Free Executive Summary

Geoscience Data and Collections: National Resources in Peril

Committee on the Preservation of Geoscience Data and Collections, Committee on Earth Resources,National Research Council

ISBN: 978-0-309-08341-6, 128 pages, 8.5 x 11, paperback (2002)

Geoscience data and collections (such as, rock and sediment cores, geophysical data, engineering records, and fossils) are necessary for industries to discover and develop domestic natural resources to fulfill the nation's energy and mineral requirements and to improve the prediction of immediate and long term hazards, such as land slides, volcanic eruptions and global climate change. While the nation has assembled a wealth of geoscience data and collections, their utility remains incompletely tapped. Many could act as invaluable resources in the future but immediate action is needed if they are to remain available. Housing of and access to geoscience data and collections have become critical issues for industry, federal and state agencies, museums, and universities. Many resources are in imminent danger of being lost through mismanagement, neglect, or disposal. A striking 46 percent of the state geological surveys polled by the committee reported that there is no space available or they have refused to accept new material. In order to address these challenges, Geoscience Data and Collections offers a comprehensive strategy for managing geoscience data and collections in the United States.

Executive Summary

Everyone in downtown Hutchinson, a city of 40,000 in central Kansas, heard or felt the explosion, Wednesday morning, January 17, 2001. Natural gas burst from the ground under Woody's Appliance Store and the adjacent Décor Shop, blowing out windows in nearby buildings. Within minutes, the two businesses were ablaze. That evening, geyser-like fountains of natural gas and brine, some reaching heights of 30 feet, began bubbling up 3 miles east of the downtown fires. The next day, natural gas, migrating up a long-forgotten brine well, exploded under a mobile home and killed two people. The city ordered hundreds of residents to evacuate homes and businesses, many of whom would not be able to return until the end of March (Allison, 2001[1]).

The Kansas Geological Survey (KGS) stepped into a situation where demand for answers was great, but information was in short supply. Fortunately, the KGS had cores preserved in its repository from a project the Atomic Energy Commission had conducted in the 1960s to investigate the geology of localities being considered for nuclear storage. Practically unused for more than 30 years, these cores contained information that could be obtained rapidly—and without the time or risk of drilling into another unknown gas pocket. Geologists examined these and other cores and samples from wells drilled in the area to get a sense of the potential paths for gas flow through the rock. Armed with this information, obtained using geoscience data and collections, the KGS gathered new seismic data around the city, from which two anomalous zones of potential high gas pressure were identified. The gas had migrated 8 miles from a leaking salt cavern used as an underground natural gas storage facility. This gas was then safely vented. Over the next two months the Kansas Gas Service consulted with the KGS about possible vent-well locations and additional vent wells were drilled to release pressure. Hutchinson was safe from further gas geysers and gas explosions—and the displaced residents finally could return safely to their homes. Understanding of the situation was initiated through the KGS's fast action—action that began with ***cores*** that had been collected for another purpose many years earlier. Having immediate access to critical geoscience data and information played a crucial role in facilitating rapid response to a local crisis.

THE NEED FOR GEOSCIENCE DATA AND COLLECTIONS

This report builds the reader's understanding of the utility of geoscience data and collections, why these were acquired initially, why many remain useful, and what should be kept. ***Geoscience data and collections***[2] (e.g., ***cores, cuttings, fossils,*** geophysical tapes, paper logs, rocks) are the foundation of basic and applied geoscience research and education, and underpin industry programs to discover and develop domestic natural resources to fulfill the nation's energy and mineral requirements. Geoscience data and collections record the history of processes that operate on the Earth today and in the past and provide insights that lead to improved prediction of hazards, both immediate and long term. The geoscience community has amassed an enormous wealth of data and collections, most of which remain potentially useful and would be costly to replace, and much of which cannot be replaced. The diversity and quantity of these geoscience data and collections continue to expand, and as

[1]Reprinted with permission from *Geotimes*, October 2001. Copyright American Geological Institute, 2001.

[2]Geoscience collections are groupings of individual geoscience items that may be related by sample type, geographic location, or scientific or applied interests (see Appendix E for more information on this and other technical terms ***highlighted*** in the text).

they have, so has need for space and funding to support their ***preservation*** and accessibility.

Archiving and maintaining data and materials collected during the course of geoscientific research carry benefits well beyond those recognized by the scientific and academic communities. Well-maintained and well-documented geoscience data and collections are storehouses of information that likely will result in better assessment and management of natural resources, better understanding of the geologic hazards with which we live, and enhanced knowledge of the history of Earth and life. Virtually every facet of our daily life is touched either directly or indirectly by geoscience data and collections—from power that lights our cities to coatings on paper in books to medicines that save lives. If you drive a car, ride a bus, walk on sidewalks, take medicine, wear synthetic fabrics, or read a magazine, you have come in direct contact with and used geoscience resources, all of which owe their origin to information gleaned from geoscience data and collections.

Both the quality and quantity of geoscience data and collections have direct bearing on the accuracy of predicting and meeting future resource and engineering needs. Moreover, geoscience data and collections provide critical information that scientists and engineers need to help inform a variety of important societal decisions, including problems resulting from increased population growth on our planet. For example, current fossil energy resource assessment and exploitation is based directly on knowledge of the subsurface geological and engineering properties of the rocks that contain the resources. Natural hazards are assessed using historical records of their occurrence, coupled with prehistoric evidence gathered using geoscience data and collections. In both cases, absence of geoscience data and collections means that interpretations will be weaker at best and erroneous at worst.

GEOSCIENCE DATA AND COLLECTIONS AT RISK

Geoscience data and collections are imperiled, even though many are potentially useful and valuable in the future. Billions of dollars have been spent to acquire them. For instance, the U.S. Geological Survey (USGS) estimates that the cost to replace the geoscience data and collections archived in its Core Research Center at Lakewood, Colorado—a facility that contains no more than 5 percent of the volume of at-risk geoscience data and collections in the United States—is on the order of $10 billion (NRC, 1999a). Other examples include federal support in excess of $500 million for the acquisition of deep-sea sediment cores by the Ocean Drilling Program between 1983 and 1998 (NRC, 2000), and the estimated $535 million value of geologic materials housed at the Kentucky Geological Survey (Kentucky Geological Survey, 2001).

The committee learned that many geoscience data and collections already have been lost, and many more are at risk. Housing of and access to geoscience data and collections have become critical issues for federal and state agencies, academic institutions, museums, and industry. Nearly two-thirds of the state geological surveys the committee polled reported that their geoscience data and collections libraries have 10 percent or less space remaining for new data and collections. Even more critical, 46 percent of those same state geological surveys either reported that there is no space available or have refused to accept new material.

THE CHARGE TO THE COMMITTEE

The dilemma over geoscience data and collections is this: more and better geoscience data and collections exist now than ever before, however planning for space and maintenance of these data and collections have not kept pace with their acquisition. Therefore, appropriate management of these data and collections has become a more critical problem now than ever before. Consequently, the overall goal of this study was to develop a comprehensive strategy to manage geoscience data and collections in the United States. Specifically, the committee was charged with the following tasks:

1) Develop a strategy for determining which geoscience, ***paleontological***, ***petrophysical***, and engineering data to preserve.
2) Examine options for the long-term archiving of and provision of access to these data.
3) Examine three to five ***accession*** and ***repository*** case studies as examples of successes and failures.
4) Distinguish the roles of public and private sectors in data preservation.

The committee concentrated its effort on the preservation and management of physical data (e.g., cores, cuttings, fossils, geophysical tapes, paper logs, rocks) as opposed to digital data (e.g., computer-stored information). Nevertheless, the committee addressed the use and importance of digital information to enhance cataloging and dissemination of information about the physical materials (i.e., metadata about the geoscience data and collections). Digital access to information about geoscience data and collections is a key ingredient to their use by the widest range of clients possible.

WHAT SHOULD BE PRESERVED?

Geoscience data and collections are valuable national resources, some of which should be preserved and made available for scientific and strategic use. Despite their importance, utility, and value, substantial amounts already have been lost. For example, the record of the deepest ***well*** cored in the United States has been lost. The present-day cost to acquire a similar core is estimated at $12.3 million to $16.4 million (Michael Padgett, EEX Corporation, personal communication, 2001).

TABLE ES-1 Criteria for Determining Which Geoscience Data and Collections to Preserve

Criteria	Well Documented[d]	Irreplaceable[e]	Potential Applications[f]	Accurate	Quality/ Completeness	Non-Replicative
Collections:						
Cuttings	X	x	x	X	_	X
Engineering[a]	X	x	x	X	x	_
Fossils	X	x	x	X	x	_
Geophysical[b]	X	x	x	_	x	X
Maps/Notes[c]	X	x	x	_	x	X
Mining Cores	X	x	x	X	x	_
Other Rock Cores	X	x	x	X	x	X
Sediment & Ice Cores	X	x	x	X	X	_

X=present or necessary for preservation (i.e., absence = candidate for deaccession).
x=may be present and may be a factor for preservation (i.e., absence may not be a factor for deaccession).
_=not present and not necessary for preservation (i.e., absence is not a factor in deaccession).
Criteria are arranged from left to right in approximately decreasing order of importance (but see text for further explanation and elaboration).
Collections are arranged alphabetically.

[a]Includes drill stem tests, completion records, site reports, and other engineering data/reports on CD, computer disk, fiche, paper, tape, or some other quasi-stable medium.

[b]Includes seismic data, down-hole geophysical data, fly-over geophysical data, and other geophysical data on CD, computer disk, fiche, paper, tape, or some other quasi-stable medium.

[c]Includes unpublished materials on CD, computer disk, fiche, paper, tape, or some other quasi-stable medium, whether or not they were used in the production of published products.

[d]All collections must be well documented before any other assessment of their utility and future can be done. Indeed, whether or not a rock, fossil, core, or other item is replaceable is completely unknown in the absence of adequate documentation to assess uniqueness. That said, if part of a collection is not replaceable, but only documented well enough to know that it is unique, it probably should be kept anyway. Documentation includes, but is not limited to, information about age, location, depth, collector or author, date acquired, and associated materials.

[e]Impossible or highly unlikely to collect a similar sample (e.g., a mine core from a completely mined-out locality; a sample from a politically inaccessible part of the world; a sample requiring great time and effort to recollect such as a deep ice core from Antarctica or Greenland).

[f]This category in particular should be weighed judiciously by a science advisory board comprised of members of the user community.

Potential causes of loss of existing materials are numerous. Examples include lack of space in repositories, changing interests of some companies away from domestic production, company mergers, deterioration of materials and accompanying information over time, changes in staff and staff research interests, and reductions in work force at government facilities. Based on information presented to us and gathered over the course of the study, *the committee concludes that many geoscience data and collections are currently in peril*. Therefore, **the committee recommends that priority for rescuing geoscience data and collections be placed on those that are in danger of being lost. The committee recommends that the highest priority for retention and preservation be directed toward data and collections that are well documented and impossible or extremely difficult to replace.** Other factors to consider when setting priorities for preservation are potential applications, accuracy, quality and completeness, and redundancy. Table ES-1 summarizes the committee's assessment of overarching factors pertinent to the decision to retain or discard (***deaccess***) geoscience data and collections.

Assessing potential applications of geoscience data and collections is an important step in prioritization, and is a challenge that should not be left to a single individual. Assessing basic and applied potential of any physical data is a task that requires vision, imagination, and broad experience. Such guidance should be sought through external science advisory boards that represent a broad range of scientific, government, and business communities (collectively, the user community). Examples of the user community advising on priorities for preservation include those for the National Ice Core Laboratory and the Ocean Drilling Program. Such advisory committees are in a position to provide realistic recommendations (as opposed to the unrealistic recommendation of "keep everything") about what to keep using criteria suggested above against a backdrop of what might be needed in the future.

Enormous volumes of geoscience data and collections are held by a large number and variety of institutions. Museums, state geological surveys, universities and colleges, federal agencies, and industry all hold geoscience data and collections that have been amassed over as many as several hundred years. The committee estimates that more than 15,000 miles of cores and cuttings, well over a quarter of a billion line-miles of seismic data, and more than 100 million boxes of fossils are in geoscience repositories today. Furthermore, *the committee concludes that sufficient geoscience data and collections in the United States are at risk of loss to fill at least 20 times the*

USGS Core Research Center in Lakewood, Colorado. These figures are estimates that reflect minimum values.

Assessing the complete breadth and depth of geoscience data and collections that exist was just one of the challenges the committee faced. Simply stated, the quantity, variety, and quality of the nation's geoscience data and collections are largely unknown. The committee found that information on geoscience data and collections that have been lost or discarded is elusive because of their proprietary nature, the unwillingness to admit to discarding such data and collections, and the challenges and costs of donating them to a public facility (for example, the ongoing 6-year-old negotiations between Shell Oil Company and the Internal Revenue Service) versus discarding them.

Consequently, the committee became keenly aware that an understanding of the wealth of geoscience data and collections available to the public and private science and technology sectors is imperative. Without that understanding, we cannot make the best use of what already exists, or understand what is now at risk of being lost or discarded. Gathering comprehensive information on existing data and collections is essential for their future use. Therefore, **the committee recommends funding cataloging efforts to gather comprehensive information about existing geoscience data and collections. The committee recommends that access to these funds be on a competitive basis, and that preference be given to institutions with holdings that meet the same priorities as those outlined above for preservation.** The Institute of Museum and Library Services and the National Science Foundation are two federal agencies with experience and demonstrated effectiveness at distribution of funds to the museum, library, and science communities on a competitive basis. The inventory process should proceed simultaneously with development of a geoscience data and collections management system, and, to stimulate knowledge and use of the data and collections, the resultant institutional catalogs should be available online.

The number of universities, colleges, museums, institutes, state agencies, and other geoscience-oriented entities that need support for these cataloging efforts is certainly in the hundreds. Therefore, **the committee recommends that this initial catalog funding effort target 5 to 10 institutions each year until the nation's geoscience data and collections are adequately assessed.** The committee estimates that this effort would be effective if supported at the level of $5 million to $10 million per year.

OPTIONS FOR LONG-TERM ARCHIVING AND ACCESS TO GEOSCIENCE DATA AND COLLECTIONS

Managing Geoscience Data and Collections in the United States

Because the volume and variety of geoscience data and collections are great, the goal of achieving long-term archiving of and access to geoscience data and collections must be achieved sequentially. **The committee recommends the establishment of a distributed network of regional geoscience data and collections centers, each with an external science advisory board.** Each center would be a consortium of government, academic, and industry entities within the region, and would likely build off existing infrastructure and expertise. Among their various roles, the centers would foster cooperation among existing repositories, encourage adoption of uniform standards, and coordinate outreach. The committee found that successful (i.e., supported, maintained, and used) geoscience data and collection centers served relatively focused communities of interest (most often geographically defined areas). An excellent example of such centers, with external science advisory boards, broad community involvement, and regional distribution can be found in the current core repositories for the Ocean Drilling Program.

There was consistency among those testifying to the committee, and consensus within the committee itself, that one model of a single, national geoscience repository was impractical. Four barriers stand out to such a model: the untenable cost of moving all geoscience data and collections to a single location, the enormity of scale that such a center would entail, the impracticalities of expecting many users to come to the center, and the unwillingness of many existing repositories to part with their collections. Regional centers, on the other hand, are large enough to achieve economies of scale, but small enough to encourage local interest and support. Distributing the centers would permit sponsors to nurture regional networks of dedicated volunteers, content donors, and financial benefactors.

The committee concludes that immediate action is needed to stop the loss of irreplaceable geoscience data and collections in areas containing the greatest volume of at-risk material. Criteria for assessing risk include those outlined earlier. In terms of sheer volume of data, shifting priorities of those holding data, and merger activity, those regions with long histories of resource extraction stand out. **The committee recommends establishing three centers (one each in the Gulf Coast, Rocky Mountain, and Pacific Coast regions). Furthermore, the committee recommends that additional regional centers, as merited, be established over the next 5 to 10 years, and that preference be given to centers that meet three main criteria: 1) need for such a center in the region (i.e., active clientele, identified collections of high priority, at-risk data in the region), 2) broad involvement and support among various regional geoscience and other entities (government, academia, and industry), and 3) active participation of an independent, external science-advisory board. The committee recommends that the centers build upon existing expertise and infrastructure—such as state geological surveys, museums, universities, and private enterprises—and that, where practical, more efficient use of existing space be**

encouraged before expansion. Furthermore, the committee recommends that access to the center-establishment and improvement funds be on a competitive basis.

For reasons stated above, the National Science Foundation is a logical distributor of the funds. The committee estimates that each center would cost between $35 million and $50 million to establish.[3] Additional support would be needed for operations costs. Therefore, **the committee recommends additional maintenance and operations expenditures, which would be re-evaluated regularly on a competitive basis, to ensure maximum utilization of each center (i.e., to encourage public outreach and awareness, use, and cost-sharing activities).** The committee estimates these costs to be in the range of $3 million to $5 million per year for each center.

A Strategy for Managing Federal Geoscience Data and Collections

Federal agencies responsible for geoscience data and collections in the United States should lead the way by setting examples of good practices in preservation and use of geoscience data and collections. Such examples serve to promote public good, increase the visibility of the federal side in a leadership role, and increase the likelihood of federal partnerships with the private sector.

The committee learned that inadequate levels of support for cataloging and archiving of geoscience data and collections exist within many federal entities. For example, at the National Museum of Natural History (NMNH), which houses the nation's largest publicly available geoscience collection, only 10 percent of the holdings currently are electronically cataloged. Moreover, both the NMNH and the USGS have experienced reductions in force that have compromised their ability to care for their collections. Therefore, **the committee recommends that federal agencies be supported to the same extent as non-federal institutes and consortia with respect to cataloging and repositories, and with regular review**. **The committee recommends that priorities for federal agency support should closely follow those recommended for the regional centers: 1) need for such a repository in the agency, 2) broad or active involvement within and among various federal geoscience agencies (e.g., BLM, DOE, EPA, NASA, NOAA, NSF, USACE, USGS, USNM), and 3) active participation of independent, external science-advisory boards.** The committee envisages (where appropriate) federal agencies as potential members of the proposed regional consortia, with funding for federal and non-federal entities in this instance converging within these consortia. Such arrangements between state and federal agencies are already in place in Colorado and Alaska, for example. Lastly, federal agencies should be permitted to offset some costs with appropriate charges for selected services.

While it exists, coordination among federal agencies that collect or archive geoscience data and collections could be improved. Such improved coordination would optimize sharing of business practices and consumer use of related data collected by various agencies or establishing priorities among agencies so that limited funds can be used to the best overall effect. Adoption of consistent and good practices, along with a clarification of roles, would, at a minimum, increase efficiencies for federal agencies and the user community, comparable in some respects to the goals of the National Spatial Data Infrastructure (NRC, 1993) and the Geospatial One-Stop initiatives.[4] In addition, such collaboration would render the whole of government holdings more complete, enhance the value of individual components, and permit a significantly (and, eventually, measurable) increased benefit to diverse communities.

To optimize federal coordination, **the committee recommends establishing a federal geoscience data and collections coordinating committee.** Such a committee could be established and funded through the Office of Management and Budget, as the committee would oversee coordination and increased efficiency among a range of federal agencies. This federal geoscience data and collections coordination committee should be broad-based, reaching between and within all federal and quasi-federal agencies involved in geoscience research or geoscience data and collections acquisition. The committee's charge should focus on coordination of federal agencies' roles with regard to geoscience data and collections preservation, access, and use. **The committee recommends that the federal geoscience data and collections coordinating committee should appoint several federal external science advisory boards to advise on priorities for federal holdings, with respect to preservation, cataloging, and access among and within federal and quasi-federal agencies.** Previous NRC reports (e.g., NRC, 2001) already have noted the value for federal agencies of having direct external community involvement and advice to help set internal priorities for funding, monitoring, and research efforts. Examples of federal external science advisory boards that deal with collections are those within the operating structure of the National Ice Core Laboratory (coordinated jointly by the USGS and NSF) and the Smithsonian Institution.

The federal, external science advisory boards would focus on holdings within the federal government, but would

[3]The committee bases its estimates on building anew, and recognizes that costs could be less if a center were to build off existing infrastructure.

[4]These two initiatives are useful models in several respects. First, they seek to render data from many federal, state, and local agencies both convenient to access and easy to use together. Second, they must address diverse missions, user communities, producer concerns, data definitions, and data formats. Information providers may themselves produce the data, or they may obtain data from external sources. Coordination of U.S. geoscience data and collections will involve all of these issues.

coordinate with the science advisory boards recommended for the regional geoscience data and collection centers. The federal, external science advisory boards, which could be discipline-based, would advise on establishment of consistent practices across agencies with respect to preservation of and access to geoscience data and collections acquired from public lands or using federal funds. In addition, the federal, external science advisory boards would advise on what geoscience data and collections should logically fall within the purview of various federal agencies. Monitoring of conformance to agreed-upon practices, as a question of how rather than what, would reside within the charge of the federal geoscience data and collections coordinating committee.

The federal geoscience data and collections coordinating committee would have other responsibilities related to how the federal effort should be streamlined, coordinated, and improved. One such responsibility would be monitoring implementation of electronic reporting for all exploration, exploitation, and research reports currently submitted to the federal government. The committee believed that electronic reporting was a necessary step to minimize the burden of cataloging newly collected geologic data and samples, while maximizing their potential use. As noted, the challenge to catalog existing geoscience data and collections is already immense. Therefore, **the committee recommends that electronic reporting be implemented as soon as possible, with additional funding as required to accelerate it.** Examples of programs of electronic reporting can already be found at the provincial level in Canada and Australia, and in the state of Wyoming.

The cataloging effort recommended for non-federal institutional holdings is of equal importance for future use of federal geoscience data and collections. Therefore, **the committee recommends that the federal geoscience data and collections coordinating committee monitor and facilitate progress of cataloging efforts across the federal government.** Here, the federal geoscience data and collections coordinating committee should work closely with the federal, external science advisory boards to determine which cataloging efforts warrant the highest priority. In addition, the federal geoscience data and collections coordinating committee should facilitate and coordinate Internet access to all federal geoscience data. This would include (but not be limited to) reports and catalogs of holdings, location and availability of similar geoscience data and collections, and contact information (where appropriate) for onsite use of geoscience data and collections. Success of this effort will be enhanced by coordinated adoption of digital data standards to improve interoperability of interagency information.

Regular review of the roles of the National Science Foundation and Institute of Museum and Library Services as distributors of funds for non-federal cataloging and repository efforts is essential. If existing external review mechanisms (e.g., committees of visitors, external steering committees) are inadequate for this task, new ones should be devised.

The Roles of Public and Private Sectors

From the testimony of those who use geoscience data and collections (see Appendix B) *the committee concluded that incentives (and even some mandates) for preservation of geoscience data and collections would encourage preservation efforts, and that partnerships and consortia are the most appropriate means by which to maintain long-term security for the various regional repositories.* Therefore, **the committee recommends establishing a combination of federal, state, regional, and local government incentives and requirements for geoscience data and collections donations and deposition. Establishing such incentives should be an immediate priority to stem the tide of lost and discarded geoscience data and collections, many of which remain useful.** Such incentives would encourage private donations of geoscience data and collections by providing credit for shipping costs and fundamental recognition that fossils, rock, sediment, and ice are unique and have donation value. When such data and collections are used to enhance recovery of resources, federal support for these incentives could pay for itself many times over (see DOE, 2002). An incentive for the research community would be a requirement that geoscience data and collections amassed during federally funded research (i.e., funded by agencies such as DOD, DOE, EPA, NASA, NSF, USGS, USNRC) be archived appropriately, cataloged, and made accessible to the public (e.g., NSF guidelines in Appendix G, and in USGCRP, 1991). Federal support for research should be, in general, contingent upon the public availability of these geoscience data and collections within a reasonable time.

The geoscience community itself must take more responsibility for preservation and use of geoscience data and collections. Although the necessity and importance of these data for research and interpretations are broadly accepted, adequate curation and long-term care for them take time and consequently fall through the cracks. The geoscience community should do more than just acknowledge the importance of geoscience data and collections—it should establish incentives, rewards, and requirements for their care and accessibility. **The committee recommends that the geoscience community adopt standards for citation in scientific and other publications of geoscience data and collections used.** Citation histories enhance credibility and importance to well-organized, often-used data and collections. In addition, **the committee recommends that institutions and professional societies establish (where appropriate) awards and other forms of recognition for outstanding contributors to the preservation and accessibility of geoscience data and collections.**

THE TIME IS NOW

Well-maintained and well-documented geoscience data and collections have both immediate and long-term value. The nation has assembled a wealth of geoscience data and collections. Some of these already have been lost, and many more are in imminent danger of being lost—through mismanagement, neglect, or outright disposal—if immediate action is not taken. The recommended solutions that this committee proposes represent a strategy for such immediate action. Future generations deserve the opportunity to build upon existing successes and avoid repetition of our failures. Geoscience data and collections are national resources, and are a part of our nation's heritage. Preservation of geoscience data and collections is a comparatively small investment in our past, our present, and our future, with both immediate and long-term benefits.

Geoscience Data and Collections

NATIONAL RESOURCES IN PERIL

Committee on the Preservation of Geoscience Data and Collections

Committee on Earth Resources

Board on Earth Sciences and Resources

Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

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 study was jointly sponsored by the American Association of Petroleum Geologists, American Association of Petroleum Geologists Foundation, American Geological Institute, Department of Energy–Fossil Energy (DE-AP75-00SW48036), Department of Energy–Yucca Mountain (DE-FG09-97NV12056), Geological Society of America, National Science Foundation (EAR-0071061), Paleontological Society (00-22-US-3548), Petrotechnical Open Software Corporation, Schlumberger, Ltd, Smithsonian Institution, and U.S. Geological Survey (00HQAG0145). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number: 0-309-08341-9

Library of Congress Control Number: 2002106850

Additional copies of this report are available from:

The National Academies Press
500 Fifth Street, N.W.
Box 285
Washington, DC 20055
800-624-6242
202-334-3313 (in the Washington metropolitan area)
http://www.nap.edu

Front cover: Geoscience data and collections examples and storage facilities. *Background left:* Flexible-space shelving at Bureau of Economic Geology, University of Texas at Austin. SOURCE: David Stephens, BEG, University of Texas at Austin. *Background right top:* Inside the National Ice Core Laboratory, at the Denver Federal Center in Lakewood, Colorado. SOURCE: Geoffrey Hargreaves, NICL. *Background right bottom:* Interior of the Ocean Drilling Program's Gulf Coast Repository (GCR) at Texas A&M University in College Station. SOURCE: Ocean Drilling Program. *Foreground left to right:* Fossil fish and trilobite. SOURCE: ExxonMobil Upstream Research Company; Foraminifera microfossils. SOURCE: ExxonMobil Upstream Research Company; Rock and mineral specimens. SOURCE: ExxonMobil Upstream Research Company; and Tapes containing data from boreholes. SOURCE: Phillipe Theys, Schlumberger, Ltd., Sugarland, Texas.

Backcover: *Background top:* Inside the National Ice Core Laboratory, at the Denver Federal Center in Lakewood, Colorado. SOURCE: Geoffrey Hargreaves, NICL. *Background bottom:* Interior of the Ocean Drilling Program's Gulf Coast Repository (GCR) at Texas A&M University in College Station. SOURCE: Ocean Drilling Program. *Foreground:* Fossil fish and trilobite. SOURCE: ExxonMobil Upstream Research Company.

Cover designed by Van Nguyen

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The ***National Academy of Sciences*** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized 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 advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON THE PRESERVATION OF GEOSCIENCE DATA AND COLLECTIONS

CHRISTOPHER G. MAPLES, *Chair,* Indiana University, Bloomington
WARREN D. ALLMON, Paleontological Research Institution, Ithaca, New York
KEVIN T. BIDDLE, Exxon Mobil Corporation, Irving, Texas
DONALD D. CLARKE, Department of Oil Properties, City of Long Beach, California
BETH DRIVER, National Imagery and Mapping Agency, Reston, Virginia
THOMAS R. JANECEK, Florida State University, Tallahassee
LINDA R. MUSSER, Pennsylvania State University, University Park
ROBERT W. SCHAFER, Mineral Exploration and Business Development Consultant, Salt Lake City, Utah
ROBERT M. SNEIDER, Robert M. Sneider Exploration, Inc., Houston, Texas
JOHN C. STEINMETZ, Indiana Geological Survey, Bloomington
SALLY ZINKE (until 8/2001), Ultra Petroleum, Inc., Englewood, Colorado

NRC Staff

PAUL CUTLER, Program Officer
MONICA LIPSCOMB, Research Assistant
TERESIA WILMORE, Project Assistant

COMMITTEE ON EARTH RESOURCES

SUSAN M. LANDON, *Chair,* Thomasson Partner Associates, Denver, Colorado
JAMES C. COBB, Kentucky Geological Survey, Lexington
VICKI J. COWART, Colorado Geological Survey, Denver
MURRAY W. HITZMAN, Colorado School of Mines, Golden
JAMES M. McELFISH, JR., Environmental Law Institute, Washington, D.C.
JOHN MURPHY, University of Pittsburgh, Pennsylvania
DIANNE R. NIELSON, Utah Department of Environmental Quality, Salt Lake City
THOMAS J. O'NEIL, Cleveland-Cliffs, Inc., Cleveland, Ohio
DONALD PAUL, ChevronTexaco, San Francisco, California
RUSSELL STANDS-OVER-BULL, Arrow Creek Resources, Pryor, Montana
R. BRUCE TIPPIN, North Carolina State University, Asheville
MILTON H. WARD, Ward Resources, Inc., Tucson, Arizona
LAWRENCE P. WILDING, Texas A&M University, College Station
PHILLIP MICHAEL WRIGHT, Idaho National Engineering and Environmental Laboratory, Idaho Falls

NRC Staff

TAMARA L. DICKINSON, Senior Program Officer
KERI H. MOORE, Staff Officer
KAREN L. IMHOF, Senior Project Assistant

BOARD ON EARTH SCIENCES AND RESOURCES

NRC Staff

Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We thank the following individuals for their participation in the review of this report:

Peter Crane
Royal Botanic Gardens
Kew, Richmond, Surrey, United Kingdom

Gordon Eaton
USGS *emeritus*
Coupeville, Washington

Stan Eschner
Trio Petroleum Inc.
Bakersfield, California

William L. Fisher
The University of Texas
Austin

Alexander H. Flax
Consultant
Potomac, Maryland

Scott Hector
Carneros Energy, Inc.
Bakersfield, California

Robert Laing
ChevronTexaco
Pleasanton, California

David Simpson
IRIS Consortium
Washington, DC

Milton A. Wiltse
Alaska Geological Survey
Fairbanks

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by Raymond A. Price, Queen's University, Kingston, Ontario, Canada. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Preface

On September 20, 1999, the National Research Council (NRC) received a letter from Dr. Philip D. Vasquez, Deputy Assistant Secretary for Natural Gas and Petroleum Technology, conveying the request of the U.S. Department of Energy (DOE) that the NRC establish a committee to determine the options and develop a strategy for the preservation and management of subsurface geoscience data. Because of the broad concern on this matter across the geoscience community, a wide range of sponsors supported the activities of the committee. These sponsors were American Association of Petroleum Geologists, American Association of Petroleum Geologists Foundation, American Geological Institute, Department of Energy–Fossil Energy, Department of Energy–Yucca Mountain, Geological Society of America, National Science Foundation, Paleontological Society, Petrotechnical Open Software Corporation, Schlumberger, Ltd., Smithsonian Institution, and U.S. Geological Survey.

The committee operated under the aegis of the Committee on Earth Resources, a standing committee of the Board on Earth Sciences and Resources. It carried out its work through 4 meetings, 6 site visits by the full committee, 6 site visits by subsets of the committee, and distribution and analysis of a questionnaire. A total of 39 state geologic surveys and 17 other entities responded to the questionnaire. A list of oral and written contributions to the committee is provided in Appendix B. The full committee visited the following sites: the Smithsonian Institution in Washington, DC; the U.S. Geological Survey in Lakewood, Colorado; the Denver Earth Resources Library in Denver, Colorado; the National Geophysical Data Center in Boulder, Colorado; the Bureau of Economic Geology, University of Texas at Austin; and C&M Storage Inc. in Schulenberg, Texas. Subsets of the committee visited the Colorado School of Mines Geology Museum; DOE's Yucca Mountain project in Nevada; the Energy Information Administration in Washington, DC; the National Archives and Records Administration in College Park, Maryland; the Northern Rockies Geologic Data Center, in Billings Montana; and the U.S. Army Corps of Engineers in Washington, DC.

In responding to DOE's request to determine the options and develop a strategy for the preservation and management of geoscience data, the committee paid particular attention to the preservation and management of physical data (e.g., cores, cuttings, magnetic tapes, paper logs, rocks) as opposed to digital data. It is beyond the charge of the committee to focus on digital data. However, in keeping with the original intent of several funding agencies, the committee task was expanded beyond the original DOE request of "subsurface geoscience data" to include collections, especially those of a paleontological nature. It is important to clarify what is encompassed by the phrase "geoscience data and collections." "Geoscience" is a term for the collective subdisciplines of the geological (solid Earth) sciences, including geobiology, geochemistry, geohydrology, geophysics, sedimentology, and stratigraphy, among others. "Data" and "collections" were distinguished from each other on the basis of whether the physical item originated naturally (a rock, mineral, or fossil) or was produced from some other medium (a paper log, a magnetic tape, a picture); the former fell under the definition of collection and the latter fell under the definition of geoscience data (see Appendix D). The committee recognizes that the terms "collections" and "data" mean different things to different sectors of the geosciences. For example, the petroleum and mining industries consider rock cores and cuttings as "data," whereas the museum community considers them "collections." The definitions of these terms as used herein reflect the need for internal consistency within the report. In terms of geographic scope, the committee focused on geoscience data and collections of unconstrained geographic origin, but housed in the United States.

DOE's request to determine the options and develop a strategy for the preservation and management of geoscience data carries with it the implication that not everything can or should be preserved. To do otherwise is unrealistic and re-

quires no determination of options—everything is kept. Consequently, the committee entered into this project with the assumption that not everything could or should be kept. However, the diversity and variety of geoscience data and collections are so vast that no specific set of protocols for obtaining or discarding geoscience data and collections applies in all cases. To that end, the committee has produced a set of guidelines under the premise that those who work with the appropriate geoscience data and collections (i.e., the user community) are the ones who are in the best position to assess which items to keep and which to discard.

The committee is indebted to the support and hard work of NRC staff. Teresia Wilmore (NRC Project Assistant) was very helpful in making sure the committee got to the right places and helped us with NRC travel and reimbursement. Monica Lipscomb (NRC Research Assistant) was instrumental in tracking down information and assisting with editorial copy after editorial copy. Paul Cutler (NRC Study Director) kept the committee on track, provided extremely useful summaries of complex discussions, reminded us of our tasks and obligations, and did the initial writing for many parts of the written document. Anthony de Souza (BESR Director) and Tamara Dickinson (NRC Senior Program Officer) provided very useful feedback and comments on rough drafts. Winfield Swanson (NRC Editorial Consultant) edited the first and last drafts.

Christopher G. Maples, Chair

Contents

Figures, Tables, and Sidebars

FIGURES

FIGURES FOR SIDEBARS

TABLES

SIDEBARS