APPENDIX 3C

Excerpts from NAS Reports

Opportunities for Research in the Geological Sciences

Committee on Opportunities for Research
in the Geological Sciences

Board on Earth Sciences

Commission on Physical Sciences,
Mathematics, and Resources

NATIONAL ACADEMY PRESS
Washington, D.C. 1983

RECOMMENDATIONS

At no time in history have the geological sciences offered more potential for mankind than today. Developments and breakthroughs are increasing our fundamental understanding of earth processes and structures, affording many practical applications in the search for resources and the use of land and water.

The geological sciences are bristling with excitement, and stimulated by new concepts with related controversy and debate. Previously accepted models of geology are being tested regularly with new measurements and new analytical capabilities. Accepted concepts of the earth's crust are being challenged in ways that have ramifications for resource exploration as well as for the basic understanding of geological processes. Models used to estimate resources and to aid in land-use planning are in need of continuous revision with improved understanding of the dynamics of the earth. The global plate tectonic model, the availability of improved instrumentation, and the importance of understanding the earth to give substance and safety to an increasing number of inhabitants of the earth combine to provide unique challenges and opportunities for geological research in the 1980s.

This report examines those research opportunities that are pertinent to the programs of the National Science Foundation's Division of Earth Sciences. Although the committee wrote most of the report, many of the ideas and statements were provided by a wide circle of earth scientists in the United States. Research priorities for the NSF's Division of Earth Sciences have been developed for the decade of the 1980s using all available information.

The recommendations presented in this chapter are of two types: those related to policy, research requirements, and budget; and those related to research strategy and priorities.

A. PROGRAM AND BUDGET REQUIREMENTS

1. Research Policy

The NSF's Division of Earth Sciences should pursue a philosophy of synergistic programs. A large mix of research is recommended in this document because a narrow approach to priority research topics (Chapter 1, section B) would fail to reach the desired goals. Increased funding for basic research is fully justified by the potential discoveries, and the additional effort should be an essential part of national science policy (p. 81).

1. Priority should be given to the scientific programs that will make the most progress in the light of scientific breakthroughs and technical developments.
2. Planning for research programs should include the realization that advances from a proper mix of efforts can be greater than the sum of the research results from individual grants and projects.
3. Funding should emphasize basic research. The activities of other sectors, including mission agencies, industry, and foreign investigations, should be taken into account. The implicit relationship between scientific research and societal needs must be recognized.

2. Research Goals

The NSF's Division of Earth Sciences research programs should be directed at obtaining a new level of understanding of the structure, composition, energetics, and evolutionary history of the earth with emphasis on the continental lithosphere and its margins. The highest priority research in this report attacks the continental lithosphere (see Chapter 1, sec. B). Improved methods of measurements, interpretation, and synthesis present opportunities for large advances in understanding the earth, including natural hazards and mineral resource distribution.

3. Research Requirements

The recommended research goals will require expanding support for some programs and taking new initiatives for others. Requirements (pp. 79-81) for attacking the research priorities (Chapter 1, section B) are as follows:

1. Standard Research Grant Program. The individual research grant program is the backbone of the Foundation's funding. The individual grants are now inadequately funded and have low graduate student participation. Grant allocations should be more than doubled (p. 77-79). The research program should be biased toward the research priorities given in this report.

2. Major Experimental Projects and Facilities. Medium- and large-scale consortium-type projects have special requirements because they need support exceeding that obtainable by individual research grants and because special management and services are required to sustain their operations (p. 73-74).

2a. New Initiative: A large array of portable digital seismographs is needed to define heterogeneities of the lithosphere by variations in the propagation paths and character of seismic waves from controlled and earthquake sources. This method and the deep reflection seismic method provide the most direct data on variations of geologic features with depth in the earth. The initial program will cost NSF about $1-2 million annually (p. 75-76).

2b. Augmented Initiative: Deep seismic reflection studies are providing new insights about the details of geologic structure at depths to 40 km. This is a major new tool for looking at the deep crust, which is a research frontier. This very successful NSF project is currently supported at $3-4 million per year. The funding for seismic reflection studies should be doubled (p. 73).

2c. New Initiative: Deep continental drilling provides the only means of directly measuring rock properties, establishing geologic structure, and observing earth processes at exceptional depths. It provides the observational data to confirm or deny inferences made from near-surface observations. We recommend that a program of continental drilling be initiated at $4 million per year (p. 74-75).

2d. New Initiative: The global system of permanent seismographs, of which the U.S.-supported Worldwide Standardized Seismograph Network (WWSSN) is a major part, provides data on the internal structure and composition of the earth as well as the characteristics of earthquake sources. The large majority of stations have analog-recording, which constrains data analysis. The development of digital seismograph equipment and the consequent increased capability in processing and analysis of digital data offer the potential for large advances in our knowledge of the earth. Some expansion of seismic research using current facilities is allowed in the recommended budget of Table 1, but no funds were allocated for the development of an advanced global digital seismograph network. Planning for such a network should be initiated (pp. 75-76; p. 45).

3. Laboratory Instrumentation. The inadequacy of laboratory instrumentation in American universities has been documented. For essential widely required instrumentation, such as electron microscopes, high-pressure and temperature equipment, etc., the NSF's Division of Earth Sciences is allocating about $5 million per year. This amount is inadequate and should be increased severalfold (pp. 70-71; p. 42).

3a. New Initiative: A program needs to be established to put expensive equipment such as ion microprobes and accelerator-based mass spectrometers in a few locations that are committed to be available for many users and that have dedicated support staff. A pilot program will require $1-2 million annually (p. 73-74; p. 42).

4. Data Management and Facilities. Rapid advances in digital computers and their extensive use in acquiring, manipulating, and storing data have made it possible to address problems that previously could not be attacked. Data management must be addressed by the scientific community, and facilities to handle large quantities of data must be made available. Many complex problems involving comprehensive data sets are amenable to solution only through using large computers. These investigations are usually at the cutting edge of science, and computer time should be made available for the required analysis of data. Table 1 implicitly allocates funds for work of this nature but does not allow funds for a dedicated facility though one may eventually be needed (p. 45; pp. 72-73).

4a. New Initiative: NSF's Division of Earth Sciences should support the development of dedicated data centers as they are needed. Problems in data handling arise from large volumes of data: need to access, needed storage capacity, need for cataloging, and needed dissemination. Geophysics and to a lesser degree geochemistry have the major needs. Seismologists have assessed the data management problem and are recommending a National Center for Seismological Studies. This is an example in a field where $1-2 million a year will be required from several federal agencies to fund such a center, probably through a consortium (p. 73).

Scientists should be strongly encouraged to invest time in data management. NSF's Division of Earth Sciences should allow a small portion of research grant funds to be so directed and should request that original data with informative headings be stored in a retrievable manner. A catalog of multiuse data should give the specifics of the data so that other users have easy access to information (pp. 72-73).

4. Research Budget

Progress toward the solution of the high-priority research problems discussed in Section B of this chapter requires a strategy using various modes of operation and a mix of techniques, with financial support greater than that now provided. The requested budget for FY 1984, $42 million, is used here as a base on which to project anticipated additional costs. It is assumed that the base budget will continue to support individual grants, but the focus of those grants should shift toward the research priorities in the next section. The funding table that follows allocates increments to this base budget among the eight research priorities presented in the table. The funding increments recommended for FY 1985 represent estimates of the amounts needed to take full advantage of immediate opportunities for fruitful research in each priority area. The funding recommendations for FY 1990 are naturally more speculative, but they represent projections for anticipated funding requirements. Work by individual investigators supported by grants awarded in response to unsolicited proposals will remain as the chief mechanism for carrying out the priority research. Examples include isotope and trace-element geochemistry, structural geology, laboratory measurements of physical and chemical properties of rocks, biostratigraphy and climatology, paleontology, volcanology and magma genesis, and mantle structure and plate tectonic theory. All of the funding increments in Table 1 include appropriate increases in the basic grants program. Other costs for research requirements are identified in Table 1 in the research initiative and the footnotes.

B. RESEARCH STRATEGY AND PRIORITIES

1. Research Strategy

We believe an effective strategy for research in the geosciences over the coming decade can be built around the following main thrusts, pursued in a coordinated fashion:

- Coordinated study of the evolution of the continents, including: (a) systematic testing and extension of relevant aspects of plate tectonics as it applies to the continents; (b) probing of phenomena not readily explained by plate tectonic theory in terms of either time or behavior; (c) widespread application of new methods and ideas in established fields such as geologic mapping, isotopic geochemistry, geophysical profiling, geochemical analysis, experimental petrology, geochronology, modern paleontology, and satellite imagery; and (d) investigating the physical and chemical character of the crust by deep-scientific drilling.
- Development, testing, and suitable modification of the plate tectonics model using the best available technology.
- Utilization of existing and improved techniques for ship-based and other investigations of problems pertaining to the evolution of ocean basins and the sea waters they contain.
- Investigation of geological processes in order to better evaluate and predict geological hazards, environmental effects, and resource bases.

Within these four thrusts are many specific approaches that have been discussed in detail in reports prepared for the National Research Council, the NSF, and other groups. These reports were prepared by groups of experts in each of the several areas. In this report we have

TABLE 1 Recommended Funding for Research Priorities and Initiatives

		PROPOSED ANNUAL INCREASES ABOVE NSF'S DIVISION OF EARTH SCIENCES 1984 REQUESTED BUDGET (In Millions of 1984 Dollars)	
RESEARCH PRIORITIES*	FOOTNOTES	FY 1985	FY 1990
Structure and composition of continental lithosphere	1,2	$ 7.0M	$ 22.0M+
Sedimentary basin development	1	2.0	3.5+
Magma generation and and emplacement	1,3	3.5	7.0+
Physical and chemical properties of rocks	3	2.0	6.5+
Physics of tectonic processes	2,3	3.0	5.0+
Convection of earth's interior	2	1.0	2.0
Evolution of life	-	1.0	2.0
Surficial processes	-	1.5	5.0
Proposed total annual increase		$ 21.0	$ 53.0
Total 1984 requested budget		$ 42.0	$ 42.0
Total proposed budget for years shown		$ 63.0	$ 95.0
RESEARCH INITIATIVE			
Continental drilling (including planning, site surveys, and background studies for dedicated drilling and implementing holes of opportunity).		4.0	20.0
	GRAND TOTAL	$67.0	$115.0

* Data management needs to be considered in all priorities.

\+ In addition to the figures shown for research priorities, the research initiative for continental drilling will complement and supplement a number of scientific priorities but will be of most value to these five.

1. Includes funding for deep seismic reflection (Research Requirement 2b) and refraction profiling (Research Requirement 2a).
2. Includes funding for limited maintenance and upgrading of existing seismic networks and deployment of some new networks for specific studies (Research Requirement 2d and 4a).
3. Includes funding for replacement or upgrading of existing inadequate instrumentation, as well as development of key frontier instrumentation (Research Requirement 3 and 3a).

attempted to integrate their ideas into a coordinated approach.

2. Research Priorities

Chapter 3 summarizes the most challenging opportunities facing the geological sciences today. Not all of these can be addressed simultaneously or with equal emphasis, and choices must be made. We have based our choices on the following criteria:

- Will research on the subject fill an important gap in the intellectual fabric of the geological sciences?
- Is the work feasible in terms of both technology and the present state of geological knowledge?
- Is there a perceived need for the answers to research questions in basic or applied science?

Following these precepts, we have singled out eight topics or areas as having the most promise for advancing geology in the next decade. Suggestions for promising approaches to the solution of these major problems are offered in Chapter 3. We believe most groups of earth scientists would agree with these topics but that their rank ordering would vary from group to group, depending on the personal preferences of individual members. Therefore, with the exception of the item on continental lithosphere (Topic 1), which we consider of overriding importance and priority, no significance is intended by the ordering of the topics.

1. A MORE DETAILED AND ACCURATE DEFINITION OF THE STRUCTURE AND COMPOSITION OF THE CONTINENTAL LITHOSPHERE, INCLUDING THE CONTINENTAL MARGINS

Our knowledge of the continents at depth is woefully incomplete, yet the continental crust is where mankind lives and from whence we draw our resources. The emphasis in this high priority topic is on the acquisition of precise information on the details of the geometry and composition of the units comprising the continental lithosphere. Technological means are at hand to increase vastly our understanding of nearly all processes operating at or near and beneath the surface to great depths, and to study the results of these processes as they have operated through long reaches of geological time. The starting point for deciphering the history of the earth during the last 3.8 billion years is accurate knowledge of the present state. We therefore seek knowledge of the detailed structure and composition of the continents comparable to that which has been so recently achieved for the ocean basins. The greater complexity of the continents requires application of the most modern techniques, involving workers in nearly all subfields of the science, to acquire this knowledge and understanding. An expanded program of large-scale geological mapping and geochemical sampling in selected areas is a necessary component of these studies. The use of high-resolution deep seismic profiling techniques, reflection and refraction, must be increased. In addition to such surface-based investigations using geophysical and geochemical methods, direct sampling by deep drilling can provide additional required data. Although much of the needed research can be done by individual researchers with basic grant support, consortia should be formed as needed to work on large problems involving major equipment and facilities for multidisciplinary and interdisciplinary investigations (pp. 28-35).

The high priority assigned to this topic is supported by international recognition of its importance. As stated by Dr. D. A. Bekoe, President of the International Council of Scientific Unions,* "ICSU considers the study of the lithosphere to be one of the most important scientific activities for the future of mankind."

2. QUANTITATIVE MODELS FOR SEDIMENTARY BASIN EVOLUTION

The rocks and fluids of sedimentary basins are the results of tectonic, erosional, sedimentary, diagenetic and thermal processes. The main thrust in basin analysis is the development of integrated quantitative models of basin evolution. The resulting understanding of the rates and intensities of the processes can lead to predictive models for basin formation and the origin of the ground water, mineral deposits and hydrocarbons that

---

*Dynamics and Evolution of Lithosphere: The Framework for Earth Resources and the Reduction of Hazards. ICL Report No. 1, Inter-Union Commission on the Lithosphere, April, 1981.

they contain. Research can be conducted through small grants that support field and laboratory studies. Some projects may be more amenable to attack by consortia, especially those that combine academia and industry. Seismic surveys, drilling programs, and computer modeling of such processes as water-rock reactions and hydrodynamic flow in the basins will provide supplemental information (pp. 23-24).

3. IMPROVED UNDERSTANDING OF MAGMA GENERATION AND EMPLACEMENT

Investigation of the formation of magmas, their geochemistry, movements in the Earth's interior, and extrusion requires a combination of field and laboratory studies. A global data set must be assembled. A major goal is the formulation of models of magma formation that can account for the variety of magma chemistries. Understanding magma genesis will require replacement of outdated or inadequate equipment, and the purchase of innovative frontier instrumentation. The successful solution of this problem will lead to improved methods for locating mineral deposits and better understanding of volcanism and geothermal processes. Some seismic studies will be required, as well as deep drilling. Approaches for research on the associated tasks are given in Chapter 3 (pp. 37-42).

4. KNOWLEDGE OF THE PHYSICAL AND CHEMICAL PROPERTIES OF ROCKS

Laboratory studies of properties of rocks and rock-forming minerals, under controlled conditions of temperatures and pressure (up to 3000°C and 1.5 megabars) provide basic data for quantitative models of geological processes. Studies of isotopes, trace elements and trace amounts of organic compounds are included. This research requires specially equipped facilities and perhaps can best be implemented through consortia of universities and by use of governmental laboratories. The need for major new instrumentation is great. Facilities are required for studies of the flow of fluids through large samples and the interaction of fluids with rocks. Improved laboratories for investigating fracture and creep processes should be equipped. In addition to laboratory studies, in situ measurement of rock properties should be made, especially in deep drill holes (pp.45-48; pp. 23-26).

5. A BETTER UNDERSTANDING OF TECTONIC PROCESSES, THE PHYSICAL AND CHEMICAL STATES THAT PRODUCE THEM, AND THE STRUCTURES THAT RESULT

Study of relations between stress and structures is required to provide models for tectonic processes. Scales of observation must range from mineral grains to continental blocks. Models for tectonic processes can, in turn, foster investigations of the mechanical and chemical interactions at plate boundaries and the energy budget of tectonic processes. Needed research includes field observations, laboratory research, and theoretical investigations. Among the special facilities required are laboratories for studies of rock mechanics and petrophysics, strategically located digital seismograph networks, and adequate computers. Special observations include in situ stress measurement, especially in deep drill holes. Field mapping of structures and stratigraphy, as well as seismic profiling, will be needed in key areas (pp. 42-45).

6. A MODEL OF CONVECTION IN THE EARTH'S INTERIOR

Current evidence supports the concept that convection drives the dynamics of the Earth, but the shape, size, and distribution of the postulated convective cells are not known. New data, some provided by satellites, and recent theoretical developments in continuum mechanics make this problem more amenable to attack than heretofore. This fundamental problem calls for a broad interdisciplinary approach in which geophysical and geodetic observations are made using a variety of modern instruments and analyses using large computers. Although much of the needed work can be done by individual investigators, they will require access to global data sets acquired by a great variety of sensors and archived in well-managed data centers (pp. 48-53).

7. EVOLUTION OF LIFE

The origin and evolution of animal and plant life, including the causes of mass extinctions and explosive or punctuated evolution, represent a first-order intellectual challenge. The relationships between organic evolution and the changing compositions of the atmosphere and oceans, paleoclimatic variations, continental rearrangements as they change the pattern of flow of ocean currents, and impact phenomena involving extraterrestrial objects all require interdisciplinary investigation. An adequate paleontological data base coupled with paleoclimatological, paleoatmospheric, and paleo-oceanographic information obtained by geochemists, paleontologists, stratigraphers, and tectonicists is required to address these problems. Modeling of the catastrophic effects of major extraterrestrial impacts and their effects on life forms requires the integration of paleontology, cratering mechanics, cosmochemistry, atmospheric sciences and biology (pp. 21-22; pp. 27-29).

8. SURFICIAL PROCESSES

Knowledge of the dynamic processes that shape and that have shaped the surface of the earth enables geologists to understand earth history better and to predict changes in the landscape, such as those produced by landslides, floods, and volcanic eruptions. It also helps us deal with the stress that man imposes on his environment. Advances in understanding the origins of landforms and rock types require the study of present-day processes and the extrapolation of results to ancient conditions. Such research requires a large data base for statistical modeling applicable to surficial processes of the past and future (pp. 18-22).

Seismographic Networks:

Problems and Outlook for the 1980s

Report of the Workshop on Seismographic Networks
Committee on Seismology
Commission on Physical Sciences, Mathematics, and Resources
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1983

1

EXECUTIVE SUMMARY

An earthquake produces seismic waves, which radiate from its focus, traveling around and through the earth, with size and persistence proportional to the dimensions of the source. Only a very few of the thousands of earthquakes cataloged annually affect mankind directly. Most are perceptible only to seismographs--the scientific eyes and ears of the seismologist. From the seismograms, which display ground motion associated with the passing waves, comes our knowledge of the global distribution of earthquakes, of the internal structure of the earth, and of the earthquake source process. Interpreting the recorded seismogram requires sophisticated analysis procedures. Recent advances in analytical methods and instrumentation have increased dramatically the information to be gained from seismograms, but acquisition of adequate seismological data requires wide coverage by seismographs, globally, nationally, and regionally. Instruments must be maintained and upgraded regularly with the latest technology. Effective management is crucial for operations and data handling. All of these needs require adequate financial support over long periods of time.

Seismographic networks provide data essential to programs such as the mitigation of earthquake hazards, the definition of geological structure on the margins and within tectonic plates, the safe siting of dams, power plants, and other critical facilities, and the investigation of dynamic processes of the earth. Operating a typical seismographic network is not overly expensive, but it does require dedication of time and talent by seismologists who run the stations. In many cases the major rewards are in providing data to help solve problems of national and global significance.

The large number of questions on seismographic networks brought in recent months to the Committee on Seismology is strong evidence that there are critical problems with network operations. At the Workshop on Seismographic Networks, prompted by these questions, participants considered global, regional, and national networks collectively as an integrated system and also as entities with specific problems. This report discusses each component of the system in terms of rationale and problems, giving recommendations for solutions. A brief statement follows of major problems and major recommendations for the global, regional, and national networks.

Global Networks. Global networks are expected to provide for the scientific community a data base that continues indefinitely. Unfortunately, managing agencies find it difficult to recognize this long-term scientific importance. The service function of the networks, i.e., providing data for other users, must be considered in funding decisions by the managing agency. Global networks require continuing financial support at an adequate level. It is recommended (1) that consideration be given to transferring management responsibility for the global network from its present organizational base to another location within the U.S. Geological Survey or even to another agency, if such a change seems clearly advantageous; (2) that stable funding for global networks be sought from normal budgetary requests from within the U.S. Geological Survey, from the Defense Advanced Research Projects Agency, and from other agencies that use data from the networks; (3) that access to digital data and use of those data be improved while networks continue to meet fully the demand for and the global coverage provided by analog (i.e., visible) data at the present time; and (4) that procedures be established and funding be provided for the orderly and continuing interagency transfer of the most recent instrumentation and technology.

Regional Networks. Regional network operations are beset with problems falling into three categories: functional definition, funding difficulties, and operational problems. Functional definition is the planned lifetime of a network, and a realistic estimate of it needs to be provided. Funding difficulties are of two types: a lack of stability on a year-to-year basis, and the vulnerability of research funding being decreased to maintain network operations in times of fiscal stress when research funding is mixed with basic operational costs of the network. Operational problems are seen in a lack of coordination among networks, the need for a more standardized data base management system, and a growing obsolescence of network equipment. These problems are interrelated and difficult to order in importance.

Recommendations are (1) that networks of planned, limited lifetime be reviewed every three to five years with respect to objectives and performance; (2) that the provision of data fundamental to research on seismotectonic processes and earthquake occurrence in the region be acknowledged by funding agencies as the main purpose of regional networks; (3) that an adequate data set from all regional networks be archived; (4) that data formats be standardized; and (5) that operations of networks be coordinated.

National Network. The concept of a national network lacks general acceptance and widespread support by the U.S. seismological community, within which there is at present little coordination of network operations. The concept is sound, and support will grow with formulation of a suitable plan for implementation.

Working Group on Seismic Networks. It is recommended that a Working Group on Seismic Networks be set up under the Committee on Seismology to provide continuity and uniformity in consideration of the various policy matters arising in network seismology. This group will provide the review functions recommended throughout this report for global, regional, and national networks. It should evaluate continually the health and status of regional networks, and advise on the development of a national network.

The contributions to the earth sciences from seismic networks of all types have been substantial in the past two decades. We have entered the 1980s with major advances in data acquisition, management, and processing techniques now available to seismology. The challenge is to build effectively on the present structure of networks, creating a new capability for addressing the next level of difficulty in the exciting problems of geoscience.

The Lithosphere

Report of a Workshop

U.S. Geodynamics Committee
Board of Earth Sciences
Commission on Physical Sciences, Mathematics and Resources
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1983

SUMMARY OF SCOPE AND PURPOSE

This document is the report of a week-long workshop on problems relating to the interpretations of the composition and dynamics of the lithosphere. The workshop was convened by the U.S. Geodynamics Committee in Austin, Texas, in March 1982. Approximately 75 scientists, working actively on such problems, were invited to attend. A special effort was made to include active, young earth scientists. Forty-four of the invitees and a few visitors attended all or parts of the workshop.

The intent of this workshop was to emphasize areas of unresolved controversy and to identify assumptions underlying proposed models and hypotheses. A wide range of topics was discussed, dealing not only with the lithosphere itself, but also with possible interactions between the lithosphere and underlying mantle, down to and including the core-mantle boundary zone. Diverging, often mutually exclusive hypotheses and assumptions, were identified and discussed.

Quite obviously, the topics treated were those of most interest to the participants. Emphasis, very broadly, was on the physical and chemical properties of the lower crust and the subcrustal lithosphere: the physical and chemical characteristics of the prominent seismic discontinuities down to the core-mantle boundary; the nature and patterns of possible convection within the mantle and its relation to the generation, subduction, and intermixing of lithospheric and mantle material; the location and nature and evolution of reservoirs supplying magmas to the crust; and the various models that have been proposed to account for the location, nature, and geological history of these magma reservoirs. The general applicability of the plate tectonics model was assumed, but virtually every widely accepted explanation for the dynamics of that model and of possible unrelated phenomena such as deep-mantle plumes and hot spots was brought into question.

The last day of the workshop was devoted to the preparation of a report by groups of participants. Divergent models and hypotheses were emphasized; no attempt was made to reach consensus and it was decided that subject matter would not be updated or expanded significantly. This report should be of particular interest to scientists and students actively pursuing research in areas related to lithospheric composition and dynamics, but the committee hopes and expects that it will stimulate both discussion and research throughout the earth science community. This report is not intended to provide a plan of action or to serve as a thorough and coherent review of the vast literature related to investigations of the lithosphere.

Data Management and Computation

Volume 1: Issues and Recommendations

Committee on Data Management and Computation
Space Science Board
Assembly of Mathematical and Physical Sciences
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1982

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the Councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the Committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the federal government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

Available from
Space Science Board
2101 Constitution Avenue
Washington, D.C. 20418

Contents

1

Executive Summary

> *Since the first satellites had orbited, almost fifty years earlier, trillions and quadrillions of pulses of information had been pouring down from space, to be stored against the day when they might contribute to the advance of knowledge. Only a minute fraction of all this raw material would ever be processed; but there was no way of telling what observation some scientist might wish to consult, ten or fifty, or a hundred years from now. So everything had to be kept on file, stacked in endless air-conditioned galleries, triplicated at the three centers against the possibility of accidental loss. It was part of the real treasure of mankind, more valuable than all the gold locked uselessly away in bank vaults.*
>
> ARTHUR C. CLARKE, *2001*

The Space Science Board (SSB) of the National Research Council has had a continuing concern with questions relating to data management and computer utilization in space science. There is concern over problems with early planning of systems for scientific data acquisition, reduction, and distribution; the quality, timeliness, and accuracy of sensor data; the allocation of processing functions on the spacecraft and the ground; programmer productivity; software compatibility and portability; and the cost to the scientific user of acquiring the data. The large amount of data that have been acquired in the past, currently being acquired, and planned to be acquired in the next decade presents a challenge that will require the establishment of principles and organizational, technical, and scientific solutions.

Although future science data management will be strongly influenced by advances in technology, from both cost and performance viewpoints, this committee believes that the majority of the current data problems are not due to technological barriers. Furthermore, projected data problems can also be solved through employing projected advances in technology, providing that the management of data operations is properly organized.

I. SUMMARY OF PROBLEMS RELATED TO ACQUISITION, ANALYSIS, AND DISTRIBUTION OF SPACE-SCIENCE DATA

In the course of its deliberations, CODMAC identified a number of data-management and computation problems associated with space-derived data. These problems, organized in sequence according to the CODMAC charge, are listed below. It must be emphasized that these problems do not apply uniformly across all missions or all disciplines. Different missions, data centers, and disciplines have had varying degrees of success in approaches to data-management and computation problems. In later sections of this report, we identify specific approaches and attempt to determine the factors that lead to the success or failure of the approach. These determinations form the basis of our recommendations, summarized in Section IV below.

1. In the area of data-system planning, three problems have been identified:

(a) There is commonly a lack of scientific involvement in data-system planning during early mission planning and during the system development phase. Typically, the interdisciplinary nature of data is not fully recognized, and, therefore, data systems are frequently not properly implemented for their actual use.

(b) Generally, data-system and data-analysis activities are not adequately funded. Underfunding results from at least three related causes: when there is insufficient planning in the early mission phases, the required funding will often be underestimated; overruns that occur during mission system development may absorb the funds allocated for data handling and analysis; and because of imperfections in the flight and ground hardware and software, the data processing may be more extensive than originally estimated.

(c) Often, a responsible scientific group for data management during and/or after missions is not clearly identified.

2. Identified problems related to preprocessing (the processing that converts data as received to the form required by the user) are the result of the huge volume of data collected by NASA missions. During 1978, almost 10^{15} bits were returned from spacecraft, and this number is expected to increase further during the 1980's. The problems are as follows:

(a) Preprocessing of all the data is possible with current computer technology, but inadequate planning and funding have prevented this from happening in most cases.

(b) Current capabilities for on-board preprocessing are insufficiently understood and on-board processors are insufficiently developed to reduce significantly the amount of data that must be transmitted from spacecraft.

3. In the area of data distribution, many problems have been identified:

(a) Commonly, there are long delays between the receipt of data on the ground and the delivery of preprocessed data to the user.

(b) Costs of data for some disciplines are frequently so high that the "small" science user cannot afford to acquire all data needed. Unfortunately, it appears that costs of some data (e.g., from Landsat) may increase by a factor of 5 as the program becomes operational.

(c) In many cases, Principal Investigators (PI's) who have been supplied with raw data that they are obliged contractually to return in corrected or reduced form to a data center after some specified period of time do not do so. In those cases where data are returned to the data center, the documentation is often incomplete, resulting in the processed data being unusable for other investigators.

(d) The user community has great difficulty in determining what data are available. The contents of some data centers (whose mission is to provide a data resource to the user community) are not widely publicized and not widely known.

(e) As a result of problems (b), (c), and (d) above, some data centers have become unable to respond to requests for data in a timely fashion.

(f) As a result of the lack of standardization of data formats (discussed below), it is difficult for users to correlate interdisciplinary data and data obtained from multiple missions or sources.

4. Problems related to data standardization and fidelity affect the capability of the scientist to use the data once they have been located and acquired. Many of the needed data are in widely distributed locations and are difficult to access. The identified problems are as follows:

(a) A wide variety of formats are employed. Users typically must devote considerable effort to understanding and/or modifying the formats of data received from data archives. This problem is particularly serious when a scientist needs to correlate data from multiple sources.

(b) Data archives generally contain insufficient or inaccurate information concerning the quality and limitations of the archived data.

(c) Data archives generally contain insufficient ancillary data such as time, attitude, orbit, or sensor calibration data.

(d) Some data are only resident at a PI location and are difficult to obtain.

5. The identified problems related to software development are as follows:

(a) Software is frequently not adequately documented.

(b) Software is not transportable as a general rule. Lack of transportability arises in part from inadequate documentation, but also because insufficient attention is given to transportability during software development.

(c) Current software development methods are costly. Lack of software transportability contributes to this because software is sometimes independently developed several times.

(d) All too often the software development is incomplete at the time of launch of a mission.

6. A problem has been identified with respect to the distribution of computational capabilities:

(a) Currently many scientists must travel to remote locations in order to obtain adequate computational resources to perform their analyses. This is an inefficient system, which reduces scientific productivity because of the inability of scientists to have access to working data files. Scientists perform best in their own environment and must have the capability to perform needed computations locally through either the use of their own computer, a distributed network, or a combination of their own computer with a distributed network.

7. In the area of mass data storage and retrieval, several problems have been identified:

(a) Often data must be purged from the archives in order to make room for current data. In many cases, data have been purged without adequate consultation with the scientific community.

(b) Catalogs associated with data archives frequently do not provide enough information for the interested user to determine whether the archived data will be useful for a particular research project. Also, catalogs are not widely available in many instances.

(c) Usually, data archives do not include an adequate browse capability. Such a facility would allow the interested user, at his home institution, to locate and inspect data sets rapidly and to select those that will be useful for further analysis.

(d) Current mass storage technology is inadequate to store at sufficiently low cost all the data returned by NASA missions. In addition, magnetic tape, the storage medium for the vast majority of the science data in archives currently, has a serious deterioration problem with time, and many of the newer technologies either have known deterioration problems or have not been available long enough to permit an assessment of their potential for storing science data.

8. With respect to interactive processing, several problems have been identified:

(a) There are a wide variety of man-machine interfaces with little standardization in hardware, operations, languages, and algorithm definition. Each time a user employs a new system, considerable effort must be expended in learning the characteristics of that system.

(b) Interactive terminals for software development, program execution, and scientific data analysis are not widely employed because of the continued use of old computer technologies.

(c) Little thought has been given to a dynamic man-machine interface with regard to scientific real-time and interactive control of flight experiments.

(d) The use of artificial intelligence and robotics for Earth-orbit and deep-space missions has not been fully exploited.

9. Finally, two problems have been identified that are applicable to several elements of the charge:

(a) Scientific users of space-acquired data frequently need the same data sets as do operational or commercial users. In many instances this results in high costs to scientists and long delays in obtaining the data.

(b) In many cases current technology is not exploited or implemented in present data systems.

(c) A number of NASA-sponsored programs (NEEDS, ADS, SSDS, JPEEIS, for example) are currently under way that are designed to alleviate many of the problems discussed above. Centralization of data systems seems to be a common theme of these programs. Such centralization has the potential to reduce active involvement by the scientific community significantly. Furthermore, these programs seem to be uncoordinated within the agency, and they seem to be proceeding without regard to developments in the industrial community.

II. FINDINGS

Since the overall objectives of this report are to identify problems associated with the management and manipulation of space-acquired data and to provide technological, programmatic, and organizational recommendations that will result in more scientific return from the data, we offer the following general conclusions. Our recommendations are given in Section IV, below.

1. There are problems with the way data are currently managed. The distribution, storage, and communication of data currently limit the efficient extraction of scientific results from space missions.

2. Technological barriers are not the major impediment to improved data handling. While certain areas of technology will need continued development (notably, on-board spacecraft systems), most of the technology required for

successful science data management either exists at present or will be available in the near future. Nevertheless, although economic factors will continue to impose technical limitations on data management, the current problems are due mainly to the structures and limitations of our institutions and management operations.

3. Data-handling problems can be significantly reduced by restructuring the data chain (from acquisition to analysis) to adhere to principles for successful science data management, as discussed in this report.

III. PRINCIPLES FOR SUCCESSFUL SCIENTIFIC DATA MANAGEMENT

In this section we state several principles on which successful scientific data management must be based. These principles were derived from the experiences described in the case studies of Chapter 3.

1. *Scientific Involvement* There should be active involvement of scientists from inception to completion of space missions, projects, and programs in order to assure production of, and access to, high-quality data sets. Scientists should be involved in planning, acquisition, processing, and archiving of data. Such involvement will maximize the science return on both science-oriented and applications-oriented missions and improve the quality of applications data for application users.

2. *Scientific Oversight* Oversight of scientific data-management activities should be implemented through a peer-review process that involves the user community.

3. *Data Availability* Data should be made available to the scientific user community in a manner suited to scientific research needs and have the following characteristics:

(a) The data formats should strike a proper balance between flexibility and the economies of nonchanging record structure. They should be designed for ease of use by the scientist. The ability to compare diverse data sets in compatible forms may be vital to a successful research effort.

(b) Appropriate ancillary data should be supplied, as needed, with the primary data.

(c) Data should be processed and distributed to users in a timely fashion as required by the user community. This responsibility applies to Principal Investigators and to NASA and other agencies involved in data collection. Emphasis must be given to ensuring that data are validated.

(d) Proper documentation should accompany all data sets that have been validated and are ready for distribution or archival storage.

4. *Facilities* A proper balance between cost and scientific productivity should govern the data-processing and storage capabilities provided to the scientist.

5. *Software* Special emphasis should be devoted to the acquisition or production of structured, transportable, and adequately documented software.

6. *Scientific Data Storage* Scientific data should be suitably annotated and stored in a permanent and retrievable form. Data should be purged only when deemed no longer needed by responsible scientific overseers.

7. *Data System Funding* Adequate financial resources should be set aside early in each project to complete data-base management and computation activities; these resources should be clearly protected from loss due to overruns in costs in other parts of a given project.

IV. RECOMMENDATIONS

Recommendations are presented under three categories: policy, technology, and general.

Policy Recommendations

1. Principles for successful data management have been defined above; they address scientific involvement and oversight, the availability of data, suitable processing facilities, software procedures, data storage, and funding. We believe that adherence to these principles will significantly improve the extraction of scientific information from space-acquired data. *We recommend that these principles become the foundation for the management of scientific data.*

2. The most successful cases of extraction of information from space-science activities have had the vigorous and continuing involvement of scientists in planning and implementing the acquisition, processing, archiving, and distribution of data. *We recommend that such active involvement be strongly encouraged and supported in the future.*

3. We define a Scientific Data Management Unit as a group of active scientists and support staff with suitable computational resources at a particular institution. These units can be implemented via a variety of organizational structures—the Principal Investigator (PI) unit, the project unit, or other interdisciplinary units that may transcend either the PI or project units in scope. *We recommend that these units be organized in accordance with the principles and guidelines presented herein. The requirements of individual disciplines, however, must be the prime concern in organizing such units.*

4. *Data-analysis funds should be adequate and should be protected against*

reprogramming as the result of such occurrences as hardware overruns and mission time delays.

Technology Recommendations

5. *We recommend that NASA have an ongoing technology management activity encompassing all areas of data systems.* The activity should be independent of any specific program and should take an overview of technology in order to establish whether NASA's needs for space data systems are being developed adequately by industry, universities, or other government agencies. The program should formulate research and development efforts in those areas where NASA and science or applications users would benefit from new developments. In the course of this determination, possibilities for technology transfer and utilization from and by industry, universities, and other government agencies should be explored to the fullest extent feasible. This program must be broader in scope than current NASA programs and must involve scientific users in order to determine the real requirements.

6. *NASA's approach to technology developments at the component level through the systems level should emphasize the capability to implement new technologies rapidly as these technologies evolve.* Modular architectures with standardized interfaces offer one approach that enables such implementation in response to requirements for growth and flexibility. One specific area that deserves special NASA attention and perhaps funding is the potential for flight use of commercial processors that have been hardened for space and military applications.

7. As new technologies emerge and/or evolve in the areas of processors, memories, computers, communications, and data handling, *NASA should have a continuously ongoing program to test, adapt, and qualify for space application those elements that will advance data management for science and applications.*

8. *Specific implementations of technology that are needed in support of science and applications data management are*:

Asynchronous data-handling systems capable of priority-controlled data acquisition, large buffer capacity, packetized data management, and retransmission capability on demand. (The NEEDS program is addressing this issue. CODMAC concurs that this is a worthy effort.)

High-capacity, all-electronic, on-board data storage with storage capabilities of up to 10^8 bits available to individual PI's, and 10^{11} bits available as a centralized mass storage device by 1985. The capacity available in each category should be capable of increasing by two orders of magnitude by 1995.

Data compression algorithms with adequate documentation that permit scientists to select and use them for special applications.

Centralized on-line mass storage devices with capacities to 10^{14} bits by 1985 for data storage within NASA.

A standard archival storage system that is compatible with NASA, NOAA, and DOI archival data requirements.

Generalized data-base-management software that emphasizes NASA scientific and applications data-base requirements.

9. *NASA should become more active in the area of satellite communications technology* in order to enable the acquisition and distribution of wideband scientific and applications data. Particular emphasis must be given to low-cost, two-way communications, to the handling of multiple wideband satellites in multiple access modes, and to low-cost receiving stations for individual scientists and applications users. As a part of this effort, *NASA should set a goal of increasing uplink capabilities to satellites by an order of magnitude during the 1980's.*

10. *Electronic transfer of data to the investigators should be implemented where economically feasible.*

11. *NASA should begin studies of man-machine interactions, requirements, and needed developments for all phases of scientific and applications data management*, beginning with real-time data acquisition and proceeding to final analysis and interpretation of data. Specific emphasis must be given to the user interface for communication with computers, including voice interaction; to data presentations, both optical and nonoptical; and to analysis and interpretation.

12. *NASA should examine its current technology development efforts to determine whether they are duplicating developments now under way in industry, DOD, or other government agencies.* This activity can be accomplished in conjunction with Recommendation 5.

General Recommendations

13. *We recommend that greater emphasis be given to documentation of space-science and application data to make them interpretable and useful to scientists not directly associated with initial acquisition of data.* Included in such documentation should be information and software to extract physical units from the raw data. Those who gather the data should also be responsible for assessing their validity as part of the documentation.

14. In some cases, data acquired from past missions have not been properly archived. *We recommend that under scientific overseers NASA vigorously pursue the archiving and preservation of such space-science data that should be permanently stored, and data no longer required for future scientific uses should be purged.*

15. *We recommend that more emphasis be given to production of user-oriented catalogs and browse files for space-science data.*

16. Software and related issues are a continuing source of problems to NASA and the science community. Consistent standards for documentation, development methodologies, languages, protocols, libraries, and portability do not exist. *NASA should establish a software organization, possibly within a structure with broader data-management activities, with responsibilities to create software policies and guidelines, to generate technical standards, to monitor enforcement of policies and standards, and to assure the availability of information related to existing software programs.* The organization should address the software issues as a joint effort among scientists within NASA, government agencies, industry, and universities. Also, the resultant standards must be compatible with the activities of international standards organizations. Specific activities within this organization would include:

(a) *The development of software acquisition management guidelines.*

(b) *The establishment of a unified software library* to minimize multiple developments of standard software.

(c) *Concentrated research on software metrics.*

(d) *The establishment of a practice of software discipline* for software developed by NASA and scientists, including, but not limited to, such practices as structured programming requirements and design languages.

17. Space-science processing requires a variety of computational capabilities. Current systems in use consist of both centralized and decentralized facilities, with centralized facilities using primarily large computers and decentralized facilities consisting primarily of minicomputers. At the same time, the sophistication of scientific models are requiring ever-increasing processing power, and the interdisciplinary nature of scientific processing is expanding the needs for access to multiple, remote data sets. Fortunately, technology advances in computers and distributed computing networks are compatible with these requirements. *We recommend that NASA work closely with the scientific community to assure access to adequate computational capabilities, communication facilities and protocols, information directories, and software and format capabilities.*

The Applications Data Service (ADS) and Space Science Data Service (SSDS) programs offer some of the required capabilities. At the same time, these programs may duplicate existing commercial capabilities. The scientific community should participate in the definition and development of these programs.

18. Computing facilities at a number of NASA Centers are a decade behind state-of-the-art systems. *We recommend that computational hardware, software, and interactive terminals be updated more frequently in order to*

keep up with the space-science data loads and developments in computer technology.

19. On-board processing of data will be important in future planetary missions, where telemetry rates constrain the total amount of data that can be returned. It will also be important in future Earth observation missions, where the amount of data collected will be large. *We recommend that greater consideration be given to preprocessing and data compaction schemes and to artificial intelligence and robotics. We further recommend that some degree of on-board preprocessing should also be experimentally implemented and evaluated in selected applications missions, but raw data should be accessible whenever possible.*

20. *We recommend that scientific investigators have access to the raw data from scientific, applications, and operational missions.*

21. *We recommend that NASA evaluate and develop the concept of "electronic browse" capability, allowing users to explore data files via communications links. We recommend that "quick-look" low-resolution data be included for use in electronic data browsing.*

8

Types of Scientific Data-Management Units

> *The open society, the unrestricted access to knowledge, the unplanned and uninhibited association of men for its furtherance—these are what may make a vast, complex, ever growing, ever changing, ever more specialized and expert technological world, nevertheless a world of human community.*
>
> J. ROBERT OPPENHEIMER, *Science and Common Understanding* (1953)

I. INTRODUCTION

The principles of science data management that are discussed in Chapter 7 were derived, in part, from an analysis of examples of present data-management methods, ranging from Principal-Investigator-oriented, through project-oriented, to systems that involve management and processing of data from a number of missions or sources (units of broader scope). Most likely, all three types of data-management units will continue to exist in the future. In this chapter we summarize the problems encountered in the past for each type, and we then show how the application of the principles would have led to a much greater science return. Our discussion does not include the major archival facilities, although these facilities would also benefit from application of the principles.

We end this chapter with a discussion of the Discipline Data Management Unit. We see this type of unit as a natural evolutionary goal. It would serve as a focus for collecting, processing, archiving, and distributing data pertinent to a particular scientific community or for a specific scientific program.

II. PRINCIPAL-INVESTIGATOR DATA-MANAGEMENT ACTIVITIES

Many NASA projects operate with the Principal-Investigator (PI) format, whereby a scientist is selected to provide an instrument or to oversee the development of an instrument to be flown on a spacecraft and to process and reduce the data. The data are usually provided directly to the PI, who then processes them and forwards raw or reduced data to an appropriate data center facility. As discussed, the success of this type of data-management unit has been variable. The best way to illustrate the problems is to summarize in a general way the extent to which each of the principles has been followed:

1. *End-to-End Science Involvement* The PI, usually a scientist, has generally been involved from the beginning through the end of a mission. At times, however, problems arise because too little thought or resources have been given to data management in the planning of the mission.
2. *Oversight Mechanisms* Peer review of the data-management tasks has largely been nonexistent. Peer pressure, to some extent, has been applied. Little thought has been given to archival storage of data.
3. *Production of Usable Data* The main task for the PI has been to reduce the data for his own needs. As a consequence, the data are sometimes not useful to, or not interpretable by, the rest of the scientific community. There have sometimes been long delays between the PI's receipt of the data and the deposition of the data into an archive. The merging or comparison of different data sets (from different experiments) has suffered.
4. *Computational Capabilities* This requirement has been met to some extent, although lack of early planning for data-reduction activities, combined with limited resources, has led in some cases to a poor match and the inability of the PI to process all the data needed.
5. *Software Considerations* The extent to which structured, transportable, documented software has been developed is hard to measure, although usually the software has been developed within the PI's own research group, where little incentive or few resources exist to make software transportable.
6. *Archival Services* The degree to which the PI has felt an obligation to help the user community has varied largely with the personality and philosophy of the PI.
7. *Adequate Financial Resources* Lack of early planning to establish adequate resources for data management and computation has often plagued the PI data-management unit.

III. THE PROJECT DATA-MANAGEMENT UNIT

In more complex missions, teams of investigators are often selected and a number of instruments are flown. In most cases, the project has assumed responsibility for processing and managing the returned data within a common data base. Either raw or processed data sets are then distributed to the teams of investigators. In some of the applications missions, even the teams are nonexistent in early phases of the mission, and the processing, storing, and sometimes the distribution of data are the concern of the project. The success of past Project Data Management Units can be measured against how well the principles were followed:

1. *End-to-End Science Involvement* The degree of science involvement has been variable, ranging from incorporation of scientists in the preplanning through the data-reduction phases to having no scientific involvement at all.
2. *Oversight Mechanisms* In some cases, no oversight mechanism has existed. In other cases, a science steering group existed whose charter has included oversight of data-management tasks. Even in the latter case, too little attention has been given to data-management activities.
3. *Production of Usable Data* For science missions, the Project Data Management Unit has usually led to a greater degree of control of data production than in the PI case. In applications missions, the lack of science involvement has led to severe problems in production of data useful to the scientific community. Problems still exist within Project Units in terms of producing interpretable, documented data with usable catalogs. Again, the incentive for the teams has been largely to reduce, interpret, and publish results for their own needs. Even with science steering groups there has sometimes been a lack of appreciation of the utility of comparing diverse data sets.
4. *Computational Capabilities* There frequently has been less than adequate funding for proper scientific computation capabilities. In addition, computational facilities for processing the data at NASA centers have usually been 5 to 10 years behind in computing technology.
5. *Software Considerations* Numerous examples exist of both poor and good practices with regard to software development. In some cases software written for a given mission has not been documented or used after the mission terminated. In other cases, older software has been inherited and implemented in new missions.
6. *Archival Services* Usually, investigators involved with the project, either directly or indirectly, have been fairly well served by the project data facility. Usually, outside investigators are not serviced (or not well served) by Project Data Management Units.
7. *Adequate Financial Resources* Even with early science involvement, major problems exist with planning and protecting financial resources, mainly

because of cost overruns related to hardware development and because of a lack of early thought in regard to needed financial resources.

IV. DATA-MANAGEMENT UNITS OF BROADER SCOPE

There are a number of examples of Data Management Units that either transcend or are broader than PI and Project Data Management Units. In some cases, Project Data Management Units have grown to a broader scale to include data from other sources and to provide a focus for consortia activities. The Atmospheric Explorer and Einstein Observatory missions are good examples of such cases. In general, units of broader scale consider diverse data sets, have continuing scientific involvement, conform more to the principles for science data management, and thereby maximize the scientific return for the investment. An analysis of how the principles are adhered to follows:

1. *End-to-End Science Involvement* Units of broader scope are usually initiated and implemented by scientists. As such, these units commonly have the highest and most vigorous degree of science involvement in all phases of implementation.
2. *Oversight Mechanisms* There are a variety of structures, both formal and informal, existing within these units.
3. *Production of Usable Data* This task has been the main goal of many of these units. Examples of groups that have this goal include the Lunar Consortium and the UCLA magnetospherics group. Data produced or handled by this category of management unit have been documented to varying degrees, depending on the extent to which the data have been produced for consumption by the user community. In some cases, advanced capabilities such as an electronic browse capability to interrogate the data sets exist. In general, there is a greater degree of concern for timely processing and availability of data; funding limitations have been the primary limitation to data management.
4. *Computational Capabilities* The match between computational requirements and capabilities has been variable. Usually, the minimum capability needed to complete the task is provided because of funding limitations. Systems have been developed that closely match the needs of the units, although existing computational resources are frequently used because of their availability.
5. *Software Considerations* The degree to which these units have taken the lead in proper software development techniques is usually a step above either the PI or the project units. This includes both software for scientific processing and for data management. The skills and needs of multiple users

have resulted in the development of more capable systems. There is more concern with software transportability and documentation.

6. *Archival Services* The service orientation is reasonably well handled by this type of management unit, if only because of informal contact between the scientists involved and the user community. Funding limitations frequently limit the degree of service to a user community, since many of these units are supported on yearly renewable grants and contracts and do not have a funded service obligation.

7. *Adequate Financial Resources* Problems often exist in this area because of insufficient early planning, and in some instances, lack of a long-range financial commitment.

V. THE DISCIPLINE SCIENTIFIC DATA-MANAGEMENT UNIT

It is clear that many of the problems associated with maximizing the science return per investment cost for the PI, the project, and the broader scope units could have been alleviated by application of the principles described in Chapter 7. Science involvement in all phases of the data chain may be the most crucial element. This statement is supported by the observation that the data-management units of broader scope, which have largely been initiated and run by scientists, have usually produced data of higher quality and reliability and of more general use to the user community (see Table 7.1). As noted, however, none of the units of broader scope have really met all the principles.

Although we recognize the need of retaining the PI and the Project Data Management Units, we also recognize that, in many cases, the broader scope management units seem to provide greater science return. We can envisage a data-management unit that is largely based on examples of Units of Broader Scope—the Discipline Data Management Unit (DDMU).

These units would evolve as the need for them arises. For example, a unit of broader scope might start collecting additional data sets because of interest in particular scientific problems and eventually find that its collection has become the primary data and information source for an entire scientific community. Such a unit could even provide the major computational resources and software programs for its discipline. Perhaps the Space Telescope Institute, which will have charge of planning, reducing, archiving, and distributing Space Telescope science data, is close to our concept of a Discipline Unit, though only Space Telescope data and not all astronomical data would be included in the data base. We could also envisage the creation of a Geodynamics Data Management Unit as another example of a DDMU. Such a Unit might house geophysical information (data on crustal movements or potential field data, for example) and geologic information (geologic map data or digital topography, for example) pertinent to tracking and understanding crustal movements and information patterns.

Although each scientific discipline must formulate its own specific requirements for a DDMU, there are several generic requirements that should be applicable:

1. The DDMU employs archive scientists as well as technical and administrative support staff.
2. The DDMU provides the interface between investigators in a scientific discipline and the NASA data-collection process.
3. The DDMU archives and distributes data relevant to its discipline.
4. The DDMU provides scientific oversight for data processing when required.
5. The DDMU exercises scientific oversight of any data purging and is prepared, if necessary, to "rescue" selected data for scientific use.
6. The DDMU provides coordination among disciplines in support of interdisciplinary analysis efforts.
7. The DDMU develops software for general use in a discipline and coordinates software development with other data-management units and with the user community.
8. The DDMU assumes a leadership role in developing software and data standards and formats.
9. The DDMU provides computation facilities for investigators who desire to use the data in the DDMU archives but who do not possess their own dedicated computational facilities.
10. The DDMU advises the user community of the availability of relevant data and assists the user community in obtaining access to these data.
11. In some cases, the DDMU operates the principal archives for its discipline.
12. In some cases, the DDMU creates specialized data bases for general disciplinary use.
13. In some cases, the DDMU creates and maintains models representing the data in its discipline.
14. The DDMU is responsive to advice from a committee of active investigators in its discipline—a users committee.
15. The activities of the DDMU are periodically reviewed by an oversight committee.
16. The DDMU, its users committee, and its advisory committee are jointly responsible for advising NASA on the scientific requirements for data collection, processing, archival, distribution, and analysis.

In addition to the specific technical facets of a DDMU operation, there are several other requirements for a DDMU to operate successfully.

Since a DDMU for handling space data will likely be funded largely by NASA, there should be a formal overview of the unit by the appropriate NASA office. A formal statement of work should be negotiated, which includes regular reporting. In addition, a proposal review committee should be established to determine which scientific proposals are to be supported and to resolve potential duplications or conflicts. The proposal review committee should include NASA representatives (other government agency representatives, if they are involved), DDMU scientific representatives, and science discipline representatives; it might be more broadly constituted to include an overall advisory function concerning general objectives and performance of the DDMU. In the case of larger units, the advisory committee should probably be separate from the proposal review committee.

Still another committee, namely a users committee, is required. This committee should meet frequently to review and advise the DDMU on detailed matters involved in the day-to-day activity of the unit. As the name suggests, the users committee should be composed of individuals actively involved in the use of the data and services provided through the DDMU.

In addition to the above oversight activities, the DDMU will require adequate funding, a long-term commitment, an appropriate technical support staff, and a high-quality scientific staff if it is to be successful. The funding inadequacies for the data phase of many of the programs described above cannot be dissolved by simply creating a DDMU. The DDMU provides a mechanism for optimizing planning and performance to maximize scientific yield, but it clearly requires adequate funding. At the same time, the software development, data distribution, and data-archiving functions of the DDMU all imply a long-term commitment. If the scientists in a given discipline are convinced that a DDMU is required and the appropriate review committees and NASA offices agree, then a long-term commitment is required to establish the unit and to attract a qualified staff. The DDMU should not necessarily be limited to a single program but, in fact, should be involved in all relevant programs from the earliest planning stages. This again implies continuity and longevity.

Although it is necessary that a DDMU be provided with sufficient continuity to perform its functions, we do not envisage that a DDMU is a "permanent" commitment. It is essential that there be established procedures for terminating a DDMU as well as for establishing one. In many cases the function of a DDMU will be for a relatively short time period; in other cases a DDMU may fail to perform satisfactorily or may lose the key personnel necessary for its function. We recommend that a DDMU be established for a prescribed period of between three and five years. At the end of this period a formal review of the unit should be carried out to determine whether it should be continued, discontinued, or revised.

Although the DDMU concept is expected to enhance significantly the scientific utilization of space-derived data, it must also be recognized that there are several limitations.

DDMU's should not be expected to replace the major data centers. The success of the unit depends on having a manageable group of scientists with an active interest in the data being handled. An operation of the size necessary to perform the functions of the data centers could become so large that the scientific objectives would be lost.

On the other hand, it would be inappropriate to establish a DDMU dedicated to a single, short-term mission. Such a unit would not be of sufficient size and longevity to assemble the personnel required for a successful operation.

Finally, it must be remembered that the requirements of each discipline are unique. Although the generic requirements for the DDMU have been outlined above, each discipline must take the responsibility of tailoring the concept to its specific requirements and determining those areas within the discipline for which it is appropriate.