of the group for about a decade, through their interest in remote sensing. The Earth Observation Centre program leader has extensive experience (also more than a decade) with this group. We believed it likely that the contact list included all current CSIRO scientists with a working interest in Earth observation data.

Contact with the study group was maintained through the research period, by e-mail, telephone, letter, and at seminars and reviews. Contacts were logged in an e-mail directory and by hard copy in notebook journals. In the initial questionnaire, researchers were given the option of declining further communication about this study. Eight scientists elected this option.

The e-mail address list maintained by the Science Program leader of the CSIRO Earth Observation Centre is an example of a scientific "invisible college" communication network. It comprises individuals, within a single organisation but in different operating units, who acknowledge their topical common interest in Earth observation. Patterns of communication among such groups were explored in the
literature survey (Chapter 2), and were further explored during interviews of the survey group.

3.7.2 Questionnaire

A Questionnaire was used in this research for several reasons:

• to identify specific individuals amongst the client group who were prepared to discuss the topic in more detail;

• to encourage the provision of information from a larger group than could be feasibly interviewed in depth; and

• to help identify issues requiring further examination.

Preparation for the data-gathering phase of the research included studying questionnaire methods. A number of references proved to be valuable; these included Burton (1990); Blackmore (un pub.); Moore (1983); and Wurzburger (1987).

Prior to distribution, the questionnaire was tested in a trial situation, with two participants from the survey group. As a result of the trial, the questionnaire was slightly modified for improved clarity, prior to distribution by mail.
The questionnaire (see Appendix E) was accompanied with a stamped self-addressed envelope and a covering letter (Appendix D) giving background information, describing the purpose of the research and explaining briefly how the results of the study could eventually benefit respondents.

The default reply was anonymous. However, those who wished to discuss the topic in more depth were invited to identify themselves. This sub-set of fourteen respondents, and an additional scientist who had only recently joined the contact group, became interviewees in the next stage, discussed below. All respondents were assured that information they provided would be treated confidentially.

The Questionnaire took about 20 minutes to complete. After assessing the replies, I discussed the responses at seminars of the Earth Observation Centre, gaining additional feedback on the study prior to interviewing those who had agreed to further discussion.
The main issues explored through the Questionnaire were frequency of use of Earth observation information systems; views on the degree to which CSIRO ought to be concerned with such systems; their most sought-after functions; and respondent’s views on priorities in Earth observation data management in CSIRO.

3.7.3 Interviews

The interviews were arranged by e-mail or telephone, and were taped. Taping was not completely successful for two of the interviews. In one case this was due to recorder battery failure; on the other occasion, only the interviewers’ part of the dialogue was recorded because a speaker phone was not available. In these two cases, especially detailed notes were kept and the interview pace was slowed to facilitate note taking.

About one or two days prior to interviews, a copy of the questions (Appendix F) was supplied to the interviewee. Permission for taping and for use of the information in research was also obtained at the commencement of the interview. When feasible, the interviews were carried out in person. Available travel funds were not sufficient to
allow face to face interviews with three scientists in remote locations.

In three other cases, interviewees were on leave or were travelling at the time I visited the relevant location to interview other researchers. These six interviews were carried out by telephone.

Prior to commencing the series of interviews, a trial interview was used to test the format and content, with a focus of the clarity of questions. The subject for the trial interview was a volunteer from the survey group. The trial led to several modifications to the final interview.

Each interview took about 40 minutes to 75 minutes, the time variation being caused primarily by the length of the replies, and whether these led to ancillary questions.

The interview consisted of both open and closed questions. Closed questions (such as "In the past 12 months have you used or explored the CEOS International Directory Network, IDN?") were used to obtain quantified responses. Depending on the replies to the closed questions, some questions were omitted or modified, as appropriate.
Open questions (such as “In your experience, does CSIRO encourage the development of information systems and the delivery of information?”) were designed to elicit qualitative information, personal views, and insight into the organisational culture. Replies to these questions frequently led to supplementary questions.

3.8 Handling and evaluating the data

Case study data were obtained by literature survey and by attendance at technical workshops and peer reviews during the period September 1993 to August 1998. Citation details and notes from this material were maintained as a Reference Manager™ database, which was also used to generate the reference list.

The data on the user environment in CSIRO was obtained from literature review, questionnaire/survey, technical workshops, site visits and interviews. The original material - questionnaire returns and interview tapes - were kept in native format (that is, hard copy and magnetic tape, respectively). Details of the work journals and interview tapes are given in Appendix G.
Copies of e-mails were retained. Copies of correspondence, and notes taken during interviews, were kept in workbook journals.

Key points made by the interview subjects were mapped onto word processor tables, so I could more readily compare responses across the survey group. The responses to the questions were grouped into loose categories, according to the aspect of the study most directly addressed:

1. frequency of use;
2. ease of use;
3. utility of systems; and
4. fit of systems to CSIRO information culture.

Relevant direct quotes from interviewees were cross-referenced to this list of issues and to the interview question which prompted the comment. Quotes were transcribed from the tapes, except in the two instances when taping failed. In those two cases, particularly detailed notes were taken and were used as the primary data.
3.9 Aims

This research has several objectives. The first is to discover whether information systems used elsewhere for managing Earth observation data were also suitable for use in CSIRO. To reach this objective, I reviewed a number of possible systems, selecting three examples which demonstrate a range of capabilities. These three systems were installed on a trial basis, and I sought comments on their suitability from a group of CSIRO scientists whose work significantly involved Earth observation data.

In evaluating the fitness of these systems for CSIRO purposes, I examined the features of the technology and the role of information within the organisation that employed the potential users. One reason for doing this was that each of the systems, if used operationally, could be implemented in at least two distinct fashions. One style is that of a centralised information system, with a single management entity responsible for determining the database content, the delivery policies, and other operational aspects. In the other approach, a number of entities determine these matters in a negotiated or federated manner. By exploring how information is regarded in CSIRO, I hoped to
determine which of these approaches would be more appropriate in
the event that one or other of the selected information management
technologies was implemented on an operational basis.

A second major ambition was to critically review the history of the
information system chosen by NASA to manage Earth observation
information from its large scale space program. This system, known as
EOSDIS, is one of the largest information systems ever developed. Of
the reasons for examining EOSDIS, among the most important for
this research are:

1. EOSDIS, or an element of it, is a potential candidate for future
   operational use in CSIRO; and

2. as a long-lived and complex project, EOSDIS illustrates significant
devvelopments in the information science world. These developments
need to be recognised if CSIRO information systems are to be
efficiently implemented and future developments successfully
anticipated.

The detailed results of these investigations follow in Chapter 4, with
the conclusions of the study appearing in Chapter 5.
4. RESULTS

"You can't have a trickle, because it's going to hold up the flood".

4.1 Introduction

In this chapter, I detail results of research into user acceptance of Earth observation data systems in the context of organisational culture and user requirements in the Australian research agency CSIRO. Technical characteristics of a range of data management systems designed for use Earth observation applications were given in Chapter 2, where I also examined key concepts in information science which are now being applied to Earth observation data.

One of these key concepts was "Federation", a decentralised style of decision making in which substantial autonomy is exhibited by the component or collaborating parts of an information or political system. When applied to information management, "Federation" usually connotes a high degree of negotiation, leading to the adoption of some common practices, protocols and standards but with substantial local control over data products remaining in the hands of the data producers. This chapter examines the application of this
concept in relation to one of the world's most ambitious data
management technologies, the Earth Observing System Data and
Information System (EOSDIS) of the United States' National
Aeronautics and Space Administration, NASA. EOSDIS has evolved
over the past decade from a centralised to a federated system
principally because of the concerns and actions of the U.S. scientific
community. This research pays special attention to this development,
because of the similarity of the EOSDIS user community to the subject
group in CSIRO, and also because the EOSDIS data management
software IMS (Information Management System) is a logical candidate
for potential routine use by CSIRO.

The organisational culture of CSIRO, especially as it relates to the
management of research data, is also explored through critical analysis
of internal literature and selective interviews of key personnel,
including the head of corporate information technology services.

Proceeding from this overview of information use in the organisation,
the study moves to the particular example of Earth observation
information systems. I describe the selection and installation of three
representative data management technologies. These "working prototypes" were web-based mirror sites of information systems used for operational purposes elsewhere in the world. They exhibited all the features of operational models with the exception that for this study, no attempt was made to populate their databases with locally-derived data.

This approach was partly a result of resource limitation, and partly a function of the objective of this research to evaluate, before long-term resource commitments were made, the suitability of specific data systems for use in CSIRO. Increasing the number of Australian datasets described in the information systems would be a prerequisite for their operational use, but it is clear from the experience of the lapsed AEON initiative that this step would be time-consuming and potentially expensive. The approach taken here is to first establish, through user survey, whether the framework of the information systems was suitable enough to warrant the effort in installing details or copies of local data. It is possible that the relative absence of Australian-sourced data discouraged responses to the questionnaire survey. Issues of devoting time and funds to adding CSIRO-held data
and metadata to operational information systems were examined during the survey and interviews.

Finally, the response to the three prototype information systems is explored through analysis of questionnaire returns and interviews results. The subjects were from the cadre of 76 CSIRO Earth observation scientists and data managers; thirty four of these responded to the questionnaire, of whom fourteen (plus another who joined after the questionnaire) were interviewed.

The research included prolonged consultation with CSIRO Earth observation scientists employed at around twenty eight laboratories throughout Australia. The action research process included seminars and pilot project/demonstrations of technology; an initial questionnaire survey; preliminary feedback to the cadre; and follow up interviews with those of the questionnaire respondents who agreed to do so. The foci of the initial questionnaire were elucidating attitudes of the survey group to a set of spatial data and Earth observation policy issues facing CSIRO; and establishing broad user requirements in relation to Earth observation data systems. The focus of the follow-
on interviews was the user response to the prototypes, from a technical perspective (did they do what the user required?) and from a cultural perspective (did these systems fit the norms of the organisation?).

4.2 Critical review of EOSDIS

4.2.1 NASA's Earth Observing System and its Information Management

NASA's Earth Observing System Data and Information System (EOSDIS) was conceived more than a decade years ago as a highly systematic, rapid-turn-around central service for delivering advanced information products from the Earth Observing System (EOS), or Mission to Planet Earth. Initially expected to account for $US 3.9 billion or about 23% of the EOS budget of $US 17 billion, EOSDIS was in part a response to criticism by the General Accounting Office that the U.S. space agencies NASA and NOAA had neglected their archives of raw data from planetary and Earth-exploration missions, often spending as little as 1-3% of mission or operating costs on data management (Office of Technology Assessment, 1994; General Accounting Office, 1990a, 1990b).
EOSDIS represents an almost heroic attempt to re-balance space program expenditure by greater emphasis on the "ground sector" from which data and data products flow to users of space information systems. This principle has contemporary advocates, such as Gabrynowicz who argues that space and ground segments are interdependent and comprise the total information system, but that the ground segment has traditionally been starved of resources even though it is the element most important to the end-user (Gabrynowicz, 1997).

Although primarily designed for the scientists who specified the performance of the EOS sensors and who conduct the research programs dependent upon the resultant data products, EOSDIS has been criticised by elements of the U.S. scientific community almost from its beginning. This criticism culminated in a proposal (National Research Council, 1995a) that the "conceptual design" of EOSDIS be fundamentally changed in order to reduce the cost of EOSDIS and broaden its organisational base beyond NASA. In the NRC model, which it described as a federation, NASA would retain responsibility for the operations of the EOS spacecraft; for data capture; and for basic
processing. A competitively selected collection of entities from
government, academia and the private sector would be responsible for
generating advanced or value added products (VAP), and for providing
user services.

NASA responded to the NRC recommendation, though on a small
scale and on an experimental basis. In 1997 it announced the selection
of the first 24 competitively selected data product providers, which
NASA termed “Earth Science Information Partners” (King, 1997).
Through the combined effects of conceptual re-design, development
delays, and budget cuts and despite the continuing delays to the launch
of the first EOS spacecraft, it is clear that EOSDIS will now have
much less capability, at the time the first EOS data are transmitted,
than was originally expected (Berger, 1998). Further, the functions
originally planned to be conducted by an information system wholly
under NASA’s control will now be spread through a heterogeneous
collection of entities, and may change more rapidly and in a less
coherent manner than was foreseen when EOSDIS was first proposed
Budget reductions to EOSDIS during the last decade have occurred at a proportionately greater rate than decreases to the EOS space segment (Office of Technology Assessment, 1994). Nevertheless at an estimated cost of about US$1.6 billion between the years 1991 and 2000, EOSDIS is the world's most expensive and complex civil information system (Jaworski, 1993). In the sections below, I examine the history of EOSDIS in more detail, with particular emphasis on the evolution of the conceptual design towards the federated structure proposed by the scientific community that was earlier expected to be the greatest beneficiary of the EOSDIS central service. This evolution may serve as a model for the eventual implementation of Earth observation information systems and services in an organisation such as CSIRO, which has a similar clientele.

4.2.2 Background to EOSDIS

The U.S. National Aeronautics and Space Administration (NASA), the world's largest space agency, is currently carrying out as one of its major tasks the "Mission to Planet Earth", MTPE, which uses space-, ground-, and aircraft-based quantitative measurements to increase

---

1 MTPE was renamed "Earth Science Enterprise" in early 1998 (King, 1998).
scientific understanding of the global climate (NASA, 1995). MTPE is said to be the largest scientific experiment in the world (Vetter et al., 1995), and has as its centrepiece the Earth Observing System (EOS), a series of complex Earth observing satellites and their attendant data systems. The first satellite in the EOS series is now scheduled for launch in 1999, although related satellites taking part in the ten-billion dollar experiment were in operation before that date. During Phase 1 of EOS (1990-1998), a comprehensive infrastructure was developed in order to lay the basis for the archiving, retrieval and exploitation of information arising in the course of the decades-long program (NASA, 1998a). In particular, NASA is overseeing the establishment of the EOS Data and Information System, EOSDIS, which is intended to manage all data—whether originating from aircraft, land or space measurements—arising in Mission to Planet Earth. When EOS is fully deployed, its measurement systems are expected to generate more than one terabyte ($10^{12}$) of data per day (Vetter et al., 1995). EOS is intended to last more than two decades, so the data ingest system must cope with a total volume of approximately $\{20 \times 365 \times 10^{12}\}$ bytes or about $10^{16}$ bytes (ten petabytes). Devising a single data management system to safeguard these data and make them accessible for use across a variety
of scientific disciplines, for a period of at least 10-15 years, represents a major development challenge, especially in software technology (Jaworski, 1993).

When the program was first formally proposed for funding, EOS was seen as:

"...necessary to develop a comprehensive understanding of the way the Earth functions as a natural system. This includes the interactions of the atmosphere, oceans, cryosphere, biosphere, and solid Earth, particularly as they are manifested in the flow of energy through the Earth system, the cycling of water and biogeochemicals, and the recycling of the Earth's crust driven by the energy of the interior of the Earth"


Of importance to later events was the scale and ambition of NASA's goal, and the almost pantheistic terminology employed to describe it. The depiction of Earth as a living "system" seems influenced by Lovelock's "Gaia" hypothesis, in which the Earth is depicted as a self-regulating super-entity, almost like an organism that can adjust to perturbations in its constituent systems (Lovelock, 1991). Lovelock
was a former NASA scientist, and his theories possibly influenced the approach to EOS. The EOS objective of describing the totality of biological, chemical and physical processes on Earth as if they were components of a single entity differs radically from the more orthodox reductionist and incremental discipline-based approach to scientific research.

4.2.3 Initial architecture

EOSDIS was initially planned as a machine, funded and controlled by NASA, in which acquisition and primary processing of Earth observation satellite data will be performed centrally, or at a small number of locations. Within a few hours of receiving the satellite data transmitted to ground receiving facilities at rates of hundreds of megabytes per second, specified data products were to be available to research investigators around the world via the "EOSDIS External Network", comprising NASA Science Internet and connections to the U.S. National Science Foundation Internet private networks (Asrar & Dokken, 1993).
It is fairly clear that the initial design approach adopted by NASA and its contractors focused more on the characteristics of the data rather than the habits or requirements of users. Harberts (1993), for example, describes a systems-design approach in which the primary driver is the data flow and volume.

The space segment to "feed" information to EOSDIS was initially conceived as comprising four very large Polar Orbiting Platforms (POPs). One was to be built by the European Space Agency ESA; one by the Science and Technology Agency of Japan, and two more, the EOS (U.S.-funded) space segment, by NASA and NOAA. EOSDIS was to archive and manage data from the numerous instruments on these platforms, and was also to issue commands for operating the spacecraft (Computer Technology Associates, 1988).

The NASA-led EOSDIS project completed its conceptual design at Goddard Space Flight Center in October 1986. At that time, the first EOS satellite, the U.S. POP, was scheduled for launch in 1995. The equivalent European POP, later re-named "Envisat" (Dornier Deutsche Aerospace et al., 1994) and the U.S. POP-2 were to be launched 1997,
and the Japanese POP in 1998. The satellite fleet for EOS has since been redesigned several times, notably in 1992 when a "mixed fleet" concept of large and medium satellites was adopted (NASA, 1993). These changes did not alter the essential character of EOS as a long-term and costly satellite series carrying highly sophisticated instruments for multi-disciplinary based studies of the global environment.

EOS was proposed and successfully presented for funding at the U.S. Congress in 1990, at a time of intense international concern at the threats science and technology appeared to pose to the environment. In the mid 80s, the world first grew familiar with the terms "Greenhouse gas" and "Ozone hole". It seemed to many at the time that the key to allaying these concerns was more research, or at least, more measurement ("...more data, more data, from Poles to Equator....").

With grand and poorly defined goals - thoroughly in keeping with space program practice at the time - it is not hard to understand why EOS has barely commenced a decade, billions of dollars and
uncountable scientist-years later. Similarly, Space Station "Freedom", announced by President Reagan at about the same time (1984), has not yet been built and will not enter service for at least four years. Like EOS and EOSDIS, the Space Station has been redesigned many times (Logsdon, 1998). In the process, its cost has grown by many times the initial price of $US 8 billion, even though it is now to be much smaller and to have much less power than originally expected (Space News, 1998c).

The Polar Orbiting Platforms and EOS were a spin-off from the Space Station - they were originally called "polar orbiting space station elements" (Richards, Kingwell, & O'Sullivan, 1987). Some space scientists opposed to the Space Station were inclined to support the Polar Platforms, possibly reasoning that given the President had promised billions of dollars for confronting what he termed "the Evil Empire", at least some of the proceeds should be spent on something halfway useful.

The initial budget for EOS was $US 17 billion, decreased following a review instigated by the Office of Management and Budget in 1991 to
$US 11 billion through to 2000, and reduced again the following year to $US 8 billion. The EOSDIS budget was reduced from $US 3.9 billion (1988) to $US 2.141 billion (1991), and cut again to about $US 1.6 billion in 1992. The planned set of data products from a single EOS spacecraft, EOS-AM 1, was reduced from 600 to 160 during the same period (Office of Technology Assessment, 1994, p.71).

Commencing in 1988, NASA established twenty nine “Interdisciplinary Science Teams”, in conjunction with research organisations round the world, with responsibility for analysing information from EOS and, importantly, to anticipate problems in the exploitation of this data wealth and advise on ways to alleviate these difficulties (NASA, 1995; Asrar and Ramapriyan, 1995). One of the responsibilities of these scientific teams was to define the performance targets of EOSDIS (Barkstrom, 1994; Schwall & Andrews, 1993). An interpretation of the delays and cost in developing EOSDIS is that the specification of system requirements by end users gave little or no weight to what was practical or cost-effective in terms of processing EOS sensor data into desirable scientific products (Glover, 1994, 1997).
NASA's concept for EOSDIS was that it would be a "one stop shop" for Earth observation data. NASA's scientific clients on the EOS science teams would specify what products they wanted, and would nominate and test the mathematical processes or algorithms to be applied to the raw EOS data (Jaworski, 1993). NASA's network of Distributed Active Archive Centers would implement the algorithms and would distribute the resultant "EOS Standard Products" to the science teams and to other users (Schier & Way, 1990). From time to time, further research might uncover more appropriate algorithms: these would be selected and documented by NASA's EOS science teams and then implemented at the DAACs, with the option that already archived EOS data would be re-processed according to the new peer-selected algorithms. This proposed iterative approach to EOS dataset management was essentially the same as that developed by NASA and NOAA for the "Pathfinder" projects which were supported by NASA as part of Phase 1 of EOSDIS (Booth & Maiden, 1993).

NASA estimated there would be 10 000 scientific users of EOS data: at the initially expected cost of $US 3.9 billion for EOSDIS, this
represents an investment of some $390,000 per researcher in data system costs alone. In fact, EOSDIS and EOS were designed around the stated needs, established about ten years before launch, of the 29 Principal Investigators of NASA's Interdisciplinary Science teams (Schier & Way, 1990): on this basis, the intended investment was more like $US 134 million per investigator. This investment appears to have been made with little analysis of alternative data sources, or of the relative effectiveness of different approaches to Earth system science research.

I now examine how the reaction of the scientific community to EOSDIS changed as the system developed.

4.2.4 Scientific community response to EOSDIS

EOSDIS was originally visualised as a public enterprise, fully serviced by government agencies. The academic community soon demanded to participate as service providers, claiming they could do at least some of the work faster, more cheaply and better. I was surprised at the virulence and hubris with which this viewpoint was expressed at a
NASA-sponsored review of EOSDIS held in Goddard in September 1993 (Kingwell, 1993).

Some U.S. scientists - even some who were employed by or whose research was financed by NASA - viewed EOSDIS as a top-down, datacentric and centralised information service which was too large to be efficient. One alternative proposal was the concept of “shared nothing”, under which a collection of services would be provided by about 40 autonomous agencies, each operating their respective database(s) containing both metadata and end products (De Witt & Naughton, 1994). This proposal was essentially a client-server architecture in which a user could access multiple databases via queries posted through the Internet. In this particular arrangement, each database operator retains full control of transactions, while there is little or no systemic interaction between database providers.

Other criticisms of EOSDIS were based on cost grounds. While expressing satisfaction that EOSDIS contractor Hughes Applied Information Systems had listened to and acted upon previous
criticisms by scientists, the EOSDIS external review panel in mid 1994 observed:

- a lack of awareness by the contractor of important developments in commercial software;

- that the proposed centralised architecture appeared to be unaffordable; and

- neither the contractor nor the NASA project supervisors appeared to show any ingenuity in reducing costs.

(Glover, 1994).

The same review concluded that NASA preferentially funded researchers who agreed to define EOS "standard products", for which the demand was at best uncertain and probably minimal, and for which EOSDIS was assumed to be the host and distributor. This process resulted in an over-specification, and therefore an excessive expense, for EOSDIS (ibid., pp.22-3).

Scientists worried that as well as inflating the cost of EOSDIS, specifying rigid "standard" products in a research environment may
stifle creativity and scientific advancement, especially because one would expect many advances in thinking, and consequently revisions to data and information products, in the 15 to 20 year operational lifetime of EOS (Glover, 1997, pp.47-8).

4.2.5 Australian scientific criticism of EOSDIS

One of NASA's original EOS Interdisciplinary Principal Investigators and then head of CSIRO's Office of Space Science and Applications presciently argued in 1992 that formulating the scientific questions to be asked was more significant, in terms of understanding anthropogenic change to the Earth's biosphere, than simply enumerating a set of measurements to be taken from space:
"When major satellite-based Earth Observing Systems were proposed it was assumed that it would not be difficult to define the correct questions and determine what should be monitored. It is not that easy. As Deep Thought concluded in 'The Hitchhiker's Guide to the Galaxy', if the question is not well formulated, the answer to Life, the Universe and Everything is merely 42. It is now being questioned whether or not it will be possible to detect global change given the 'state of the art' in global science and remote sensing in the 1990s and given that change is normal in the global system."

(Harris, 1992, p.275).

Harris pointed out that the world is much more complex than the "box models" shown in space agency brochures, and questioned whether the conceptual understanding of global change was well developed enough to efficiently utilise a deluge of new space data. Harris noted the data volume problems of EOSDIS, being a quantum leap from existing standards that even in 1992 had exceeded capacity to process and utilise information effectively. For example, EOS would generate about 1-2 terabytes of data per day, over a 15-20 year period, or some 11,000 terabytes. He pointed out that previous satellite sensors, such as the Coastal Zone Color Scanner, produced data sets in
the range 1-5 terabytes and those archives were not fully processed after a decade (ibid., p.278).

4.2.6 Structural criticism of EOSDIS

A major review of EOS by the peak U.S. science body NRC (National Research Council, 1995a) severely criticised the "conceptual model" for EOSDIS and called for its re-design as a "Federated" concept in which universities, public agencies, and the private sector would contend for selection as value-adding service providers.

NASA's contractor, Hughes Applied Information Systems, had by this time already designed the EOSDIS Core System (ECS), the flight operations, science data processing and system management heart of EOSDIS. ECS was to be a self-contained system to operate the EOS satellites according to schedules determined by scientific investigators; receive raw (Level 0) sensor signals; derive value added products (Level 2 and higher) according to procedures selected and tested by the principal investigators; distribute products direct to users; and archive the Level 1 data (sensor data plus time, geo-location and calibration readings).
Instead, the NRC review called for value-added product generation to be separated from the satellite operations, data reception, and Level 1 archiving. Level 2 and higher processed products would be generated and distributed by a "federation" of competitively-selected organisations from government, universities and the private sector. The federation elements would also provide user services, including access via Internet or physical media (Figure 1).
FIG 1 Proposed Federated EOSDIS (after NRC, 1995a).
About three-quarters of the review team were academics engaged in
global change research in the Earth sciences. Few if any of the review
team were computer or network specialists. The review team’s focus
was to reduce the cost of EOS, and to expand the role of the private
sector (ibid., pp.vi-vii). However, they were reluctant to recommend
changes to the space segment, fearing that any further re-design would
cause intolerable delays in the launch dates for the EOS fleet; the main
recommendations of the review thus focussed on EOSDIS.

The review group argued that EOSDIS should become more accessible
and open, allowing a higher degree of participation by data producers
outside NASA. They claimed that this step would result in substantial
cost reduction, through diminishing the engineering and management
“superstructure” needed to maintain a large suite of centrally-
controlled computing centres - the NASA data warehouses or DAACs
(ibid., p.78).

It is arguable that the U.S. scientific community, through the NRC,
was calling for “outsourcing” of parts of EOSDIS, rather than the
creation of a true "federation" of autonomous, voluntarily collaborating units in the sense the term is used in information science (for example, Thuraisingham, 1997) or in the corporate and political arenas (Handy, 1992). The terminology used by the NRC ("bid", "competitive", "contractual obligation", for example) implied that the NRC expected NASA to fund the successful entrants to the federation. Irrespective of semantics, it was clear that organised scientific opinion in the USA doubted the wisdom of letting a single organisation manage EOSDIS.

NASA responded to the NRC report by funding a number of public and private sector organisations selected by the space agency as "Earth Science Information Partners" or ESIPs. NASA however retained the bulk of EOSDIS information distribution and processing in its own data centres (Maiden, 1998). The first 24 competitively selected ESIPs were announced by NASA in 1997 (King, 1997; NASA, 1997). NASA defined three classes of federation partner (NASA, 1997): Type 1 (those providing services currently supplied by NASA facilities); Type 2 (those providing alternative "innovative and creative" scientific
information for research users) and Type 3 (those whose products address a broader user community).

NASA selected twelve Type 2 and twelve Type 3 ESIPs, in what it described as a pilot project to build a “Working Prototype Federation”. By the end of the pilot scheme, to cost about $US 50 million over 3-5 years, the Type 3 ESIPs (only) were expected to be self-funding (NASA, 1997). Of the twelve research-oriented value-added producers selected by NASA (Type 2 ESIP), institutions that had participated in the 1995 NRC review were well represented, capturing half the contracts.

Given the fundamental nature of the change proposed, the NRC review of EOSDIS in 1995 expressed a remarkably sanguine belief that implementing the new approach would present few difficulties, provided the underlying assumptions were sound:

“If we create and commit ourselves to the right model, all of the details related to the design and technology will fall into place readily”

(NRC, 1995a; p.77).
In practice, there are a number of risks associated with federated approaches to information management. For example, in federated structures, three schemata are simultaneously necessary: namely import, export and negotiations schema. By contrast, a "logically centralised" database architecture uses a global schema, applied through a "global manager", to address queries to all the constituent databases (Papazoglou, 1991). Because retention of autonomy is one of the main features of true federations, in practice a high degree of negotiation - a management overhead - is required to prevent the elements of the federation losing, or not developing, coherence. For example, the ability of a user to interrogate the federated databases and to obtain services in a consistent manner may require constant policing. Maiden (1998) and the NRC (1995a) noted risks and difficulties specific to a federated approach to EOSDIS, in particular:

- reconciling collaborative intent with a competitive environment;
- maintaining standards for metadata;
- expanding governance to include additional stakeholders;
- ensuring the interests of pre-existing EOSDIS partners such as non-U.S. space agencies;
- synchronising with the launch schedule for EOS satellites; and
• reliance on the ability of the Internet to supply, at affordable rates, sufficient bandwidth to distribute EOS products.

There may be additional risks in the proposed federated approach. For example, if a Type 2 ESIP loses its contract, what happens to the value-added products it has developed? How is the continuity of EOS data to be guaranteed over its fifteen-year lifetime as membership of the federation changes? Will private sector ESIPs come to expect or require a monopoly on EOS data, or in a particular market? Will effort be concentrated on discipline-specific information products, at the expense of the interdisciplinary studies that EOS was created to service?

Later in this chapter I explore the appropriateness, in the Australian research agency CSIRO, of either a logically-centralised approach, such as that initially adopted by NASA for EOSDIS, or alternatively of a logically-decentralised or federated approach as advocated by the U.S. scientific community. First, however, I examine the nature of CSIRO and of its use of information.
4.3 Information management: its place in CSIRO

4.3.1 Strategic management of information

Strategic information resources management connotes the recognition of information as one of the essential raw materials for the success of a modern organisation. In this section, I examine through literature analysis and interview the extent to which Australia's largest scientific research agency exhibits a strategic approach to information management.

Alongside staff, finance and physical assets, information is now regarded by some management theorists as the "fourth resource" which is essential for survival in an increasingly competitive and changeable world (D. Best, 1996). The United States General Accounting Office (1994) analysed patterns of information management and the use of technology in 19 organisations which it considered leaders in information practices. That study - a contribution to the theme of "reinventing government" - identified a number of techniques, issues and strategies for strengthening use of information management and technology in order to improve delivery of goods or services. The principal element in the information approach of these
leading organisations was their ability to relate information practice to
the group's primary goal or mission. An important factor in achieving
this end was the existence of a champion in senior management, a
"Chief Information Officer" responsible for all aspects of information
flow (General Accounting Office, 1994).

Information can be categorised as extrinsic or intrinsic (D. Best, 1996,
pp. 9-11). In the first case, it plays a supporting role in an organisation:
payroll data are extrinsic to most organisations, but are intrinsic to a
company which supplies payroll management services to client
companies. Intrinsic information is the underlying object of the
business process. For CSIRO, Australia's principal research and
development organisation, information is intrinsic, because the
organisation's principle function is the creation and exploitation of
knowledge. However, contemporary publications, official statements,
and staff profiles in CSIRO indicate that information management is
regarded as an ancillary function in relation to the "real" work of
scientific research. This core belief is illustrated by the 1996 decision to
abolish the Information Management branch - which, among other
things, had been participating in an international research program on
knowledge management - and to merge the remnant with the Information Technology branch. According to a statement in the CSIRO annual report (CSIRO, 1997b, p.20), "....This resulted in a redistribution of funds to support research, and a smaller corporate group focusing on activities that support the library and information community across CSIRO". The theme of "redistributing" funds for research by downsizing service areas was frequently repeated by CSIRO managers in the 1990s, and was given as a principal rationale for the re-organisation which saw the number of Divisions reduced from 33 to 26 in the period 1995-97 (McIntosh, 1996). However, in a personal commentary on a staff survey he conducted under contract to CSIRO, Falls (1998) found that the mono-cultural and internal focus on research excellence was often at odds with the expectation of clients that CSIRO will deliver good service. The latter objective is secondary for CSIRO staff for whom the primary objective is the science itself.

CSIRO's primary corporate information focus in the past has been on extrinsic functions such as finance and personnel data, sometimes in combination with "work in progress" reports; collectively, these categories comprise "management information". Recent changes have
seen a greater emphasis on CSIRO’s corporate management of one form of intrinsic information, in the form of its Intellectual Property (IP) portfolio. This change was neither spontaneous nor a result of strategic analysis, but was rather the result of a drastic external stimulus: specifically, the loss of millions of dollars as a result of court decisions against CSIRO in the mid 1990s arising from inadequate internal information flow relating to business deals between CSIRO Divisions and private sector. At least two cases of litigation in relation to intellectual property were settled out of court in: the terms of settlement were not made public, although the scale can be judged from CSIRO financial reports showing provisions of $9.5 million and $2.012 million, respectively, for “legal settlements” in 1993/94 and 1994/95 (CSIRO, 1994a; CSIRO, 1995).

Much of the effort in the management of research data itself (as opposed to IP arising from research outcomes) has traditionally been handled at the unit, laboratory or project level. On face value, based on the strategically-oriented information management approaches noted by the General Accounting Office (1994), D. Best (1996), Orna (1990, 1996), and Webb (1996), there may be merit in greater corporate
emphasis on research data management, with possible organisation-wide efficiency dividends, greater access to CSIRO research results by external clients, and increased community appreciation of CSIRO's work. However, this study suggests that the organisation values of CSIRO are such that in the absence of severe external forcing it is unlikely that the organisation will devote the resources required to rectify deficiencies in its handling of research data, including its Earth observation information holdings.

I examine the nature of the organisation in more detail in the following section.

4.3.2 The business of CSIRO

The Commonwealth Scientific and Industrial Research Organisation is an independent statutory body established through the Science and Industry Research Act 1949 and succeeding the Council for Scientific and Industrial Research, created in 1926.

The organisation is managed by a Chief Executive (CE), appointed by the Governor-General following consultation between the relevant
Minister (currently, the Minister for Science and Technology) and the CSIRO Board, to which the Chief Executive reports (Parliament of Australia, 1949).

According to the Act, CSIRO's *primary* functions are:

- to carry out scientific research relevant to Australian industry, the community, national objectives, national or international responsibilities, or for any other purpose determined by the Minister; and
- to encourage or facilitate the application or utilization of research results.

In addition, CSIRO has explicit *secondary* responsibilities:

- to carry out services and make available facilities in relation to science;
- to liaise with other countries in scientific research matters;
- to train researchers;
- to award grants, fellowships and studentships relevant to the Organisation's research;
- to recognise, cooperate with and make grants to industrial research associations;
- to establish and promote the use of standards of measurement;
- to collect, interpret and disseminate scientific and technical information; and
- to publish scientific and technical reports.

(CSIRO, 1994b, p.1). I have added underlining to the secondary responsibilities which have explicit information management connotations.

The Organisation is also subject to Ministerial guidelines (CSIRO, 1997c) which modulate the functions described in the Act. The first guideline emphasises the priority of CSIRO’s activities:
- CSIRO’s main task will be the conduct of strategic and applied research in support of national economic, social and environmental objectives.

In recent years, this emphasis on the “core” function of research and development has led to a deliberate and explicit policy of redirecting organisational resources into scientific projects and from “support” areas such as management, communication, administration and library services. The Chief Executive, Dr Malcolm McIntosh, made this
approach quite clear when announcing recent re-organisation measures:

"A goal of the restructuring of CSIRO has always been to make our administration more efficient and hence free up more resources for research".

(McIntosh, 1996).

However, since the late 1980s, CSIRO has been required by government direction to attain about 30% of its total operating budget through means other than direct appropriation. A difficulty for the organisation here is that discretionary purchasers of research and development services often require more than "good science", frequently expecting high standards of legal, commercial, advertising and other information-related services. To the extent that performance in these areas are not rewarded commensurately with scientific performance, CSIRO may find it difficult to build professional teams comprising all the skills and disciplines needed to deliver marketable products derived from research and development.
4.3.3 Current issues in CSIRO information management

The executive body in CSIRO comprises the Chief Executive and four Deputy Chief Executives (DCE), all of whom are scientists. Each DCE has line responsibility for a collection of 6-11 CSIRO business units (research Divisions) known as an Alliance, and for one or more functional areas of Corporate support. The Information Management function is overseen by Dr Bob Frater, the DCE who chairs the “Information Technology, Infrastructure & Services Alliance” (Table 1). Although this implies that information management has a champion at the most senior level, this is only one of a large number of responsibilities of the Deputy Chief Executive concerned. In CSIRO, there is no real equivalent to the “Chief Information Officer”, an individual whose primary responsibility is for corporate information flow (General Accounting Office, 1994).
Table 1. Deputy Chief Executive responsibilities at 19 March 1996 (CSIRO, 1996).

<table>
<thead>
<tr>
<th>Deputy Chief Executive</th>
<th>Chair</th>
<th>Divisions and Corporate Support Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr Colin Adam</td>
<td>Minerals and Energy Alliance Manufacturing Alliance (Alternate)</td>
<td>Building, Construction &amp; Engineering, Coal &amp; Energy Technology, Exploration &amp; Mining, Materials Science &amp; Technology, Minerals, Petroleum Resources, Commercial Group, Corporate Property, Legal Network</td>
</tr>
</tbody>
</table>
There appears to be little obvious synergy between information science research and corporate information service requirements in CSIRO. For example, in 1995-96, CSIRO's Information Technology Division commercialised information tools such as the Spatial Database Manager™ for the U.S. telecommunications company, Convergent Group Asia Pacific. The same Division also developed techniques, in conjunction with BHP and Datacraft Technologies Pty Ltd, for establishing global electronic directories which were interoperable with existing digital data stores; and helped establish a Land Information System used by the South Australian Department of Environment and Natural Resources (CSIRO, 1996). None of these research breakthroughs appears to have been utilised within CSIRO for operational use.

During the same period, the Organisation’s Corporate Information Management Unit was turning to overseas-based research programs in “knowledge management”, in order to achieve its objectives of developing information management systems and processes (CSIRO, 1996; de Gooijer, 1997).
Whereas CSIRO’s research in information science occurs in a work environment determined by the organisation’s dominant scientific culture, its information technology services and operations is developed in an environment influenced primarily by non-scientific staff. For example, in a major review of information systems for top-level decision making (CSIRO, 1992a), only 13 of the total of 100 submissions came from practicing scientists, the remainder coming from administrative and other non-research employees.

The U.S. National Research Council (1995b) noted a general preference, on the part of scientific organisations, for carrying out new research instead of re-analysing scientific data that had already been gathered. One manifestation of this preference was the low level of resources allocated by most agencies to the systematic management of existing experimental data.

To better understand whether and to what extent this situation applied to CSIRO, I interviewed the General Manager of the organisation’s Information Technology Services Branch, who expressed the view that “There is no corporate will to spend money on the dissemination of
experimental information" (J. Potter, personal communication, May 21, 1997). He noted a number of factors in relation to CSIRO's research capacity in information science and its apparent reluctance to employ this to resolve corporate information management deficiencies. The list below was compiled from interview notes. Quotations are direct:

- history (aversion of research managers to carrying out an operational service role; perceived poor past performance of researchers in such roles);

- policy choices (for example, deciding to adopt off the shelf "industrial strength" commercial software rather than to develop in-house solutions for corporate service applications);

- the association of "information" with administrative or routine tasks such as records management, finance, and legislative reporting (Freedom of Information, Archives Act and Privacy Legislation, for example). This association mitigates the attention of senior CSIRO managers whose interests are predominantly scientific;

- the belief that experimental data belongs to individual scientists, rather than to the organisation which employs them;
• the cultural perception that information management is a "service" area, subsidiary to the "real" task of research; building a research program in a service area would be regarded as running a "separate agenda";

• conflict between the objective of commercialising research results (a process which may require concealment of information, at least during key periods such as patenting process) and the objective of sharing research data with colleagues or the community; and

• the intrinsic difficulty in co-ordinating activities across a large-scale organisation comprising scientists whom are trained to think and act independently.

(J. Potter, personal communication, May 21, 1997).

If any of these potential influences have prevailed for significant periods, they may be reflected or institutionalised through staff profiles. In the next section, I examine what those statistics may reveal about information management in CSIRO.
4.3.4 Core responsibilities and who performs them

During economic downturns, many organisations cut back their “discretionary” activities in order to concentrate on their “core” areas. Commonly in these circumstances organisations sacrifice longer time horizon activities such as research and development in order to stabilise activities, such as sales or marketing, which are more profitable in the shorter term. CSIRO is nearly unique as a large Australian enterprise in having R&D as its core function, instead of as a discretionary area. In times of insufficient cash flow, CSIRO appears to deliberately reduce expenditure on management and services in order to sustain or enhance its core commitment to research.

This approach is consistent with the organisation’s charter, in which information, communication, publishing and other data services occur explicitly only as secondary objectives. Perhaps as a result, these tasks are primarily performed by staff groups outside the dominant research culture.

“Front-line” information service staff (which includes the categories of librarians, library technicians, data processing operators, and
receptionists) are overwhelmingly female and low-ranking. The research scientist group is overwhelmingly male and high ranking (Table 2).

**TABLE 2 CSIRO STAFF PROFILE IN 1993 and 1996**

Note: Mode classification = seniority level of largest single group within category. Detailed figures for this column not available from 1993 onwards. The nomenclature of employment categories has changed slightly but the comparisons refer to like areas.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>72</td>
<td>73</td>
<td>CSOF3</td>
</tr>
<tr>
<td>Communication &amp; information</td>
<td>64</td>
<td>&quot;Other professional&quot;</td>
<td>Librarians [&quot;other professionals&quot; CSOF4]; library officers GSOF3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55; library officers 91</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Research management</td>
<td>4</td>
<td>0</td>
<td>CSOF9</td>
</tr>
<tr>
<td>Research scientist</td>
<td>10</td>
<td>7</td>
<td>CSOF8</td>
</tr>
<tr>
<td>Senior specialist</td>
<td>9</td>
<td>9 (&quot;senior executives&quot;)</td>
<td>CSOF8 (&quot;senior executives&quot;)</td>
</tr>
</tbody>
</table>

Although some 39% of CSIRO’s staff are scientists, this group has been extremely under-represented in corporate information management. At the time when this functional area of CSIRO was last extensively reviewed (CSIRO, 1992a) the three units responsible for supplying the information needs of CSIRO’s executives and other senior staff— the Management Information Systems unit; Headquarters Library; and Information Services Branch—accounted for 173.5
“equivalent full time” staff, of whom only one had a job classification “scientist” (CSIRO, 1992b). At the same time some 259 scientists were employed by CSIRO to carry out research in Information Science and Engineering (CSIRO, 1992b). Many of these scientists were employed in developing decision support systems, high speed networks, data mining, spatial data systems, and other tools which one may have expected to prove useful for managing research data and information.

In a knowledge-based activity such as R&D, information is intrinsic to organisational performance. Failure to exploit new capabilities in information science and data management could result in loss of earnings, low efficiency and competitiveness, and adverse perception by clients. In this context, it appears odd that CSIRO does not use its scientific strengths in information science to service at least part of its strategic information needs, and to bring this task into the purview of the dominant scientific subculture.

On the other hand, scientific data are viewed by some scientists as “personal”, because it usually arises through the efforts of individuals or small teams. Experimental data may be the path to group approval,
fame, promotion, and income for the research team to continue its work. Some CSIRO archivists and record managers believe that scientific information collected by CSIRO researchers is sometimes hoarded or is otherwise difficult to recover, either by design or through inadequate tracing procedures (CSIRO, 1997d; Gray, 1996; Sunter, 1996).

Many of these observations may be equally valid for other research organisations. However some, such as the Australian Geological Survey Organisation, have recently implemented systematic “whole of organisation” methods for cataloguing metadata, or dataset descriptions, in order to better safeguard and benefit from their collective information resources (Root, 1997). The National Research Council (1995b) in the United States studied the problem of uncoordinated and inefficient custodianship of experimental information, and proposed a union or “federation” of co-operating data managing organisations, linked with common policies but with individual responsibility to apply these within separate constituencies.
4.3.5 From Management Information to Information Management

During 1996-97, an extensive re-organisation in CSIRO resulted in a 30% decrease in the number of Divisions, in an effort to forge units with higher efficiency - as measured by proportionately greater numbers of scientists compared to “support” staff.

In parallel with this business unit re-organisation, CSIRO moved towards a focus on the subjects of its research, rather than on its administrative structure. This move is epitomised by a new matrix management structure in which the business units may address stakeholder interests in up to 22 Sectors, representing industry or environmental markets. The Sectors are intended to become the focus for priority setting and planning, while the Divisions will provide the scientific discipline focus and the means of delivering research outcomes to the markets (CSIRO, 1996).

In order to align information management operations with the changing work environment, CSIRO Corporate Information Management group commenced around 1995 to develop an information strategy based upon the concept of ecological modelling of
information, devised by Thomas Davenport at the Ernst and Young Center for Business Innovation in Boston (de Gooijer, 1997). The strategy’s objective was to align “human processes” and information systems, instead of aligning information management with “...managing systems and the information resources delivered by those systems” (ibid.).

A Working Group was established to develop this strategy, and this uncovered several impediments to effective use of information management systematically across CSIRO:

- wide geographic distribution (about 100 sites in Australia and in limited overseas locations);
- multi-disciplinary nature;
- poor inter-discipline communication exacerbated by inappropriate management structures;
- contradictions between commercial orientation and scientific co-operation; and
- emphasis on technology rather than information process.

(de Gooijer, 1997).
The subsequent absorption of the Information Management Branch into Information Technology Services Branch (ITSB) and the "devolution" of some functions to individual divisions resulted in the termination of this theoretical work, and also saw a reduction in corporate information staff by about 50% in 1996/97 (J. Potter, personal communication, May 21, 1997). This reorganisation (CSIRO, 1997e) followed the departure of the former General Manager (de Gooijer, cited above) who initiated the knowledge management research and the information strategy noted above.

At about the same period (1997) an Information Access Group was set up to investigate the acquisition by CSIRO of commercial databases. The focus here was on access by CSIRO business managers and program leaders to commercially oriented information such as the World Patents Index. However, with the exception of certain map data, there has not to date been a significant corporate move towards systematic sharing and/or tracking research data. Scientific computing services and the storage and retrieval of scientific information remains primarily decentralised, at the site or project level.
At Divisional level, it is extremely uncommon to encounter systematic approaches to the custodianship of experimental data in CSIRO.

Possibly the first and only Division to develop a strategic plan for research data is reported by Finney (un pub.). This data manager, who had only recently (1996) joined CSIRO with a background in spatial data management, proposed a data management plan for the new CSIRO Division of Marine Research (formed 1997 by merger of former Divisions of Fisheries and Oceanography) which would "...allow researchers to capitalise on the multi-million dollar investment both the Divisions of Oceanography and Fisheries have made over the past 60 years" (ibid., p.2).

She analysed prior performance in six key areas of data management (data policy; data registration; data archiving; data processing; databases; and dissemination) and concluded that with one exception (processing of data collected by ship cruises), this performance had been "patchy to poor at the Divisional level" (ibid., p3). Research projects with well developed data management plans were in the minority and there is "...little incentive at present for these projects to
share their experience with the rest of the Division" (ibid., p3). The plan proposed the establishment of a Divisional Data Centre, in which the emphasis would be on the documentation of existing and newly-created data sets, and the systematic listing of these metadata on local and national directories in order to improve access to scientific data by the Division's natural constituency, including external clients.

In this section, I explored aspects of the operation of a large research organisation in terms of the Strategic Information Management precepts noted by business analysts such as Orna (1996). In CSIRO's case, I conclude that cultural aversion, as well as structural impediments, make it unlikely that a centralised system to manage experimental data, including its voluminous Earth observation information, could be successfully introduced without significant external stimulus. However, in at least one case, on a decentralised basis there appears to be potential for a more structured approach to spatial data management using similar methods and systems to those employed elsewhere for storing and retrieving Earth observation data. In the next sections, I explore the extent to which these organisational
characteristics also apply in the more particular case of Earth
observation data management.

4.4 Evaluation of prototypes

In this section, I describe the selection and installation of three
prototype information systems for managing Earth observation data in
the Australian research agency CSIRO. I explore user requirements in
this environment, and reactions to the prototype systems. Also
examined are the status of information management and of
information professionals in the organisation, and preferences in
relation to either a centrally controlled or a federated structure for
information management.

4.4.1 Selection of prototypes

I narrowed down the range of potential prototypes for testing,
through consultation with CSIRO scientists involved in the Earth
Observation Centre. I also interacted frequently with knowledgeable
peer groups such as the CEOS Working Group on Information
Systems and Services and the Commonwealth Spatial Data Committee.
Systems examined included:

- the International Directory Network (IDN) of the Committee of Earth Observation Satellites (CEOS);
- NASA’s Information Management System (IMS) from the EOSDIS Core System;
- the Intelligent Satellite Information System (ISIS) of the German aerospace research agency, DLR;
- the proposed Australian Earth Observation Network, AEON;
- the CEOS Information Locator System, CILS, developed by DLR and the European Community’s Centre for Earth Observation on behalf of the German space agency, DARA; and
- the Australian Spatial Data Directory of ANZLIC/NRIC.

Apart from technical suitability, several factors influenced the choice:

- IDN and IMS were already needed for a planned data networking demonstration at the CEOS Plenary hosted by CSIRO in Canberra in November 1996;
- CSIRO resources were available to help conduct this demonstration;
• Australian agencies, including NRIC and CSIRO, had previously expressed a desire to implement IDN in Australia;

• NASA was committed to IMS maintenance and improvement over a long term, and a considerable level of technical advice was available to CSIRO through teleconferences of CEOS task teams;

• through bilateral contact and because of its participation in the CEOS Working Group, CSIRO would be able to access improvements to IMS initiated by NASA;

• IMS is the operating system for EOSDIS, and several CSIRO scientists have been participating since about 1988 in EOS investigations;

• the CEOS IDN was used and supported by numerous international space agencies and Earth observation data organisations;

• as a well-established system, the IDN was already populated with significant metadata; and
AEON's development ceased during 1996 as a result of the termination of the sponsoring agency (the former Australian Space Office).

CSIRO increased its level of participation in CEOS from about the time of DARA's chairmanship of that organisation in 1994-5; it was a natural step for it to support the CILS pilot project led by DARA, offering an Asia-Pacific node for the system and giving a practical follow-on to a feasibility study of an Earth-Space data system, conducted by the UN regional commission ESCAP and supported by CSIRO (Kingwell et al., 1995).

DLR wished to commercialise ISIS and the issues of intellectual property, future product support and the possibility that a user agreement would have to involve commercial parties made use of ISIS unattractive, for the purposes of this study. Further, development of a web-compatible version of ISIS was later than expected, meaning the product not available at the optimum time.
DLR was however able, through the support of DARA, to send personnel to Canberra to install CILS and to provide (limited) operating training to CSIRO staff. DLR was also committed to maintaining the CILS software and the global infrastructure for the duration of the pilot scheme.

Because only limited in-kind resources (skilled personnel, adequate computer disk space and network connections) were available when this research was carried out, commercial systems, in-house development, or systems requiring extensive modification or maintenance were ruled out as candidates for evaluation.

Finally, NRIC’s spatial data directory lacked a web interface and contained reference only to Australian data sets, limiting its attractiveness to those CSIRO scientists seeking desk-top access to details about regional or global data sets.

On the basis of these general factors of technical feasibility; low or zero cost; ease of availability; upgrade path; and support or use by kindred organisations, I chose to install and evaluate IDN; IMS; and
CILS (Kingwell, un pub.). This selection gave a nice symmetry, including an entry-level information locator and extension service (CILS); an intermediate level and reasonably populated directory (IDN) containing information about data holdings in participating agencies' data centres world-wide; and a comprehensive and evolving data management system (IMS) that can be used for a range of data tasks including inventory control. Colleagues from the CSIRO Earth Observation Centre installed the Australian IDN node at CSIRO Office of Space Science and Applications in August 1996. The use of IMS was formalised in early 1997 through a Memorandum of Understanding between CSIRO and NASA (Appendix C). The actual installation of IMS was performed by EOC staff in September 1996. CILS was installed through courtesy of the German space agency DLR, which funded Dirk van Gulik's travel to Canberra to complete this in August 1997. All three systems are accessible on a public access server, firewalled from internal applications (Australian Cooperating Node, 1998; CSIRO-EOC Installation, 1998; CSIRO, 1998a).
4.4.2 National or international data systems?

Australian scientific organisations, including CSIRO and the National Resources Information Centre (NRIC) of the Commonwealth Department of Primary Resources and Energy, had been considering from at least 1993 the establishment in Australia of a host site (or "node") of an international Earth and space information directory (Kingwell, 1994).

Unfortunately, lack of resources precluded NRIC taking this step: an additional complicating factor was that group's commitment to establishing a purely national spatial data directory, for which purpose a national (strictly speaking, a bi-national) standard metadata format had been developed by the Australian and New Zealand Land Information Council ANZLIC (Baker & Finney, 1995). The ANZLIC metadata standard was drawn up after an extensive process of consultation with surveying and geoscience organisations in Australia, and it was intended to form a template for recording, in a national spatial data directory, descriptions of geographically-referenced datasets held by Australian government agencies.
The ANZLIC standard was developed specifically for Australian and New Zealand's spatial data industries, and it differed from emerging international metadata standards, such as NASA's Directory Interchange Format (DIF) used by most of the world's space agencies to compile directories of Earth observation data available globally.

The industry consultation leading to the ANZLIC metadata standard concentrated on land information disciplines: surveying, geoscience, and terrestrial renewable resources. At around the same time, maritime disciplines were compiling a national directory of ocean and coastal zone spatial data, the Marine and Coastal Data Directory of Australia (MCDD). By late 1996 this contained details of some 3,000 sets of data held by several dozen Australian agencies (Blake, 1998). The MCDD used a metadata template specially developed for the purpose, but containing some information fields identical with the ANZLIC standard (ERIN, 1998).

Assuming that CSIRO wished to "advertise" the existence of its spatial data holdings by publishing these in an electronic directory system, the question arises as to whether it should proceed with a national focus,
using the ANZLIC metadata standard or discipline-based derivatives such as the MCDD; or with a global focus, using an international metadata standard such as DIF. One aspect of this choice was whether, on balance, CSIRO researchers were more likely to exchange Earth observation data with international space agencies, or with agencies in Australia. Feedback from CSIRO Division chiefs in 1996 showed aversion to "double handling" of data, meaning that resources would not be available to enter CSIRO data details onto international and national data directories if these used non-compatible formats. One option - adopted by the Australian Antarctic Division in 1996 - was to write a conversion program so that data directory entries in DIF format could be mapped into ANZLIC format, and vice versa.

I employed a user survey to find out more about attitudes of CSIRO researchers to this question, as well as to explore their reactions to typical on-line Earth observation information systems.
4.4.3 Survey results

4.4.3.1 Overview of survey

The survey comprised two parts: a questionnaire mailed to 76 CSIRO Earth observation researchers in February 1997, and interviews with fifteen researchers conducted between February and August 1998.

Thirty four researchers in the target group replied to the questionnaire, eight expressing no interest in the subject of Earth observation information systems. Of the twenty six researchers who completed the questionnaire, four described their main role in Earth observation data as “provider or manager”, twenty as “users” and two as “both”. Fourteen of the respondents, including all four managers, agreed to be interviewed, as did an additional Earth observation scientist (also a data manager) who joined CSIRO after the questionnaire was completed. In the fifteen interviews performed, the question emphasis was on user response to the three prototype information systems and on user preferences in information retrieval.

The questionnaire and its covering letter are shown in Appendices D and E, with details of the responses appearing in Appendix H. The
interview questions are reproduced in Appendix F, and a list of
interview tapes is given in Appendix G.

4.4.3.2 Information needs and responses to prototypes

Use of the prototype systems by the group of fifteen interviewees is
summarised in the table below.

Table 3. Use and preference in Earth observation information
system prototypes

<table>
<thead>
<tr>
<th>USE OF SYSTEMS</th>
<th>PREFERRED SYSTEM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>intension</td>
<td>Intending</td>
</tr>
<tr>
<td>User</td>
<td>0</td>
</tr>
<tr>
<td>User/manager</td>
<td>0</td>
</tr>
<tr>
<td>Data managers</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
</tr>
</tbody>
</table>
Consistent with earlier surveys of Earth observation information needs in CSIRO (Simpson et al., 1995; Fandry et al., 1993), the twenty six questionnaire respondents generally believed that a more systematic approach to Earth observation information management was required. Only one respondent believed that a well-populated on-line directory would not be useful; this researcher was one of six who felt that their work area already had access to an inventory management system suiting their requirements. Interview results tended to suggest that where respondent's had access to a suitable information system, this was usually operated by a non-CSIRO research collaborator. A much larger number (17) felt that their work area was not currently served by an efficient information management system in Earth observation.

Of the twenty five who believed that a metadata -based directory would be useful, seven felt that this would be “very useful” and a “high priority”. In several cases, particularly when the interviewee was a data manager, the principal perceived advantage of these systems was to organise and more efficiently utilise existing local Earth observation data archives. The “discovery” function, through which a user could locate datasets held elsewhere, was generally seen as less important. As
became clear during interviews, a significant number of these scientists worked for long periods, or even exclusively, with well-defined, highly specialised and relatively small volume datasets. In many cases these researchers had devised local consortium arrangements or had built human networks (“invisible colleges”) of colleagues who shared with them information and data. In those cases, unless the researcher changed field, they could be reasonably confident of keeping informed about new data sources through word of mouth.

“If you work in a particular area, you tend to know where the data is and you have access to the data anyway...so those systems are only really useful if you need new data which is secondary to your main area of research, or unless you are starting a new research area and need to find data you don’t normally deal with.”

(Researcher, ID #7).

There appears to be a slight preference for data sets covering Australia, but this was not as pronounced as may have been expected. Four questionnaire respondents would “never” require access to international data; two would need to search for this about every week; and nineteen at less than monthly frequency. By contrast, all
twenty six would seek Australian data - four about weekly and the remaining twenty two less than once a month.

Seven researchers preferred a single database giving details of both Australian and international data, while a nearly equal number preferred separate databases. Thirteen were not concerned whether this information was on one or two directories.

The geographic coverage of required data depends strongly on the discipline area, with environmental applications often exhibiting a greater need for global or large scale data sets. This reflects a preoccupation with global issues (such as the Greenhouse effect and impacts of climate change) or perhaps the use of tools such as global climate models. On the other hand, in the geological sciences where there was a close relationship with applied research in conjunction with the exploration industry, the geographic area under consideration tended to be localised, with smaller numbers of satellite images required. In most cases, such low-volume datasets were intensively processed to a high degree of accuracy, and customised for use in particular projects. These few images or small-scale industry-specific
databases were accessed frequently, and were generally retained in a local data warehouse operated by a non-CSIRO group.

"When ever we [CSIRO research group] get a tape [of Earth observation data], we give it to them [state government mapping group] to look after...because they have the resources and staff and know what they’re doing. If I need a directory search I can ask them."

(Research project leader, ID #13).

Eleven of the fifteen interviewees had used one or more of the three prototypes. One had done so “hundreds of times”, some for dozens of occasions, while yet others had done so only once, in preparation for the interview. Of the four who had not used the systems, three intended to do so in the next month. The most preferred prototype (6 choices) was IMS. Of the six scientists preferring IMS, three were data managers, two had dual roles as managers and users, and one was a user. Next, with one preference, was the IDN. No one in the interview group preferred CILS, although at least two intended using it again for its information on sensors and its extensive glossary. Four of the five data managers expressed a clear preference for either IMS or
IDN, as did both user/managers, while only one of the eight users expressed a clear preference for any particular system.

Two interviewees had equal preference for each system, noting that different applications or end purpose may require access to different databases.

The information elicited in interview did not always tally with this preference hierarchy. For example, several interviewees who preferred the IMS on the grounds it had more utility subsequently cited either CILS or IDN when asked to give an example of any benefit they had experienced in using the prototypes. Possible reasons for this inconsistency include:

- Confusing the similar acronyms;
- Relying on memory;
- Assuming that the most complex and expensive system would have greater relevance or power; and
- "Brand name recognition" resulting in the assumption that the NASA product would be preferable.
Each prototype was felt to be difficult or dissatisfying to use by several interviewees: “clunky” and “non-intuitive” were terms applied by two scientists to describe, for example, the IMS web interface query facility.

“For a total novice, they are very opaque. You just don’t know where to start: if you push buttons, the system tends to collapse.....you have no indication of what you ought to be doing......They are all much of a muchness in this respect”

(Researcher and business development manager in geoscientific information systems, ID #4)

By comparison, one user commented that CILS had an elegant feel which many users would find attractive, especially for those not familiar with Earth observation data systems. At 16 August 1998, CILS had 127 registered users, of whom 14 were Australian. Of these fourteen, ten are in CSIRO.

Nearly all the interviewed scientists had become accustomed to using Internet tools such as web crawlers for their research. Some contrasted the intuitive nature of proprietary web software with the prototypes, which these individuals felt were poorly designed.
"I don't want to have to read a manual for an hour to learn that I have to press a button".

(Research business manager, ID #4).

The IDN allows searching through selecting variables (such as sensor type) from a controlled vocabulary of terms ("valids") in pull-down lists. Two scientists found this clumsy, neither having discovered that free-text searches were also supported. Two scientists suggested that all systems, but particularly IDN and CILS, could be improved through "worked examples" showing typical searches and retrievals. The picture gallery in IMS went some way in demonstrating, to first users, the kind of information that could be recovered.

Seven of the fifteen interviewees felt the prototypes were valuable as templates for databases they were developing for their group's spatial and Earth observation data. They felt that by adapting existing information management software such as IMS, CSIRO Earth observation data systems could be grown in an extensible or scalable manner, through gradual addition of facilities such as on-line archives.
The ability to browse sub-sets of archived data was cited by four scientists as an essential requirement for a data system. An equal number regarded online data delivery - especially to their clients - as a critical feature needed in operational information systems, with one researcher preferring online data purchasing. However, neither faster data delivery nor electronic commerce was an advantage if the product quality was not known and consistent. Unless an information system contained the essential attributes of error flagging and processing history, many of the scientists would have little faith in relying on data products it generated:

"I don't think data availability is the problem. I think data consistency and standardisation is [sic]. Some people say that net delivery is the solution [to growing the EO market], but I think that, basically, good data is the solution".

(Research leader and data manager, ID #20).

Other features that individual interviewed scientists regarded as desirable included:

- distributed high power processing;
- collections that can be searched by theme or application, as well as by geography; and
• algorithms that would allow extrapolation of data to different scales.

4.4.3.3 Attitudes to information management

4.4.3.3.1 Federation and centralisation

The preferences of the study group for either centralised or federated styles of information architecture were tested in two ways. First, the questionnaire recipients were asked to select one of five models of information service they would prefer in CSIRO. The choices ranged from top-down centralised to, at the other end of the scale, completely decentralised. On this scale, a federation was most closely represented by the third option, featuring a common product catalogue, agreed access and pricing policies, and decentralised management of high level products. This option was preferred by the largest block of respondents (12), with all but one of the 12 other respondents opting for even less coordinated models. None of the 26 who completed the questionnaire - even the data managers - thought that a centralised option was an appropriate choice for CSIRO.

The second test of attitudes to federated or centralised structures was through interview. This approach yielded some extremely illuminating comments which clearly indicated a high degree of scepticism about
"outsiders" (those not involved in one’s project or local group) providing an information system or service. This scepticism extended even to units in CSIRO that specialise in corporate information technology, and to the CSIRO research group that markets its ability to design and implement information systems for Earth observation.

Ironically, this division has developed for the Australian Centre for Remote Sensing an on-line data delivery system for SPOT satellite products, and is working with the Australian Wheat Board to supply on-line information on wheat futures, based on remotely-sensed images. A federated architecture was assumed for this pilot project, in which value-adding data providers can share the network infrastructure with primary data providers and end users (Research project manager, ID #4). One interviewee expressed the hope that such approaches would “grow” Australian Earth observation enterprise by deepening its market reach, noting however that there would be increasing competition from other value-adding companies based overseas.
A number of researchers and data managers recalled past bad experiences of purchasing highly processed data from suppliers specialising in data services, such as the Australian Centre for Remote Sensing. All too often, errors introduced by standard processing mask the small signals that these researchers were seeking. As one put it, "...you have no hope of finding the 1% level of meaningful information in a satellite scene if this has been smoothed or processed out..." [by the supplier]. For these scientists, whose work depends on exceptionally intense scrutiny of a small number of satellite scenes, complete knowledge and confidence in the processing chain is vital. This group represents "leading edge users" having specific data requirements that are unlikely to be met by "standard" products generated for established markets. This is an important factor contributing to distrust between research users and data suppliers outside the research project or discipline.

Some scientists also expressed scepticism, based on experience, of the commitment and reliability of those with corporate or central responsibility for maintaining datasets:
"In my experience, when people try to combine [project databases] into one big database, the wheels tend to fall off. The technoheads sometimes want to do this but it becomes cumbersome to maintain and structurally this is not efficient."

(Research project leader).

For this scientist, and several others, it was more credible to believe that researchers and project teams who derived value-added products would be more likely to retain a commitment and assign a higher priority to maintenance of those data:

"Most people I work with want to have a reasonable access to archived data. The difficulty you always get with a centralised system is priorities. I guess the reason to put your own system in place is that you know it will work. ...Most people's experience tends to be that centralised systems aren't as transparent as you would like. Why should they be? As an organisation-wide priority, why should they jump when we say jump? Whereas with your own system, if it's holding up, you put resources into it, because you know that otherwise you can't deliver on your projects. That's fundamental."

(Research project leader and data manager).
Handy (1992) found that an important feature of self-correcting or learning organisations was the value they placed upon their high-achieving professional staff. This acted as a stimulus for federated structures in corporations, for many talented, mobile and innovative professionals preferred to work in autonomous small groups where their freedom of thought and their influence on outcomes was maximised. A similar tendency was observed in the interview group: most intuitively favoured federated paradigms in data management, although none used the term unprompted.

"Probably best it is done that way [leaving custody of datasets to projects/scientists] because they've collected it, they understand it the best...they're probably the best source of information about that data, they are probably the ones who have played around with it most....as long as you can change people's attitude [to] get them to adopt good data management practices that's probably where the data should lie."

(Data manager).

Another scientist, who was often consulted on data management, expressed a similar view:
"The model we have to be getting towards is where you have a research group working on a piece of data. You need to have that [processed] data on site. Then they might produce their own little product - it might be sea surface temperature maps, or cloud cover maps. ... Then they have a vested interest in making sure that product is up to date and correct. ... In ... making available more processed datasets, the only way is on a local basis...because if you try to centralise this, you lose the association with the product and as a result, nobody bothers to maintain them and you end up in a mess."

Despite the enthusiasm for localised value adding, there is recognition that there is a place for concentrating resources of lightly processed data. Irrespective of recent reductions in the cost of data storage hardware, economies of scale and the quantum of funds required still dictate that it is sensible to locate voluminous data in a few warehouses, so these data can service a number of dispersed value-adding producers. The scientist quoted above continued:

"There's nothing wrong with having a couple of data centres to protect your raw datasets...but [product development] should be local, because then there is a sense of ownership."
4.4.3.3.2 Value of CSIRO's research information

Most interviewees expressed the view that CSIRO does not systematically utilise its body of collected scientific data, preferring instead to focus on new research projects. This situation is of course not limited to CSIRO: the U.S. National Research Council noted that this preference is prevalent in all scientific disciplines in the United States, even though the scientific return from re-analysis of existing data may be greater than that obtained from new projects (National Research Council, 1995b, p.2).

That same study proposed a new, federated approach to the management and preservation of experimental data in the USA, citing four primary reasons:

1) because some data or events are unrepeatable;

2) to extend the baseline for time-varying events;

3) because "a data record may have more than one life"; and

4) because the expense of acquiring data means that the small additional cost of preserving it is justified.

(ibid., p.1).
One scientist expressed a similar vision for a systematic compiling of a database of experimental observations held by CSIRO:

"Many datasets in CSIRO would be extremely valuable. Technology is moving towards enabling multiple datasets to be accessed in a cohesive environment. This [should not be]...limited to Earth observation but also biodata: CSIRO’s data inventory must be absolutely huge. The value must be incredible if it was to be utilised in some way."

(Researcher and business development manager in information science, ID #4)

4.4.3.3.3 CSIRO servicing other users

Apart from cases where collaborative arrangement were in place through sponsored research contracts, there was remarkably little sense of CSIRO having a special responsibility, as the nation’s premier scientific organisation, to “husband” or exercise custody of data on behalf of other users. In fact, there was little sense of an imperative to manage data effectively on behalf of other CSIRO users.

When asked if CSIRO should explicitly contribute metadata to a national directory such as NDAR (in addition to logging CSIRO data onto an international database), only one of twenty five respondents
thought this task sufficiently important to devote more than a tenth of a staff year to.

In interview, several researchers - particularly those whose primary task was "managing" rather than "using" data - were emphatic that data management was a low priority in most CSIRO projects, with low probability of being allocated resources consistent with "best practice" standards. Several respondents viewed this as an example of generally low esteem for support and service providers in the organisation. The theme of organisational attitudes to information tasks resulted in the highest levels of animation by interviewees. For those respondents, low levels of accessibility to CSIRO's stores of Earth observation data indicates wider problems: neglect of the organisation's collective body of experimental data and low status accorded to information professionals.

One data manager expressed this with perhaps unconsciously vivid imagery:
“It's an issue of resources...the Divisions are finding it hard to find a balance between putting their money into research areas as opposed to putting it into research support. There's a strong corporate push to really minimise the amount of, let's say, administrative and technical support and [to instead] invest in more research scientists. And the drive for the scientist is to produce a paper at the end of the day, and once that paper has been produced there is little thought to the data that went into it...There isn't [sic] sufficient resources available to 'scoop'...behind the scientist and try and make some sense and order of that data for reuse within the Division later, and the scientist isn't given many rewards for doing so. In fact, probably penalised, because if you spent a fair bit of time trying to adequately manage data, I guess you are not spending that time writing papers and pursuing external money...”

This perception was not limited to data managers, however. Several researchers who were primarily “users” of data also felt that information management was not regarded in CSIRO as “real work”, in contrast to the actual writing of research papers. This attitude resulted, several said, in poor communication of data within a Division or unit, as well as inefficiency in information use across the organisation.
“CSIRO is definitely not a place where one gets promoted for managing data effectively. Information services [to general users of scientific data] are a low priority, with low status. ...The competitive atmosphere results in little sympathy or understanding toward scientists outside one’s own discipline...

[working in fields such as] data management”

(Researcher, ID #7).

Similar views were expressed by a scientist in an entirely different discipline:

[laughter] “We’re bloody hopeless [as a service provider]”. As a general rule, we have difficulty in balancing scientific leadership with client needs.....CSIRO has a hell of a long way to go in improving its act in delivering on client requirements. [There are exceptions, the subject noted]...There is a perception in the field that we do not deliver information well to clients”.

(Research project leader, ID #23)

This respondent felt that poor performance in delivering information to clients was one of the major issues which the organisation needed to address with urgency. This person strongly believed that the organisation’s mores and reward processes emphasised scientific
achievement but failed to recognise that teams comprising a range of personnel, including information service providers, were essential to deliver science-based services to external clients. Like several others making similar comments, this respondent had prior to joining CSIRO worked for other organisations, in which significantly more regard was shown for information management and support functions:
"These [non-PhD staff] are essential teams members in terms of ability to deliver, but CSIRO culture tends to ignore their contribution.... This is one of about 6 issues we have to address or CSIRO is going to go down the toilet... [My research group] is about delivering outcomes. None of us can do it on our own... [the team members] who provide a support role are just as important as the lead research scientist on the project. They are all providing critical input to ensure that we deliver... [but] the [CSIRO] reward process is totally biased towards bloody publications... There is a lot more we have to do in order to be able to perform large multi-disciplinary projects and retain key support staff. They are not valued. It is difficult to attract the good ones in the first place. You know, I can’t blame these blokes (or women) - what incentive is there? The system doesn’t recognise them... It’s a very 1940s attitude to the way you run an organisation. [In reality] everyone’s important, but [in CSIRO’s approach] everything is skewed to one sector or demographic... [If you need] highly skilled and motivated [support/non-research] people, you have trouble attracting them, because our track record is bloody awful."

(Research project leader, ID #23)

An exception to the aversion to data and information management occurs when there is a financial incentive. One example was with a
project jointly funded by CSIRO and extractive industries. In this instance the contracted services provided by CSIRO included data delivery to the industry partner, and that partner’s payments provided strong leverage for the supply of the information service. The researcher concerned felt that when the research was funded by a client, CSIRO investment on information systems was stimulated to a degree not usually observed in appropriation-funded research.

A less direct type of financial incentive occurs where an industry sector or government funding agency mandates data practices, and these become part of the conditions for obtaining grants or industry contracts. Certain sector-wide data management activities may be subsidised and this can have a cascading effect on the sector’s constituents, who feel that unless they too invest greater effort they may be left behind their peers. For example, the Department of Environment established the Marine and Coastal Data Directory, MCDD (Blake, 1998), and this stimulated the CSIRO Marine Research Division to appoint a Divisional Data Manager who was responsible for coordinating the Division’s effort in collating metadata for the national marine directory (Finney, un pub.).
In an externally stimulated approach of this kind, the major emphasis in the first instance may be simply the classification of existing data - compiling descriptions of datasets in a consistent metadata format.

In other cases, CSIRO groups may attempt more general on-selling of Earth observation data or information products - for example, sea-surface temperature charts to the fishing industry; vegetation index for fire-fighters and farmers (Simpson et al., 1995). In such cases, different business units of CSIRO sometimes found themselves in conflict, by addressing the same market. In at least one case this resulted in “client poaching”, with one Division undercutting another to attract a customer (ibid.). Clearly, this sort of competition militates against the easy flow of data across CSIRO administrative or project boundaries:

“There's not much incentive [to share information within CSIRO] and it has probably got worse now there is more competition and less goodwill between different programs and projects because of the [financial] crunch, the squeeze…”

(Research scientist).
The proprietary nature of some data was cited by several interviewees as a reason for poor data sharing practice within CSIRO:

"I guess you have to differentiate between what are CSIRO holdings, and what are basically holdings of commercial datasets. If CSIRO have archives of AVHRR data, well they are in the public domain...[but] CSIRO with its collaborators will have holdings of calibrated and rectified data...I think maybe it's worth [releasing these to the public], I'm not sure....The problem I have with a lot of these issues is the fact that ownership doesn't mean availability. You've still got dollars involved, copyright issues. Letting people know we have...data...doesn't mean that those data are available..."

(Research leader, ID #20).

Contrary to this apparent problem of managing copyright issues and associated royalty payments, the CSIRO library service routinely and with apparent ease handles issues of copyright law for printed material or electronic documents. Most metadata standards (including DIF used in the IDN, ANZLIC as used in NDAR, and MCDD used in the Marine Data Directory or “Blue Pages”) have a facility for flagging whether data have special access or use conditions attached. Only two of the fifteen interviewees expressed familiarity with these metadata
standards. In both cases, their projects had adopted an industry- or sector-wide standard, mainly due to top-down decisions from stakeholder peer groups or councils. For those two respondents, familiarity with industry best practice reduced concern about their ability to screen requests by other users for CSIRO data, and generated a sense of obligation and even enthusiasm for providing the community with CSIRO spatial data:

"I think in the public interest there is [an obligation for CSIRO to populate national data directories], and if we were sufficiently organised then that ...should be relatively easy...."

This interviewee, who had joined CSIRO about two years ago, explained the approach that the particular Division had adopted in order to keep abreast of data management trends within their industry sector:

"What we had to do first was to get our house in order, so that in a structured manner we were capturing that information on a frequent basis and in a specific format. We have just begun to do that and I think that a year down the track when we have that internal infrastructure in place, I see it as relatively simple for us then to be able to contribute to or link into other [external] directories, because we can automate that."
Conversely, thirteen of the fifteen CSIRO researchers interviewed appeared to be unfamiliar with the “conditions of use” field in spatial data directories. Several of this majority group expressed concern that “advertising” the existence of proprietary or poorly documented research data would generate requests for access to the data by non-CSIRO users. They worried that servicing these requests would divert resources from research; be embarrassing to refuse; or could threaten existing data sharing arrangements.

In other cases, a reluctance to share data was attributed to the extra work required to document it, and especially the work involved in explaining to other users the drawbacks, deficiencies, and inaccuracies in the data. This should not be an onerous task when the data products were “standard”, but as several researchers noted, that situation is rare for CSIRO where most experimental data sets are specialised and possibly volatile.

Often data custodians are aware that their data contains instrument artefacts which a little bit more attention and time will allow them to remove. In the meantime, potential users are anxious to get their hands
on the data, either for their own research or for applied use. Resource limitations often result in data custodians having to choose between supplying lower quality data products in the near term, or concentrating on improving data product quality, at the cost of not servicing current demand.

"Quality control is what it’s about…that’s one of the problems - …[when a research] ship comes back, everyone says ‘I want the data’, … the managers say ‘no you can’t have it yet, it’s not right, if I give it now it will have some errors…’ [The users will] Blame us for the errors. [You have] Got to get everything fixed up, doubled checked …so the data sits for months. I’m sure I must be guilty of that, too, in terms of satellite data…what I do is to say [to clients] ‘you can have the data, it will take me a couple of days because I have to do it [data processing] by hand. But if you wait a few weeks, I’ll have that [processing system] all automated, and then I’ll not only be able to satisfy you but also half a dozen other customers’. So I’m always tending, you know, not to deliver. I say to them, ‘eventually you’re [the customer] going to have a flood of it, but you can’t have a trickle, because it’s going to hold up the flood.’”

(Researcher and data manager, ID #13).
5. CONCLUSION

5.1 Reiteration

This research was undertaken to test service models and user needs relative to a class of scientific information derived from satellites observing the Earth, and the particular working environment of a group of Australian scientists employed by CSIRO and sharing a need for these data.

The main aim of the work was to understand whether efficiency in this working environment could be improved by adopting special information management systems devised to manage Earth observation data. To answer this question, I looked at the nature of the technology itself, but more importantly, I study the nature of the organisation and the role information plays in helping CSIRO to meet its objectives.

In the process, I examined general trends in Earth observation and, more widely, in information science. In particular I examined in detail the evolution of a large information system designed by NASA to manage Earth observation data from the ambitious "Mission to Planet
Earth”, exploring what this evolution reveals about client preferences in relation to either “top-down” or “federated” information systems in a scientific environment.

5.2 Trends in Earth observation data systems

The ground segment of space technology, including information processing and delivery systems, is the component closest to the end customer but is also less glamorous and often neglected in comparison to the space segment. Gradual maturation of space technology has been accompanied to some degree by a declining or static level of public sector investment, at least in the “traditional” or first wave of space-faring nations. The restricted flow of public funds has stimulated space system operators to seek greater return on investment. Two manifestations of this tendency are rising levels of private investment in Earth observation programs and greater effort on the part of the world space agencies in respect of information management. In both cases, the focus of effort has been upon delivering suitable data products to the client in a more efficient manner. This focus has had the effect of making information management in Earth observation more “user-centred” in comparison to the past when data systems were
primarily, in the terms of Allen (1996), “data-centred” in that they were designed around the peculiarities of the sensor and managed according to the convenience and priorities of the operators.

Although investment in space has reached a plateau in much of the western world and has markedly declined in the former Soviet Union, an increasing number of nations in Asia are increasing their efforts to leverage social advancement from the use of space technology. In addition, both here and in the western world, private sector investment in Earth observation and other space applications has increased, even though evidence to date indicates that this field is not yet absolutely profitable and requires considerable public subsidy, as pointed out by Mansell et al. (1993). Regardless, users of Earth observation are experiencing diversification of suitable data sources. Similarly, rising interest in multi-disciplinary approaches to the study of global environmental change has increased user demand for Earth observation data and information products from a growing number of information systems and operators.
User focus on the requirement to access, compare, and merge Earth observation data from a variety of satellites and sensors has encouraged the rise of "interoperable" information systems. This mirrors wider trends in information science, particularly the emphasis on distributed heterogeneous database management as a means of helping geographically-dispersed enterprises make most efficient use of their collective information resources, as observed by Thuraisingham (1997).

Recently, space agencies have devoted considerable effort to build interoperable information systems for managing and providing user access to their Earth observation data holdings. The international Committee on Earth Observation Satellites (CEOS) has been an important protagonist in these developments. Typically, these information systems comprise a query facility; a means of displaying search results; and one or more databases containing data and/or metadata. Three representative examples are the CEOS Information Location System, CILS; the CEOS International Directory Network, IDN; and NASA's Information Management System, IMS. On the basis of availability; supporting services; technical suitability; price and familiarity to CSIRO users, these three systems were chosen in this
study for evaluation as prototypes of potential operational Earth observation information systems in CSIRO.

Rapid take-up of the World Wide Web and related technologies by sophisticated information users, such as the scientific community, has led to the rise of the notion of the Internet "virtual marketplace" (Sarrat et al., 1995; C. Best et al., 1996) for Earth observation data and information products. This development, replacing a set of point to point connections between user and supplier, has been abetted by the advent of intelligent middleware. These are device independent and non-proprietary standards and tools such as ODBC and CORBA, which make it easier for client queries to be brokered or mediated between the user and multiple and disjointed databases.

Theoretical and practical considerations of workplace specialisation and management efficiency gave rise to the notion of "data warehouses", where non-volatile, fundamental and lightly processed (or "atomic") information is stored away from the production environment (Inmon, 1992). An offshoot of the data warehouse concept is found in the "data mart", specialising in a small number of
products or databases (Tanler, 1997). The widespread availability of commercial off the shelf (COTS) Web utilities, including JAVA and browse tools, mean that it is becoming easier for end users to interface with data marts and other forms of distributed information systems serving the Earth observation market.

An important infrastructure issue that has assisted this process is the gradual acceptance of metadata standards, which make more uniform the way Earth observation data sets are described. DIF, FGDC, ANZLIC and other international or national metadata standards are fundamental to the ability of custodians to more readily advise others, in a form easily searchable by standard web tools, about the existence, whereabouts and, most importantly, the quality of archived spatial and Earth observation data. Basic characteristics of Earth observation data which set these apart from information used in many other fields of scholarship are geo-location and time-dependence: all Earth observation information refers to a specific location and time. There is some hope that, in the development of universal metadata standards for information posted on the Internet, inclusion of geographical and temporal metadata fields will eventually allow users to locate granules
of Earth observation information independently of specialised information retrieval systems (van Gulik, 1996b).

5.3 Prospects for EOSDIS

NASA's Earth Observing System Data and Information System, EOSDIS, represents an almost heroic effort to redefine the priority between the space and ground segment in space applications. Conceived during a period when NASA and its counterpart NOAA had been strongly criticised for not devoting enough attention to data preservation and management, EOSDIS was awarded an order of magnitude more resources, relative to total mission costs, than had been typical for previous space projects. EOSDIS is probably the largest civil information system ever attempted, requiring a hundred million lines of source code and aiming to accumulate dozen of petabytes of data over a fifteen to twenty year lifetime (Jaworski, 1993).

Almost from the outset, however, EOSDIS laboured under serious handicaps that may still frustrate its completion. The first handicap was the woolliness of the thinking behind the intention and design of
the Earth Observing System. With no less an objective than to
"understand the Earth as a system", EOS's principal method seemed to
be to attempt to "measure everything, everywhere, all the time". As
Harris (1992) has pointed out, the strategy for EOS under-emphasised
the role of modelling and process study in framing the exact questions
to which detailed and long-term measurements would then perhaps
give the answer. If the fundamental scientific issues are not stated with
clarity, then the measurement or experimental strategy is likely to be
flawed and unsuccessful.

The second major handicap inherent in EOSDIS was the process used
to define its capabilities. Here, the most influential inputs came from
the twenty four instrument teams and the twenty nine Principal
Investigators selected by NASA to carry out multi-disciplinary studies
of the global environment using data from the EOS spacecraft (NASA,
1993). This was an extremely small pool of users upon which to base
the design specifications for a multi-billion dollar, long-term
information system. Apart from being unrepresentative of the broader
community of potential users of EOS data, this group produced
demanding specifications which were interpreted by system designers
with little regard for practicality and efficiency in production. Further, by preferentially supporting researchers prepared to elucidate "standard" EOS data products, NASA ensured that EOSDIS would become an "amplifier" of data products, despite the fact that many of these would by their nature have a small and possibly temporary market (Glover, 1994, 1997).

The third major handicap of EOSDIS was the immensely long planning cycle. The product specifications had been developed nearly a decade in advance of commencement of operations (Schier & Way, 1990). The intervening period has seen rapid and profound developments in information science and technology, with phenomenal growth in the use of broadband public switched networks for data dissemination; dramatic reduction in the cost to performance ratio for data mass storage; and accelerated development of architectures, theoretical frameworks and practical utilities for distributed heterogeneous data systems.

The fourth handicap for EOSDIS's development was finance: funding for the data component fell at a much greater rate than that for the
EOS program as a whole, with the result that EOSDIS decreased from around 30% of the intended cost of EOS to less than 15% over the first half of the decade (Office of Technology Assessment, 1994). What probably made matters worse was NASA's persistent inability to expend, in a given fiscal year, the resources allocated to EOS and EOSDIS (Berger, 1998; Oler, 1998).

These four handicaps, combined with the trends in information science and in Earth observation data systems described above, gave impetus and substance to attempts by the U.S. scientific community to re-define EOSDIS as a managerially- and logically-decentralised, or "federated" system. These attempts culminated in the influential report by the peak U.S. scientific body (National Research Council, 1995a), which recommended that NASA should competitively select, from universities and the private sector, "Information Partners" to supply value-added services and data products based upon raw EOS data archived by NASA in its data warehouses.

Although it does appear that NASA is moving (albeit cautiously) towards a federated architecture for EOSDIS, the future of the
enterprise is still highly uncertain. Papazoglou (1991) has pointed out both advantages and disadvantages of federated approaches to information services. One of the primary advantages is the implicit commitment of the partners in the federation, while one of the principal disadvantages is the loss of coherence resulting from a preservation of autonomy. In the model NASA has adopted for the "prototype Working Federation" (King, 1997), the federation partners exhibit a strong asymmetry, with NASA funding much of the other participant's costs. This approach threatens to preserve the worst disadvantages of a federated approach, while minimising the potential advantages. I conclude that while EOSDIS probably has sufficient momentum to survive in the short term, its longer term future is highly uncertain. Given criticism of NASA for unfairly competing with private sector Earth observation information suppliers, it is possible that EOSDIS will ultimately be bypassed or absorbed by market-oriented commercial information systems displaying much greater emphasis on analysing the requirements and work habits of the potential user pool. The risks for NASA in big-ticket but low-demand items like EOSDIS and the space station are high and not even the continuation of NASA in its present form can be taken for granted.
5.4 Fitness for purpose in CSIRO

Seventy six CSIRO researchers were asked if typical Earth observation data management systems would be useful to them. Most of those who replied were later interviewed about their response to three representative working prototypes. All but one of twenty six respondents and each of fifteen interviewees believed there was benefit to be gained through systems such as CILS, IDN and IMS. Seventeen of the 26 questionnaire respondents did not currently have access to a suitable information management system for Earth observation data.

Eleven of the fifteen interview subjects had used one or more of the prototype systems during the eighteen-month evaluation period. The fact that only three used any of the systems frequently, and four had not used them at all, suggests that these systems had drawbacks, or were desirable rather than essential, at least for this working environment. Analysis of the interviews of fifteen CSIRO Earth observation researchers showed that both conclusions applied.
The fifteen interview subjects comprised eight who regarded themselves primarily as “users” of Earth observation data; five who regarded themselves primarily as “managers” of data; and two whose responsibilities were divided. Most of the interview group had intimate knowledge, through their networks of peers, about the existence and custodians of the Earth observation data sets they were likely to need. At least three had made arrangements, at a local or project level, for more operationally-inclined kindred organisations to manage these data archives: several of these non-CSIRO groups had developed online Earth observation information systems during the study period (for example, ACRES, 1998; Department of Land Administration, 1998). Barring a change in discipline or geographic area of interest, or an influx of new researchers, the requirement for “search and find” information systems such as IDN and CILS was limited in this particular group of CSIRO researchers.

Those in the study group with responsibility for managing Earth observation data expressed the clearest preference and support for comprehensive information management systems such as IMS. This system has the potential for local implementation as a data base or
inventory management tool, and has the added advantage of being familiar to those researchers already collaborating with NASA's EOS research teams.

Six of the CSIRO scientists hoped to use IMS as a paradigm or template for local Earth observation inventory management, and another proposed to use it as an exemplar for larger-scale distributed information processing and distribution networks.

Of a spectrum of information needs in Earth observation, the ability to browse individual examples from a dataset before selection was most frequently cited by the interview group. It follows that systems like IMS, which can manage data at the granule level as well as at the dataset or directory level, were generally preferred. However, some of the study group noted particular features of either the IDN (breadth of dataset entries) or CILS (ease of use) which made these systems more suitable or worthwhile for particular applications or searches. On-line ordering and on-line delivery were also seen as advantages of the IMS. For each on-line information system, utility was generally regarded as being strongly coupled with the extent and reliability of stored
metadata, especially that relating to the processing history, accuracy and consistency of the products.

Areas in which all three systems were felt to be deficient in comparison with proprietary web products such as Netscape or Outlook were counter-intuitive or obtuse operation, and absence or “worked examples” or tutorials demonstrating the system capabilities.

5.5 Information culture of CSIRO

Based on evidence from contemporary corporate literature and from staff interviews conducted during this study, CSIRO lacks a culture of service necessary for an effective remote sensing information system that delivers knowledge products to the wider community.

Compared to many other “knowledge-based” enterprises, it displays an inward looking culture in which the preference of scientists to carry out “good science” is supreme. Information provision is a low-caste activity in the organisation, performed primarily by marginalised groups (women, lower ranked scientists, and those at the low end of the resource pecking order).
Because information professionals and services are viewed as having a supporting rather than a fundamental or intrinsic role in CSIRO, it is unlikely in the absence of a major external forcing stimulus that CSIRO would devote the resources necessary to successfully implement one or other of the three prototype Earth observation systems on an operational and organisation-wide basis. However, the lack of commitment to information services is not evenly distributed across the organisation, and it may vary over time at different rates in different sections. Furthermore, decision making in CSIRO is relatively decentralised, reflecting the preference, observed by Handy (1992) as "the pull of the professionals", that highly innovative professional staff often show towards working in small autonomous groups. CSIRO also demonstrates a tendency, described by Handy as characteristic of federated political structures, of subsidiarity, in which responsibilities are left with the component units unless there is an explicit consensus that those responsibilities could be better managed centrally.

Projects and units within CSIRO are subject to varying degrees of stakeholder pressure to supply information services in accordance with
industry standards or expressed client needs. The leverage applied by these (often) paying customers is forcing some portions of CSIRO, at least, to devote more attention to the systematic management of spatial data, Earth observation information, and other forms of experimental data.

These factors of decentralised decision making, uneven impact of client expectation, and the vagaries of small group dynamics mean that it is probable that without drastic external forcing, modern Earth observation management systems such as CILS, IDN and IMS would only be implemented within CSIRO on a voluntary and federated basis.

The most probable outcome in CSIRO in terms of Earth observation information systems is for each scientist or group to continue their own ad hoc arrangements, for their own immediate purposes. There is insufficient management interest or understanding at present to enforce coherent adaptation of universal system or systems, and there are no strong external drivers for such a step.
5.6 Further research

Evidence arising from this study broadly supports previous examination of deficiencies in Earth observation data management in CSIRO (Simpson et al., 1995); the view of information as a strategic resource in capable modern organisations (D. Best, 1996; Orna, 1990; GAO, 1994); and of the opportunities that advances in information science have lent to decentralised or federated modes of data management and user services (Thuraisingham, 1997; Allen, 1996; National Research Council, 1995a, 1995b).

Clearly, the continued evolution of one of the world's largest information systems, EOSDIS, will provide a source for further observations of the practical difficulties in devising large-scale information systems in a rapidly changing technological environment.

Examination of results from NASA's attempt to devise a "Working Prototype Federation" for EOSDIS may reveal whether this asymmetric partnership can avoid the fundamental disadvantage of federated approaches, namely the difficulty of maintaining coherence.
and stability in the face of the desire of participants for autonomy (Papazoglou, 1991).

The extent to which generic web crawlers and other knowledge-harvesting tools can eventually replace or supplement special-purpose information systems devised to manage Spatial or Earth observation data will also be interesting to study further. Those who planned dedicated information dissemination networks for EOSDIS will be vindicated if Internet bandwidth fails to expand to meet the requirements for real-time Earth observation markets. Similarly, if Internet costs rise, or changes to international database copyright regimes make Earth observation and kindred scientific information too expensive for most users, the concept of a publicly subsidised information network dedicated to scientific data may appear more attractive.

Finally, I note that CSIRO is one of Australia’s best known but least-studied icons. Butler and Bourke (1997) remarked that “CSIRO ...represents a natural site for the conduct of systematic enquiry into many of the major international puzzles in research and science
policy". It will be both interesting and instructive to observe whether federated information systems are sufficient to re-orient the organisation towards a more strategic approach to information management.
APPENDICES
## APPENDIX A: Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADEOS</td>
<td>Advanced Earth Observation Satellite of NASDA, launched Aug 1996</td>
</tr>
<tr>
<td>ACRES</td>
<td>Australian Centre for Remote Sensing (of AUSLIG)</td>
</tr>
<tr>
<td>AEON</td>
<td>Australian Earth Observation Network (proposed ~1995, not built)</td>
</tr>
<tr>
<td>ANZLIC</td>
<td>Australia and New Zealand Land Information Council</td>
</tr>
<tr>
<td>ATBD</td>
<td>Algorithm Theoretical Basis Document</td>
</tr>
<tr>
<td>ATM</td>
<td>Asynchronous Transfer Mode</td>
</tr>
<tr>
<td>AUSLIG</td>
<td>Australian Surveying and Land Information Group (in Commonwealth Department of Industry, Science and Resources)</td>
</tr>
<tr>
<td>BoM</td>
<td>Bureau of Meteorology, Australia</td>
</tr>
<tr>
<td>browse</td>
<td>A preview, snapshot or sample of data in a dataset, allowing a user to more quickly interrogate the dataset, to select particular data, or to assess their potential usefulness</td>
</tr>
<tr>
<td>calibration</td>
<td>The act of making an instrument accurate through converting instrument data to geophysical or biophysical quantities</td>
</tr>
<tr>
<td>catalogue</td>
<td>An extended directory, containing guides and an inventory, and supporting searches</td>
</tr>
<tr>
<td>CCRS</td>
<td>Canada Centre for Remote Sensing</td>
</tr>
<tr>
<td>CCSDS</td>
<td>Consultative Committee for Space Data Systems</td>
</tr>
<tr>
<td>CCT</td>
<td>Compute Compatible Tape</td>
</tr>
<tr>
<td>CEC</td>
<td>Commission of the European Communities</td>
</tr>
<tr>
<td>CEO</td>
<td>Centre for Earth Observation, at JRC, Ispra, Italy.</td>
</tr>
<tr>
<td>CEOS</td>
<td>Committee for Earth Observation Satellites</td>
</tr>
<tr>
<td>CILS</td>
<td>CEOS Information Location System</td>
</tr>
<tr>
<td>CINTEX</td>
<td>Catalogue Interoperability Experiment of CEOS</td>
</tr>
<tr>
<td>CORBA</td>
<td>Common Object Request Broker Architecture</td>
</tr>
<tr>
<td>CSA</td>
<td>Canadian Space Agency</td>
</tr>
<tr>
<td>CSIRO</td>
<td>Commonwealth Scientific and Industrial Research Organisation, Australia</td>
</tr>
<tr>
<td>DAAC</td>
<td>Distributed Active Archive Center, series of data warehouses funded by NASA</td>
</tr>
<tr>
<td>DARA</td>
<td>Deutsche Agentur für Raumfahrtangelegenheiten, former (to 1997) German Space Agency</td>
</tr>
<tr>
<td>DAS</td>
<td>Department of Administrative Services (abolished 1998)</td>
</tr>
<tr>
<td>dataset</td>
<td>a collection of logically-related data</td>
</tr>
<tr>
<td>DAT</td>
<td>Digital Audio Tape</td>
</tr>
<tr>
<td>DBMS</td>
<td>Data Base Management System</td>
</tr>
<tr>
<td>DEM</td>
<td>Digital Elevation Model</td>
</tr>
<tr>
<td>DIF</td>
<td>Directory Interchange Format (for CEOS' IDN)</td>
</tr>
<tr>
<td>directory</td>
<td>A collection of metadata: a guide to the existence, location, access conditions and (sometimes) the reliability of datasets. Usually gives information, in a uniform fashion, about a large number of datasets</td>
</tr>
<tr>
<td>DLR</td>
<td>Deutsche Forschungsanstalt für Luft und Raumfahrt, German flight and space research agency (later [1997], Deutsche Zentrum für Luft und Raumfahrt, German Aerospace Research Center, the German Space Agency</td>
</tr>
</tbody>
</table>
DTM  Digital Terrain Model
ECS  EOSDIS Core System
ENVISAT Environmental Satellite of ESA (planned)
EOIS Earth Observation Information System, NASDA Japan
EOS  Earth Observing System of NASA
EOSDIS Earth Observing System Data and Information System. The information management portion of NASA's Mission to Planet Earth.
ERIN Environmental Resources Information Network of Australian Department of Environment
ERS-1 European Remote Sensing Satellite-1
ERS-2 European Remote Sensing Satellite-2
ESA European Space Agency
ESF European Science Foundation
ESOC European Space Operations Centre
ESRIN European Space Research Institute of ESA, in Frascati Italy
ESSC European Space Science Committee of the ESF
ESTEC European Space Technology Centre, in Noordwijk, The Netherlands
EUMETSAT European organisation for the exploitation of Meteorological Satellites
festschriften Volumes published in honour of a scholar – usually comprising compilation of work by students or colleagues.
FTP File Transfer Protocol
GB Gigabytes, 10^9 bytes
GCOS Global Climate Observing System
GDS Guide and Directory Service of ESA at ESRIN
GENIUS Global Environment Network Information User System of ESA at ESRIN
geolocation to precisely fit a measurement or datum with a point on the Earth
GIS Geographic Information System
GLIS Global Land Information System of USGS
GOOS Global Ocean Observing System
granule An atom of information: the smallest logical example of an item of data in a collection
GTS Global Telecommunication System of WMO
GUSS Gateway for Users to EO Services of ESRIN
GUI Graphic User Interface
guide A detailed description of one or more datasets, allowing a potential user to determine the extent and location of each and their potential usefulness for specific applications
HDF Hierarchical Data Format
HEASARC High Energy Astrophysics Science Archive Research Center of NASA
HTML Hyper Text Mark-up Language
HTTP Hyper Text Transfer Protocol
IAU International Astronomical Union
IDL Interface Definition Language
IDN International Directory Network of CEOS
IEOS International Earth Observing System, the comprising remote sensing missions of CSA, ESA, NASDA, and NASA
IMS | Information Management System of NASA, used for mediating requests/searches between different components of EOSDIS. Provides users services and manages information at data centres linked to the EOSDIS Core System.

information system | Discrete set of resources and processes, organised for the collection, processing, maintenance, use and sharing of information.

inventory | Uniformly presented record of granules from one or more data sets, giving user information required to select a subset of those granules. A list or other representation of data, allowing particular datum to be selected.

IP | Intellectual Property


ISDN | Integrated Services Data Network

ISIS | Intelligent Satellite Information System of DLR

ISRO | Indian Space Research Organisation

JERS-1 | Japanese Earth Resources Satellite -1 (launched 1992)

JRC | Joint Research Centre of CEC, in Ispra Italy

LAN | Local Area Network

MB | Megabytes, $10^6$ bytes

Mbps | Megabits per second

MCDD | Marine and Coastal Data Directory of Australia

metadata | Data about data: a description or summary or classification of the data set under consideration

MMBS | Multi-Mission Browse Service of ESRIN

MMDD | Multi-Mission Data Distribution of ESRIN

MMIS | Multi-Mission Inventory Service of ESRIN

MMRA | Multi-Mission Reference Archive of ESRIN

MPEG | Motion Picture Experts Group

MUIS | Multi-mission User Information Services of ESRIN

NASA | National Aeronautics and Space Administration, USA

NASDA | National Space Development Agency of Japan

NOAA | National Oceanic & Atmospheric Administration, USA

ODBC | Open Data Base Connectivity

OLE | Object Linked Environment

OQL | Object Query Language

PAF | Processing and Archive Facility for ERS

QA | Quality Assurance

QC | Quality Control

RDMS | Relational Data base Management System

resolution | Least linear separation able to be discriminated by a sensor

SAR | Synthetic Aperture Radar

SCF | Science Computing Facility (of or affiliated to NASA)


SQL | Structured Query Language

surrogation | Substitution (replacing older knowledge). Also, conversion of primary to secondary sources.

TB | Terabytes, $10^{12}$ bytes
TDRSS  Tracking and Data Relay Satellite System of U.S. government
TRMM  Tropical Rainfall Measuring Mission of NASDA and NASA (launched November 1997)
UIT   User Interface Terminal
URC   Universal Resource Characteristics
URL   Universal Resource Locator
URN   Universal Resource Name
USGCRP United States' Global Change Research Program, a multi-agency effort aimed at understanding the causes and consequences of global and climate change.
USGS  U.S. Geological Survey
validation  Checking processing algorithms for self-consistency by comparing information products with field observations
VAP   Value-Added Product (or Producer)
verstehen  Intuitive or experiential understanding
WAIS  Wide Area Information Servers
WAN   Wide Area Network
WCRP  World Climate Research Programme
WGISS Working Group on Information Systems and Services of CEOS
WIPO  World Intellectual Property Organization
WMO   World Meteorological Organization
WWW  World Weather Watch of WMO; or World Wide Web
**APPENDIX B: Milestones**

| Literature survey |  |  |  |  |  |  |  |  |  |  |  |  |
| Negotiate software availability |  |  |  |  |  |  |  |  |  |  |  |  |
| Data management systems installed | IDN | IMS |  |  |  |  |  |  |  |  |  |  |
| Questionnaire |  |  |  |  |  |  |  |  |  |  |  |  |
| Research proposal |  |  |  |  |  |  |  |  |  |  |  |  |
| Interviews |  |  |  |  |  |  |  |  |  |  |  |  |
| Draft |  |  |  |  |  |  |  |  |  |  |  |  |
| Revise |  |  |  |  |  |  |  |  |  |  |  |  |
| Submit |  |  |  |  |  |  |  |  |  |  |  |  |
APPENDIX C: Memorandum of Understanding with NASA

CSIRO OFFICE OF SPACE SCIENCE AND APPLICATIONS
Institute of Natural Resources and Environment

ANU Tech Court
Corner North and Daley Roads
ANU Campus
Acton ACT 2601, Australia

Head: Dr Brian J. Embleton

Postal Address:
GPO Box 3023
Canberra ACT 2601 Australia
Telephone: 06 216 7200
International: 61 6 216 7200
Facsimile: 61 6 216 7222

Dr Lisa R Shaffer
Director,
Mission to Planet Earth Division
NASA
300 E Street, SW
Washington DC 20546 USA

Dear Lisa

I am responding to your letter of 27 December 1996, in which you propose an agreement between the National Aeronautics and Space Administration (NASA) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO), through this office. The agreement will be to establish an interoperable interface between NASA's Earth Observing System (EOS) Data and Information System (EOSDIS) and CSIRO's system for earth observation data.

The terms and conditions contained in your letter are acceptable to CSIRO, and document our joint understanding as to the implementation of this cooperative effort.

Please note however that the last paragraph on page 6 of your letter should be corrected to read “Head, CSIRO Office of Space Science and Applications”, not “Director of CSIRO”.

Yours sincerely

Brian Embleton
Head,
CSIRO Office of Space Science and Applications
3 February 1997
APPENDIX D: Covering letter for questionnaire

Managing Earth observation data

This survey is for CSIRO staff who have identified themselves as stakeholders in the Earth Observation Centre. The survey's purpose is to find what sort of Earth observation (EO) data management services are useful to you, and which of these services (if any) should be provided through the EOC. From this user input, we hope to help arrange the approach which best suits your requirements. This work forms part of Project 5 ("Data base and access tools for environmental time series data") of the EOC Science Plan.

Data from the survey will also be used in a research project on the balance between user needs and system design considerations, relative to Earth observation data management systems. I am carrying out this study, which will also include case histories of EO data management approaches, as part of a MSc (Info Sci) course at Edith Cowan University, Perth.

In reports and publications based on this survey, responses will be anonymous. Individuals will not be identifiable from the data.

For some of you, data management may be neither interesting nor relevant. In that case, please answer only the first question, and return the form.

Some of you with particular interest or responsibility in data management and use, may wish to have longer term or more specific input into the evolution of data systems in the EOC. The survey form invites you to identify yourself, so that I can contact you for this purpose. During May and June I plan to visit a number of CSIRO laboratories to follow up the survey and to obtain additional user perspectives. In particular, I will then be seeking user responses to evaluate the prototype dataset Directory (the CEOS International Directory Network) and inventory manager (the NASA Information Management System), both of which are accessible on the EOC homepage, http://www.eoc.csiro.au

Please contact Murray Wilson on 06 216 7197 or murray.wilson@cossa.csiro.au if you have any questions about the survey, or the EOC research project to which it contributes.

The survey is very simple and should take less than 20 minutes to complete. I would appreciate your response by 1 April 1997. Thank you for your time and interest.

Yours sincerely

Jeff Kingwell
Data base and Access Tools Project.

28 February 1997
APPENDIX E: Questionnaire

Survey on Earth observation data systems

General instructions

Please tick the answer which is most nearly correct. Choose only one response per question.

Section A: Role of respondent

This section addresses your role in relation to Earth observation.

1. Data systems for Earth observation are relevant to my work or are of interest to me

   - Yes [please continue with survey]
   - No [please write your name here and return form without answering further questions: Your name: .................................................................]

2. My function in relation to Earth observation data is primarily that of a

   User

   Provider/manager

Section B: Directories

This section addresses what kind of Directory services you require.

A Directory gives information about the existence and location of datasets or thematic collections. It specifies a custodian of the dataset and usually provides a basic description of the collection (for example, the period and geographic extent covered, revision frequency, accuracy and access constraints applicable). Such descriptive and identifying parameters are called metadata.

An example of a Directory system containing references to international space and Earth science collections around the world is the CEOS International Directory Network (IDN). A prototype Australian node for this is available through http://eoc.csiro.au

NB This is not yet well populated with Australian EO data sets, but the EOC is considering this step.
The IDN uses a simple international metadata template or standard, known as DIF, and used by many or most space agencies and data centres.

An example of a Directory containing reference to national datasets is the National Directory of Australian Resources (NDAR), operated by the National Resources Information Centre of the Department of Primary Industries and Energy. NDAR is available through http://www.nric.gov.au:80/nric/data/data.html

NDAR uses a simple, unique Australian metadata template, devised by the Australian and New Zealand Land Information Council and known as the ANZLIC metadata standard. This has been formally adopted by many (most) Australian agencies affiliated with ANZLIC, including state government land management organisations.

3. For my work in Earth observation

I already know the whereabouts and characteristics of the datasets I require

I sometimes need to trace datasets or find their characteristics

4. For my work, a functioning, well-populated on-line Directory containing metadata about Earth observation datasets would, if available, be:

<table>
<thead>
<tr>
<th>1. Not useful</th>
<th>2. Useful, but low priority</th>
<th>3. Very useful, high priority</th>
</tr>
</thead>
</table>

5. If it was available, I would probably use a well-populated on-line Directory containing metadata about Earth observation datasets held outside Australia:

<table>
<thead>
<tr>
<th>1. Never</th>
<th>2. Sometimes - less than once a month</th>
<th>3. About once a week</th>
<th>4. More than once a week</th>
</tr>
</thead>
</table>

6. If it was available, I would probably use a well-populated on-line Directory containing metadata about Earth observation datasets held in Australia:

<table>
<thead>
<tr>
<th>1. Never</th>
<th>2. Sometimes - less than once a month</th>
<th>3. About once a week</th>
<th>4. More than once a week</th>
</tr>
</thead>
</table>
7. Bearing in mind that different Directories may (will probably) use different metadata elements and user interfaces, I would:

| 1. Prefer to use a **single** Directory containing information about **both** Australian and non-Australian datasets | 2. Not care whether I used **two separate** Directories, for Australian and non-Australian datasets | 3. Prefer to use **two separate** Directories, for Australian and non-Australian datasets |

**NOTE FOR QUESTION 7:** assume that each Directory contained the same level of detail about the datasets.

8. Suppose CSIRO was to place information about Australian/CSIRO datasets on an easily accessible Directory which already contained information about non-Australian datasets. Should CSIRO then also put the same information on a purely national Directory?

| 1. No, this is **not necessary** | 2. Yes, if the relevant information can be transferred automatically | 3. Yes, if the relevant information can be transferred automatically or with only a small manual effort | 4. Yes, this is **important and should be done even if it required more than 0.1 staff years per year effort in CSIRO.** |

9. I believe that CSIRO should

| 1. Encourage Divisions and laboratories to use whichever metadata fields or standard that best suits their own purposes | 2. Encourage Divisions and laboratories to adopt a widely used international metadata standard | 3. Encourage Divisions and laboratories to adopt a widely used national metadata standard | 4. Not care about metadata or its standards |
**Section C: Inventory management**

This section addresses what kind of inventory management services you require. "Inventory management" includes the storage, selection and retrieval of individual data elements or granules - such as a specific AVHRR, Landsat or SPOT scene, or a particular calibration measurement.

An example of an inventory management system is the Information Management System (IMS), developed by NASA for its Mission to Planet Earth. A version of this system is available for trial on: http://eoc.csiro.au

NB This is not yet well populated with Australian EO data granules, but the EOC is considering this step.

10. I already have an efficient inventory management system for the EO data I need to use.

- [ ] Yes
- [ ] No

11. For my work, the most important feature of an inventory management system is:

<table>
<thead>
<tr>
<th>1. Ability to file or store a data item</th>
<th>2. Ability to locate a particular data item</th>
<th>3. Ability to browse a quick look of an item</th>
<th>4. Ability to place a request for a data item</th>
<th>5. Ability to perform features 1-4</th>
</tr>
</thead>
</table>

12. Which of the following is most nearly correct?

1. Most of the EO data I need is stored and managed locally. I only need inventory management tools for local data.

2. I do need to refer to EO data held elsewhere, but other agencies (ACRES, Bureau Met., NASA, AGSO, others) already operate the on-line inventory management systems I need.

3. I do need to refer to EO data held elsewhere, but other agencies (ACRES, Bureau Met., NASA, AGSO, others) should develop the on-line inventory management systems I need.

4. I do need to refer to EO data held elsewhere. I need an inventory management system which will help me locate EO information within CSIRO, and which will connect me to databases of other agencies.
Section D: CSIRO management of commercially-acquired data

This section addresses approaches to commercial EO data acquired by CSIRO.

Background:

Each year, CSIRO spends in excess of $100,000 to purchase commercial EO data from ACRES. Until 1996, COSSA and ACRES operated a discount and data share agreement, allowing a 20% discount for any CSIRO purchases of ERS, TM and MSS data. These data could be used throughout CSIRO. New arrangements now apply, under which a 10% discount is payable for individual Landsat or Spot purchases up to $50,000. The rate progressively increases to 35% for individual orders of $500,000 or more.

13. In relation to commercially EO data acquired by CSIRO:

1. Each Division or laboratory should purchase data as required, and manage their own commercial data archive independently.

2. Metadata details on product acquisitions should be available through an inventory management system accessible to anyone in CSIRO, so units can check, prior to purchase, if a particular scene is already in CSIRO. They can then negotiate an access arrangement with the unit holding the data.

3. Data should be purchased, catalogued and managed much like books in the current CSIRO library system. Divisions and units should buy what they require, but the data should then be available for use by other CSIRO groups for a nominal sum.

4. Commercial EO data should be purchased, catalogued and managed corporately, like the system for licensed Off the Shelf software. A CSIRO group wanting a copy would pay a discounted sum, offsetting the initial purchase price.

Section E: Products and services

This section addresses approaches to EO products and service provision. In the following questions, “market” also includes non-commercial transfer (for example, for internal scientific use).

14. Lack of standardisation in product format limits the market to which CSIRO EO products can address

   Agree
   Disagree
15. Lack of standardisation in product specifications or accuracy limits the market to which CSIRO EO products can address
   
   Disagree
   Agree

16. Lack of ready information about what EO data and products exist limits the market which CSIRO can address
   
   Agree
   Disagree

17. In general, when EO data products are originated by a scientific research group, and those products may have a market outside this group;

   1. an independent, possibly commercial entity should manage, distribute and improve the product

   2. a group of information management or marketing specialists from the same organisation should manage and distribute the product. The scientific group which originated the product should be acknowledged or rewarded, and should be involved in product improvement

   3. for as long as the product continues to require R&D, the originators should retain full control over product distribution, documentation, and product improvement

18. In CSIRO, the best model for using, distributing and improving EO products is:

| 1. Single data management infrastructure, access policy, standards, products and pricing. Product improvement by negotiation between central structure and users. | 2. Single data management infrastructure, access policy, standards, products and pricing. Structured consultation with users and originators on product improvement. | 3. One product catalogue. Agreed access policy, standards and pricing for basic products and limited higher level products; decentralised decisions for other products. Structured consultation with users and originators on improvements to certain products. | 4. Interoperable inventory systems/ product catalogues. Some common basic products. Ad hoc consultation with users and originators on product improvement. | 5. Decentralised and independent infrastructure, access policy, standards and pricing. Product improvement determined by negotiation between each data centre and its clients. |
Section F: Priorities.

This section addresses your perception of priorities for EO data management in CSIRO

19. The **single** most urgent area for improvement is:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

20. Thank you for your time. If you would like to discuss in more detail some of these data management issues, please give your name below:

YOUR NAME: ............................................................................................................
APPENDIX F: Interview questions

Interview questions
Earth Observation Data Systems: Interview questions and format

Interviewee ID #:............................

Format

a. Introductions, purpose, permission. (3 mins)
b. Reiterate history of initial questionnaire; follow-up questions and clarifications. (10 mins)
c. Further questions (below) 30 Mins
d. Can you show me how you use EO data and information in your work? 30 mins.

Note: Question D. omitted from telephone interviews or when otherwise not appropriate.

Part A. Purpose and Permission

This research is being carried out as part of the requirements for a Master of Science course. It is also intended to contribute to the improvement of Earth Observation data and information systems and services, as part of an Earth Observation Centre research project.

Comments from interviewees will remain anonymous.

Results will be available to interviewees through EOC reports and the thesis.

1. In your questionnaire reply of March 1997, you indicated that you would be prepared to discuss Earth Observation data system requirements in more detail. May I ask you some further questions on this field?
2. May I record this interview?
3. May I use your anonymous comments for this research and work based on it?
Part B Follow up to Questionnaire
Comments or unclear answers to the initial survey may be clarified or expanded during this section of the interview.

Part C Further Questions

1. In the past 12 months have you used or explored:
   - the CEOS International Directory Network, IDN
   - the NASA Information Management System, IMS
   - the CEOS Information Location System, CILS?

2. How many times have you used these systems? Is your use increasing, decreasing or remaining the same?

3. Can you comment on whether these systems are useful to you in your work?

4. Which of the three have most value, and why?

5. How can they be improved?

6. Please briefly describe your work.

7. What sort of features or services in Earth Observation data management systems would bring most benefit to your work?

8. Apart from Earth observation, for what other purposes do you use electronic information systems?

9. Do you think electronic information systems for your use should be integrated, or stand-alone for different purposes?

10. In your opinion, what are the most significant current data and information requirements of the community of Earth Observation researchers?

11. For CSIRO researchers, what are the most significant current data and information needs of any kind?
12. Are your Earth observation data requirements primarily for **raw data** or for **processed data products**?

13. (For those who answered “processed data products” to the previous question): what are the most important things you would need to know about these products?

14. Who in your work area makes decisions about what data systems and services you use?

15. In your opinion, is it likely that IDN, IMS and/or CILS will be used in your area in the future?

16. Of any system, what is likely to be the preferred or prime system for your Earth observation data management needs in the next three years, and why?

17. In your experience, does CSIRO encourage the development of information systems and the delivery of information?

18. Do you have any other questions or comments?

**Part D (if appropriate)**

Can you show me how you obtain and use Earth Observation data for your work?
**APPENDIX G: Interview details**

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Tape No.</th>
<th>Journal notes</th>
<th>Date</th>
<th>Interview duration</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>T1</td>
<td>v2 p85-88</td>
<td>4.2.98</td>
<td>70 minutes</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>T2</td>
<td>v3 p1-3</td>
<td>13.3.98</td>
<td>45 minutes</td>
<td>Telephone. Side B (last third of interview) blank.</td>
</tr>
<tr>
<td>7</td>
<td>T3</td>
<td>v3 p25, v4 p31</td>
<td>19.5.98</td>
<td>75 minutes</td>
<td>Side B blank</td>
</tr>
<tr>
<td>9</td>
<td>T4</td>
<td>v3 p46-48</td>
<td>16.6.98</td>
<td>45 minutes</td>
<td>Telephone. Interviewer only on tape.</td>
</tr>
<tr>
<td>10</td>
<td>T5</td>
<td>v4 p5-8</td>
<td>4.8.98</td>
<td>40 minutes</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>T6</td>
<td>v3 p51-53</td>
<td>30.6.98</td>
<td>30 minutes</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>T7</td>
<td>v3 p36-38, v4 p31</td>
<td>15.6.98</td>
<td>55 minutes</td>
<td>Telephone. Tape bad after 280</td>
</tr>
<tr>
<td>14</td>
<td>T8</td>
<td>v3 p33-35</td>
<td>11.6.98</td>
<td>55 minutes</td>
<td>Telephone</td>
</tr>
<tr>
<td>15</td>
<td>T9</td>
<td>vol 2, p90-91</td>
<td>11.2.98</td>
<td>25 minutes</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>T10</td>
<td>vol 2, p4-5</td>
<td>13.3.98</td>
<td>33 minutes</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>T11</td>
<td>vol 4, 26-28</td>
<td>24.8.98</td>
<td>47 minutes</td>
<td>Telephone</td>
</tr>
<tr>
<td>20</td>
<td>T12</td>
<td>vol 3, 43-45, vol 4, 36-</td>
<td>16.6.98</td>
<td>43 minutes</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>vol 3, p16-18</td>
<td>17.4.98</td>
<td>40 minutes</td>
<td>Tape corrupt</td>
</tr>
<tr>
<td>23</td>
<td>T13, T11</td>
<td>vol 4, 22-26</td>
<td>24.8.98</td>
<td>68 minutes</td>
<td>Telephone</td>
</tr>
<tr>
<td>X</td>
<td>T14</td>
<td>vol 3, p39-41, vol 4, 33-36</td>
<td>15.6.98</td>
<td>55 minutes</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX H: Questionnaire analysis

Introduction

In March 1997, a short survey was undertaken of CSIRO 76 researchers who identified themselves as stakeholders of the Earth Observation Centre (EOC). The purposes of the survey were:

1. to find out what sort of Earth observation data management services and systems this group required;

2. to identify individuals interested enough to discuss these matters in more detail;

and

3. to elucidate the views of CSIRO researchers to various practical and policy matters facing CSIRO in Earth Observation data management.

The survey was an initial step in making available data systems that suit the objectives and work patterns of researchers in the field.

Before its distribution to the sample group (with return-address, prepaid envelope), the survey was “road tested” by a knowledgeable volunteer, and modified slightly for greater clarity. The Questionnaire and covering letter are shown in Appendix E.

This section summarises the response to this initial survey.
Summary

1. 34 replies (45%) were received from 76 requests, which is about an average response level. Of the replies, 8 indicated that the respondent had left CSIRO or had no interest in Earth observation data systems, leaving 26 useful responses (34% of survey population).

2. Of the 26 ‘useful’ respondents, 20 categorised themselves primarily as data users and four primarily as data managers/providers. Two responded as being ‘primarily’ both.

3. The survey results give useful guidance on pressing issues such as the level and speed of CSIRO’s response to national spatial data directory initiatives.

4. The survey appears to indicate a greater emphasis, on the part of CSIRO researchers and data managers, on standards and software systems, as opposed to hardware and networks.

5. The survey has helped to better identify a) who is interested b) what issues need closer attention

6. About half of the respondents indicated interest in further discussion.

Follow up

More detailed responses were obtained through during fifteen interviews carried out during February to August 1998. The Interview questions are given in Appendix 6 and the results are reported in Chapter 4.

This interview stage focussed on the on the specific information needs of the users of Earth observation data and of Earth observation data systems; and on the relationship between EO data systems and other information networks/systems in CSIRO.
Responses to each question

Question 1:
Data systems for Earth observation are relevant to my work or are of interest to me

- Yes [please continue with survey]

- No [please write your name here and return form without answering further questions: Your name: .......................................................

Responses 1:
26 yes
8 no

Comment 1
Identification of the “not interested” group allowed us to refine our mail list, reducing the possibility of wasting researcher's time with further irrelevant communication.

Question 2
My function in relation to Earth observation data is primarily that of a

User

Provider/manager

Responses 2:
Users 20
Providers/managers 4
Ticked both 2
Comment 2
The purpose of the question was to allow testing of several hypotheses relating to attitudes of different staff categories towards information systems. It further allowed a better understanding of the roles of users of data management systems.

Question 3
For my work in Earth observation

I already know the whereabouts and characteristics of the datasets I require

I sometimes need to trace datasets or find their characteristics

Responses 3
Six usually know where to find data: eighteen sometimes need to search for it while two agreed with both statements.

As might be expected, a greater proportion of the "managers" knew where to find information. Four of the 20 "users" and two of the 4 "managers" usually know where to find the data they need, and knew the data characteristics. One of the two "dual" respondents had to search out data or data characteristics. Two respondents usually knew where to get required data but sometimes had to search.

Comment 3
As expected, the proportion of "managers" or data custodians in this position was higher relatively than that of users. On the face of it, a high proportion of the sample could benefit from a data management system with search and metadata functions. In principle, this proportion may be higher since even some of the "satisfied" could benefit from improved retrieval performance.
Question 4
For my work, a functioning, well-populated on-line Directory containing
metadata about Earth observation datasets would, if available, be:

<table>
<thead>
<tr>
<th>1. Not useful</th>
<th>2. Useful, but low priority</th>
<th>3. Very useful, high priority</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Responses 4
One person (a "manager") said "not useful".

Twenty five (96%) said an on-line Directory would be useful, but only 7 (all "users") thought it was "very important and a high priority".

Comment 4
A high proportion of replies (73%) indicated that an on-line Earth science data directory would be either "not useful" or "useful but low priority". This response warrants closer attention: had previous experiences with older, low-functionality directories coloured the reply? What were some of the higher priorities - did they involve higher level data management systems?

Some of these questions should be followed up by interview.

Question 5
If it was available, I would probably use a well-populated on-line Directory containing metadata about Earth observation datasets held outside Australia.

<table>
<thead>
<tr>
<th>1. Never</th>
<th>2. Sometimes - less than once a month</th>
<th>3. About once a week</th>
<th>4. More than once a week</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Responses 5
One did not reply, while four expected never to need a Directory of "external" data. Twenty (77%) thought they would use it less than once a month while two (8%) would use it about weekly.

Comments 5
There was little difference in the response of users and providers (presumably, the "providers" would be seeking information on behalf of end-users, in any event).

A possible follow up question is "do users only rarely need foreign-sourced data, or do they need it frequently but already know exactly where to find it?" A subsidiary question is "do researchers rely on the invisible college [e.g. colleagues' word of mouth], in preference to automated searching, to find information they require?"

It is possible that some respondents thought "data held outside Australia" meant the same thing as "data not of Australia". Further questioning may be required to confirm whether researchers have discounted, or otherwise accounted for, little-known EO data about Australia archived elsewhere.
Question 6
If it was available, I would probably use a well-populated on-line Directory containing metadata about Earth observation datasets held in Australia -

1. Never
2. Sometimes - less than once a month
3. About once a week
4. More than once a week

Responses 6
All twenty six would probably make use of a Directory of Australian data, of whom 22 (85%) would use it less than once a month; while four (15%) would use it about weekly.

Comment 6
There was little difference between responses of users and providers. A Directory containing references to "Australian" data would appear to have slightly utility that a Directory containing only external data.

Was this because respondents assume it would be easier/quicker to GET data from Australian sources (or metadata from distant web sites with intervening low band width path segments), or does this in fact reflect accurately the usage pattern?
Question 7

Bearing in mind that different Directories may (will probably) use different metadata elements and user interfaces, I would:

| 1. Prefer to use a single Directory containing information about both Australian and non-Australian datasets | 2. Not care whether I used two separate Directories, for Australian and non-Australian datasets | 3. Prefer to use two separate Directories, for Australian and non-Australian datasets |

NOTE FOR QUESTION 7: assume that each Directory contained the same level of detail about the datasets.

Responses 7

One person did not reply. Seven (five of the users and two of the providers) preferred a SINGLE directory with both Australian and non-Australian datasets. Thirteen (nine of the users, two of the providers and both dual roles) didn't care if there were one or two Directories. Five (all users) preferred to have SEPARATE directories for Australian and non-Australian data.

Comment 7

One of the reasons for this question was to determine the acceptability, to the EOC community, of participating in national Earth science/spatial data directory initiatives. An example is the National Spatial Data Directory of NRCC (National
Resources Information Centre of the Commonwealth Department of Primary Industries and Energy1, on
<http://www.nric.gov.au/nric/data/anzlic_search2.html>). An alternative or additional option was participating in global initiatives, such as the International Directory Network of CEOS, on
<http://atlas.eoc.csiro.au:8000/~md5devel/>

This issue is also considered in the questions that follow (8-9). It would seem that given only 27% preferred a single directory, it would be acceptable for CSIRO to contribute metadata to both systems (provided the effort/cost was limited, see below).

If there are clear differences in search, order and use patterns between researchers employing mainly foreign or mainly Australian data, it may turn out that if relevant CSIRO data are supplied to both national and international directories, most researchers need only consult a single directory.

Why provide CSIRO metadata to an international directory at all? The answer lies in the collaborative nature of the information sharing. Most CSIRO researchers searching for data on a global directory will be looking for externally-sourced data, though there is always a chance that an unexpected "resource discovery" trail will lead to a colleague or another CSIRO laboratory. However, unless those who benefit from an information catalogue refresh and extend it, there is no incentive for other providers to do likewise, and the system becomes moribund.
Question 8
Suppose CSIRO was to place information about Australian/CSIRO datasets on an easily accessible Directory which already contained information about non-Australian datasets. Should CSIRO then also put the same information on a purely national Directory?

<table>
<thead>
<tr>
<th>Responses 8</th>
<th>1. No, this is not necessary</th>
<th>2. Yes, if the relevant information can be transferred automatically</th>
<th>3. Yes, if the relevant information can be transferred automatically or with only a small manual effort</th>
<th>4. Yes, this is important and should be done even if it required more than 0.1 staff years per year effort in CSIRO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses 8</td>
<td>1. (No) 10 (38%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-3. Yes, if automated or little effort required 15 (58%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Yes, even if more than 0.1 staff-year pa required 1 (4%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comments 8
This was one of the few issues which respondents volunteered additional comments on:

“There are significant issues in translating between metadata specifications, unless there is a very restricted scope and a minimal set of descriptors”

From a respondent who DID NOT think CSIRO should put its metadata on a purely national directory:

“CSIRO should establish its own inventory which can then be made available externally”

From another who also DID NOT think CSIRO should put its metadata on a purely national directory:

“Provided one can immediately search [the global directory] on Australian data, and response time is good (or minimal).”

And again “No, if the single one [global] had similar level of detail.”
Question 9

I believe that CSIRO should

| 1. Encourage Divisions and laboratories to use whichever metadata fields or standard that best suits their own purposes | 2. Encourage Divisions and laboratories to adopt a widely used international metadata standard | 3. Encourage Divisions and laboratories to adopt a widely used national metadata standard | 4. Not care about metadata or its standards |

Responses 9

Everyone cared! Four thought that Divisions and laboratories should choose according to their own purposes; 23 (88%) thought a widely used standard should be adopted: of these, 14 supported an international, and 9 a national (one person suggesting that the “national” standard should be the same as the international one).

I was a little surprised that as many as four felt that a “laissez-faire” approach was preferable. This could indicate:
- a low premium placed on data sharing with colleagues

- a belief that the laboratory concerned possessed unique data whose management required a unique approach

- a low demand for integration of local data with that derived elsewhere.

Further investigation may uncover whether any of these factors apply: if so, this would reduce the internal market for unified EO information systems and, therefore, the incentive to promote or develop them.
Question 10
I already have an efficient inventory management system for the EO data I need to use.

Yes

No

Responses 10
Yes 6 (4 users, 1 provider, 1 “both”)

No 17 (14 users, 3 providers, 1 user/provider)

Equivocal 1

No reply 2.

Comment 10
Lessons from the “satisfied six” may be useful to others. The “equivocal” responder noted that their existing system was adequate for some data types/sets, but not for others: if the number of data types and sources increases for other users, then similar inadequacies may eventually appear in the existing data management systems.

Question 11
For my work, the most important feature of an inventory management system is:

<table>
<thead>
<tr>
<th>1. Ability to file or store a data item</th>
<th>2. Ability to locate a particular data item</th>
<th>3. Ability to browse a quick look of an item</th>
<th>4. Ability to place a request for a data item</th>
<th>5. Ability to perform features 1-4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Responses 11

The largest group (13) wanted all features. Eight (combined total of those answering 1 and 2) regarded storage and locating to be the most important features, while 3 regarded browse as the most significant. No-one regarded ordering as the most critical. Three did not respond on this issue.

Comment 11
I had expected “managers” to be more concerned with “filing and finding” than users. In fact, of the four data managers, one placed this first; two regarded the full range of functions as necessary; and the other did not respond. Of the twenty users, eight would apparently be satisfied with storing, locating and browsing while ten required the full function range given in the question. Of the two with dual responsibilities, one placed “locating” as the first priority and the other wanted all features.

Question 12
Which of the following is most nearly correct?

1. Most of the EO data I need is stored and managed locally. I only need inventory management tools for local data.

2. I do need to refer to EO data held elsewhere, but other agencies (ACRES, Bureau Met., NASA, AGSO, others) already operate the on-line inventory management systems I need
3. I do need to refer to EO data held elsewhere, but other agencies (AGRES, Bureau Met., NASA, AGSO, others) should develop the on-line inventory management systems I need.

4. I do need to refer to EO data held elsewhere. I need an inventory management system which will help me locate EO information within CSIRO, and which will connect me to databases of other agencies.

Responses 12
A significant number (9) of the 26 respondents require only data held locally.

Eleven needed inventory information both from within CSIRO and from other organisations. Three felt that other agencies already operate the on-line retrieval systems they required, and another two felt that those agencies should do so.

Comment 12
There would appear to be a significant need for internal data management systems, especially if these could be connected to or operate in conjunction with management systems of kindred agencies. The survey bifurcated into two nearly equal groups: those requiring only local data management services, and those needing data from throughout CSIRO and in other agencies.

Question 13
In relation to commercial EO data acquired by CSIRO:

1. Each Division or laboratory should purchase data as required, and manage their own commercial data archive independently.
2. Metadata details on product acquisitions should be available through an inventory management system accessible to anyone in CSIRO, so units can check, prior to purchase, if a particular scene is already in CSIRO. They can then negotiate an access arrangement with the unit holding the data.

3. Data should be purchased, catalogued and managed much like books in the current CSIRO library system. Divisions and units should buy what they require, but the data should then be available for use by other CSIRO groups for a nominal sum.

4. Commercial EO data should be purchased, catalogued and managed corporately, like the system for licensed Off the Shelf software. A CSIRO group wanting a copy would pay a discounted sum, offsetting the initial purchase price.

Responses 13
Only one respondent preferred scenario 1 (the existing situation).
Sixteen (including all four data managers) preferred option 2 - transparent possession and negotiated sharing. Eight preferred scenario 3, similar to current arrangements for resource sharing across CSIRO libraries.
Six preferred centralised management of commercial EO data.
Four gave multiple answers: one person equally preferred options 2, 3 and 4.
Two people equally preferred options 2 and 3.
One person equally preferred options 2 and 4.
Comment 13
Clearly there is dissatisfaction with the present wasteful and disorganised acquisition of commercial EO data. In this case, a significant number (nearly half) are attracted to a corporate service, whether "centralised" as in software licensing or distributed as in the CSIRO library network. Perhaps this unusual readiness indicates the researchers believed this was a simple and non-intrusive task for "headquarters", or they possibly felt the case for a corporate approach was so strong that the traditional preference to local control should be overlooked. However, all four responding data managers preferred a semi-autonomous system in which knowledge of data's existence was shared, but the "custodian" could negotiate on a case by case basis with other "internal" users.

While this approach mitigates against consistency or organisation-wide policies, it probably accurately reflects attitudes to other forms of resource sharing, especially in the context of sharing experimental data. It allows the custodian to assess each approach on its merits, and to determine the relevant factors in each case - these could include past level of collaboration and competition; subjective factors (gut feeling, personality matches/mismatches); practicality (e.g. degree of proximity).

Question 14
Lack of standardisation in product format limits the market to which CSIRO EO products can address

Agree

Disagree
Comment 13
Clearly there is dissatisfaction with the present wasteful and disorganised acquisition of commercial EO data. In this case, a significant number (nearly half) are attracted to a corporate service, whether “centralised” as in software licensing or distributed as in the CSIRO library network. Perhaps this unusual readiness indicates the researchers believed this was a simple and non-intrusive task for “headquarters”, or they possibly felt the case for a corporate approach was so strong that the traditional preference to local control should be overlooked. However, all four responding data managers preferred a semi-autonomous system in which knowledge of data’s existence was shared, but the “custodian” could negotiate on a case by case basis with other “internal” users.

While this approach mitigates against consistency or organisation-wide policies, it probably accurately reflects attitudes to other forms of resource sharing, especially in the context of sharing experimental data. It allows the custodian to assess each approach on its merits, and to determine the relevant factors in each case - these could include past level of collaboration and competition; subjective factors (gut feeling, personality matches/mismatches); practicality (eg degree of proximity).

Question 14
Lack of standardisation in product format limits the market to which CSIRO EO products can address

Agree
Disagree
Responses 14
An unusually high total of six did not reply. Fourteen agreed and six disagreed. A majority of both users and providers (ten from fourteen, and three from four, respectively) agreed with the proposition.

Comment 14
The Simpson report, and EOC projects such as the common AVHRR processor, are firmly premised on the belief that lack of standardisation of products is a major market disincentive. On this basis, the level of disagreement (one third of respondents) is surprising.

Six did not reply, with several annotating a query against this question. Perhaps the word “format” was unclear - some may have thought of presentation format (form of the data) - as was intended - whereas others may have interpreted it as exchange format (ie media). Interview questions should clarify this issue, which is quite important in terms of research directions and commercialisation strategy.

Question 15
Lack of standardisation in product specifications or accuracy limits the market to which CSIRO EO products can address

Disagree

Agree

Responses 15
18 agreed and five disagreed. All four managers agreed with the proposition. Three users did not respond.

Comment 15
Ratio of yes to no was higher than for previous question (3:6 cf 2:3).
Question 16
Lack of ready information about what EO data and products exist limits the market which CSIRO can address

Agree
Disagree

Responses 16
19 agreed and three disagreed, with four not replying.

Comment 16
There seems a clear consensus, by a ratio of six to one, that CSIRO EO products have poor visibility in the market.

Question 17
In general, when EO data products are originated by a scientific research group, and those products may have a market outside this group;

1. an independent, possibly commercial entity should manage, distribute and improve the product

2. a group of information management or marketing specialists from the same organisation should manage and distribute the product. The scientific group which originated the product should be acknowledged or rewarded, and should be involved in product improvement

3. for as long as the product continues to require R&D, the originators should retain full control over product distribution, documentation, and product improvement
Responses 17
Only two people supported the “outsourcing” or commercial spin-off model 1. Fourteen supported a partly centralised, partly federated model 2; ten supported the fully decentralised model 3.

Comment 17
It was suggested by one respondent that the phrase “for as long as the product continues to require R&D” skewed positive responses to 3. Even so, a majority wished to put some distance between the researcher and the product (by preferring options 1 or 2). This may indicate a desire not to have scientific time “wasted” in operational (i.e. information service) tasks.
Question 18

In CSIRO, the best model for using, distributing and improving EO products is:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>infrastructure, access policy, standards, products and pricing.</td>
<td>infrastructure, access policy, standards, products and pricing.</td>
<td>Agreed access policy, standards and pricing for basic products</td>
<td>Some common basic products.</td>
<td>Ad hoc consultation with users and originators on product improvement determined by negotiation between each data centre and its clients.</td>
</tr>
<tr>
<td>Product improvement by negotiation between central structure and users.</td>
<td>Product structured and limited Ad hoc consultation with users and originators on product improvement.</td>
<td>Consultation and limited improvement by negotiation between originators on central product decisions for product improvement.</td>
<td>Some common basic products.</td>
<td>Product improvement determined by negotiation between each data centre and its clients.</td>
</tr>
<tr>
<td>Product improvement</td>
<td>Product improvement</td>
<td>Product improvement</td>
<td>Product improvement</td>
<td>Product improvement</td>
</tr>
<tr>
<td>by negotiation with users and originators on central product improvement.</td>
<td>by negotiation with users and originators on central product improvement.</td>
<td>by negotiation with users and originators on central product improvement.</td>
<td>by negotiation with users and originators on central product improvement.</td>
<td>by negotiation with users and originators on central product improvement.</td>
</tr>
<tr>
<td>298</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Responses 18
Centralised Option 1: nil

Centralised Option 2: 1

Federated Option 3: 12 (10 users, 1 provider, 1 both)

Decentralised Option 4: 7 (three users, three providers, one provider equally preferred option 5)

Fully independent Option 5: 4 (two users, one provider who equally preferred option 4, and one user/provider)

Nil response: 2.

Comment 18
The question, although complicated, indicates little support for Option 1, resembling a fully implemented, organisation-wide information management system (such as the original conception of NASA’s EOSDIS). There was also little support for a fully decentralised, laissez-faire approach (Option 5). A federated structure (Option 3) had most support, with the next most popular “lightly federated” option 4. What is perhaps surprising is the apparent preference of managers for loosely coupled or independent systems (4 and 5). Could this indicate the belief that cooperation/federation is “too hard”? Does it show that managers prefer more degrees of freedom?
Question 19

The single most urgent area for improvement is:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Responses 19
Networks: 2
Mass storage: 2
Inventory software: 6
Standards: 13.

Comments 19
Interesting from several aspects: 1. The relative lack of emphasis to technological fix either in networks, storage or operating systems; 2. The unanimity of the four managers, who all pointed to "standards" as the priority issue; and 3. The almost equal split of users choosing operating systems and standards as the first priority.

Question 20
Thank you for your time. If you would like to discuss in more detail some of these data management issues, please give your name below:

YOUR NAME: ...........................................................................
Responses 20

More than half of the respondents (15 from 26) indicated interest in further discussion of these issues. A further name was added after the initial presentation of results at the EOC Workshop on 29 July 1997.

Comment 20

I interviewed the fifteen ‘interested parties’ in the next stage of the study. The principal issues for this stage 2 were:

1. Evaluation of the IMS and IDN systems
2. Individual workplace requirements for data management systems
3. Relationship/commonalities between EO data management approaches and other CSIRO data systems.
REFERENCE LIST


archiving and distribution of space science data. Strasbourg: European Science Foundation.


MacDonald Dettwiler Pty Ltd. (1994a). *Pre-feasibility study for the development of an Australian Earth observation network.* Sydney, Australia: MacDonald Dettwiler.


McIntosh, M. (1996). [e-mail: CSIRO staff]. *Minister's Statement on staff numbers*. Canberra: CSIRO Chief Executive's Office.


