

DOE/BC/14833-5
(DE98000458)

NATIONAL GEOSCIENCE DATA REPOSITORY SYSTEM
PHASE II

Final Report
January 30, 1995 to January 28, 1997

April 1998

Performed Under Grant No. DE-FG22-95BC14833

American Geological Institute
Alexandria, VA

RECEIVED
MAY 11 1998
OSTI



**National Petroleum Technology Office
U. S. DEPARTMENT OF ENERGY
Tulsa, Oklahoma**

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government.

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield VA 22161

DOE/BC/14833-5
Distribution Category UC-122

National Geoscience Data Repository System
Phase II

Final Report
January 30, 1995 to January 28, 1997

April 1998

Work Performed Under Grant No. DE-FG22-95BC14833

Prepared for
U.S. Department of Energy
Assistant Secretary for Fossil Energy

Chandra Nautiyal, Project Manager
National Petroleum Technology Office
P.O. Box 3628
Tulsa, OK 74101

Prepared by:
American Geological Institute
4220 King Street
Alexandria, VA 22302

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

**Portions of this document may be illegible
electronic image products. Images are
produced from the best available original
document.**

Table of Contents

EXECUTIVE SUMMARY	v
ABSTRACT	xi
INTRODUCTION AND BACKGROUND	1
SUMMARY OF PHASE I FEASIBILITY AND ASSESSMENT STUDY	7
Goals and Objectives	
Data Contributions	
Establishing a Repository System: Recommendations from Phase 1	
RESULTS OF PHASE 2 PROJECT: PLANNING, DIRECTORY AND PILOT STUDIES	11
PROJECT OVERVIEW AND SCOPE	11
Project Oversight and Support	
Report Audience	
Project Approach	
THE OVERALL NGDRS MODEL	15
Design Principles	
Strategic Objectives	
The Stakeholders	
Business Model	
MODELS FOR DATA TRANSFER AND ACCESS	19
Data Audit and Prioritization	
Data Transfer General Steps	
Clearinghouse	
Financial Models for Data Transfer	
Geotrek Data Browser Software Development	
Metadata Definition and Related Standards	
Metadata, Catalogs, and Synchronization	
NGDRS Data Ownership and Entitlement	
CREATING THE REPOSITORY SYSTEM	38
Overall Architecture of the Repository System	
Architecture for the Source Nodes	
Role of Standards in the Creation of the Repository	
Metadata Repository Physical Components	
UTILIZING THE REPOSITORY SYSTEM	51
Using the Metadata Supercatalog	
Registering Users and Their Organizations	
NGDRS User Support: Help-Desk and Technical Support Groups	
OPERATING THE REPOSITORY SYSTEM	58
Organization, Resources and Competencies Needs	
Data Center Operations Guidelines	
Quality Monitoring Process	

Configuration Management	
Systems Maintenance	
Capacity Planning	
Disaster Recovery	
NATIONAL DIRECTORY OF GEOSCIENCE REPOSITORIES	65
STEERING COMMITTEE OPERATIONS	66
CONCLUSIONS AND RECOMMENDATIONS	67
REFERENCES	71
GLOSSARY	73
APPENDIX A: PHASE II INFORMATION EXCHANGES	83
APPENDIX B: METADATA EXAMPLES FOR A VARIETY OF DATA TYPES	87
APPENDIX C: NATIONAL DIRECTORY OF GEOSCIENCE DATA REPOSITORIES	
QUESTIONNAIRE	93
APPENDIX D: SUMMARY OF DATA FROM THE NATIONAL DIRECTORY OF GEOSCIENCE	
DATA REPOSITORIES	97
APPENDIX E: AGI NATIONAL GEOSCIENCE DATA REPOSITORY SYSTEM	
STEERING COMMITTEE	119

TABLES

Table 1	Organizations Providing Input and Support to NGDRS Project	12
Table 2	Data Audit and Prioritization Survey Visits	20
Table 3	General Procedural Steps for Transfer of Major Data Types	23
Table 4	Transfer of Tapes to Public Domain	27
Table 5	Transfer of Core to Public Domain	29
Table 6	Matrix of Possible Standards Involved in the Creation of NGDRS	48
Table 7	NGDRS Steering Committee Meetings	66

FIGURES

Figure 1	NGDRS Funding and Revenue Model	18
Figure 2	NGDRS Clearinghouse Coordination Activities	25
Figure 3	Logical Linking of Data Records	33
Figure 4	NGDRS Metadata Supercatalog	35
Figure 5	NGDRS Overall Systems Architecture	40
Figure 6	NGDRS Overall Information Flow	43
Figure 7	Physical Components of NGDRS Metadata Supercatalog	50
Figure 8	NGDRS Metadata Supercatalog User Perspective	52
Figure 9	NGDRS Metadata Supercatalog Data Browser	54
Figure 10	NGDRS Metadata Supercatalog Organization	59

Executive Summary

PROJECT OVERVIEW

The American Geological Institute (AGI) has completed Phase II of a project to establish a National Geoscience Data Repository System (NGDRS). The project's primary objectives are to preserve geoscience data in jeopardy of being destroyed and to make that data available to those who have a need to use it in future investigations. These data are available for donation to the public as a result of the downsizing that has occurred in the major petroleum and mining companies in the United States for the past decade. In recent years, these companies have consolidated domestic operations, sold many of their domestic properties and relinquished many of their leases. The scientific data associated with those properties are no longer considered to be useful assets and are consequently in danger of being lost forever. The national repository project will make many of these data available to the geoscience community for the first time.

To address this opportunity, AGI sought support from the Department of Energy (DOE) in 1994 to initiate the NGDRS Phase I feasibility study to determine the types and quantity of data that companies would be willing to donate. The petroleum and mining companies surveyed indicated that they were willing to donate approximately five million well logs, one hundred million miles of seismic reflection data, millions of linear feet of core and cuttings, and a variety of other types of scientific data.

Based on the positive results of the Phase I study, AGI undertook Phase II of the program in 1995. Funded jointly by DOE and industry, Phase II encompasses the establishment of standards for indexing and cataloging of geoscience data and determination of the costs of transferring data from the private sector to public-sector data repositories. Pilot projects evaluated the feasibility of the project for transfer of different data types and creation of a Web-based metadata supercatalog and browser. Also as part of the project, a national directory of geoscience data repositories was compiled to assess what data are currently available in existing facilities.

The next step, Phase III, will focus on the initiation of transfer of geoscience data from the private sector to the public domain and development of the web-based Geotrek metadata supercatalog. Contributed data will be placed in existing public sector facilities, such as state geological surveys, state oil and gas commissions, and association libraries. Users across the country will be able to access the data through a Web browser interface. DOE funding will support startup of the implementation phase and the transfer and cataloging of data. The project will also be supported through commercial subscriptions and is expected to become self-sustaining.

BENEFITS OF A NATIONAL REPOSITORY SYSTEM

The NGDRS will serve as an important and valuable source of information for a variety of applications, such as environmental protection, water resource management, global climate change studies, reduction of risks from earthquakes and other geologic hazards, and basic and applied research. Critical data contained in the repository system will enhance the ability of domestic energy and minerals companies to find new reserves and increase the recovery of proven reserves. Users

across the country will be able to access a variety of information sources via a Web-based metadata catalog that incorporates industry-endorsed standards to facilitate user interaction with data and ensure consistency across all participating repositories. There are many potential users of the system: petroleum companies, consultants, services companies, engineering companies, data brokers, universities, government agencies, lending institutions and the public at-large.

The NGDRS would also play a critical role in achieving key objectives of the President's Domestic Natural Gas and Oil Initiative, as well as technology transfer and environmental protection goals. Improved access to geoscience data will stimulate exploration activities that will help to reduce U.S. vulnerability to oil supply disruptions by increasing domestic oil production and reducing our dependence on unstable sources of foreign oil. It would help improve the competitiveness of the domestic energy and minerals industries, create high-quality jobs for American workers, and reduce the U.S. balance of trade deficit.

A growing number of professional, industrial and state organizations have gone on record in support of establishing a National Geoscience Data Repository System. In the past year, the NGDRS effort received key endorsements from the Association of American State Geologists, the Department of the Interior's Outer Continental Shelf Policy Committee, the American Petroleum Institute's Exploration Affairs Committee, and a National Research Council study.

The cost of transferring data to the public domain is very small compared to the original costs of acquiring data. Once the data goes to a landfill, is burned, or is otherwise destroyed, it is gone forever. The federal funds provided to support the preservation and access of geoscience data will maintain a national treasure available to future generations of users.

PHASE I: FEASIBILITY AND ASSESSMENT STUDY

The objective of the first phase of the project was to assess the feasibility of establishing the NGDRS. It focused on two major issues. First, it documented the types and quantity of data available for contribution to the NGDRS. Second, it documented the data needs and priorities of potential users of the system. The results of the Phase I feasibility study revealed that major oil companies, large independent petroleum producers, and minerals companies were indeed willing to contribute large amounts of data to the NGDRS. The acquisition cost of these data alone is on the order of \$35 to 40 billion. Moreover, a large percentage can never be re-collected due to environmental and economic restrictions and increased urbanization. Many state geological surveys, federal agencies, and a number of existing public repositories also indicated they would participate in a national repository system. The Phase I study showed that there is an excellent match between the types of geoscience information that companies are willing to contribute to the NGDRS and the types of data that are of greatest interest to independent petroleum producers and other potential users of the repository system.

PHASE II: PLANNING AND PILOT PROJECTS

The positive findings of the Phase I study led to the initiation of Phase II, which addresses specific organizational and operational requirements for establishing and implementing the NGDRS. It is a joint industry, academic, and government project with funding provided by DOE and industry. A

steering committee established in the Phase I feasibility study provided the oversight and guidance for Phase II as well. There are three major components of Phase II:

Planning and Specification. The planning and specification component of the Phase II study identified the requirements and operational business plan for implementing the NGDRS.

National Directory of Geoscience Data Centers. This directory was produced in order to show what data are available in existing repositories before implementation of the NGDRS, where they are located, and how the data can be acquired. Four tasks were involved in the development of this directory. They included selective site visits to both public and private data centers across the United States, development of a survey questionnaire sent to all identified data centers, follow-up contacts, and development of the directory itself, which was published in April, 1997 and is available from the American Geological Institute.

Pilot Projects. Pilot projects involved the indexing, cataloging, and transfer of various types of geoscience data from the private sector to the public sector. The results contributed critical information for the design and development of an efficient and cost-effective repository framework.

The results from the pilot projects yielded a model for the NGDRS and specific plans for implementation of individual system functions as well as for creating, utilizing, and operating the system as a whole. These are outlined in the following paragraphs.

The NGDRS Model

From an overall functional point of view, the repository system consists of two major components. One is the prototype metadata (data about the data) catalog of data residing in the individual regionally distributed data repositories. The other is the distributed set of repositories containing the physical data such as cores, well logs, tapes, and related information. A typical repository is made up of multiple data stores organized around specific data types. Although the data stores are operated independently and are geographically distributed, they can be accessed through a single interface connected to the metadata catalog developed for each data store. NGDRS, therefore, is a distributed system of data stores encompassing a common metadata supercatalog that can be accessed electronically through a single point of entry with a geographical oriented data browser software application.

A model for the NGDRS was developed around the following key design principles:

- Establish a program involving industry, academia, and government entities to preserve large volumes of geoscience data in danger of being discarded or destroyed.
- Reduce costs of the repository system by sharing information and repository facilities, thereby taking advantage of economies of scale.
- Develop a centrally managed metadata supercatalog and a decentralized network of geographically dispersed data repositories.
- Accept only data contributions that conform to agreed-upon industry standard data models.
- Improve accessibility of data already in the public domain and make publicly accessible all private data that is donated and transferred to the public domain.

- Encourage and facilitate the movement of all participating repositories to the POSC industry standards environment in order to eliminate unwarranted data-formatting operations.

Implementing Clearinghouse, Data Transfer, and Cataloging Functions

The objective of Phase II was to translate the strong interest in data transfer and preservation demonstrated in Phase I into a series of action plans in order to determine the cost and benefits, technical viability, and procedural requirements involved in the transfer of large volumes of data from the private to the public sector. The overall process included:

- Information exchanges and discussions with petroleum companies, data services vendors, data standards organizations, and project managers of similar repository projects worldwide.
- Workgroups, made up of experienced personnel in the data management area, were created to define detailed financial (business) models for the transfer of cores and cuttings, well log tapes, seismic data tapes and hardcopy data.
- Pilot projects were also defined and carried through to test and scale-up many procedural aspects of data collection, preparation, loading, indexing/cataloging, and accessing.

Creating the Repository System

Phase II was used to transform the vision and broad strategies identified in Phase I into specific objectives for creating a repository system. The principal vehicle for doing so was the implementation of pilot metadata repositories that permitted potential users and contributors to have a clearer understanding of the system's future capabilities. These pilot implementations permitted the project team to start creating more specific architecture for the design of the overall repository system network. The pilot implementations supported the installation of more than twenty distributed data servers, including a major source node at the Bureau of Economic Geology (BEG), the installation of major software components (MobilView and GeoTrek data browser prototypes), and the determination of the viability of using the Internet as the predominant choice for wide area networking by NGDRS users.

In parallel with the technology-oriented activities, the project team continued their efforts to find the most desirable business model for creating, utilizing and operating the repository. Detailed information on technology, current economic factors affecting the oil and gas industry, state-of-the-art data management practices, and the status of current industry standards were combined to refine the business model to be proposed for creating the repository during Phase III.

Utilizing the Repository System

Also during Phase II, user requirements were abundantly tested via appropriate pilot projects, and many discussions and demonstrations using the GeoTrek software browser. The pilot implementations permitted the project team to start creating the specific architecture of the system and user interface. The practical experience acquired during the installation and running of the pilot metadata repositories was used to create the final specifications for GeoTrek and to define the support services to be provided as part of the system's utilization. Documentation, some training aspects, marketing the NGDRS functionality, and receiving feedback from users were all tested in

Phase II using Internet capabilities. Detailed information on the Internet, current economic factors affecting the petroleum industry, and user requirements were all combined to refine the business model to be adopted for utilizing the NGDRS.

Operating the Repository System

A key Phase II objective was to collect and organize detailed administrative and operational procedures to operate the NGDRS, including staffing requirements. The principal vehicles for this phase were the implementation of a pilot project and direct contacts with dozens of both private and public-sector organizations. From these contacts, the project team obtained manuals and other documentation to facilitate and streamline the future operations of the NGDRS. The pilot project allowed the project team to start developing a more specific understanding of the complexities associated with managing a large repository and dealing with data issues revolving around privacy, security, backup, and disaster contingency. It permitted the design of user services to avoid or mitigate operational difficulties.

THE NEXT STEP: PHASE III IMPLEMENTATION

The Phase III project will use the knowledge and experience obtained during Phase II to scale up the pilot projects and implement the repository system. The objective of Phase III is to provide for a coordinated and orderly transfer of significant volumes of geoscience data from the private sector to the NGDRS and to provide improved access to the data. The business model and network support base, defined in the Phase II study, will assist in assuring sufficient support, efficient utilization, and day-to-day operations of the NGDRS. The Phase II study also provided valuable assistance in identifying major data contributors and regional data repositories that can house the data transferred from the private sector to the public domain.

The Phase III activities are organized around three entities. The first entity is a steering committee which will provide guidance on policy issues. The second is AGI, which will establish a clearinghouse that manages the relationship between the data contributors and the receiving data repositories. AGI will also be responsible for the creation, utilization and routine update of the metadata catalogs. Data contributors and receiving data repositories together constitute the third entity.

Three major components are identified for Phase III, building on the plans developed in Phase II:

- The **establishment of the clearinghouse** organization and the processes and procedures needed to coordinate all activities relating to the acquisition and transfer of data and data transfer priorities.
- The **creation of the Geotrek metadata supercatalog** -- a master electronic catalog of information about the data housed in source repositories -- including hardware, software, database and support service components.
- The **utilization of the Geotrek metadata supercatalog**, including the general day-to-day data center operations and support services.

It is expected that Phase III will span several years. During the first year, the major components of the Geotrek metadata supercatalog will be established and the data acquisition and transfer activities will be initiated.

Abstract

This report summarizes the results of Phase II of the American Geological Institute's project to establish a National Geoscience Data Repository System (NGDRS). The project's primary objectives are to transfer private-sector geoscience data in jeopardy of being destroyed to public-domain repositories and to make these and other geoscience data easily accessible to those who have a need to use it in future investigations. This project will make many of these data available to the geoscience community for the first time.

The NGDRS will serve as an important and valuable source of information for a variety of applications, such as environmental protection, water resource management, global climate change studies, reduction of risks from earthquakes and other geologic hazards, and basic and applied research. Critical data contained in the repository system will enhance the ability of domestic energy and minerals companies to find new reserves and increase the recovery of proven reserves. Users across the country will be able to access a variety of information sources via a Web-based metadata catalog that incorporates industry-endorsed standards to facilitate user interaction with data and ensure consistency across all participating repositories. There are many potential users of the system: petroleum companies, consultants, services companies, engineering companies, data brokers, universities, government agencies, lending institutions and the public at-large.

In 1994, the Department of Energy (DOE) funded the NGDRS Phase I feasibility study to determine the types and quantity of data that companies would be willing to donate. The petroleum and mining companies surveyed indicated that they were willing to donate approximately five million well logs, one hundred million miles of seismic reflection data, millions of linear feet of core and cuttings, and a variety of other types of scientific data.

Based on the positive results of the Phase I study, AGI began Phase II of the program in 1995. Funded jointly by DOE and industry, Phase II encompasses the establishment of standards for indexing and cataloging of geoscience data and determination of the costs of transferring data from the private sector to public-sector data repositories. Pilot projects evaluated the feasibility of the project for transfer of different data types and creation of a Web-based metadata supercatalog and browser. Also as part of the project, a national directory of geoscience data repositories was compiled to assess what data are currently available in existing facilities.

The objective of Phase II was to translate the strong interest in data transfer and preservation demonstrated in Phase I into a series of action plans in order to determine the cost and benefits, technical viability, and procedural requirements involved in the transfer of large volumes of data from the private to the public sector. The overall process included information exchanges, workgroups, and pilot projects to test and scale-up procedural aspects.

Implementation of pilot metadata repositories permitted potential users and contributors to have a clearer understanding of the system's future capabilities. These pilot implementations included a major source node at the Bureau of Economic Geology (BEG), the installation of major software components (MobilView and GeoTrek data browser prototypes), and the determination of the viability of using the Internet as the predominant choice for wide area networking by NGDRS users.

Pilot projects were also used to test user requirements for the GeoTrek software browser and to collect and organize detailed administrative and operational procedures to operate the NGDRS, including staffing requirements.

The next step, Phase III, will focus on the initiation of transfer of geoscience data from the private sector to the public domain and development of the web-based Geotrek metadata supercatalog. Contributed data will be placed in existing public sector facilities, such as state geological surveys, state oil and gas commissions, and association libraries. Users across the country will be able to access the data through a Web browser interface. DOE funding will support startup of the implementation phase and the transfer and cataloging of data. The project will also be supported through commercial subscriptions and is expected to become self-sustaining.

Introduction and Background

GOALS AND OBJECTIVES

All scientific and technical investigations require access to basic fundamental data. The capture and long term preservation of data are required to address a wide range of scientific issues. The National Research Council released a report entitled *Preserving Scientific Data on Our Physical Universe (1995)*, that took a broad look at the challenges of scientific data preservation and management in Federal Agencies. The report concluded that "a general problem prevalent among all scientific disciplines is the low priority attached to data management and preservation by most agencies. Experience indicates that new research projects tend to get much more attention than the handling of data from old ones, even though the payoff from optimal utilization of existing data may be greater." No discipline is in greater need of an increased focus on data preservation than the geosciences, where private-sector downsizing and public-sector budgetary constraints have combined to jeopardize vast quantities of valuable geoscientific data critical to our understanding of the Earth's environment and natural resources.

The American Geological Institute's (AGI) National Geoscience Data Repository System (NGDRS) was initiated in response to the fact that billions of dollars worth of domestic geoscience data are in jeopardy of being irrevocably lost or destroyed as a consequence of the ongoing downsizing of the U.S. energy and minerals industry. Preservation and access to domestic geological and geophysical data are critical to the energy security and economic prosperity of our nation. There is a narrow window of opportunity to act before valuable data are destroyed. The data truly represent a national treasure and immediate steps must be taken to assure their preservation.

The NGDRS would serve as an important and valuable source of information for the entire geoscience community for a variety of applications, including environmental protection, water resource management, global change studies, and basic and applied research. The repository system would also contain critical data that would enable domestic energy and minerals companies to enhance their exploration and production programs in the United States for improved recovery of domestic oil, gas, and mineral resources.

A model for transferring data from the private to public sector is provided by the 1994 transfer of Shell Oil's core facility in Midland, Texas to the University of Texas at Austin. Shell deeded its collection of 2.2 million linear feet of core and cuttings to the university along with its warehouse and a \$1.3 million endowment to cover annual operating expenses. All of these data entered the public domain for the first time.

The NGDRS project is highlighted in the recently released *Oil and Gas R&D Programs* report developed by the DOE Office of Natural Gas and Petroleum Technology. The report describes the current R&D programs and provides a roadmap for future oil and gas technology development by DOE. That report states a near-term goal, by 2000, of "complete integration of a national network of geoscience data repositories, assuring industry access to more than 2 million record sets of information and wellbore samples that would otherwise be lost."

In 1996, the NGDRS effort received a number of key endorsements. The Association of American State Geologists passed a resolution in support of the effort as did the Department of the Interior's Outer Continental Shelf Policy Committee, which advises the Secretary of the Interior on issues related to offshore resource development overseen by the Minerals Management Service. The American Petroleum Institute's Exploration Affairs Committee endorsed the NGDRS effort to preserve rock core holdings. Last year, the National Research Council released a report entitled *The Dynamics of Sedimentary Basins* that described the NGDRS project and recommended "continued funding for efforts to preserve, archive, and disseminate data on sedimentary basins." The report goes on to state that such data, if preserved, "will sustain continued advances in basin research for many years."

These endorsements add to the list of professional, industrial and state organizations that have gone on record in support of establishing a National Geoscience Data Repository System. On a national level, the American Association of Petroleum Geologists, American Institute of Professional Geologists, Geological Society of America, Independent Petroleum Association of America, Interstate Oil and Gas Compact Commission, Society of Exploration Geophysicists, and the Society of Independent Professional Earth Scientists are among the organizations that support the concept of establishing a National Geoscience Data Repository System.

DATA RESCUE AND NATIONAL ENERGY SECURITY

In the course of their exploration and development activities over the past several decades, major U.S. oil and gas companies have acquired enormous amounts of domestic geological and geophysical data. These companies have the most comprehensive geoscience datasets and sample inventories of information related to the Earth's crust of any organizations in the world. As major oil and gas companies downsize their operations and refocused their attention on foreign ventures over the past ten years, they have less need for domestic geological and geophysical data. Literally billions of dollars worth of subsurface geoscience information stored in companies' offices, warehouses, and other repositories is in jeopardy of being lost due to the general decline in support for curation and maintenance of data repositories at individual companies.

In addition to data centers maintained by major oil and gas companies, a variety of public and private regional collections of geoscience data exist to serve the needs of industry and academia. In the past, many of these collections received financial support directly or indirectly from major oil and gas companies, but company support has decreased significantly or terminated in the last few years. As a result, some of these facilities have closed or have significantly reduced their operations. The NGDRS could also preserve data from those facilities that are near closure or are no longer operational.

In the course of Phase I of this study, AGI documented industry's interest in contributing billions of dollars worth of inactive domestic company data files to a national repository system. The data files contain unique and detailed information on numerous localities throughout the United States. The data are in a variety of formats, ranging from digital well logs and seismic reflection data tapes, to paper and film records, to rock core and cuttings samples to paleontological collections. The diversity of data types and formats poses significant data management challenges, but failure to

rescue the data would represent a major economic and scientific loss to the nation and significantly impact our ability to efficiently conduct future investigations to assess our nation's resources.

A central goal of President Clinton's Domestic Natural Gas and Oil Initiative is to reduce U.S. vulnerability to oil supply disruptions by increasing domestic oil and gas production and reducing our dependence on unstable sources of foreign oil. In announcing the Initiative, President Clinton expressed the need to undertake "a determined effort to find and produce more domestic energy, with an industry-led solution that especially recognizes the role of independent drillers and producers." According to many independent oil and gas producers, the establishment of a National Geoscience Data Repository System would do more to aid the future search and development of domestic petroleum resources than most current or proposed programs or facilities. Such a data repository system could provide critical and hard-to-obtain information that would prevent premature abandonment of producing fields, and assist domestic producers in their evaluation of geologic trends, development of new plays, and assessment of remaining resources in existing fields.

CRISIS IN THE OIL AND GAS INDUSTRY

Oil and natural gas will remain critical components of energy supply in every nation for the foreseeable future. The United States, long a dominant producer of oil for domestic consumption, has become a net oil importer. As a result, the nation is increasingly dependent on imports from foreign areas with more readily accessible oil. This shift is particularly significant because for over 20 years, costs for imported oil have amounted to more than 60 percent of our trade deficit. Between 1980 and 1995, the United States paid \$950 billion (1987 dollars) to other countries to purchase imported crude oil and petroleum products. For several years now, crude oil imports have accounted for over 50 percent of U.S. consumption.

In 1995, domestic oil and gas operators produced 2.4 billion barrels of oil and 19 trillion cubic feet (Tcf) of natural gas. According to the Energy Information Administration estimates, domestic oil production is projected to continue to decline while demand and foreign imports increases over the next 20 years. On the other hand, natural gas production is projected to increase and expected to meet domestic demand requirements, reaching 26 Tcf by 2015. The projections assume high levels of R&D and future access to advanced exploration and recovery technologies. Despite this fact, both industry and the federal government are continuing to reduce their support for oil and gas R&D. Since 1985, industry has reduced its funding for R&D by 45 percent. At the same time, the federal government has reduced its funding for the DOE's Fossil Energy Program by 40 percent. These reductions in funding present a serious impediment to maintaining future projected oil and gas levels.

The downsizing of the U.S. oil and gas industry has been accompanied by profound changes in the basic structure and operating strategies of the industry. For the first time in their history, the major US-based international oil and gas companies are committing the majority of their operating budgets to foreign ventures. The major integrated oil and gas producing companies are selling many of their domestic producing properties, relinquishing leases, consolidating their U.S. operations, and focusing their upstream investments in overseas opportunities. Future exploration and development of domestic oil and gas resources will be conducted increasingly by independent companies. The basic foundation and organization of the U.S. oil and gas industry are undergoing profound changes.

Historically, independent oil and gas producers have played a vital role in the nation's domestic energy and economic stability. After the embargo and price shocks of the 1970's, independent operators played a key role in halting the decline in domestic oil production and reserves through expanded exploration and increased development drilling. In 1990 independent producers accounted for more than 43 percent of the oil production in the lower 48 states, and nearly 60 percent of the total U.S. natural gas production. Additionally, independent operators drill more than 85 percent of all wells, both onshore and offshore, in the lower 48 states and Alaska. The role of independent oil and gas producers may become even more important as the major oil and gas companies continue to shift their emphasis to foreign investment opportunities, as the potential for discovery of large new domestic oil and gas fields declines.

TECHNOLOGY TRANSFER

Traditionally, major oil companies have operated large research laboratories that developed more efficient methods of petroleum exploration and production technologies. Advances in petroleum technology, first utilized by major companies, ultimately were transferred to small companies and independent producers. The historic symbiotic relationship between major oil companies and independents has largely disappeared. Both major companies and independent operators would benefit by continuation of research and information transfer efforts performed in major company research facilities that have been severely downsized. The National Energy Act calls on the federal government to increase its participation in technology transfer, and to be more responsive to the needs of industry and the public. Considering the important role that independent operators play in domestic oil and gas production, their needs deserve careful attention. According to a survey of over 400 independent oil and gas producers in Texas (TIPRO, 1992), one of their highest priority technology transfer needs is improved access to "critically compiled oil and gas data." This data is needed for improved play and trend evaluation. Of particular importance is the need for increased access to geological and geophysical information derived from well logs, cores, seismic reflection data, and integrated geoscientific databases to explain the heterogeneity of complex reservoirs and subsurface fluid properties.

The Petroleum Technology Transfer Council (PTTC), a national network for the dissemination of oil and natural gas technological information to domestic producers, can serve as an effective national clearinghouse for information that is made available through the NGDRS. These two efforts have been closely coordinated to take advantage of potential synergies that can benefit the U.S. petroleum industry.

APPLICATIONS OF A GEOSCIENCE DATA REPOSITORY SYSTEM

The primary motivation for establishing the NGDRS is to protect a national treasure that is in jeopardy of being irrevocably lost or destroyed. The NGDRS would serve as an important source of information for university researchers, state and federal agencies, and private companies in addressing a broad range of issues including

- Exploring and developing energy and mineral resources
- Planning for environmental protection
- Managing and evaluating water resources

- Reducing risks from earthquakes and other geologic al hazards
- Screening sites for municipal, toxic, and nuclear waste disposal
- Designing highways, bridges, dams, and utility lifelines
- Supporting the needs of university-based research

There are many potential users of the system: petroleum companies, consultants, services companies, engineering companies, data brokers, universities, government agencies, lending institutions and the public at-large. In short, the NGDRS would benefit all elements of the geoscience profession and many additional sectors, including the major oil and gas companies that would contribute much of the data. Given the cyclical history of the exploration and production business, these same companies as well as the independent producers will make use of the preserved industry data sets.

Summary of Phase I Feasibility & Assessment Study

GOALS AND OBJECTIVES

In 1994, the American Geological Institute completed the first phase of a study to assess the feasibility of establishing the NGDRS to capture and provide accessibility to valuable geoscience data. The Phase I Report (1) documents the types and quantity of data available for contribution to the NGDRS, and discusses the data needs and priorities of potential users of the system.

AGI's Phase 1 feasibility study was coordinated by a Steering Committee composed of the presidents of three national professional associations, senior executives from major oil companies, independent petroleum producers, and representatives from oil service companies, state geological surveys, the National Research Council, and the academic community. Representatives from four federal agencies served as liaisons to the Steering Committee. AGI distributed more than 1,100 questionnaires to a broad spectrum of geoscience organizations in all 50 states, and conducted 53 site visits in 10 cities.

DATA CONTRIBUTIONS

The results of the Phase I study were extremely positive. Major oil companies, large independent petroleum producers, and minerals companies have indicated they would consider contributing vast amounts of data to the NGDRS. The amounts identified are as follows:

Cores	10,000,000	Linear Feet
Cuttings	2,500,000	Boxes
Thin Sections	30,000	Slides
Seismic - Hardcopy	1,500,000	Line-Miles
Seismic - Films	1,000,000	Films
Seismic - Digital	100,000,000	Line-Miles
Related Seismic Data	25,000	Velocity Surveys
Well Logs - Paper	5,000,000	Logs
Well Logs - Fiche	1,500,000	Fiche Cards
Well Logs - Digital	600,000	Tapes
Scout Tickets	2,500,000	Fiche and paper
Geochemical Analyses	500,000	Paper

The total amount of seismic data identified is conservatively estimated to represent more than 100 million line miles, which constitute a substantial fraction (perhaps 25 percent) of all seismic data collected in the United States since 1950. The rock core and cuttings identified in the table below are estimated to represent a significant fraction (perhaps 60 percent) of the core and cuttings held by the major oil and gas companies. Companies participating in the study have indicated that they would substantially increase their data contributions once the NGDRS has been established.

The proposed industry contributions represent billions of dollars worth of geological and geophysical data that were collected at no expense to the federal government. In some cases, the data are unique

and cannot be replaced because of urban development and new restrictions that place land off limits to resource exploration and development activities. All private sector data placed in the NGDRS would enter the public domain for the first time and become available to all users. State and federal agencies have indicated that the large volumes of geoscience data already in their possession would also be made available through the NGDRS.

The Phase 1 study also found an excellent match between the types of geoscience information that companies would be willing to contribute to the NGDRS and the types of data that are of greatest interest to independent petroleum producers and other potential users of the repository system.

ESTABLISHING A REPOSITORY SYSTEM: RECOMMENDATIONS FROM PHASE I

Potential data contributors and data users agreed that the most important issues and recommendations regarding the establishment of a National Geoscience Data Repository are the following:

- Establish a national repository system composed of an integrated network of data centers, rather than a single, centralized facility.
- Build upon existing state and regional data centers as a starting point for developing a distributed national repository system.
- Ensure that researchers can effectively use the data by placing ease of data access at the heart of the data management system.
- Establish a web-based national directory of data centers that includes information about data holdings and guidance for locating and obtaining data.
- Ensure full and open access to all data in the repository system through the establishment of an integrated metadata web site.
- Establish procedures and criteria for setting priorities for data acquisition, retention, and purging to ensure preservation of the most important and highest quality data and to maximize cost-effectiveness.
- Develop a plan and mechanism for efficiently transferring massive amounts of private-sector data to a public repository system.
- Adopt national standards for storing, processing, and distributing data in order to reduce costs and increase productivity.
- Act quickly to capture large volumes of data in jeopardy of being lost.
- Develop a plan for long-term funding, including adequate support from the private sector, federal and state agencies, and the user community.

The broad, positive response to Phase I clearly indicated the need to proceed expeditiously to Phase II, which would address the specific organizational and operational requirements for establishing a National Geoscience Data Repository System.

Results of Phase II Project: Planning, Directory, & Pilot Studies

PROJECT OVERVIEW AND SCOPE

The second phase of the NGDRS project was initiated in response to the positive findings of the Phase I study. The Phase II study addresses specific organizational and operational requirements for establishing and implementing the NGDRS. Phase II was a joint industry, academic, and government project and was led by the project steering committee established in the Phase I feasibility study. The project was initiated during the first quarter of 1995 and completed in April, 1997. Phase II focused on the three major components outlined below.

1. **Planning and Specification Activities.** The planning and specification component of the Phase II study identified the business and organizational requirements and operational plan for implementing the NGDRS.
2. **Pilot Projects.** Four pilot projects involving the transfer of various types of geoscience data from the private sector to the public sector were completed, providing critical information on the development of an efficient and cost-effective NGDRS. Specific pilots were aimed at understanding and defining scale-up, indexing, and cost metrics for various geoscience data types.
3. **Directory of Geoscience Data Repositories.** The directory resulting from this effort is critical to the implementation and operation of a decentralized but networked NGDRS, as it provides a means to identify what data are currently available, their location, and how they can be acquired.
4. **Steering Committee Operations.** The operations of the NGDRS Steering Committee represents a fourth activity separate from the three principal components of the Phase II project but no less important. The 25-person Steering Committee established to oversee the Phase I study continued to provide oversight for Phase II of this project. The committee held biannual review meetings and other ad-hoc meetings as required to consider technical and policy issues. The committee and its operations are described later in this chapter.

PROJECT OVERSIGHT AND SUPPORT

Phase II was a joint industry, academic and government project with major support and input from the following organizations listed in Table 1.

Table 1. Organizations Providing Input and Support to NGDRS Project

ORGANIZATION	ROLE
U.S. Department of Energy, Bartlesville Project Office	<ul style="list-style-type: none"> • Grant funds for support of Phase II. • General guidance and oversight of project
American Geological Institute (AGI)	<ul style="list-style-type: none"> • Overall project management control, reporting, and budgeting • Liaison with DOE, Steering Committee, and project sponsors
NGDRS Project Sponsors: Conoco, Chevron, Maxus, Mobil, Pennzoil, Unocal, PGS and Gas Research Institute	<ul style="list-style-type: none"> • Provide match funding support for project • Oversight and general guidance of project objectives, metrics and progress
Mobil E&P Technology	<ul style="list-style-type: none"> • Contributed metadata browser prototype software • Contributed Seismic data set - Grand Isle South, Louisiana
PGS	<ul style="list-style-type: none"> • Provided Gulf of Mexico digital seismic dataset
The Information Store	<ul style="list-style-type: none"> • Project planning and management of sub-contractors • Assessment, development, and installation of metadata browser prototype software • NGDRS architecture specifications and procedures for utilization and operations
University of Texas, Bureau of Economic Geology	<ul style="list-style-type: none"> • Contributed digital data store of cores and cuttings • Established cost model for transfer of cores and cuttings
POSC	<ul style="list-style-type: none"> • Data modeling support • Support for the loading of POSC data stores
NGDRS Steering Committee	<ul style="list-style-type: none"> • Oversight and review of project objectives, status, and plans

REPORT AUDIENCE

This report is primarily directed at the users of geosciences data, and in particular those responsible for the data management aspects. It is intended to provide a general but comprehensive understanding of:

- The overall business, administrative and organizational, and infrastructure required to transfer and create large volumes of various types of geoscience data from the private to the public domain;
- The specifications, i.e. data standards and system architecture, required for the implementation of the NGDRS;
- The business cost model and technical procedures required to operate the NGDRS;
- The systems required for improved access to data at regional repository centers.

The intent of the report is to provide high-level information to permit a business and technical operations plan to be established. It is *not* intended to provide detailed technical specifics on the creation or administration of any singular component of a typical geoscience data repository.

PROJECT APPROACH

The overall process adopted to deliver the components required for Phase II focused on three key initiatives:

1. Establish contacts with private-sector data providers and existing regional repositories
2. Define business and technical specifications and standards for Phase III implementation
3. Identify and select data access and data browsing software tools

The first initiative is **development of relationships** with many organizations whose holdings may become part of the national network of data stores available to NGDRS users. These relationships were developed primarily through information exchanges, phone and mail contacts, and site visits. The key contacts involved a variety of organizations and events outlined in Table A.1 in Appendix A. An important output of the information exchanges is the information contained in the *National Directory of Geosciences Data Repositories*, which is a key part of Phase II. The directory contains information on both public and private data repositories that house geoscience and engineering data, and it will be available on the World Wide Web and updated on a periodic basis.

The second initiative was to conduct a series of pilot studies for the **development of business and technical specifications and data standards** that are required for the implementation and operation of the repository. The specifications are focused on the fundamental requirements to organize, audit, transfer, and access data in the NGDRS. Development of these specifications has been provided by exchanges of information with other existing repositories worldwide and by initiation of four pilot projects. These contacts involved a variety of organizations and projects described in Table A.2 in Appendix A. These Phase II information exchanges have led to the development of practical standards and methods for developing an NGDRS business model. In addition, these exchanges have provided the NGDRS with business and technical standards to be adopted for data auditing, data preparation and transcription, data cataloging and indexing, and data access and browsing.

The third initiative is the **development of prototype software** for data accessing and browsing. Access to data is a critical project requirement. Data not accessible or retrievable have no value to the community of potential data users. Knowledge of browsing software development was obtained by a software prototype database contributed by Mobil Exploration and Production Technology and through information exchanges with software service providers.

THE OVERALL NGDRS MODEL

DESIGN PRINCIPLES

The goal of the NGDRS is to preserve geoscience data and make them readily accessible to individuals and companies with an interest in such data. Achieving this goal requires the careful management of the process for transferring data from data providers to the public domain and the creation of an information system to allow data consumers to find and access the data. This section deals with NGDRS business management aspects, in particular with the definition of the business data model for establishing and utilizing the NGDRS.

The NGDRS was developed around the following key design principles:

- Develop the support of industry, academia, and government entities to preserve large volumes of geoscience data in danger of being discarded or destroyed.
- Reduce the costs of data preservation by sharing information and repository facilities, thereby taking advantage of economies of scale.
- Develop a centrally managed Metadata Supercatalog and a network of regionally distributed data repositories.
- Accept only data contributions that conform to agreed-upon industry standard data models.
- Improve accessibility of data already in the public domain and make publicly accessible all private-sector data that is donated and transferred to the public domain.
- Facilitate the movement of all participating repositories to the POSC industry standards environment in order to eliminate unwarranted data-formatting operations.

STRATEGIC OBJECTIVES

The NGDRS management team identified the strategic objectives (or action plans) for Phase II. These goals were deemed necessary to make the stated NGDRS vision become a reality:

- Encourage the widest participation of as many of the existing geoscience data repositories as possible. The larger the number of participating repositories, the more valuable the overall data content of the NGDRS.
- Share and exchange plans and knowledge with as many organizations as viable from all the of the types of identified stakeholder organizations. This openness is aimed at making the NGDRS a win-win proposition for all.
- Interface and share information with similar national repository projects worldwide to enhance the efforts related to planning, development and implementation of the U.S. repository. This collaboration will streamline the implementation time and reduce cost for everyone.
- Adopt commonly used standards for all aspects and components of the system giving strong preference to industry POSC adopted standards. In the long run, this will help control costs and schedules and will provide users with common access to data. Furthermore systems and technologies based on standards tend to have a longer life cycle and a better return on investment.

THE STAKEHOLDERS

The expected users of the data repository system run the gamut from private petroleum companies, service companies, data brokers, and consultants to public government agencies and universities. All these stakeholders require easy access to geoscience and engineering data. Certain stakeholders are primarily data providers, others are data consumers, and some are data service providers. In many cases, each stakeholder may take on all three roles as each determines their overall needs for data access, data distribution and data management. To fully implement the NGDRS, it is necessary to define a financial model that recognizes that each stakeholder has its own objective and motivation for participation in the NGDRS.

The individual stakeholders must have some incentive to participate in the NGDRS, for example:

- **Petroleum and mining companies** want to donate large volumes of geoscience and engineering data to the repositories from those geographical areas where they presently have no commercial interest. Their principal motivation is a desire to reduce costs. Since those areas, however, may again become commercially viable prospects in the future, companies want to retain access rights to the data donated.
- **State and federal agencies** would like to have the data being offered for donation by the petroleum and mining companies in order to stimulate local petroleum and mining activities for the generation of future royalties. The new data sources would also be helpful in conducting new resource assessment evaluations.
- **Data brokers** view NGDRS as a new market channel for data sales and possibly for data distribution. They are interested in displaying and advertising, selling, and possibly distributing their data using the Internet-based Metadata Supercatalog with its powerful networking and electronic commerce capabilities.
- **Universities** view NGDRS as a source of geoscience information entering the public domain for the first time. The combination of many new data sources from a variety of repositories will enhance the capabilities for support of advanced multi-disciplinary research projects within the academic environment.

All data consumers need to have easy access to information on the availability of geoscience data. They require simple, but effective, information technology to provide quick access to data using a well-structured index or catalog of the data. The data consumers' needs can best be met by utilizing cost-effective networking capabilities, such as those provided by the Internet-based Metadata Supercatalog.

The market environment created by the concentration of available information to be provided by the Metadata Supercatalog will likely offer new and expanding business opportunities supporting the long-term costs of operating the NGDRS. A plethora of ideas could be stimulated on methods of improving the value of the data through reprocessing with state-of-the-art signal processing and re-interpretation using modern workstations. A secondary market in re-selling the enhanced information may also evolve. The results created by these activities could lead to new exploration opportunities in the USA in addition to new job opportunities in the expanding service business.

BUSINESS MODEL

The business model must balance the requirement and incentives for data contributors, data stores, and data consumers. The model must be able to provide:

- Strong incentives for contributors to donate data to existing geoscience data repositories.
- Reasons for existing geoscience data repositories to share their data via the NGDRS Metadata Supercatalog Internet database.
- Incentives for all users to make use of the updated metadata catalogs.
- Commercial incentives for certain users, such as independent petroleum companies, to access and trade data using the NGDRS infrastructure.
- The financial means to support the NGDRS infrastructure and operations, beyond the initial installation and start-up phase.
- A market place for selling and trading data for petroleum companies, data brokers, service contractors and financial institutions.

In addition the business model needs to consider these additional factors :

- The natural economies of scale of data transfer projects; large projects will permit infrastructure costs to be shared at a lower unit cost, such as with core repositories; small projects, not economically viable by themselves, may be implemented if a cost-effective infrastructure is already in place.
- The role of industry standards such as a standard data model and data exchange system can reduce the overall cost of creating and managing the repository system.
- The role of a shared, common information technology infrastructure, such as Internet and data access software can decrease the overall cost for creating and managing the repository system.

As part of Phase II, the following financial components were formulated and proposed. They will take final form during the implementation of the NGDRS in Phase III (Figure 1).

- Short-term operational funds to support on-going data transfer, including indexing and cataloguing, will be provided by federal government and industry support. This is in line with DOE plans and current budget as well as the identified national need to preserve geosciences data for the public good.
- The one-time cost to create and start-up the Metadata Supercatalog will come from DOE, conforming with existing plans and budget already in place. Shared infrastructure will be managed by AGI.
- Long-term utilization and operational support for the repository system will be generated by modest user fees and revenues from subscribers for companies or agencies that will make use of the NGDRS to advertise and promote the existence of their data, in private or public repositories.

NGDRS

Funding and Revenue Model

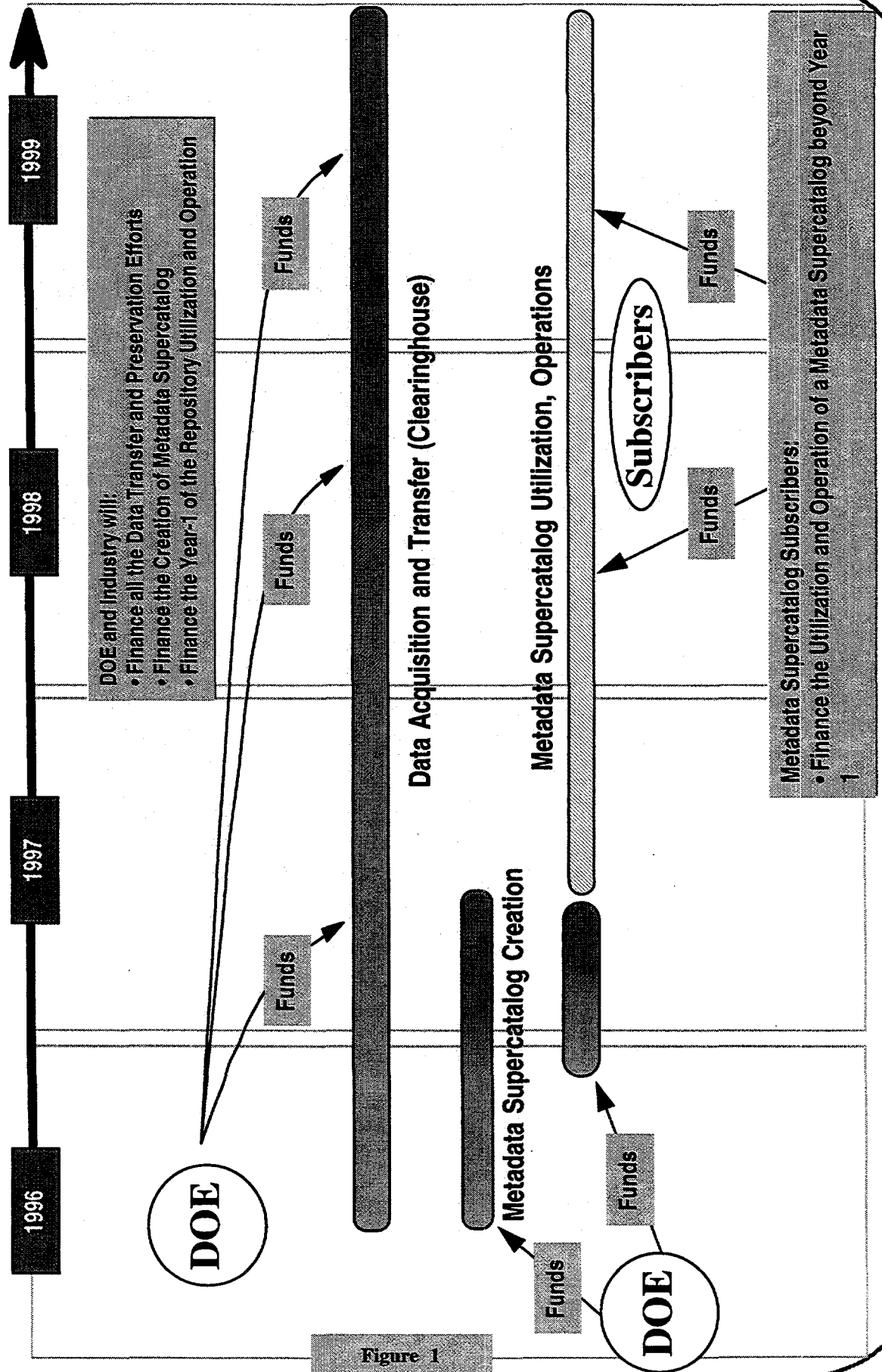


Figure 1

MODELS FOR DATA TRANSFER AND ACCESS

The Phase I feasibility study identified and provided the initial inventory of the geoscience data available to the NGDRS. Initial priorities were also established by type of data (e.g., well or drill-hole, geophysical) to be offered to the NGDRS. At the same time, discussions were initiated regarding the financial models required for the transfer of the different data types. In 1994, the transfer of Shell Oil's core facility in Midland, Texas to the University of Texas at Austin offered an example of a business model for transferring core and cuttings data from the private to public sector. Shell deeded its collection of 2.2 million linear feet of core and cuttings to the university along with the warehouse facility and a \$1.3 million endowment. The endowment provided long-term financial support to cover annual operating expenses of the repository. All of these core and cuttings data entered the public domain for the first time. It was the strong positive initial commitment from many private-sector organizations willing to donate data to the NGDRS that led AGI to seek support to initiate Phase II of the project.

The primary objective of Phase II was to translate this strong interest in data transfer and preservation into a series of action plans in order to determine the cost and benefits, technical viability, and procedural requirements necessary for the transfer of large volumes of data from the private to the public sector. The overall process included:

- Information exchanges and discussions with petroleum companies, service vendors, data standards organizations, and managers of similar national repository system projects worldwide.
- Workgroups made up of experienced data management personnel were convened to define the detailed financial models required for the transfer of cores and cuttings, well log tapes, seismic data tapes and hardcopy data.
- Pilot projects were undertaken to test and scale-up many procedural aspects of data collection, preparation, loading and index/cataloging.

DATA AUDIT AND PRIORITIZATION

In preparation for the data transfer activities, the project team conducted a series of visits to those companies that had demonstrated interest during Phase I to donate data to the NGDRS. Table 2 below lists the visits made during Phase II that were used to define the data transfer priorities for Phase III implementation. Essentially, the outcome of the visits was quite similar to the one obtained during the comprehensive survey performed in Phase I, suggesting the following two key priorities for data transfer. The companies indicated that their first priority was the transfer of rock cores and cuttings, and secondly geophysical data (e.g., 3-D and 2-D Seismic, and well logs). Donors view rock cores and cuttings as their key priority because of the potential cost savings associated with the usage of shared-core warehousing and service facilities. Furthermore, since samples provided by cores are not repeatable, it is critical to avoid the disposal or destruction of cores. Hence the project team has made the data transfer and preservation of cores the first priority of NGDRS.

Table 2. Data Audit and Prioritization Survey Visits

COMPANY	DATE OF VISIT(S)	COMPANY	DATE OF VISIT(S)
Amoco	July 1996	Marathon	July 1996
Anadarko	July 1996	Maxus	April and July 1996
ARCO	May 1996	McMoran	November 1995
Ashland	July 1996	Meridian	July 1996
BP	May and August 1996	Mitchell Energy	August 1996
Chevron	July 1996	Mobil	May and August 1996
CNG	November 1995	Murphy Oil	November 1995
Coastal	July 1996	Oryx	August 1996
Conoco	July 1996	Pennzoil	July 1996
Enron	July 1996	Texaco	May and July 1996
Enserch	August 1996	TGS	July 1996
Exxon	July 1996	Unocal	May and July 1996
Fina	May 1996	Vastar	July 1996

DATA TRANSFER GENERAL STEPS

The effective and efficient transfer of large volumes of data from the private sector to the public domain requires a considerable amount of planning and effort. The following list of general procedural steps needed for the transfer of the major data types involved in the NGDRS project is summarized in Table 3.

Step 1 involves identifying a donor interested in providing a specific dataset and willing to make the necessary investment in time and money to inventory, audit, and prepare the data for transfer.

Step 2 requires identifying, contacting and, securing a recipient repository for the dataset. For physical data sets, such as cores and cuttings, this step is a major one because the space requirements are a major consideration. For digital data sets, the typical strategy will be to transcribe and compress the original tapes to a media that requires two or three orders of magnitude less physical space such as from 9-track magnetic tape to DD-2 cartridges.

Step 3 is needed to define the business model and includes determining costs and benefits (tangible and intangible), reviewing legal and tax issues, and creating and signing the legal documents required for the transfer.

Step 4 requires an inventory of the data considered for transfer. For digital data sets and electronic databases, it will be necessary to define the number of datasets involved, their content and supporting field data. The donor organization should also value the information and ascertain their exclusive rights to the data. Data that are jointly owned need clearance from all parties before transfer can proceed. The data inventory serves as the basis for selecting and prioritizing the transfer of datasets.

Step 5 is used to select specific datasets and the associated metadata from the existing inventory. In the case of well or seismic tapes, for example, the observer's logs, acquisition parameters, processing parameters, and position data must be collected, organized and quality-controlled. Datasets should be as complete as possible, and the donor must be willing to spend the time ensuring the integrity of the dataset. Typically, at this stage, digital datasets are in some low-density media, such as 9-track tapes.

Step 6 involves the definition and selection of vendors for the services required in the following steps. This step ends with the signing of service agreements and contracts with the selected vendors.

Step 7 involves the preparation of the datasets for transfer. This step varies considerably with data type. Core and cuttings require careful identification before boxing, which may require bar codes or other identification procedures. Digital tapes are generally identified using a single unique tape label number. Because the contents of these low-density tapes may not be known (such as what wells or what seismic lines are on a tape), there is a need, either during the inventory step or during this step, to list the tape contents in preparation for the transcription step.

Step 8 involves packing the dataset for the physical movement of the dataset. This is most critical for all physical datasets, particularly cores and cuttings.

Step 9 involves the transportation of the physical datasets to their new location.

Step 10 involves unpacking and storing the physical datasets at their new location.

Step 11 involves transcribing the digital datasets to higher-density media. Tape transcription involves moving the data to high-density media, for example DD2 tape, and using an industry-established standard such as the POSC Exchange Format (PEF). Hardcopy datasets are also commonly transcribed, a process involving scanning or possibly digitizing of the data. The primary reasons for transcribing the datasets are: 1) longer life of the new high density media, 2) lower operating cost for the receiving repository, and 3) ease of access by the data consumers.

Step 12 involves converting and reformatting electronic databases. This step typically requires defining the data model of the contributed dataset and converting it to a database management system supported by the designated receiving repository.

Step 13 is necessary to define the metadata elements that are included in a particular data set. For example, when transferring core information from a particular well, the data may not contain the exact coordinates for the well. If the data catalog of the receiving repository expects that these data elements are included, then data loading problems will occur. It is necessary therefore to map the metadata of the dataset to correspond with that of the standard catalogs. It is imperative that the metadata be captured before the data is transferred to a data repository to be sure the data is accessible to future data consumers.

Step 14 requires the definition of the standard metadata catalog elements for the receiving repository.

Step 15 involves the loading of the records into the receiving repository's metadata catalog. Physically, computer-processed information is organized in records and the records into files (or tables in relational database terminology). Each receiving repository treats the definition of records differently. A primary record may be a well, and the cores may be sub-records connected by a key to the primary record. In other systems, however, each core may be the primary record. Under the NGDRS project, however, all data stores of the same type (for example cores) will have a common standard metadata catalog.

Step 16 is required when the installation of an electronic database is involved.

Step 17 involves defining and implementing the necessary procedures to dispose of the dataset. This is particularly important for the large digital datasets accumulated by the exploration and production companies. When photographed and/or scanned, paper and in some case cores may also be candidates for disposal. The costs associated with this step are defined as part of the financial model.

Step 18 is the last major step and requires that metadata catalog records and database "records" be accessed to assure the transfer has been completed and correctly done.

Each repository in the NGDRS will have its own basic process for data purging, consolidation, and retention. It is strongly suggested that before any data is purged that it be advertised on the NGDRS

Metadata Supercatalog to see if others may have a need for the data.

Table 3. General Procedural Steps For Transfer of Major Data Types

No.	Step Name	Applicable to			
		Physical Data	Digital Data	Electronic Databases	Hardcopy Data
1	Identify donor	X	X	X	X
2	Identify recipient	X	X	X	X
3	Define business model for transfer	X	X	X	X
4	Inventory data set	X	X	X	X
5	Assess dataset integrity	X	X	X	X
6	Select service vendor	X	X	X	X
7	Prepare dataset	X	X	X	X
8	Pack dataset	X			X
9	Transfer physical dataset	X			X
10	Unpack dataset	X			X
11	Transcribe dataset		X		(X)
12	Convert dataset			X	
13	Define metadata elements	X	X	X	X
14	Define catalog	X	X	X	X
15	Load catalog and index records	X	X	X	X
16	Install dataset			X	
17	Define dataset disposal procedure		X		X
18	Test access to metadata and dataset	X	X	X	X

(X) Denotes optional step

CLEARINGHOUSE

Once the data is prepared, the metadata will be cataloged and indexed at the NGDRS metadata repository. Once cataloging is complete, the data and metadata will be sent to the receiving data repository. The process of prioritizing data contributions and matching data donors with receiving data repositories is going to take a considerable amount of effort, as the receiving repository is taking on the long-term obligation of preserving, managing, and making accessible the data they receive. The Clearinghouse function (Figure 2), operated by AGI, manages this relationship between the data contributors and the receiving data repositories. AGI is responsible for facilitating the Clearinghouse review and prioritization for donated data and matching donors with recipients.

The role of the AGI Clearinghouse will include:

- Identifying specific datasets for acquisition and transfer, and also receiving repositories.
- Prioritizing those datasets based on business, logistics and other criteria.
- Allocating the resources for data transfer, transcription, and preservation.
- Coordinating data transfer from the donor to the receiving repository.
- Selecting and supervising service providers to handle the movement of physical data and conversion services e.g. tape transcription and hardcopy scanning and digitizing.
- Setting and enforcing basic quality-control standards to be adopted by data donors and recipients. The Clearinghouse management may reject the transfer of data sets that do not meet the established quality control standards.
- Ensuring that information of new acquisitions is widely distributed to regional data repositories and potential users.

NGDRS Clearinghouse Coordination Activities

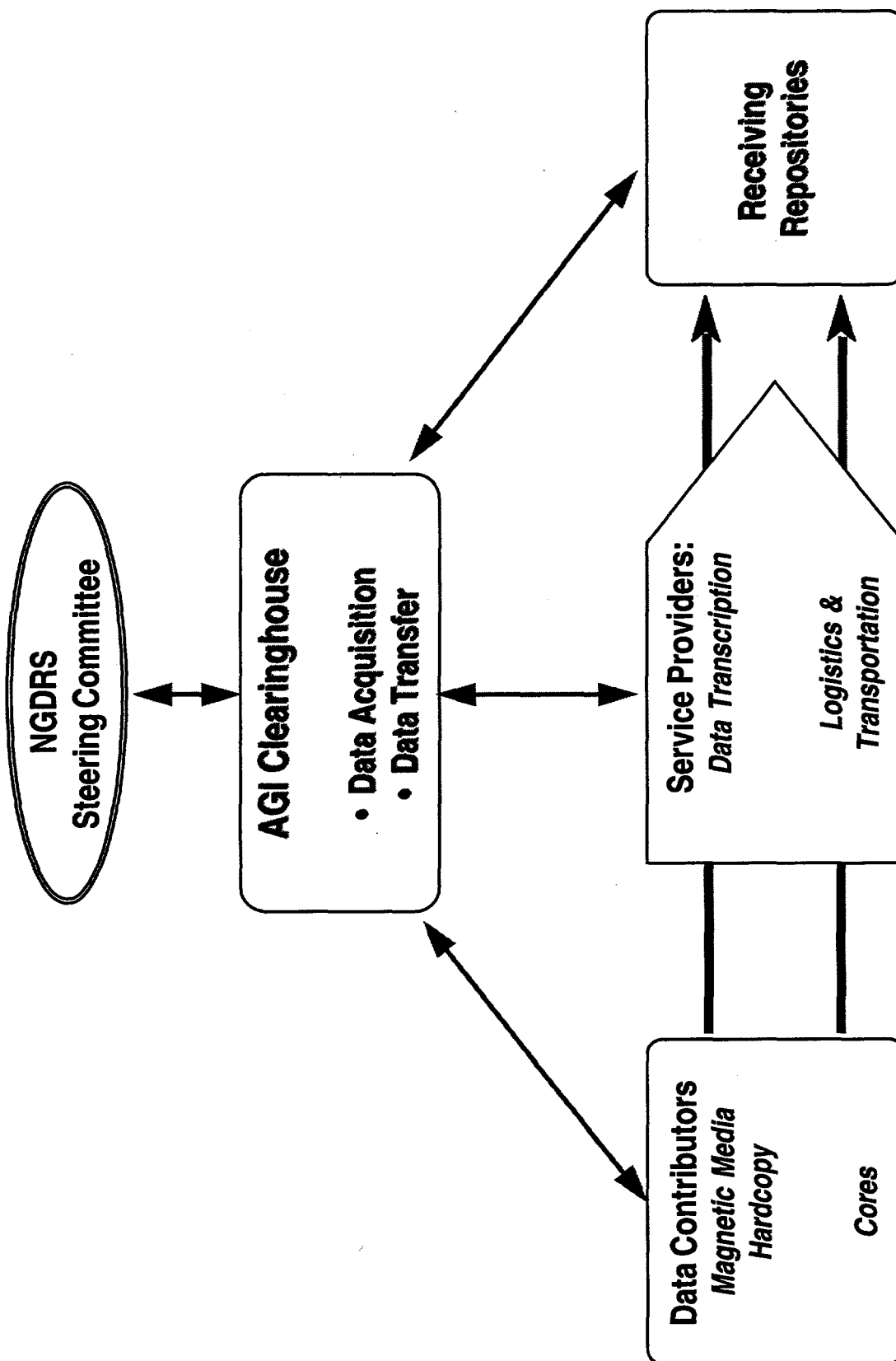


Figure 2

FINANCIAL MODELS FOR DATA TRANSFER

Four pilot projects were conducted to study the cost of implementation of the NGDRS and permit scale-up factors to be determined. Two of the pilots involved transferring data (e.g., seismic, core and cuttings) from the private sector to the public domain and were used to evaluate the time and cost involved in transferring the data. Tables 4 and 5 below present the financial models derived from the cost pilot projects.

Transfer of Seismic Data

The following business objectives were identified as part of this financial model:

- Preserve seismic data on digital media
- Reduce the cost of storage and managing seismic data
- Enhance access capabilities to digital media

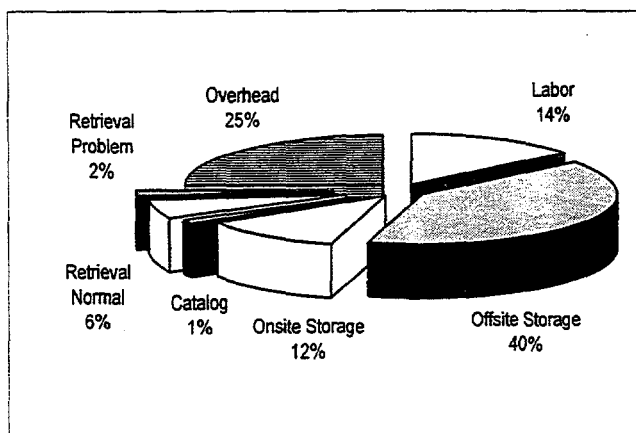
The following basic assumptions were adopted in the development of the model:

1. The majority of the current seismic tape media is 1/2 inch 9-track magnetic tapes, and no robotic units are available for data access.
2. The number of tapes involved in the transfer is 800,000, the total for Company X.
3. The technical strategy adopted is to transcribe and preserve digital seismic data on DD2 high-density tape media.
4. The metadata will need to be collected before the data is transcribed to DD2. The metadata includes, but is not limited to, observers logs, acquisition parameters, processing parameters, and navigational position data.
5. Three (3) years will be required for transfer of the tapes.
6. Planning time frame for the model is twenty years.
7. No tax write-offs were assumed.
8. Future value of money was not included - constant dollars are assumed.
9. DOE funds are available for the data transfer activities.

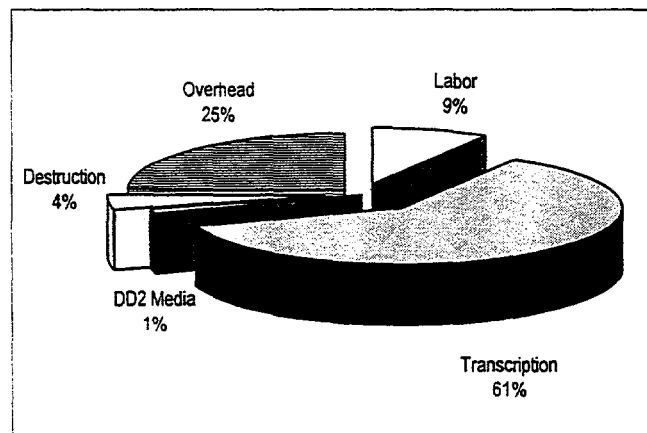
Table 4. Transfer of Tapes to Public Domain

	1	2	3	4	5	6	7	8	9
Typical Costs For Maintaining Existing Repository									
	Items	Volumes	Units/Year	(\$ Cost/Unit - Range		(\$ Cost/Year - Range		% of Total	Comments
				Low	High	Low	High	Cost (Ave)	
1	Labor	3.0 Persons		40,000.0	50,000.0	\$ 120,000.0	\$ 150,000.0	14.1%	
2	Offsite Storage	700,000 Reel		0.50	0.60	\$ 350,000.0	\$ 420,000.0	40.4%	At Salt Mine
3	Onsite Storage	100,000 Reel		1.00	1.20	\$ 100,000.0	\$ 120,000.0	11.6%	At Co. warehouse
4	Catalog	1 System		10000.00	15000.00	\$ 10,000.0	\$ 15,000.0	1.3%	
5	Retrieval Costs	10,000 Tapes		5.00	7.00	\$ 50,000.0	\$ 70,000.0	6.2%	Normal Tapes
6	Retrieval Costs	1,000 Tapes		15.00	20.00	\$ 15,000.0	\$ 20,000.0	1.8%	Problem Tapes
7	Sub-Total					\$ 645,000.0	\$ 795,000.0		
8	Overhead		Percent	30.0%	35.0%	\$ 193,500.0	\$ 278,250.0	24.5%	Legal, Human Resources...
9	TOTAL	800,000 Tapes				\$ 838,500.0	\$ 1,073,250.0	100.0%	
10	Cost Per Tape					1.05	1.34		
11	20-Year Cost	20 Years				\$ 16,770,000.0	\$ 21,465,000.0		

Existing Cost Breakdown



Transfer Cost Breakdown



Typical Transfer Costs

	Items	Target Volumes	Units/Year	(\$ Cost/Unit - Range		(\$ Cost/Year - Range		% of Total	Comments
				Low	High	Low	High	Cost (High)	
12	Labor	12.0 Persons		80,000.0	90,000.0	\$ 960,000.0	\$ 1,080,000.0	8.7%	4 persons for 3 years
13	Transcription	800,000.0 Tapes		8.0	10.0	\$ 6,400,000.0	\$ 8,000,000.0	61.1%	
14	DD2 Media	400.0 Tapes		400.0	400.0	\$ 160,000.0	\$ 160,000.0	1.4%	Reduction is 2000 to 1
15	Destruction	800,000.0 Tapes		0.5	0.8	\$ 400,000.0	\$ 640,000.0	4.3%	
16	Sub-Total					\$ 7,920,000.0	\$ 9,880,000.0		
17	Overhead		Percent	30.0%	35.0%	\$ 2,376,000.0	\$ 3,458,000.0	24.5%	
18	TOTAL	800,000.0 Tapes				\$ 10,296,000.0	\$ 13,338,000.0	100.0%	
19	Cost Per Tape					\$ 12.87	\$ 16.67		
20									

Summary - 20 Years

		Low	High	Comments
21	Current Cost	\$ 16,770,000.0	\$ 21,465,000.0	20 Years
22	Transfer Cost	\$ 10,296,000.0	\$ 13,338,000.0	To be paid with DOE funds
23	Transitional Cost	\$ 2,515,500.0	\$ 3,219,750.0	Running existing repository for 3 years
24	Future Cost			
25	Savings	\$ 14,254,500.0	\$ 18,245,250.0	
26				
27	Payout	3.00	3.00	Years

The financial model highlights the following points:

- About 50% of existing costs are related to storage -- physical space and other items. An additional 40% of the existing costs are related to direct and indirect labor to manage the tapes.
- Existing annual data management cost per tape range from \$1.05 to \$1.34.
- About 60% of the transfer costs are related to tape transcription, an activity requiring skilled labor and specialized technologies.
- Transfer costs per tape range from \$12.87 to \$16.67.
- Assuming Company X will pay only for the running of the existing repository during the transition (3 years) and that there are no unforeseen future costs, the payoff time for the project is three years.

Transfer of Core and Cuttings Data

The following business objectives were identified as part of this financial model:

- Preserve core and cuttings for future use
- Reduce the cost of managing core and reduce future liability
- Improve access to core and cuttings data

The following basic assumptions were adopted during the development of the model:

1. The number of cores managed by Company Y is 220,000 boxes. Each box contains, on average, nine (9) linear feet of core.
2. The technical strategy adopted is to transfer core to a repository already in existence and with sufficient space to house the collection of core being donated.
3. No core will be disposed and there is no cost provision in the model for the disposal of cores.
4. No core will be photographed before or after transfer. Model has no provision for transferring hardcopy data associated with the core, such as photos or laboratory analysis.
5. All cores are properly identified. The core metadata will need to be collected before the data is transferred.
6. One year will be required for transfer of the core.
7. Planning time frame for the model is twenty years
8. No tax write-offs were assumed
9. Future value of money was not included - constant dollars are assumed.
10. DOE funds are available for the data transfer activities.

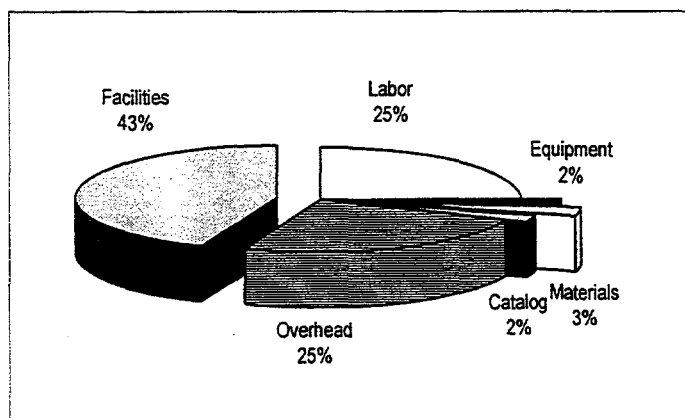
The financial model highlighted the following points:

- About 44% of existing costs are related to storage facilities. An additional 50% of the existing costs are related to direct and indirect labor to manage the cores.
- Existing data management costs per box range from \$2.20 to \$2.80.
- About 40 % of the transfer costs are related to transportation of cores and the remaining 60% are related to direct and indirect labor to manage the core transfer process.
- Transfer costs per box range from \$2.50 to \$3.60.
- Assuming Company Y will pay only for running the existing repository during the transition (1 year) and that there are no unforeseen future costs, the payoff time for the project is 0.6 years.

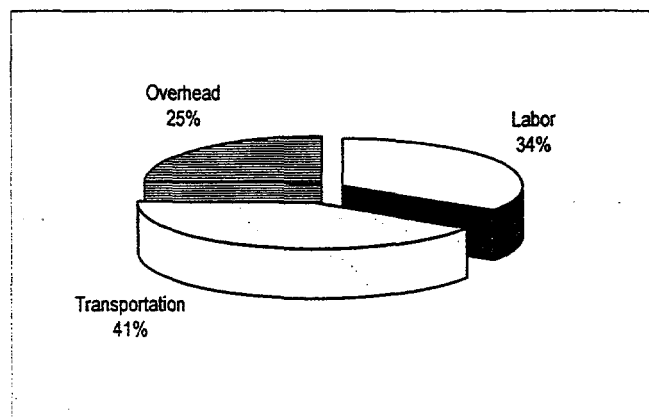
Table 5. Transfer of Core to Public Domain

	1	2	3	4	5	6	7	8	9
Typical Costs For Maintaining Existing Repository									
	Items	Volumes	Units/Year	(\$ Cost/Unit - Range		(\$ Cost/Year - Range		% of Total Cost (Ave)	Comments
				Low	High	Low	High		
1	Labor	3.0 Persons		40,000.0	50,000.0	\$ 120,000.0	\$ 150,000.0	24.7%	
2	Facilities	40,000.0 Sq. Ft		5.5	6.5	\$ 220,000.0	\$ 260,000.0	44.0%	
3	Equipment	1.0 Forklift		8,000.0	10,000.0	\$ 8,000.0	\$ 10,000.0	1.6%	Forklift operating costs
4	Materials	220.0 Boxes		50.0	100.0	\$ 11,000.0	\$ 22,000.0	2.9%	Boxes, etc..
5	Catalog	1.0 System		10,000.0	15,000.0	\$ 10,000.0	\$ 15,000.0	2.3%	
6	Sub-Total					\$ 369,000.0	\$ 457,000.0		
7	Overhead		Percent	30.0%	35.0%	\$ 110,700.0	\$ 159,950.0	24.5%	Legal, Human Resources...
8	Total Cost	220,000 Boxes				\$ 479,700.0	\$ 616,950.0	100.0%	Equivalent to 660,000 lin. ft.
9	Cost Per Box					\$ 2.2	\$ 2.8		
10	20-Year Cost	20 Years				\$ 9,594,000.0	\$ 12,339,000.0		
11									

Existing Cost Breakdown



Transfer Cost Breakdown



Typical Transfer Costs

	Items	Target Volumes	Units/Year	(\$ Cost/Unit - Range		(\$ Cost/Year - Range		% of Total Cost (High)	Comments
				Low	High	Low	High		
12	Labor	5.0 Persons		40,000.0	50,000.0	\$ 200,000.0	\$ 250,000.0	34.3%	6 months of 10 persons
13	Transportation	220,000.0 Boxes		1.0	1.5	\$ 220,000.0	\$ 330,000.0	41.2%	
14	Sub-Total					\$ 420,000.0	\$ 580,000.0		
15	Overhead		Percent	30.0%	35.0%	\$ 126,000.0	\$ 203,000.0	24.5%	
16	Total Cost	220,000.0 Boxes				\$ 546,000.0	\$ 783,000.0	100.0%	
17	Cost Per Box					\$ 2.5	\$ 3.6		
18									

Summary - 20 Years

		Low	High	Comments
19	Current Cost	\$ 9,594,000.0	\$ 12,339,000.0	20 Years
20	Transfer Cost	\$ 546,000.0	\$ 783,000.0	To be paid with DOE funds
21	Transitional Cost	\$ 273,000.0	\$ 391,500.0	Running existing repository for 1/2 year
22	Future Cost			
23	Savings	\$ 8,775,000.0	\$ 11,164,500.0	
24				
25	Payout	0.6	0.6 Years	

GEOTREK DATA BROWSER SOFTWARE DEVELOPMENT

In the process of transferring the data during the Phase II pilot projects, complete data stores were created on a computer system. In addition to well, seismic, and core and cuttings data; cultural, political, and lease boundary data stores were created. These data stores provided the framework to initiate the third and fourth pilot projects - developing data browser software and establishing metadata standards.

The third pilot project involved designing a data browser system for implementing the AGI metadata repository that would support all the data repositories participating in the NGDRS. Mobil Oil Corporation contributed a software system that enabled the design and implementation of a prototype software browser (named GeoTrek) . The browser provides a means to test the concept of accessing distributed data repositories with different data models and data access methods. The data models tested were from the Texas Bureau of Economic Geology, United States Geological Survey, Central Intelligence Agency, Mobil Oil Corporation, British Petroleum, Elf, Total, and Petroleum Geoscience Services. The data access methods included specific access to flat files, Standard Query Language, and a Data Access and Exchange layer specified by POSC.

The data browser is a map-based user interface that provides a single point of entry to find geoscience and engineering data. A data broker software component registers and keeps track of the various data stores opened by a user during a particular working session. A data access layer provides a means to access different data models and different database management systems. The data stores are collections of geoscience and engineering data in different formats, be they physical, electronic or hardcopy data.

In all, twenty-two data stores were connected to the data browser and accessed through a single world map-based user interface. Three of the twenty-two data stores were POSC Epicentre-compliant data stores. One of the goals of the NGDRS is to facilitate the move of all participating repositories to the POSC standards environment to eliminate unwarranted data formatting operations. The POSC format will be used as much as possible for donated data, for data being distributed from the data repositories, and for automation and the synchronization of all the repositories with the AGI metadata repository. Using POSC standards will eliminate cost and provide a single point of entry for the data consumer.

METADATA DEFINITION AND RELATED STANDARDS

The fourth pilot project involved determining metadata standards and protocols. The definition and the standards for metadata have evolved rapidly in the last few years. The traditionally accepted definition of metadata is, simply put, data about data. Metadata can be thought of as the information required to make scientific data useful. As suggested by the qualitative nature of such a definition, the term means different things to different people, defying precise definition. According to the Federal Geographic Data Committee: "Metadata are descriptive information about data and information resources. Typically, they describe, point to, or otherwise complement the information content of the data to which they are related. Metadata provide concise aid in locating desired information on a variety of topics, and help make such information easily accessible. Metadata may describe a range of information resources such as digital data images, databases, and printed

materials such as books, memos, or maps (Federal Geographic Data Committee, 1997)."

Standards for metadata have evolved in support of the diverse needs of many communities. "Data and information resources are generated by diverse user communities and the metadata required to describe these resources are equally diverse. Metadata must adequately describe data in terms useful to user communities and appropriate to the data or information resource. Over time, an array of diverse metadata formats have evolved which enable various organizations, agencies, and user communities to tailor metadata to specific needs. Standardizing these metadata provides uniformity to the information presented, which facilitates information sharing among various organizations and agencies (Federal Geographic Data Committee, 1997)." During Phase II several metadata standards were identified as relevant to the implementation of the NGDRS:

- AACR2: The Anglo-American Cataloging Rules (AACR2) is a standard by which library materials (such as books, audio recordings, and films) are organized and described. Descriptions of these items typically have headings and/or uniform titles to make the items more accessible in catalog searches.
- MARC: Machine Readable Cataloging (MARC) is a metadata standard used by the library community to facilitate exchange of catalog records which employ AACR2 to describe a myriad of resources.
- DIF: The Directory Interchange Format (DIF) is a data structure for directory metadata developed by NASA and maintained by the Interagency Working Group on Data Management for Global Change (IWGDMGC). The format, originally designed to describe satellite and other remotely-sensed data, suggests metadata elements for describing data, prescribes content values for selected elements, and provides a structure for transferring metadata among information systems.
- GILS: Provides a framework for individuals and organizations to improve access to their information resources including a publicly accessible catalog of federal information sources, each described using a specific metadata format. GILS supplements other government and commercial information dissemination mechanisms, and uses international standards for information search and retrieval so that information can be retrieved in a variety of ways.
- The Content Standards for Digital Geospatial Metadata (DGM): This standard defines geospatial metadata and the minimum set of metadata elements needed to describe a spatial data resource. It was approved at the June 8, 1994 meeting of the Federal Geographic Data Committee (FGDC).

The DGM was particularly useful as a basis to understand and discuss the NGDRS metadata needs (general and specific). According to the proposed DGM, metadata should include the following content items:

1. **Identification Information** - basic information about the data set. Examples include the title, the geographic area covered, currentness, and rules for acquiring or using the data.
2. **Data Quality Information** - an assessment of the quality of the data set. Examples include the

positional and attribute accuracy, completeness, consistency, information sources, and data generating methods.

3. **Spatial Data Organization Information** - the mechanism used to represent spatial information in the data set. Examples include the method used to represent spatial positions directly (such as raster or vector), indirectly (such as street addresses or county codes), and the number of spatial objects in the data set.
4. **Spatial Reference Information** - description of the reference frame for, and means of encoding, coordinates in the data set. Examples include the name of and parameters for map projections or grid coordinate systems, horizontal and vertical datums, and the coordinate system resolution.
5. **Distribution Information** - information about obtaining the data set. Examples include a contact for the distributor, available formats, information about how to obtain data sets on-line or on physical media (such as cartridge tape or CD-ROM), and fees for the data.
6. **Metadata Reference Information** - information on the currentness of the metadata information and the responsible party. Examples include currentness and information about the organization that provided the metadata.
7. **Entity and Attribute Information** - entities represent business objects and activities. Attributes include the names and definitions of all geoscience data (surface and sub-surface) features. Information about the content of the data set must include the entity types, their attributes, and the domains from which attribute values may be assigned. This information should be consistent and compliant with the POSC defined standards for geosciences and engineering data, i.e. must be formally described on the *Epicentre* data model. The following guidelines will be followed:
 - The entity will represent, and be named after, the underlying nature of an object, not the role it plays in a particular context. Typical examples of business objects in the NGDRS are well, wellbore, plant, business entity, customer, supplier, agent, repository, and data store.
 - The attributes are characteristics-of or events-related-to a particular object; typical attributes are identifier(s), description, value(s) from related activity such as a measurement, and the date and time of object creation and termination.

Appendix B contains six tables that present examples of the metadata items required for cataloging several major data types relevant to the NGDRS, including 3-D seismic data, offshore well data, wellbore data, well logging data, well core data, and hardcopy data.

Four of the examples are logically linked approximately in the way they are defined by the POSC *Epicentre* data model. Specifically the Offshore Well Data, Wellbore Data, Well Logging Data and Well Core Data are related as in Figure 3. Technically, these logical records may be implemented in different ways in computer systems. The traditional and older way of implementing logical records would use physical records in flat files of information. Relational database management systems, currently the most common database technology, will make use of tables to implement these logical records. Emerging technologies such as object oriented database management systems, will make use

of "business objects" to implement these logical records. The technical implementation of the metadata is not discussed in this report.

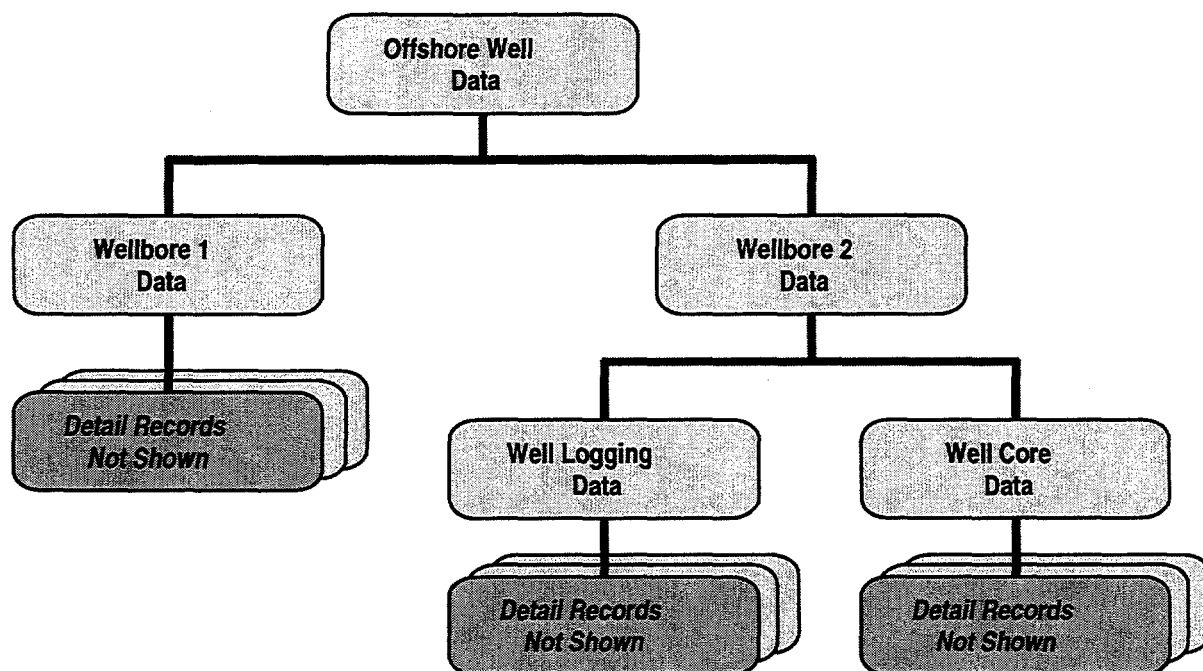


Figure 3. Logical Linking of Data Records

The example depicts the pertinent information necessary to describe the Well Data including that the particular record (or instance) has two wellbores (Wellbore 1 and Wellbore 2). Wellbore 2 is shown to have intervals that are logged and cored. The interval logged is described in the Well Logging Data record. Detailed records for individual well logging techniques (such as resistivity or SP curves) are not shown. The interval cored is described in the Well Core Data. Detailed records relative to analysis associated with cores (such as porosity) are not shown.

METADATA, CATALOGS AND SYNCHRONIZATION

The metadata examples presented in Appendix B show the information elements and a typical record of information. For practical reasons, it is very difficult to define in advance all the possible information elements required for the NGDRS metadata system. Therefore, additional implementations may be required before the NGDRS metadata definition stabilizes. Furthermore, the on-going information exchanges with other national repositories around the world are required to establish more universal metadata standards for geoscience and engineering data. Doing so will permit users to access data anywhere in the world using interfaces that are as similar as possible. It should be noted, however, that such an undertaking will probably force further adjustment and modifications of the NGDRS metadata. These modifications are viewed as part of the routine maintenance requirements associated with the **NGDRS catalog**.

Logically, the NGDRS catalog is an information file with two dimensions; one consists of the metadata elements of information discussed previously; the other consists of the records of information loaded into the file (Figure 4). For example, when dealing with core data, the first

dimension describes the metadata elements (Appendix B, Table B.5); the other dimension consists of the records of information for all cores sets in a given repository. Both dimensions are not static; the first dimension - metadata definition - as discussed above can be affected by the inclusion of new metadata elements of information. The second dimension - records of information - will be affected as soon as new core sets are added to the repository and made publicly available to all NGDRS users. .

NGDRS Metadata Supercatalog

Metadata Elements -->	Well Logs				Seismic Data							Cores				
	1	2	3	4	1	2	3	4	5	6	7	1	2	3	4	5
Records Number																
1234	x	x	x													
1235	x	x	x													
1236	x	x		x												
1238	x			x												
1249	x	x		x												
3456								x	x	x	x	x	x			
3467								x	x	x	x					
3489								x	x		x	x	x			
3589								x	x							
A10002															x	x
A10003															x	x
A10008															x	x
CA2445																
TX2345																
OK2222																

Repository A
Subcatalog
(Well Logs)

Repository B
Subcatalog
(Cores)

Figure 4

The process and procedures associated with adding new metadata elements of information or new records of information is known as **catalog synchronization**. It requires that periodic updates be performed to the master catalog for NGDRS - known as the **supercatalog**. The supercatalog contains, at any moment, the official catalog for the NGDRS since it is the collection of **all** information (metadata and records) contained and described in the **subcatalogs** created for individual data stores. As a reminder for our purposes, a data repository contains one or more data stores, typically organized around the data type given in the tables above.

It will be the supercatalog that NGDRS users will see and use as part of the data access and browsing activities. Two major synchronization activities need to take place:

- The first activity relates to **availability of information records**; simply stated, if a metadata element of information does not include the proper value, that data is not 'accessible' to the user at that moment. For example, if core XXX in a certain repository had a permeability analysis recently completed, and the analysis is either not posted in the respective subcatalog, or this same subcatalog has not been synchronized with the supercatalog, the analysis will not be available to the user. It is critical, therefore, from a user service point of view, that the catalogs (supercatalog and subcatalogs) be synchronized as often as possible. Frequent updating of the supercatalog involves both economical and technical considerations. NGDRS management will establish minimal levels of service to assure users that the supercatalog reflects as close as possible the latest data store information. This type of synchronization is expected to be performed on at least a monthly basis.
- The second activity relates to **metadata changes or additions**. As discussed previously, the full definition of NGDRS metadata may not be completed until the system has matured, i.e., until many data stores from all types of technical information are integrated into the NGDRS. Even once maturity is reached, new information elements, currently not contemplated to be part of NGDRS (such as environmental specific data) may be added to the repository. For whatever reasons, it is almost certain that periodically data elements may be added or modified to the original set of metadata. This will require changes affecting both the supercatalog and many (or all) subcatalogs for the synchronization to be achieved. This type of synchronization is expected to be performed less frequently than the record synchronization; it needs to be staged more carefully since it may require the down-loading and up-loading of all (computer) information records and thus it will, more likely, be done on a quarterly basis. To the extent these metadata extensions can be formulated using the **POSC** standards (*Epicentre*), the technical difficulties will be reduced. There are, however, many types of geoscience and engineering data not covered (or planned to be covered) by *Epicentre* such as data related to certain geophysical techniques (gravity and electrical methods).

The synchronization must be announced in advance to the users using the World Wide Web (Web) for access. New records can be promoted using the Web to alert users for new data availability. Planned extensions to the metadata can be announced well in advance to be able to gather feedback on the business value of the extensions and the possible difficulties and impacts generate by the implementation of the extensions.

NGDRS DATA OWNERSHIP AND ENTITLEMENT

From a data ownership perspective, the NGDRS will handle the following major groups of data store/repositories:

1. **Data stores already in the public domain.** This group consists of regional data stores already established and operational. Examples include many of the state geological survey core repositories in the US.
2. **Data stores in the private domain intending to remain in the private domain but wishing to announce, promote or sell their data using the NGDRS facilities.** This group consists also of data stores already organized and operational. Examples of these data stores include the seismic brokers and well-log information vendors.
3. **Data stores created as a consequence of the transfer of data from the private to the public domain.** This group consists of data stores to be organized as part of the NGDRS efforts as a result of the large volumes of data to be contributed by the private sector. An example would be the University of Texas' Midland Core Repository contributed by Shell Oil.

For all repository groups, it is assumed that *all* information provided as part of the metadata catalog will be publicly accessible in their entirety. Metadata information considered proprietary by the owners of a particular data store or repository (for group 2 above) must be withheld (not filled-in) or better filled-in with the flag 'proprietary' to alert the users the data exist but is available only from the owner of the data store.

This simple approach to ownership and entitlement will be valid while the goals of NGDRS are to promote and make accessible metadata only. If the NGDRS goals are extended to include access to technical data, not just the metadata, a complex set of business rules and associated procedures will need to be implemented.

CREATING THE REPOSITORY SYSTEM

General objectives and goals for the repository system were reviewed and evaluated through a variety of ways that included site visits to existing repositories and discussions with potential data users and contributors to the NGDRS. These initial contacts helped the Project Management Team firm up the vision and strategies for the repository, including the major options for the business model for creating, utilizing and operating the repository. Initial requirements for the technical aspects of the repository were also studied and reviewed but kept at a high level architecturally since today's information technology evolution is forcing careful annual re-consideration of the technical options. The fast pace of implementation of the Internet infrastructure, for example, is a great example of the need to keep flexible.

Phase II was used to transform the vision and broad strategies identified in Phase I into specific objectives for creating the NGDRS. The principal vehicle for this phase was the implementation of pilot projects that permitted potential users and contributors to have a better understanding of the system's future capabilities. These pilot projects permitted the Project Management Team to establish more specific architecture for the design of the overall network of "nodes" for the repository, the metadata catalog repository node, and the typical nodes where the geoscience data will reside, called source nodes. The pilot projects provided for the installation of more than twenty distributed data servers, including a major source node at the Bureau of Economic Geology (BEG); the installation of major software components (MobilView and GeoTrek data browser prototypes); and the testing of the viability of using the Internet as the choice for wide area networking by NGDRS users. The practical knowledge acquired during the installation and networking of the pilot repositories, together with the detailed architectural designs, can be used for the final configuration of the metadata catalog node and the overall network of nodes. These activities represent the scaling-up phase of the process.

In parallel with the technology-oriented activities, the Project Management Team continued their efforts to develop the most desirable business model for creating, utilizing and operating the repository. Detailed information on technology, current economic factors affecting the oil and gas industry, state-of-the-art data management practices, and the status of current industry standards were combined to refine the business model to be adopted for creating the repository during Phase III.

OVERALL ARCHITECTURE OF THE REPOSITORY SYSTEM

One of the main objectives of the Phase II project was to develop the specifications and components for the overall repository system and in particular for the NGDRS Metadata Supercatalog. Because of the complexity of the systems involved, it was necessary to define the general system architectural principles and models to assure the repository will perform as expected. These same architectural considerations were used to facilitate the development of a comprehensive business model to utilize, operate, and manage the actual repository system scheduled to be created during Phase III of the project.

The general vision statement for the NGDRS states that "potential data contributors and data users agree that the NGDRS should be composed of integrated network of regional data centers, rather

than a single, centralized facility. The system should build upon existing state and regional data centers as a starting point for developing a distributed and interconnected national repository network". This vision is depicted in Figure 5.

NGDRS Overall Systems Architecture

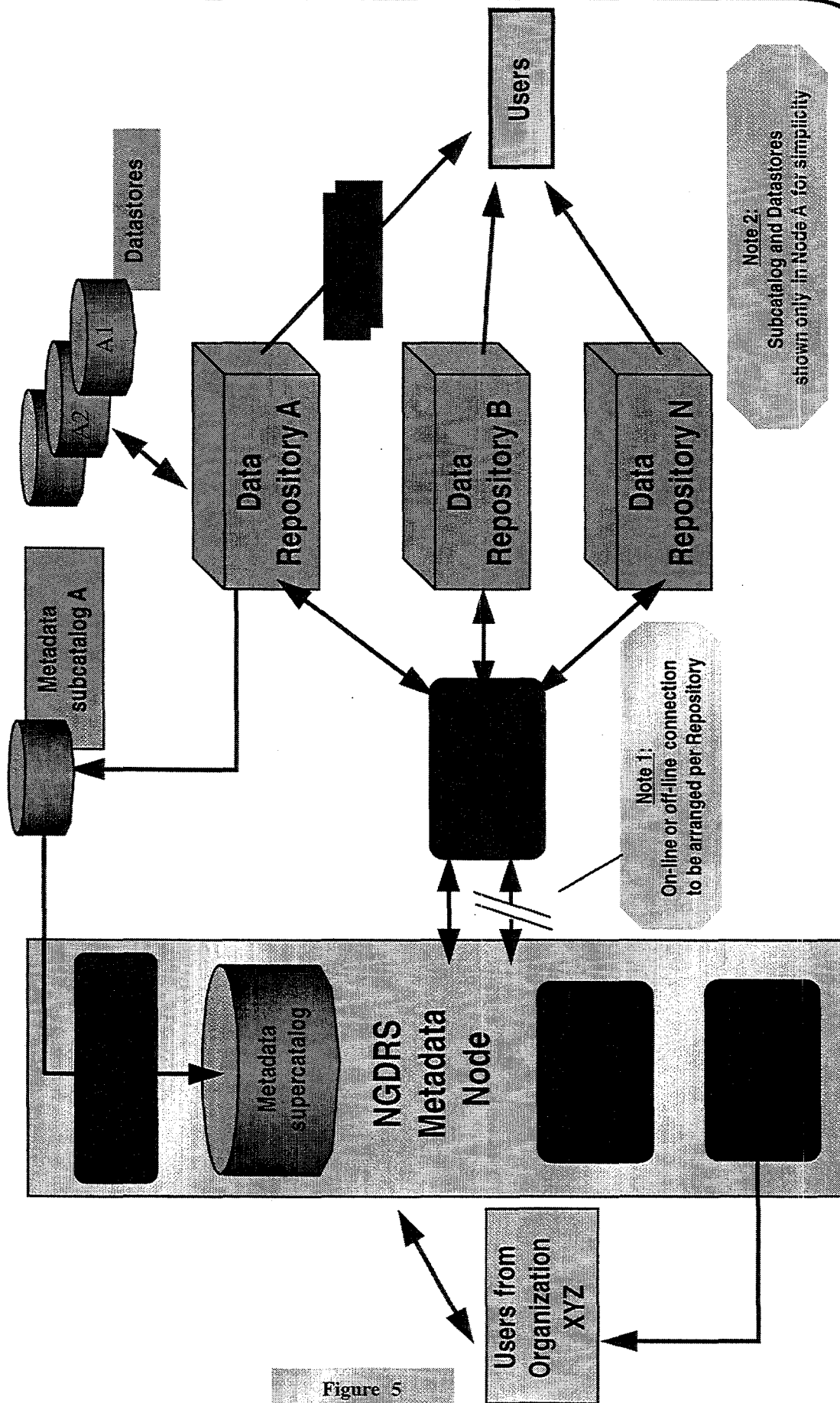


Figure 5

The System and the Source Nodes: The NGDRS will be comprised of public and private sector data repositories (which are designated as NGDRS Source Nodes or simply as Nodes) connected via the Internet. Individual Source Nodes are the responsibility of the Source Node Repository - such as for example the Texas Bureau of Economic Geology, or the University of Oklahoma, or the U.S. Geological Survey. The collection of all NGDRS Source Nodes is referred to as the System. The NGDRS will have overall responsibility for managing the System.

Data Stores: Individual Source Nodes may provide one or more data services to users. Each data service will be handled via a Data Store resident at each Node. The collection of data stores managed by a particular Source Node represents a Repository. Data stores may contain technical data (such as cores or seismic lines), reference data (such as general reference maps or tables), or metadata on either of the first two types.

The NGDRS Metadata Supercatalog Node: The System will include a NGDRS Metadata Node to provide the front-end for entry-level services required by all users of the repository. The Metadata Node will contain the master on-line catalog of data stores for the System - known as the supercatalog. It will provide users with metadata for data available across the System, its ownership, location, technology support environment, and other data characteristics.

Metadata Synchronization: Each Source Node will contain subcatalogs with the latest metadata on the data it provides. Procedures will be defined between each Source Node and the NGDRS for periodic and frequent updates (synchronization) of the subcatalogs with the supercatalog. Thus, the NGDRS Metadata Node supercatalog will contain the aggregate of the subcatalogs, and hence metadata, of all Source Nodes in the System. However, not all metadata will be synchronized simultaneously. Instead, a polling process will be established to update periodically the metadata database for the Nodes with the periodicity agreed between the NGDRS and the Source Node Repository.

User Validation Subsystem: The NGDRS Metadata Node will need to validate the user as part of a business entity (company, government organization, university, services, etc.) allowed to browse and to order **information** from the repository. Although the NGDRS is conceived fundamentally as a system open to all, because of marketing, general accounting, and billing functions it will be necessary for individual users to be validated. A single, unique password will be adopted for all individual NGDRS users.

Usage Tracking Subsystem: The NGDRS Metadata Node needs to be able to track the usage of the data within the System. The Metadata Node will employ a usage tracking subsystem which will accumulate the usage of the System resources for each user on a Working Session basis. Software packages supported by each Source Node Repository will track the usage of the resources at each Source Node.

Transfer Control Subsystem: The NGDRS Metadata Node can also connect the user directly to the Source Node that holds the target data being sought. Depending upon agreements made between the NGDRS and the participating Source Nodes, the actual data may be physically retrieved from the Source Node directly to the user or alternatively to the Metadata Node.

ARCHITECTURE FOR THE SOURCE NODES

Each Source Node in the System is comprised of a number of components to perform correctly and effectively the services specified by the NGDRS project. These components, shown in Figure 6, are described below to provide a better understanding of the general Source Node activities and to help specify the functionality for the products, services and procedures to be implemented.

Data Collection, Mapping and Preparation: Each node requires a main entry or input component for entering all the technical data. Preparing and entering the data is the most labor-intensive and time-consuming aspect of the implementation process and typically includes the following:

- Conversion of the data from the contributor to the Source Node. The conversion will likely involve logical (data protocols) and physical formats.
- Reconciliation of the data sets to assure no multiple copies of the same set are entered in the System.
- Reconciliation and selection of the data sets to be adopted and entered, such as the most modern version of a particular log suite.
- Creation of an index database listing all the documents planned to be entered. As the conversion process - from document to electronic image - proceeds, the index database will be used to help manage the process while providing the users with the latest information on the status of the documents.
- Development of specialized steps that assure data quality and data consistency for the hard copy data sets. This involves processes that address image resolution, document clarity and data completeness.
- Identify the physical location or information for all physical data sets (such as cores). This will trigger a record on the Metadata file describing the position of each major data set.
- Development of specific entry, edit, and validation systems (including software, computer files and procedures) to ensure the correctness and completeness of entry processing. The specificity is generally related with the major types of geoscience data (wells, geophysical, geological, drilling, production, engineering, etc.).

Data Modeling and Metadata Definition: The mapping of data into an established industry standard data model, such as *Epicentre* is referred to as data modeling. This step is required to rearrange the data, since each geosciences database has an implicit but different data model, to conform with industry standards, permitting both easier access to data and better means to integrate different data sets. This step typically includes manual and automated steps using data mapping software tools.

NGDRS Overall Information Flow

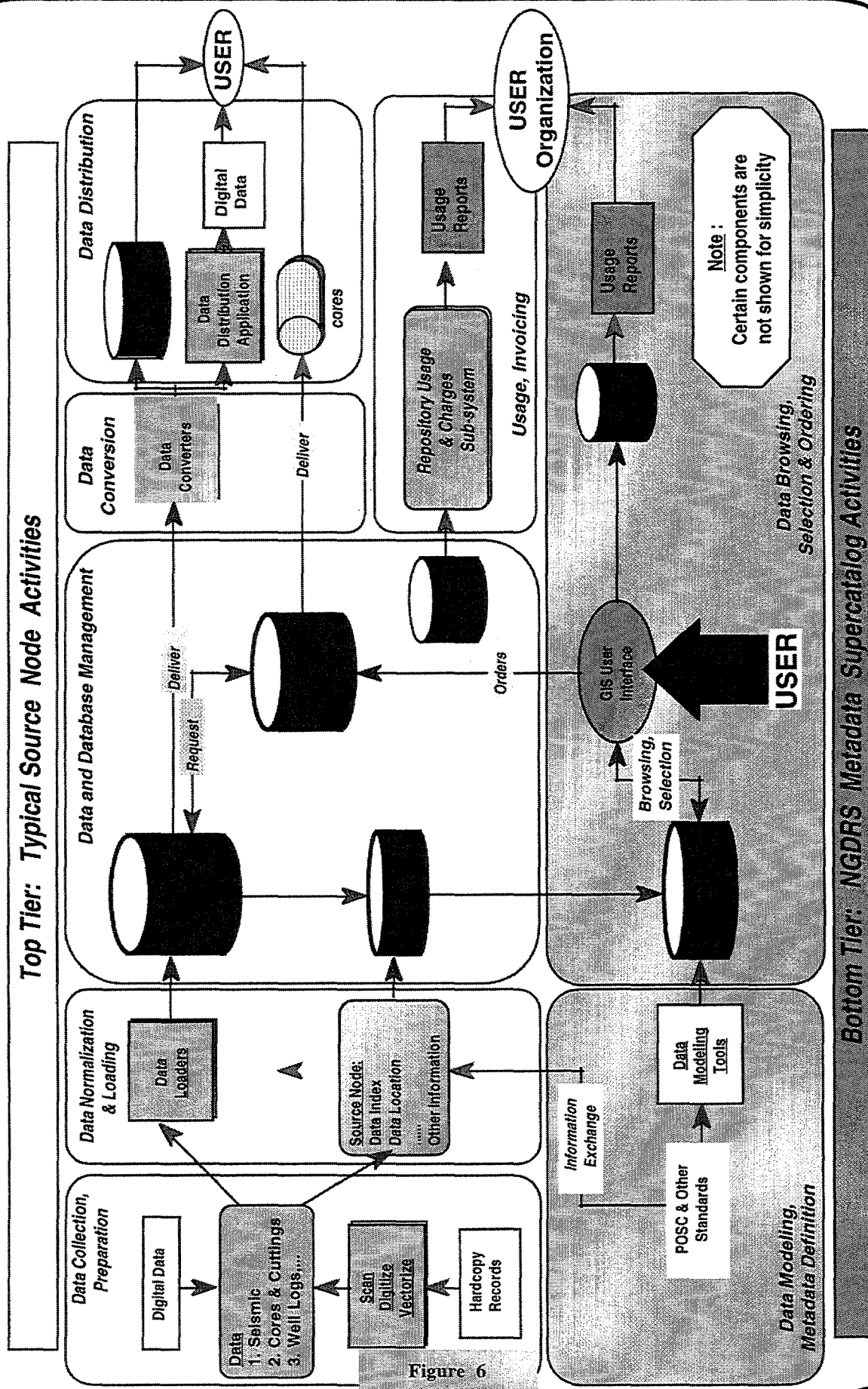


Figure 6

Data Normalization and Loading: Once the data exists in a commonly accepted electronic protocol or format - such as for example ASCII - it is loaded into a database using software data loaders. The result of this operation is the creation of databases managed under an industry-accepted database management systems (DBMS) software product. However, before loading data, some major decisions on data normalization must be made. Normalization of data involves making critical decisions on items, such as:

- Adoption of a consistent scheme for units-of-measure and the means to convert those units into other established (standard and non-standard) systems of units-of-measure; what to do with existing data fields which are not part of the adopted scheme, but historically have been abundantly used by systems.
- Development of a consistent scheme of geodetic coordinates to describe spatial and topographic data and alternate coordinate systems that may have been adopted historically. Finally it is necessary to define the algorithms for the conversion of coordinate systems.
- Representation of data is particularly critical to spatial oriented data - such as geological sample values of porosity(\emptyset) and permeability (k). These properties can be represented by a discrete individual point (value) in space attached to a certain set of coordinates (x, y, z, \emptyset , k). They can also be represented by the same coordinates in space (x, y, z) but the properties be represented by functions that describe not a single value but a set of discrete values or even an analytical function (such as relative permeability curves.)
- Representation issues occur with the time-dependent information, particularly business events (such as a decision to drill or not drill in a particular location). It is well known that the so-called commercial information systems view and treat most business transactions as if they were "instantaneous" - that is, their duration time is accepted as requiring zero units of time.

Data and Database Management: The next key component is the creation of the database management environment. Once data have been logically mapped and loaded into the database management system, it is ready to be accessed via a data browser and other application software. The major effort around the database management revolves around how physically to organize the data sets. Guidelines for how to split databases were developed in the early days of database systems and are still evolving. There are some basic factors affecting the guidelines for splitting databases. They relate primarily to system performance (how quick data is displayed) and access (how easy or simple is it for users to access data). Performance is dependent on the systems resources available (hardware, software, networking) and their expected utilization. Ease of access will be stimulated by the adoption of standard browser software.

Data Warehousing: It is expected that some of the Source Nodes will be moving to a data warehousing environment, a trend much in vogue currently in information systems. **Data warehousing** consists fundamentally in organizing and integrating data from multiple sources into a subject database where the business objects are closely related. A fundamental objective of data warehousing is to streamline data access and facilitate browsing, querying, and analysis by users. In the process of creating the data warehousing, issues such as data quality, data normalization, and data derivations are all addressed so that users can cut their time-to-access for data. A data

warehouse is basically a read-only database environment; data is extracted from the warehouse and placed in project databases for analysis or interpretation by software applications. The updates to the data warehouse are done from snapshots or frequent updates from the basic production systems and these snapshots are done often in a batch or background operating mode. By separating the data creation, loading and processing side of environment from the data warehouse (data access) side, the systems are more responsive to user access needs since there is a reduced contention for computing resources.

Data Archiving and Storage: In addition to the database management tasks, there are also data archiving and storage requirements, which are not shown in Figure 6 for simplicity. **Data archiving** is defined as on-line or near on-line storage of information in digital form. This process includes the physical management of digital data at the source node. It is particularly important in situations where magnetic media (tape, cartridges, and disks) are involved. Typically in this situation, data management hierarchy is established: data frequently used (such as the local metadata) is kept always on-line and on 'fast' systems devices (such as hard-drive disk or even memory). Less frequently used data is kept on on-line (or near on-line) tapes or cartridges. Data archiving guidelines are normally established in conjunction with database management guidelines. **Data storage** is defined here as the off-line (sometimes off-site) storage of data such as cores, paper documents, and all magnetic media created as part of backup or disaster recovery contingency plans. Information about database management, data archiving, and data storage capabilities will be part of the metadata catalog or Internet bulletin boards.

Data Conversion: This component will provide the user with the basic facilities to translate data from internal stored formats (such as LIS for well logs) to the format requested by the user during the selection of the data sets. Software for data converting will be developed or adopted as necessary. The data selected and converted may then be organized by the Source Node or the user into **Project Databases**.

Data Distribution: Once data is ordered and converted it will be up to the data distribution component to trigger the output mode and media requested by the user for the data. This will also activate the Repository Usage & Charges Subsystem and post in a file the date and time when the data was "shipped" to the user.

Usage & Charges Subsystem: Repository users may be charged according to the overall systems resource usage by each Source Node. A subsystem is typically implemented to permit:

- Collecting at the lowest level possible (as provided by the operating system) of usage for all type of system resources: CPU, memory, storage, communications, printed reports, plots, media costs, etc.
- Maintaining a file containing all data orders processed and shipped and later received by the user requesting the data.
- Generating invoices by user, by cost center or by company, on demand or for set periods.

The NGDRS will maintain general statistics by user, organization and data type to facilitate planning and promotional activities for geoscience data.

Data Browsing, Selection, and Ordering: Repository users will enter the System using a data browser installed at the NGDRS Metadata Supercatalog Node. The data browser is a map-based user interface that provides a single point of entry to find data stores. A data broker software component registers and keeps track of the various data stores opened by a user during a particular working session. A data access layer provides a means to access different data models and different database management systems. The browser at the Metadata Supercatalog will only access metadata.

The browser environment will provide facilities for user documentation and feedback, data promotion, and advertising. These facilities will support repository requirements for data purging, consolidation, and retention for the NGDRS. It will be suggested to the Source Node repositories that before any data is purged that it be advertised on the NGDRS bulletin board, to see if others may have a need for the data. The NGDRS Metadata Supercatalog will only purge metadata that has been purged at the Source Node.

ROLE OF STANDARDS IN THE CREATION OF THE REPOSITORY

Standards are a key part of the architecture; they give meaning to the higher-level business principles defined for information architecture domains (hardware networking, data, software applications). The need for standards are key to the creation of NGDRS, because of:

- The need for simple, easy access to data; users of systems are expected to come from all sectors (private and public) and many technical disciplines with various levels of technical and scientific sophistication. The simplest "common denominator" will help guarantee that the vast majority of users will get meaningful results from the utilization of NGDRS.
- Integration of the data sets; typical geoscience or engineering projects requires multiple data sets presented or represented in many different ways; standards are key to facilitating integration of data sets.
- Administrative costs of data management; individual productivity and data administration are just two examples of the need for standards.

Standards play a role not just in the definition and creation of the metadata but in all the architectural components discussed in this section. Table 6 presents a list of the principal standards that will impact each component and sub-component of the NGDRS. The list is not exhaustive and will evolve in time; new relevant standards will be defined and may need to be incorporated into the overall NGDRS architecture.

Of particular importance in the creation of the NGDRS is the role of the Petrotechnical Open Software Corporation (POSC) standards and specifications. Created and managed by the petroleum industry, POSC acts both as a provider of standards and a catalyst for the inclusion of standards created by other standards bodies. It is a prime driver and creator of E&P industry specific standards (such as the *Epicentre* data model and PEF data exchange), and it is also a prime participant in non-

E&P specific IT standards committees. POSC is poised to see the industry succeed, because it is managed by the major stakeholders in the petroleum industry. POSC standards are driven by "open thinking" and "wide access" to all users, characteristics that are the reason for the success of open systems (such as UNIX) and the World Wide Web (WWW) protocols.

NGDRS will maintain a strong POSC connection to assure the integration of all pertinent protocols and standards. The POSC and AGI memberships (which are closely aligned) will provide feedback to the NGDRS to ensure that all pertinent standards are effectively introduced. Although NGDRS does not have the power to impose standards, it will lobby very hard for their adoption. Incentives will be considered for the quick adoption and deployment of standards, ensuring that users will endorse the NGDRS services and that industry and government obtain an acceptable economic return for their investment in the creation of the NGDRS.

Table 6. Matrix of Possible Standards Involved in the Creation of NGDRS

System Component	System Sub-component	Hardware	Network	Software	Data
<i>Data Collection and Preparation</i>	<i>Document Scanning, Digitizing</i>	UNIX Servers Specialized Scanners	N/A	Specialized Tools (Several Vendors)	Image files, 600 DPI resolution LIS, others
	<i>Tape transcription and conversion</i>	DD-2, 4mm, 8mm tapes	N/A	Specialized Tools ((Several Vendors)	SEG-Y, RODE
	<i>Catalog File Synchronization</i>	UNIX Servers	TCP/IP	UNIX, Rdbms, C++, other	POSC Epicentre
<i>Data Modeling and Metadata Definition</i>		UNIX Servers	N/A	Express, CASE and special tools	POSC Epicentre, Digital Geospatial Metadata (DGM)
<i>Data Normalization</i>	<i>Coordinate Conversion, Lease Identification</i>	N/A	N/A	Express, CASE and special tools	POSC Epicentre
	<i>Spatial Data Representation</i>	N/A	N/A	Express, CASE and special tools	POSC Epicentre
<i>Data Loading</i>		UNIX Servers	TCP/IP	UNIX, Rdbms, C++, Java, special loaders	POSC Epicentre
<i>Database Management</i>		UNIX Servers	TCP/IP, Internet	UNIX, Rdbms	POSC Epicentre
<i>Data Storage</i>		Ampex DST, Other	N/A	N/A	SEG-Y, RODE
<i>Data Archival</i>		Ampex DST, Other	TCP/IP	UNIX, Rdbms	SEG-Y, RODE
<i>User Services</i>	<i>Data Browsing & Selection - Server</i>	UNIX and NT Servers	Internet protocols, TCP/IP	UNIX, Rdbms, C++, HTML, Java, GeoTrek	POSC Epicentre
	<i>Data Browsing & Selection - Client</i>	PC, X-Terminals	Internet protocols, TCP/IP	Windows, NT, UNIX, HTML, SQL, GeoTrek	POSC Epicentre
	<i>Data Conversion and Project Databases</i>	UNIX Servers	TCP/IP	UNIX, Rdbms C++, Specialized Converters	Provider standards: Landmark, Geoquest, others
	<i>User feedback & documentation, data promotion</i>	N/A	Internet WWW facilities	HTML, e-mail, bulletin board, conferencing	Internet WWW
<i>Data Output,</i>		DD-2,optical disk, 4mm and 8mm tapes	TCP/IP, X.25, Internet protocols	UNIX, Rdbms C++	SEG-Y, RODE, RP66, LIS
<i>Data Distribution</i>		N/A	Internet WWW facilities	HTML, e-mail, bulletin board, conferencing	Internet WWW
<i>Usage and Charges Sub-system</i>		UNIX or NT Servers	TCP/IP	UNIX, Rdbms, C++, SQL, Java	POSC Epicentre

METADATA SUPERCATALOG PHYSICAL COMPONENTS

The physical components of the NGDRS Metadata Supercatalog are diagrammed in Figure 7. Users will be able to enter the system using their existing desktop computers. In the Windows 95, Windows NT, UNIX, or Macintosh operating systems, users can use a standard World Wide Web browser (e.g. Microsoft Explorer or Netscape Navigator) to access the NGDRS system (for more on user access, see the following section). Once that browser is selected, the system router and firewall (security) will provide access to the Metadata Supercatalog and from there to the Source Node data. The connection between Internet services providers supporting the user and router systems will be accomplished using a T1 (high throughput) dedicated communications line. The metadata databases will reside on high speed disk arrays that are part of one or more NGDRS servers (either UNIX or NT). A robotic tape unit will be used for tape backup requirements. Workstation computing platforms (PCs or UNIX workstations) will be used to support the help-desk, tech support and general administrative requirements of the Metadata Supercatalog.

Physical Components of NGDRS Metadata Supercatalog

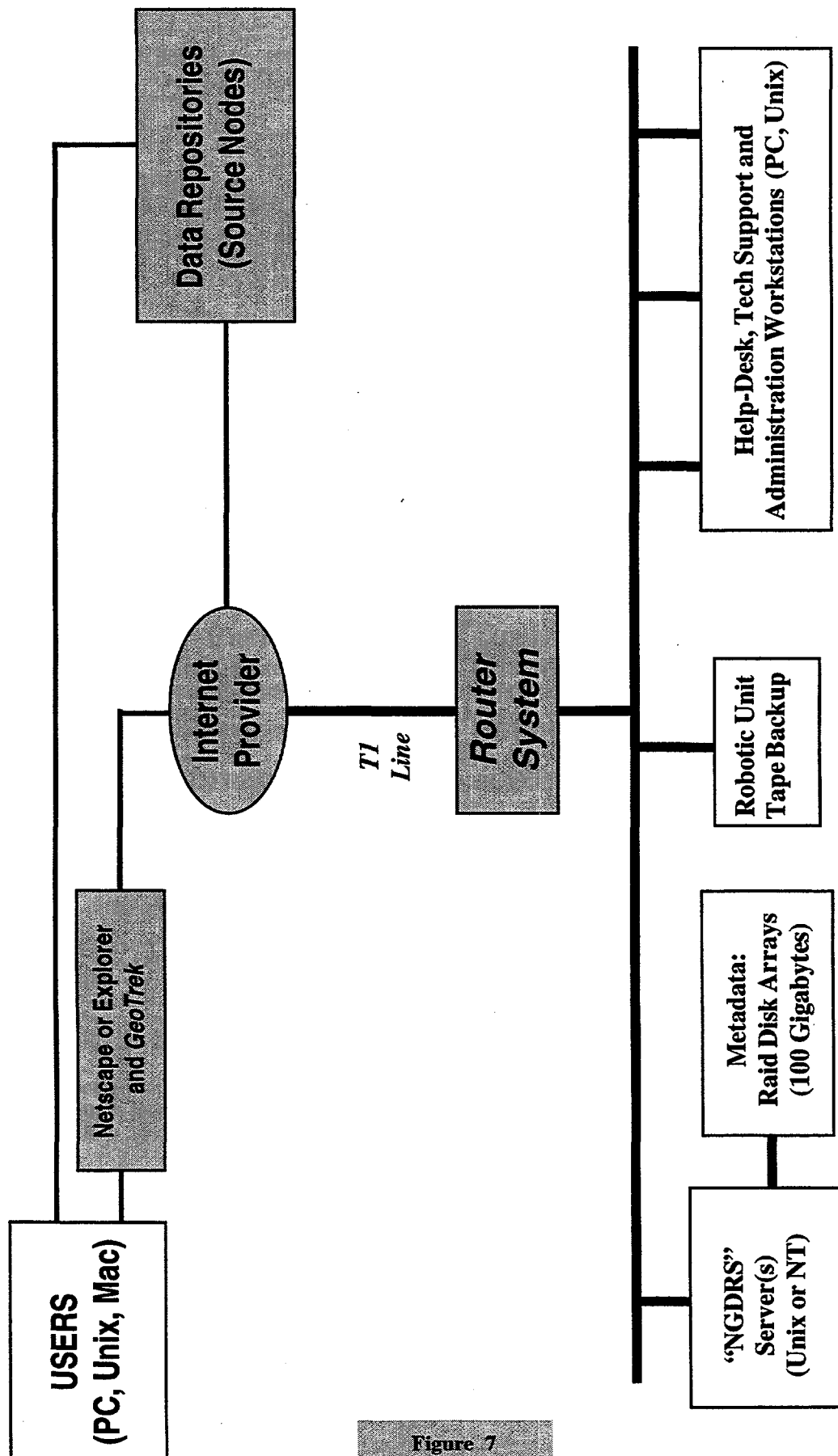


Figure 7

UTILIZING THE REPOSITORY SYSTEM

A key goal of the first two phases of the NGDRS project has been to ensure ease of use and improve access to data for users of the system. In Phase I, visits were made to petroleum companies with an interest in donating data and potential users of a national repository system. These visits helped to define the major issues and basic user requirements for the metadata catalog and other aspects of the system. Contacts with information technology suppliers and groups such as POSC led to basic utilization scenarios.

During Phase II, user requirements were tested using pilot projects and demonstrations of the prototype GeoTrek software browser developed for the NGDRS. The browser was developed from the MobilView software package, through a license for which was donated by the Mobil Technology Company. GeoTrek has the advantage of being based on a system that has many years of positive experience at Mobil handling a variety of types of data stores and data servers. The pilot implementations permitted the project team to start creating a more specific designs for the final version of a data browser and the overall architecture of the metadata catalog.

The pilot implementations supported the installation of more than twenty distributed data servers, all accessible using the GeoTrek data browser, including a major source node - the Texas Bureau of Economic Geology (BEG). The practical experience acquired during the installation and running of the pilot repositories will be used to create the final specifications for GeoTrek and to define the support services to be provided as part of the system's utilization. Documentation, some training aspects, marketing the NGDRS functionality, and receiving feedback from users were all tested in Phase II using Internet capabilities. All these experiences will be used to scale-up the utilization of NGDRS in Phase III.

USING THE METADATA SUPERCATALOG

From a user's perspective, the Metadata Supercatalog is made up of the following major components (Figure 8):

- The Metadata Supercatalog itself
- The Geotrek software browser for viewing the metadata
- The public and private data stores described in the supercatalog

NGDRS Metadata Supercatalog User Perspective

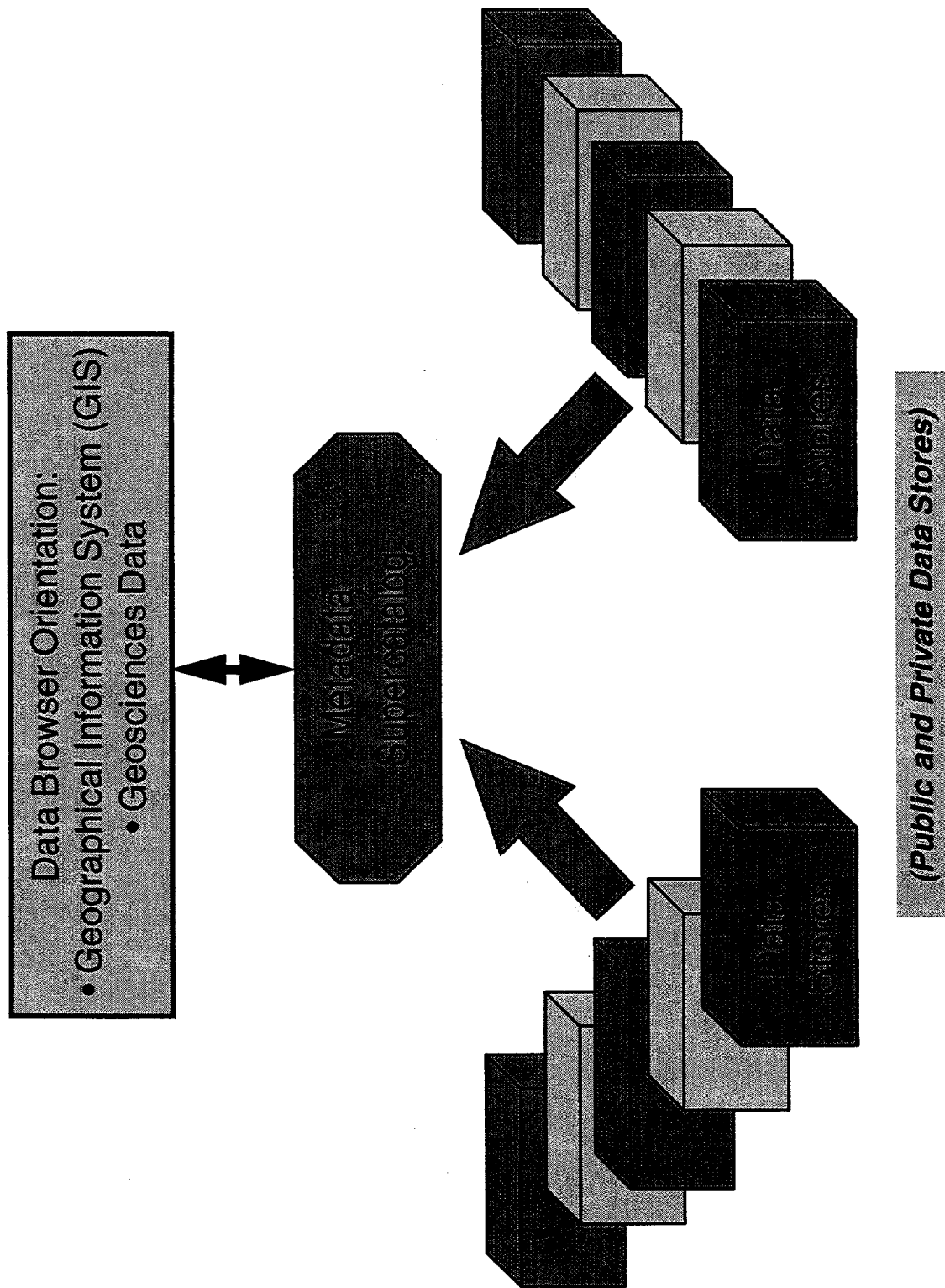


Figure 8

The first major logical component is the overall metadata catalog for NGDRS, known as the **supercatalog**. It will tell users what information is available in specific geographical areas, its ownership, location, computer systems environment, and other items of data describing the properties and characteristics of the data. The catalogs for each individual regional repository included in the NGDRS are referred to as subcatalogs. The collection of all subcatalogs will make up the supercatalog. Industry standards, such as the POSC *Epicentre* data model, will be encouraged to be used to implement all the subcatalogs. Standards will be particularly critical in the definition of a consistent metadata catalog for the whole repository since the supercatalog and subcatalogs will need to be synchronized frequently. To be effective, users need to rely on the latest possible metadata information.

The second major logical component is the **Geotrek software browser**. Users will run GeoTrek through existing World Wide Web browsers. They will not need to download any software; they will access Geotrek by pointing their own browser software to the NGDRS's Internet address.

NGDRS Metadata Supercatalog Data Browser

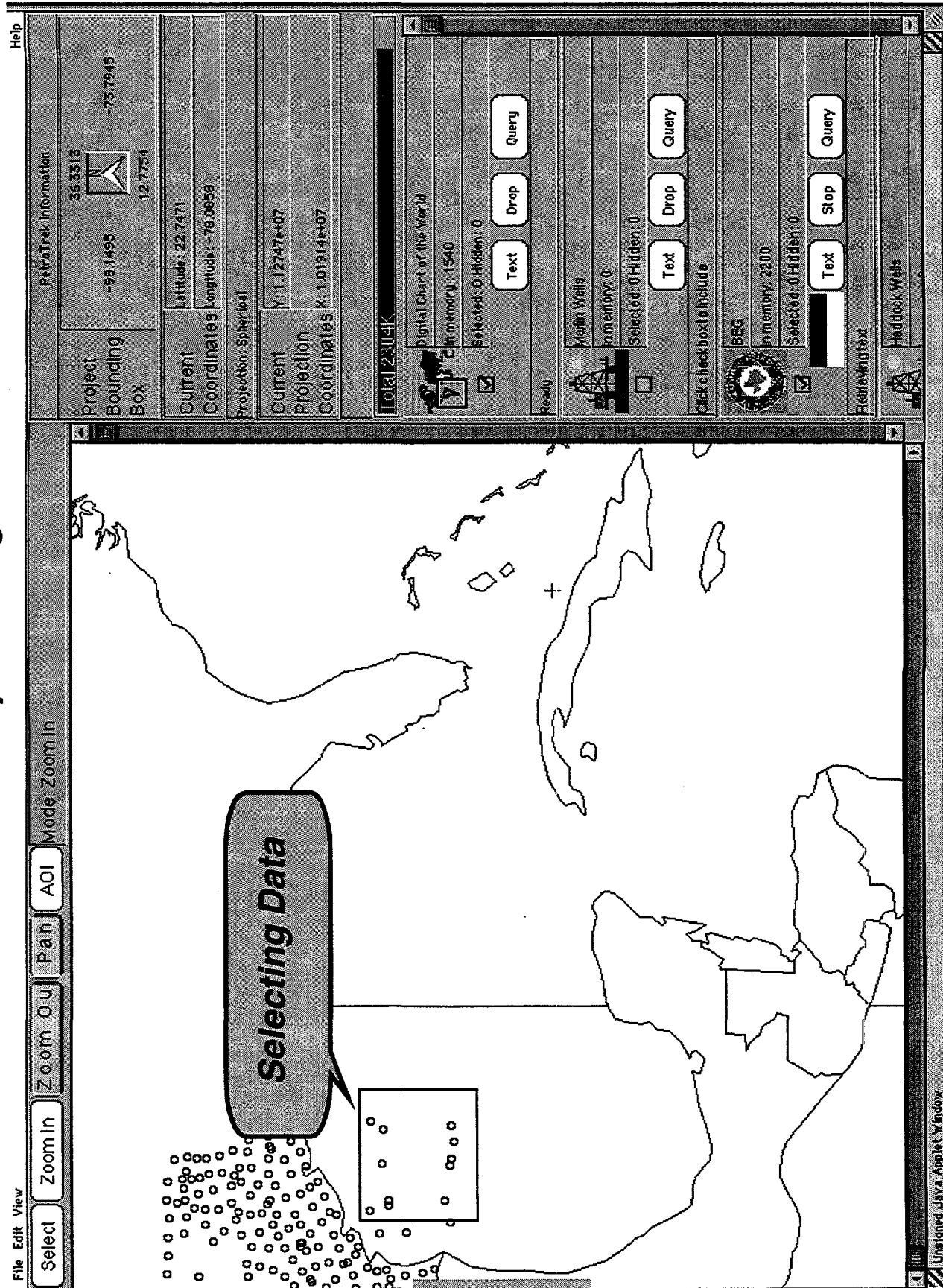


Figure 9

The system will then present the user with the GeoTrek screen, providing them with on-line capabilities to query and view metadata and data using a geographical oriented interface (Figure 9). Results of data searches and queries are displayed in a table format, allowing users to further select, filter, drill-down, and sort the contents of the data stores. More specifically, the GeoTrek browser capabilities will include:

- Finding data in a variety of database environments, including industry standard Rdbms and specific petroleum industry supported geoscience data bases (such as PetroBank, PetroVision, Finder and others)
- Selecting data by area-of-interest or via cursor; cataloging areas of interest for easier reference
- Applying to the data multiple Boolean operators, combined if necessary with sorts
- Drilling-down into several additional levels of detail
- Making data joints between distinct data stores
- Viewing data, specifically text and images
- Exporting files into standard formats, including industry standard spreadsheet formats and petroleum industry supported geosciences formats, such as POSC PEF, Geoshare, and OpenWorks.
- Saving the search and request environment created by the users as part of a working session. This will permit users to re-enter searches not completed or start new searches using the "routine" used on a previous session.

In addition this software browser environment will provide facilities for user documentation and feedback, and for data promotion and advertising. Advertising and promoting data will inform users of new available data sets in a data store or of entirely new data stores. Users will also be informed of data sets being considered for purging, consolidation or transfer to another site.

The third and final component of the NGDRS are the geographically distributed **repositories** themselves, called Source Nodes, where the actual geoscience data are housed. These repositories are either private, for example a seismic data vendor, or public, such as the many regional state government- or university-managed core repositories. A repository will be a part of the NGDRS once their metadata subcatalog is created, loaded, and synchronized with the supercatalog. This is true for existing data as well as for data that is being transferred to the recipient repositories.

The NGDRS will have no direct input into the type and qualities of services to be provided by the Source Nodes repositories. However, the NGDRS Clearinghouse will set up the criteria for the inclusion of data stores. The criteria include not only the quality of the data, but also the user services provided by the repository entity, in particular the output services or the means of distributing the data ordered by users. Output services are critical in completing the cycle started when a user searches for data availability. The types and other details associated with output services will be indexed in the Metadata Supercatalog. If users, however, are unable to get data sets recorded as available in the metadata catalog, the value of NGDRS is diminished. NGDRS support groups, the Help-Desk in particular, will gather user feedback on overall service performance of the Source Node repositories and relay this feedback to NGDRS Management.

REGISTERING USERS AND THEIR ORGANIZATIONS

The first time a user enters GeoTrek, a registration process will be required. The process will be simple, requiring only general information about the users and their organization affiliation. The need for registration is prompted by the following reasons:

- To open a vehicle for information exchanges with users. These exchanges are expected to include systems difficulties, new or special requirements, discussions of the availability of new data sets and other general topics.
- To provide input to the NGDRS data transfer plan and priorities. These priorities are expected to be affected by data usage, and the types of users (geoscientists, engineers, and others) and their organizations (petroleum companies, academia, government agencies, and others).
- To provide input to the NGDRS operational plan. The profile of user types and their data requirements will be used by NGDRS management to plan for the expansion of the computing infrastructure and support organization.

It should be remarked that the NGDRS vision is to be an environment open to all. This does not mean security precautions will not be installed and applied. However, the requirement for registration comes less from the need to secure the system against any potential misuses than from the reasons stated above.

NGDRS USER SUPPORT: HELP-DESK AND TECHNICAL SUPPORT GROUPS

Utilization of the NGDRS system is expected to increase significantly in the first full year of utilization and require the installation of two user support groups: (1) a Help-Desk to serve as a first point of contact with users for all issues, and (2) a Technical Support Staff to provide users with more advanced support. These groups will be installed during the creation of the Metadata Supercatalog and are expected to start actively supporting user requirements before the end of the first year of operations.

The primary responsibilities of the Help-Desk group will be:

- Provide general awareness of system features, options, and general utilization procedures.
- Provide NGDRS management with user feedback on problems encountered while accessing data at Source Nodes.
- Provide feedback to the NGDRS technical group on user requests for new data stores or system features.
- Define the severity of the problems encountered by users and decide the next steps.
- Resolve the first-level problems - the simplest ones - encountered by the users.
- Contact technical groups for the resolution of all other problems.
- Contact users to determine if problems have been resolved satisfactorily.

The primary responsibilities of the Technical Support group will be:

- Contact users for problem clarification.
- Define, in technical terms, the problem encountered by the user.
- Define, together with other technical team members, the solution to the problem and its implementation.
- Implement the appropriate systems changes needed to correct the problems.
- Define system improvements, hardware and software capacity requirements, and general systems procedural modifications.

OPERATING THE REPOSITORY SYSTEM

In order to implement the NGDRS, the Project Management Team developed a set of administrative and operational procedure guidelines. The pilot projects provided a better understanding of the complexities associated with managing a large repository and dealing with data issues concerning privacy, security, backup, disaster contingency, etc. It permitted the design of user services to avoid or mitigate operational difficulties. The pilot implementations included: (1) the installation of more than twenty distributed data servers, including a major source node - the Bureau of Economic Geology (BEG), (2) the installation of a major prototype software component (Geotrek data browser), and (3) the understanding of metadata catalog synchronization and related issues.

The practical experience acquired by installing and operating the pilot Source Nodes, combined with the detailed documentation provided by other organizations, was used for defining and detailing NGDRS's future operational environment. In parallel with the technology-oriented activities, the Project Management Team continued their efforts to find the most desirable business model for creating, utilizing and operating the repository.

ORGANIZATION, RESOURCES AND COMPETENCIES NEEDS

The operations and user support activities of the NGDRS metadata system require a strong technical and user support organization. All jobs viewed as non-core to the support and running of the Metadata Supercatalog operations will be sub-contracted. Figure 11 is the proposed organization chart showing some of the primary relationships of the organizational groups with the NGDRS users, the management of other repositories and service suppliers. The following are the main organizational components:

NGDRS Metadata Supercatalog Organization

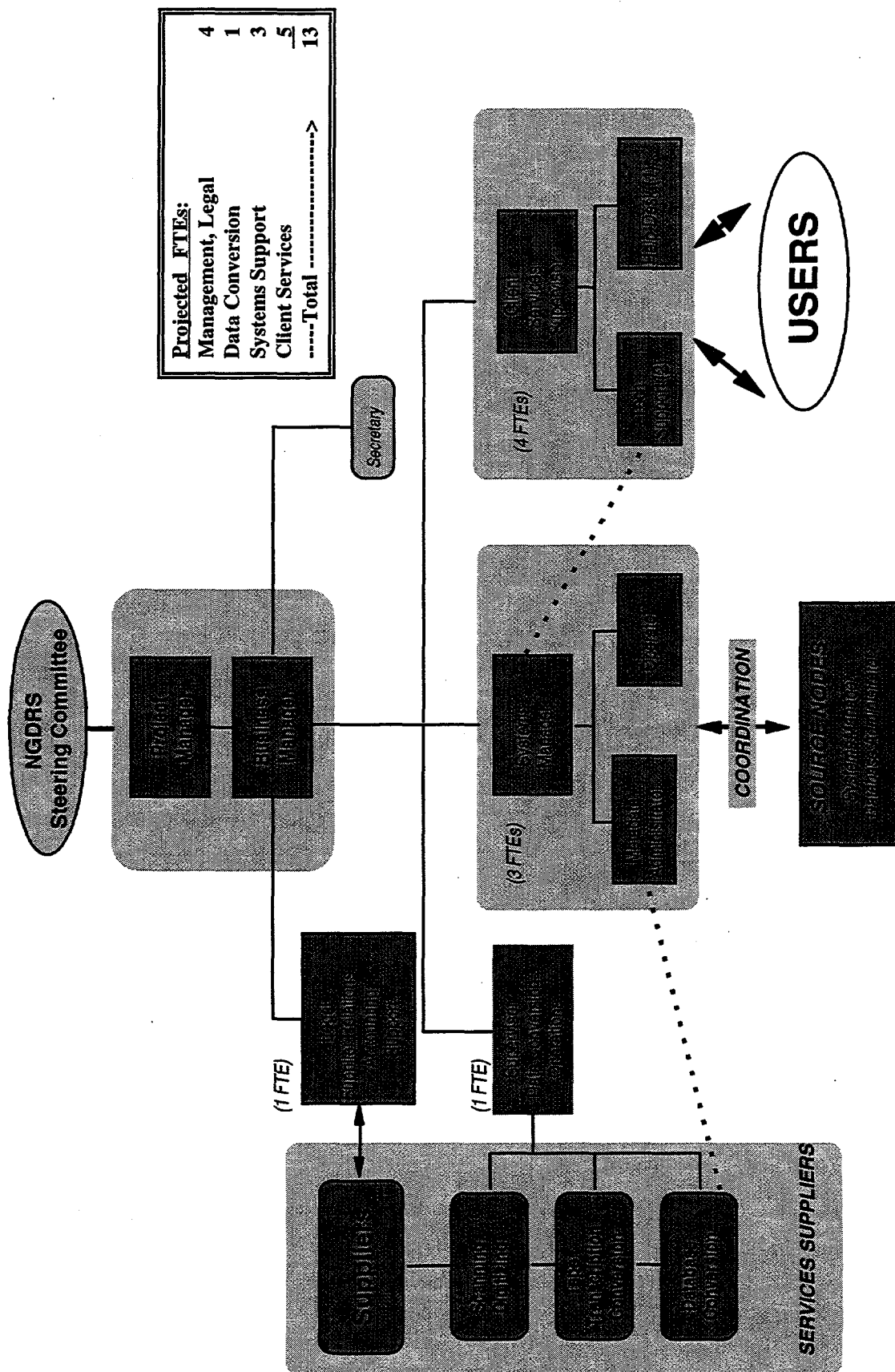


Figure 10

Project Manager and Business Management: Key to the success of the project is the overall project leadership. The organization contemplates the existence of two main administrative roles: the Project Manager and the Business Manager. The Project Manager will have overall project responsibilities with direct and permanent access to NGDRS Steering Committee and the responsibility of resolving all business issues. The Business Manager will have administrative responsibilities for the all day-to-day activities including utilization, client support services, and system operations.

Supplier Relations, Legal Support and Accounting: The relationships with service suppliers are key to the success of the data transfer activities. Both during the contract negotiations and management of the physical movement of data from data donor to recipients, this individual will be the key to assure that economics and logistics are kept in check. It requires a person with strong project management experience and negotiation skills. Accounting responsibilities for this individual relate mostly to project accounting. The person will interface routinely with Supervisor of Data Conversions Operations to assure suppliers services are being delivered as promised.

Data Conversion Operations: This activity includes a wide variety of tasks and competencies not possible to be covered with full time staff. Rather the approach here will be to have a competent senior geoscientist, familiar with the diverse data types being transferred - seismic, well logs, cores, electronic databases and hardcopy. This individual will work with Supplier Relations and jointly will procure and engage the necessary service providers. The person will also follow-up and monitor the data conversion or transfer process and assure that service vendors are in compliance with NGDRS quality standards for the data conversion.

Client Services Support: This group will be responsible for catering users issues, needs and feedback. Jointly their responsibilities include:

- Educate and coach users on the Metadata Supercatalog features, options, and search capabilities.
- Define the severity of problems encountered by users.
- Resolve user problems, engaging other parts of the organization as needed
- Implement changes required to solve the problems raised by users
- Contact users to determine if problems have been satisfactorily resolved.
- Participate with the rest of the organization on the formulation of system improvement, hardware and software capacity requirements, and general systems procedural modifications.

The jobs require skills that combine extensive interactions with many types of users with diverse levels of technical sophistication and a good understanding of technologies deployed by the NGDRS and in particular the Metadata Supercatalog.

Systems Management, Database Management and Data Administration: This group will be responsible for installation of products or procedures related to:

- Routine data center operations, including work shifts, service levels, and data backup.
- Systems software and network management; installation of software upgrades and overall systems performance.
- Database management systems (Rdmbms); installation of software upgrades and performance.

- Staff systems management, including establishing directories, accounts, and access privileges.
- Metadata administration, including data modeling and mapping, data access and privacy, and data security.
- Support of the security for the physical facilities, including protection against vandalism and accidents.

This group will also be responsible for coordinating activities with the Technical Support group, to include informing them of systems changes or upgrades, general coaching in new systems or software, and helping to diagnose and correct problems raised by users.

DATA CENTER OPERATIONS GUIDELINES

System operations will not be limited to simply supporting the operation and maintenance of the systems and communications. The goal is to support the NGDRS mission, and to do so in the most innovative, client-oriented, and cost-effective way. In this context, the following critical services were identified:

- **Access to information resources:** For the NGDRS to succeed, it must have access to technical data; new software applications and packages; and input and output hardware devices. Doing so requires providing reliable, available, maintainable, stable platforms for general computing, filing, printing, and plotting services; providing business communications services for e-mail, bulletin boards, computer conferencing, and feedback procedures; proactively monitoring and repairing failed components; maintaining solid backup and recovery processes; and defining procedures for constant testing of disaster recovery facilities.
- **Competent, responsive technical support services to all clients;** the success of the repository system is largely based on the responsiveness to users accessing reliable information systems and network resources. Technical support and Help-Desk groups will be key personnel providing the required coaching, training and assistance. Systems groups will assure operations are running smoothly.

QUALITY MONITORING PROCESS

The NGDRS Metadata Supercatalog and Source Node Repositories will convene semi-annually to review operational and service issues to:

- Consider any operational matters referred to either organization by users of the Repository or members of either organization.
- Review the NGDRS's performance against the proposed service levels.
- Review and recommend for adoption standards and guidelines for performance measurement against industry benchmarks.
- Endeavor to resolve any matters of interpretation, disputes, or disagreements arising from the services delivery. In particular ensure that there is an effective coordination and dialogue among management of both entities.

CONFIGURATION MANAGEMENT

Configuration management will provide a method for control and tracking of components being added to the existing system, or to components being removed or changed. Systems Management will implement techniques utilizing the WWW as a vehicle to communicate updates. The methodology adopted is applicable to all major components; hardware, systems software, service support database management system, and applications. The methodology will ensure:

- All repository activities will adhere to configuration management practices
- User documentation is continuously developed or updated concurrent with the installation of the new hardware, software, applications and data components.
- Configuration baselines are consistent with the respective project phases.
- Uniform policy, procedures, data, forms, and reports are utilized.
- System changes are processed in a timely and cost-effective manner and evaluated in terms of schedule impact, costs, and total performance.
- Facilities, metadata catalogs, and other data are protected, secured and kept private at all times.

SYSTEMS MAINTENANCE

The NGDRS Systems Staff will provide maintenance support for systems software, packages and custom applications at the three following levels:

- Maintenance directed at *perfecting* systems: this maintenance will be used to manage changes demanded by the systems organization and user community.
- Maintenance directed at *adapting* systems; when changes to the base software environment occur adaptive maintenance procedures will be implemented.

Systems Staff will consider such factors as the user demands, software application complexity, database management system, development environment, and hardware stability when performing maintenance activities. In coordinating system software upgrades, the Systems Staff will ensure that they do not create adverse affects to applications or data. The Systems Staff will also interact with vendors and third-party custom software providers to incorporate product upgrades in behalf of the NGDRS. In all upgrades, the Systems Staff will fully test software compatibility as a result of the integration prior to moving to a production environment.

CAPACITY PLANNING

One of the key on-going responsibilities of the NGDRS will be capacity planning for the Metadata Repository. Capacity planning is the process of determining the likely future of computing and network resource requirements. Since the growth in the number of NGDRS users and the increasing services demands are anticipated to progress steadily, it is necessary to monitor the depletion of physical computing resources and be prepared, at the correct moment, to expand the resources to fulfill the demands. Besides understanding the demand side, capacity planners have to pay attention to the supply aspects; how to take advantage of new technologies and strategies, such as client/server, to maximize the cost/benefits of introducing new infrastructure including systems platforms.

Because of the sheer amount of information required to determine capacity, both statistical and logical methods are used to simplify the process. Where a lack of methods, tools, and data to accurately model future requirements exists, human intuition is used. Often, the problem isn't a lack of information but too much low-level information. System management has to review information coming in from hubs, routers, and other devices throughout the network. This information must be analyzed, interpreted, organized, correlated, and synthesized before it can be used effectively for capacity planning. NGDRS Management will introduce appropriate automated tools to effectively measure, analyze, balance, and plan the use of the capacity they have.

Generally, the capacity planning process is made up of the following steps:

Establish a Performance Baseline. Requires assessing the current performance and service levels. The outcome of this step is typically the creation of an organized database of measurement to be used for further planning. Included in this step is the characterization of current workload, i.e. identifying the type of work people do on the network, when that work is performed, and the computing resources required to do those jobs.

Establish User Service Level. Requires definition of user services levels in terms of timeliness and availability. These terms are normally tied to the business objectives prescribed for the system and they may be stated as, for example, the NGDRS system response time cannot exceed 5 seconds and that the network availability must exceed 95% between 8 AM and 6 PM EST. Since the NGDRS users will make use of the WWW as the WAN to the Metadata Repository it will be necessary to have an understanding of the implications of the WWW architecture in the expected service levels. In particular, it will be necessary to understand the impact on performance of several software components, including:

- Specialized HTTP daemons (tuned for performance)
- TCP/IP implementation on a particular server platform
- File system or database management used to serve the data sets (in the NGDRS Metadata Repository)
- Data browser software and other CGI programs.

DISASTER RECOVERY

The steady and consistent use of the data resources provided by NGDRS is expected to grow rapidly during the first year of activity. As usage becomes part of the routine activities by possibly thousands of users, their dependence on NGDRS will increase. Although the NGDRS is open and 'free' to all, the responsibility and accountability of NGDRS towards users will grow. As part of using the system, users will learn to expect a minimum level of service, even when a disaster occurs. It is necessary, therefore, to formulate, during the initial phase of implementation on Phase III, a reasonably balanced disaster recovery plan.

A disaster recovery plan is a comprehensive set of action steps to be taken before, during and after a disaster. The plan is documented and tested to ensure the continuity of operations and availability of critical resources after a disaster. The plan must reflect NGDRS business requirements including core activities, recovery windows, and customer service levels and metrics. The plan must provide

guidance during the crisis and insure that vital issues are not overlooked. Business recovery plans are required by law in certain industries but, more commonly, when regularly tested, they are used to help organizations insulate themselves from negligence litigation.

The NGDRS is ultimately responsible for the plan since it is responsible for managing NGDRS assets, including hardware, software and information systems. However, business recovery will be a concern of all those involved in the NGDRS project. System and networking management will have the overall responsibility for the formulation and development of the emergency procedures that includes not just systems but all other aspects such as logistics and personnel movements.

A project workplan will be established to manage the tasks, deadlines and deliverables for the disaster recovery plan, which typically includes the following steps:

- **Project organization;** includes project administration, the definition of basic assumptions, and policy issuance.
- **Facilities inventory;** key to determining what needs to be done is an updated inventory of the facilities provided by the NGDRS. The inventory must include all hardware, systems software, database management software, networking (LAN and WAN), applications and systems detailed documentation and personnel critical competencies.
- **Risk assessment;** identifies the type and severity of disaster a specific location is likely to endure.
- **Business impact analysis (exposures);** involves determining the exposures, such as revenue losses, associated with not being able to transact business. It will help determine the breath of the recovery plan.
- **Recovery strategy development;** it requires determining and deciding how best to provide support following a disruption. Typical alternatives include:
 - Hot Sites - vendors in the business of providing a prepared site with hardware, some software, telecommunications, and technical support personnel.
 - Cold Sites - empty facility or leased space that is ready for occupancy. Equipment and support staff is shipped by vendors to the location.
 - Internal backup - facility owned by company in another region to provide emergency on an emergency basis.
 - Mutual support agreement - an agreement with another entity to share resources after a disaster.
- **Documentation;** plan can use traditional word processing software or specialized software applications. Critical to the documentation is change control procedures.
- **Training;** it is essential to educate and train all members of the organization and, in particular, the emergency operating team to insure each person knows their role and responsibility.
- **Testing and simulation;** the plan must be tested on a regular basis. Most companies do so at least semi-annually. The simulated disaster will exercise the plan, identify weaknesses and demonstrate interaction between the participants. The exercise typically generates improvements on the plan.
- **Periodic updates;** updates must be considered on a periodic basis. Changes on the business requirements (user needs and levels of service) and on the technology deployed must be quickly reflected on the documentation to avoid mistakes during the period right after a disaster.

NATIONAL DIRECTORY OF GEOSCIENCE DATA REPOSITORIES

One of the three principal components of the Phase II project was development of the *National Directory of Geoscience Data Repositories* to document the current holdings of data types in the public and private sector. The directory was published in April, 1997. It is available directly from AGI and will be available on the World Wide Web at <<http://www.agiweb.org>> in Fall, 1997.

The following four tasks were required for the development of the directory:

Selective Site Visits: As identified by the NGDRS Steering Committee, site visits were made to a number of public and private data centers across the U.S. These site visits provided a cross-section of critical data-center operations and services that were used to help develop the scope and content of the survey questionnaire.

Survey Questionnaire: In addition to the site visits, letters were sent to all 50 state geological surveys, to approximately 60 regional geological associations, and to members of the Preservation of Samples and Core Committee of the American Association of Petroleum Geologists. These letters requested names and addresses for all data centers known to these groups and individuals. A pilot questionnaire was then drafted and sent to approximately 25 of the identified data centers to gain insight into the questionnaire content. Feedback from that pilot test was incorporated into the final questionnaire (copy attached as Appendix C).

The questionnaire was then sent to all identified data centers as well as to private-sector data centers selected from several oil and gas directories. There were approximately 300 questionnaires mailed. The respondents were asked to return the questionnaires either via mail or e-mail via the Internet.

Follow-up Contacts: Approximately four weeks after the initial mailing, follow-up reminder postcards were sent to non-respondents. Two weeks following that mailing, phone calls were made to approximately 50 non-responding data centers. The final number of data centers included in the Directory is 124, although several additional respondents disqualified themselves as inappropriate for the Directory. Several additional questionnaire responses have been received since publication of the directory and will be included in the web version and future revisions.

Directory Development: The Directory is organized alphabetically by and within each state. The questionnaire was the model for the basic presentation of the data, which includes the following: name of center, address, contact person, with telephone and fax numbers and e-mail address, size of data center staff, type of repository, data access fee, catalog and index type, data delivery media, quantity and type of holdings, data center services provided, geographic areas of holdings, by type and media, and comments. The directory also contains a cross-referenced matrix of broad data types (geophysical data, drill-hole data, field/pool information, maps and location information, and other) by geographic areas of holdings and media. This matrix provides a fast look-up table that allows users to determine which data centers have the specific data types to meet their needs. It is included in this report as Appendix D.

STEERING COMMITTEE OPERATIONS

The NGDRS Steering Committee assists AGI in assessing the needs and opportunities for establishing the NGDRS. The committee was appointed in consultation with AGI's Member Societies and consists of leaders in the public and private sectors familiar with the issues and concerns related to long-term preservation of large volumes of geoscience data. Liaisons to the Steering Committee were appointed by a number of federal agencies that deal with geoscience data, including DOE, the Bureau of Land Management, the U.S. Geological Survey, NOAA, and the Minerals Management Service. The committee was chaired by Roger Brown of Conoco. A list of the Phase II committee members is included as Appendix E.

The responsibilities of the Steering Committee are:

- Assist in development of guidelines and scope of the repository system study.
- Establish selection criteria and requirements for organizations to participate in the repository system.
- Define organizational and coordination roles of repository network data centers.
- Establish workgroups to develop and scope requirements for data transfer, prioritization, management, and access.
- Develop a business model clearly defining the organization and structure plan for the repository system.

The Steering Committee met eight times during the course of the Phase II project. The times, locations, and topics of discussion are listed in Table 7.

Table 7. NGDRS Steering Committee Meetings

Date	Location	Topic
March 14, 1995	Houston, TX	Phase II Organizational Meeting
August 28, 1995	Houston	Review of requirements for establishment of repository system
September 29, 1995	Houston	Repository business plan and review of metadata catalog and data access
January 20, 1996	Houston	Business plan and data access update
May 20, 1996	San Diego, CA	Overall status and activities review
January 13, 1997	Houston	Requirements for establishing core repository
February 4, 1997	Houston	Repository update
March 14, 1997	Houston	Repository update

Conclusions and Recommendations

AGI's goal in establishing the National Geoscience Data Repository System (NGDRS) is to provide leadership for development of an efficient and cost-effective process to preserve geoscience data and improve its accessibility to individuals and companies with an interest in such data. By working collectively with the major private-sector data contributors and vendors and public-sector data repositories, AGI can provide a data standards framework for indexing and cataloging and achieve the required economies of scale to reduce the long-term cost of data management. AGI's vision is to establish a common network of geoscience data repositories across the United States to promote improved data management and access.

The NGDRS network will serve as an important and valuable resource of geoscience information for a variety of applications, such as environmental protection assessments, water resource management, global climate change studies, earthquake risk assessment and other geologic hazards, as well as for basic and applied research. Critical data contained in the repository will enhance the ability of energy and minerals companies to discover new petroleum and mineral reserves and increase recovery of proven reserves. Geoscientists across the country will be able to access a variety of information sources via a Web-based metadata catalog that incorporates industry-endorsed standards to facilitate user interaction with data and ensure consistency across the network of participating repositories.

Based on results of the 1994 Phase I *Feasibility and Assessment Study*, petroleum and mining companies surveyed indicated that they are willing to donate billions of dollars worth of geoscience data to the NGDRS. The Phase I study documented the data types and volume of data that companies would contribute. The companies incentive for making such contributions is to lower their data management costs for data holdings from areas in which they no longer have a proprietary interest. To date the single largest data transfer is Shell Oil Company's contribution of their Midland (Texas) Core Repository to the Bureau of Economic Geology at the University of Texas at Austin. The Shell contribution amounted to 2.2 million linear feet of cores and cuttings. Additionally, Shell provided a \$1.3 million endowment to support future operations of the repository. These data entered the public domain for the first time with Shell's contribution.

The NGDRS project has been endorsed by a growing number of professional, scientific, industrial, and state organizations that have gone on record in support of the national repository system concept. In recognition of its potential value, the NGDRS has received key endorsements from the American Association of Petroleum Geologists, Society of Exploration Geophysicists, Association of American State Geologists, the Department of Interior's Outer Continental Shelf Policy Committee, the American Petroleum Institute's Exploration Affairs Committee, and the National Research Council's Panel on the Geodynamics of Sedimentary Basins.

The Phase II Study initiated in 1995 is a joint industry, academic, and government project supported by the U.S. Department of Energy and corporate sponsors. The study provides a plan for development of the business and organizational framework, data standards, and technologies required to create the NGDRS. Long-term operations of the repository will be overseen by an advisory steering committee and are to become self-sustaining based on revenues received from vendor subscription fees and commercial user software licenses. The proposed NGDRS consists of

three strategic components — the Clearinghouse, the Web-based GeoTrek Metadata Supercatalog, and the network of regional data repositories.

Clearinghouse: The Clearinghouse will be charged with the responsibility for identification of data contributors, prioritization of data to be accepted, and identification of repositories to receive data. Comprised of representatives from industry, academy and government organizations, the Clearinghouse will be managed by AGI. Protocols will be established to ensure that data contributors, network repositories, and data users will have input into Clearinghouse policy decisions.

It is proposed that transfer of data from corporate contributors to regional repositories follow the Shell Oil model and include endowment funds to sustain long-term data management and access. DOE will be requested to continue support for data transfer, indexing, and cataloging of data accepted from private-sector contributors.

GeoTrek Metadata Supercatalog: The GeoTrek Metadata Supercatalog is a map-based browser system that allows a user to conduct data searches. With GeoTrek, users can determine the availability of various data types for specific areas being held in participating regional repositories. Because GeoTrek will be Web-based, users will not need to download any software — they can access the NGDRS system using their standard Internet browser software. The system will allow users to query and view metadata using a geographic (GIS) oriented interface. Results of data searches and queries will be displayed in a table format allowing the user to select and sort data of interest.

Regional Data Repository Network: The NGDRS will be comprised of public and private sector repositories linked via the Internet to the GeoTrek Metadata Supercatalog. The network system will be built upon existing state and regional data repositories. Individual repositories or source nodes will be responsible for indexing and cataloging their data inventories. The NGDRS' GeoTrek master on-line Metadata Supercatalog will provide users access to metadata available across the repository network. Individual source node repositories will update and synchronize their data on the NGDRS on a predetermined frequency.

The most important guiding principles to ensure successful launching of the system are:

- **National Network:** The NGDRS will be comprised of a national network of geoscience data repositories built on existing state and regional organizations and commercial data vendors.
- **Publicly Availability:** All data in the NGDRS will be available for public access at predetermined costs as designated by participating repositories and data vendors.
- **Joint Operations:** The NGDRS will be managed by AGI with oversight provided by a steering committee comprised of representatives from industry, academia, and government.
- **Data Prioritization:** The AGI Clearinghouse will identify data contributors and receiving repositories and prioritize data for acceptance by the NGDRS. The initial focus will be on cores and cuttings as well as paleontological collections. Subsequent efforts will include seismic data, paper records, and other types.

- **Data Contributors:** Private-sector contributions of cores and cuttings should be coupled with endowment funds to individual repositories to support long-term data management and access costs.
- **Data Standards:** The NGDRS will maintain a strong POSC connection to encourage individual data repositories to adopt the *Epicentre* data model.
- **Improved Access:** Through the Geotrek Metadata Supercatalog, users will have improved access to geoscience data with the ability to conduct Web-based searches on data stored at repositories across the country.
- **Repository Center:** Because many regional repositories are essentially filled and can accept data only on a very selective basis, the NGDRS will attempt to respond to additional data storage requirements such as designated regional repositories to house "spill-over" data.
- **Economy of Scale:** The NGDRS will reduce the cost of data transfer, cataloging, and indexing by operating on a national system basis.
- **Federal Support:** Continued support from federal agencies will be required to support transfer, indexing, and cataloging of data moving into the public sector.
- **Self-Sustaining:** The goal of the Geotrek Metadata Supercatalog is to become self-sustaining through commercial data vendor metadata subscriptions and private-sector license fees.

Phase III: Implementation and Operation of the Repository System

Phase III of the NGDRS project is starting up in the first quarter of 1997. Phase III will provide for a coordinated and orderly transfer of significant volumes of geoscience data from the private sector to the NGDRS and include installation of the GeoTrek Metadata Supercatalog for improved data access. The electronic metadata catalog will be Web-based and allow efficient access to the existence and characteristics of the data. The business model and required network support base defined in the Phase II study will ensure sufficient support to allow appropriate operations of the NGDRS. The data contributors and regional repositories identified in Phases I and II will be networked in Phase III. It is expected Phase III will span several years. During the first year the GeoTrek Metadata Supercatalog will be installed and data acquisition and transfer activities will be initiated. Follow-on years will be primarily focused on transfer of prioritized data from the private sector to public sector regional repositories.

The NGDRS GeoTrek Metadata Supercatalog operations are expected to be self-sustaining after year one of Phase III. Private sector corporate and federal support will be requested to cover the cost of data transfer, indexing, and cataloging in out years.

References

- American Geological Institute (1994a). *National Geoscience Data Repository System, Feasibility and Assessment Study*, Department of Energy Report DOE/BC/14968-1, 68 pp.
- American Geological Institute (1994b). *National Geoscience Data Repository System, Phase II: Planning and Pilot Study Proposal*, Alexandria VA, 48 pp.
- American Geological Institute (1996). *AGI Strategic Plan*, Alexandria VA, 16 pp.
- Department of Energy (1997). Oil and Gas R&D Programs
- Federal Geographic Data Committee (1997). *Content Standards for Digital Geospatial Metadata*, [on-line] URL: <http://www.fgdc.gov/Metadata/metahome.html>.
- National Research Council (1995). *Preserving Scientific Data on Our Physical Universe*, National Academy Press, Washington DC, 67 pp.
- National Research Council (1997). *The Dynamics of Sedimentary Basins*, National Academy Press, Washington DC, 43 pp.
- Texas Independent Producers and Royalty Owners Association (TIPRO), and the Texas Bureau of Economic Geology (1992). *Technology Transfer Needs and Requirements for Texas Independent Oil and Gas Producers*, University of Texas at Austin, 33 pp.

Glossary

AGI	The American Geological Institute is a federation of 31 geoscience societies. The AGI provides information and education services to its members, promoting a united voice for the geoscience community. AGI was the initiator of the NGDRS project.
Ampex	A product and services provider of mass storage equipment.
Architecture	(a) The set of principles or beliefs, models and standards that describe a particular domain of information systems (such as applications, data, infrastructure or organization domains). (b) Design or the way components fit together.
Boolean Operators	The set of operations usually performed on the data once selected. Typical operators are: equal, not-equal, join, intersection, greater-than, etc.
Broker	A software component that registers the data sources or data stores available for display by the data browser.
Browser	Same as Data Browser
Bulletin Board	Software (proprietary or public domain) that allows a particular entity or individual to broadcast information to others. A one-to-many type of communications software.
C++	Software programming language. One of the most used object-oriented languages, a super-set of C language developed primarily at AT&T Bell Laboratories in 1986.
CASE tools	Computer Aided Software Engineering, techniques and tools using computers to help with one or more phases of the software life-cycle, including the systematic analysis, design, implementation and maintenance of software.
Clearinghouse	A systematic approach to transferring data from the private sector to the public domain and to be adopted as part of implementation phase of the NGDRS.
Client	A software program that is used to contact and obtain data from a Server software program on another computer, often across a great distance. Each Client program is designed to work with one or more specific kinds of Server programs, and each Server requires a specific kind of Client. Web browsers are specific kind of client software.
Conferencing	Software (proprietary or public domain) that allows dispersed groups to participate in joint discussions. A many-to-many type of communications software.
CPU	Central Processing Unit, the part of a computer which controls all the other parts.
Data Browser	The means to finding and displaying data on a computer system.

Data Consumers	(a) The systems or applications that predominantly make use of data or information created by the data provider systems. (b) Refers to all the private and public entities that have committed to or are expected to make use of the NGDRS facilities.
Data Contributors	Refers to all the private and public entities that have committed to donate data to NGDRS.
Data Donors	Same as Data Contributors
Data Model	(a) A formal method of describing the behavior of the real-world entities. A fully developed data model specifies entity classes, relationships between entities, integrity rules and operations on the entities. A data model is similar in many ways to a conventional dictionary. In Epicentre, there is an alphabetical listing of names and definitions for more than 750 real-world technical and business objects pertaining to petroleum exploration and production. In POSC's data modeling terminology, these objects are called entities. A data model, however, goes beyond names and definitions. It details the characteristics that each entity may have. Some characteristics, such as the identifier or description of a Well, provide simple identifying and explanatory information about an occurrence of the entity. Other characteristics specify interrelationships between entities. For example, a Well has the characteristic of containing one or more related Well bores, which are specified as another entity in the model. (b) The result of the conceptual design process. A generalized, user-defined view of the data related to applications.
Data Node	Same as Source Node.
Data Providers	(a) The systems or applications that generate most of the data used by a computer systems. (b) Same as data contributors.
Data Server	A particular type of software server responsible for services related to data, information or knowledge management.
Data set or dataset	A named collection of logically related data items arranged in a prescribed manner.
Data Warehouse	A database, often remote, containing recent snapshots of corporate data. Planners and researchers can use this database freely without worrying about slowing down day-to-day operations of the production database.
Databank	A logical collection of interrelated information managed or not using a computer system.
Database	A logical collection of interrelated information, managed and stored as a unit, usually on some form of mass-storage system such as magnetic tape or disk.
Database Management System (DBMS)	A set of computer programs for organizing the information in a database. A DBMS supports the structuring of the database in a standard format and provides tools for data input, verification, storage, retrieval, query, and

manipulation.

Datastore or Data Store	A database typical dealing with one single type of data for a single region or area, such as well data store for the oil field X. The smallest unit of logically organized data that can be meaningfully by the business organization, such as a petroleum asset team organization.
DD-2 or DD2	A high-density tape media. It is a 19mm helical scan tape. Capacities cartridges are 25, 75, or 165 gigabytes.
DGM	Digital Geospatial Metadata (DGM) was approved in June 1994 by the Federal Geographic Data Committee (FGDC) as a standard for metadata. DGM describes the specifications for the content, quality, condition, and other characteristics of metadata. The standard provides a common set of terminology and definitions for the documentation of geospatial data. DGM establishes the names of data elements and groups of data elements to be used for these purposes, definitions of these data elements and groups, and information about the values that are to be provided for the data elements.
Drilling-down	Drilling-down is a common way of describing the fact that an inquiry into a database needs to be taken into a deeper level of detail.
E&P	Acronym for Exploration and Production. The sector of the Petroleum industry that deals with finding reserves and bringing to market basic hydrocarbons (petroleum and gas).
Entity or Source Node Entity	Every Source Node is managed and is the responsibility of Source Node Entity (or just Entity) - such as, for example, the Bureau of Economic Geology - BEG, or The University of Texas, or the USGS
Epicentre	Epicentre is a logical data model designed to support a diverse set of technical and business functions and applications associated with the operations, administration and regulation of petroleum resources, commonly known as exploration and production (E&P) activities. Epicentre is intended to facilitate the sharing of data across business functions by providing an integrated model of data. Epicentre is the data model component of the POSC Software Integration Platform.
Ethernet	A very common method of networking computers in a local area network (LAN). Ethernet will handle about 10 million bits-per-second and can be used with almost any kind of computer.
Express	A data modeling language adopted by the International Organization for Standardization (ISO) working group on STEP - Standard for the exchange of product model data. The Epicentre data model is formally defined using Express.
FDDI	(Fiber Distributed Data Interface) -- A standard for transmitting data on optical fiber cables at a rate of around 100,000,000 bits-per-second (10 times as fast as Ethernet, about twice as fast as T-3). FDDI is an emerging high-speed networking standard. The underlying medium is fiber optics, and the topology is a dual-attached, counter-rotating Token Ring. FDDI

	networks can often be spotted by the orange fiber "cable." The FDDI protocol has also been adapted to run over traditional copper wires.
Finder	Exploration and Production data management software system provided by the Geoquest-Schlumberger corporation.
Firewall or Fire Wall System	A combination of hardware and software that separates a LAN into two or more parts for security purposes. A particular example of server.
Geoshare	E&P industry standard that connects software applications to a master database and to each other.
GeoTrek	Data browser that uses a GIS user interface to display geotechnical and other data. Developed as part of the Phase II of the NGDRS project based on the MobilView software. GeoTrek is a trademark of the American Geological Institute.
GIS	Geographic Information System (GIS), is an organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information.
HTML	HyperText Markup Language, is the coding language used to create Hypertext documents for use on the World Wide Web. HTML files are meant to be viewed using a World Wide Web client program, such as Netscape or Explorer.
Hypertext	Generally, any text that contains links to other documents - words or phrases in the document that can be chosen by a reader and which cause another document to be retrieved and displayed.
Internet	The vast collection of inter-connected networks that use the TCP/IP protocols and that evolved from the ARPANET (Advanced Research Projects Agency Network) of the late 60's and early 70's. The Internet in July 1995 connected roughly 60,000 independent networks into a vast global network.
intranet	A private network inside a company or organization that uses the same kinds of software that you would find on the public Internet, but that is only for internal use. As the Internet has become more popular many of the tools used on the Internet are being used in private networks, for example, many companies have web servers that are available only to their employees.
Java or JAVA	Java is a simple, object-oriented, distributed, interpreted, robust, secure, architecture neutral, portable, multi-threaded, dynamic, general purpose programming language developed by Sun Microsystems. Java supports programming for the Internet in the form of hardware and software independent Java "applets".
Landmark	Landmark Graphics, a software provider for the Petroleum industry.
LIS	Log Information Standard (LIS) was defined by the Schlumberger Company in 1981, to provide the means for the exchange of well log information. The

format was designed to work with various types of computers and with various types of information in a real-time data acquisition environment. This information ranges from sensor response (log) data to interpretation parameters and binary bit patterns. The LIS format was designed primarily for use with magnetic tape, although it may also be used with other types of storage media. A hierarchy is used for storing and retrieving data; this hierarchy separates physical representation from logical information interpretation. Each magnetic tape is kept logically complete by recording on each tape the information required to interpret the data on that tape. The LIS format is made up of two structures: logical and physical. Logical structure refers to the type of information and the organization of the information ; physical structure refers to the physical dimensions of the information (e.g., how much space it occupies).

Local Area Network or LAN	A computer network limited to the immediate area, usually the same building or floor of a building.
Metadata	Metadata is defined as data about data or the information required to make scientific data useful; Metadata are descriptive information about data and information resources. Typically, they describe, point to, or otherwise complement the information content of the data to which they are related. Metadata provide concise aid in locating desired information on a variety of topics, and help make such information easily accessible. Metadata may describe a range of information resources such as digital data images, databases, and printed materials such as books, memos, or maps.
Metadata Supercatalog	The metadata supercatalog contains, at any moment, the official catalog for the NGDRS since it is the collection of all information (metadata elements and records) contained (and described) in the subcatalogs created for all data stores.
Metadata Synchronization	The process and procedures associated with adding new metadata elements of information or new records of information is known as catalog synchronization.
MobilView	Name of software developed by Mobil Corp. that serve as the guide for the development of GeoTrek and PetroTrek software products.
NGDRS	The National Geosciences Data Repository System (NGDRS) is a network of existing and to-be created data repositories aimed at preserving and making accessible geoscience and engineering data to a large number of users in the public and private sectors.
Node	Same as Source Node.
NT System	NT operating systems from the Microsoft Corporation.
OpenWorks	Landmark Graphics E&P data management framework
POSC	The Petrotechnical Open Software Corporation (POSC) is a not-for-profit, member-directed and member-funded organization for the Exploration and Production (E&P) Sector of the Oil Industry. POSC mission is to benefit the

E&P industry by establishing, maintaining, and promoting specifications to be used as standards for the sharing of information throughout the asset life cycle.

PEF

The POSC Exchange Format (PEF) is a way to transfer compliant E&P data to different computing environments, to efficiently manage that data, and to decode it without misinterpretation. Whether the exchange programming is done by the commercial software developer or the in-house programmer, the specifications provide a vendor-neutral data link to the POSC world that enables end users to: transfer data between sites within a company; purchase data from vendors; exchange data between companies; archive data for future reference regardless of changes in hardware; and provide data for various government regulatory agencies.

PetroBank

Exploration and Production data management systems provided by the IBM corporation.

PetroTrek

Software system that uses a GIS user interface to display geotechnical, engineering and business data. It uses web-based technology to allow single point access to data stores and repositories regardless of their base technology, vintage, or geographical location. PetroTrek provides the means for users to view and select data from a variety of different data sources. It consists of a data browser, matching data servers that interface to databases or files, and broker servers which provide indexes to the data sources available. PetroTrek is a trademark of The Information Store Inc.

PetroVision

E&P data management system provided by the PECC corporation.

Platform

(a) Specific computer hardware, as in the phrase "platform-independent". (b) It may also refer to a specific combination of hardware and operating system and/or compiler, as in "this program has been ported to several platforms". (c) It is also used to refer to support software for a particular activity, as in "This program provides a platform for research into routing protocols".

Principle

One of the component of an information architecture. In this context, principles are statements of belief, reflecting an organization perception of "what is best" and that have the following general characteristics: (a) drive behavioral change, (b) be stable over a period of time, (c) be defensible, be specific (clear), and (d) serve as a starting point for difficult decisions. The establishment of principles should precede the determination of what standards to adopt.

Protocol

A set of formal rules describing how to transmit data, especially across a network. Low level protocols define the electrical and physical standards to be observed, bit- and byte-ordering and the transmission and error detection and correction of the bit stream. High level protocols deal with the data formatting, including the syntax of messages, the terminal to computer dialogue, character sets, and the sequencing of messages.

Rdbms or RDBMS

Relational database management system. A database management system with the ability to access data organized in tabular files that can be related to

each other by a common field (item). An RDBMS has the capability to recombine the data items from different files, providing powerful tools for data usage.

Record	A logical unit of data in a file. For example, there is one record for each well bore in a well.
Repository	A facility housing a collection of geosciences data stores supported by a particular Source Node Entity.
RODE Format	Record Oriented Data Exchange (RODE) is a standard administered by the Society of Exploration Geophysicists dealing with high density formats. Digital data forms the basis of exploration geophysics. The formats used for acquiring and processing data were (traditionally) designed for very specific needs largely determined by the storage systems available. As a result these formats are discipline, system and application dependent. Through the years, the industry has archived several million seismic field tapes recorded on old media. These archives are growing rapidly with the increasing use of 3-D acquisition for exploration and appraisal of hydrocarbon plays. The introduction of new high capacity media could reduce acquisition, processing, and storage costs and support automated access to the very large volumes of digital data. The processing industry needs to be able to move to the new media in stages protecting their investment in existing software and their ability to process data acquired on both new and old media. RODE is a standard encapsulation format provides the means of migrating old data into the new environment and allows new data to be recorded efficiently on new media using existing formats.
Router System	A device which forwards traffic between networks. The forwarding decision is based on network layer information and routing tables, often constructed by routing protocols.
RP66 Format	American Petroleum Institute Recommended Practice 66 dealing with exchange formats for the Petroleum industry. An alternative title for RP66 is Digital Log Interchange Standard (DLIS). DLIS is also commonly used to refer to the schemas, semantics and codes specific to well logs.
SEG-Y Format	One of the standard tape format for processed seismic data. SEG refers to Society of Exploration Geophysicists.
Server	A computer, or a software package, that provides a specific kind of service to client software running on other computers. The term can refer to a particular piece of software, such as a WWW server, or to the machine on which the software is running.
SISMO	Mobil Corporation internal format for storing seismic data.
Source Node	The NGDRS will be made up of a set of computing and data storage regional centers (which we shall designate as NGDRS Source Nodes or simply as Nodes) connected via one or more networking means such as, for example telephone private lines or Internet public services providers.

SQL	Structured Query Language. The international standard language for defining and accessing relational databases.
Standard	One of the components of an information architecture. Standards make the architecture "real".
Subcatalog	In the NGDRS, is the metadata catalog for a the data store of a particular repository at a source node.
Supercatalog	<i>see "metadata supercatalog"</i>
System	The sum of all NGDRS Source Nodes plus the Metadata Repository.
T1 or T-1 Line	Leased-line connection capable of carrying data at 1,544,000 bits-per-second. At maximum theoretical capacity, a T-1 line can move a megabyte in less than 10 seconds. That is still not fast enough for full-screen, full-motion video, that need speeds of at least 10,000,000 bits-per-second. T-1 is the fastest speed commonly used to connect networks to the Internet.
T3 or T-3 Line	A leased-line connection capable of carrying data at 44,736,000 bits-per-second. This is more than enough to do full-screen, full-motion video.
TCP/IP	(Transmission Control Protocol/Internet Protocol) -- This is the suite of protocols that defines the Internet. Originally designed for the UNIX operating system, TCP/IP software is now available for every major kind of computer operating system. TCP/IP was developed for internetworking, encompassing both network layer and transport layer protocols. While TCP and IP specify two protocols at specific layers, TCP/IP is often used to refer to the entire (Department of Defense) DoD protocol suite based upon these, including Telnet, FTP, UDP and RDP.
UKOOA Format	There are several UKOOA (U.K. Offshore Operators Association Limited) data exchange formats: (a) P2/91 UKOOA data exchange format is designed to record positioning data for both 2-D and 3-D geophysical seismic surveys. This format allows for individual time-tagging of observations. The format uses a coded system of records so that certain record types may be omitted entirely if they are not relevant. Any physical data storage medium may be used by prior agreement between the parties involved in exchange of data. (b) P2/94 UKOOA data exchange format deals with positioning of marine raw data. (c) P5/94 UKOOA data exchange format, a revision of the P5/86 format, has been designed to facilitate the exchange of pipelines, other sub-sea flow lines, umbilicals and power cables' position data. (d) Format for Data Exchange for Seismic 3D Binning Grids, is a guideline developed to explain the difficulties inherent in the choice of co-ordinate reference systems, and supply a set of parameters which defines a 3D seismic binning system. The variety of options in current use for the definition of seismic binning will not be supplanted by these parameters, but they will provide an exchange mechanism to allow users to uniquely and unambiguously transfer data. From the defined parameter set, users will be able to determine a subset of parameters which satisfy their present definitions. UKOOA is the representative organization for the British offshore oil and gas industry. Its

Members are the companies licensed to explore for and produce oil and gas in UK waters.

UNIX

A computer operating system (the basic software running on a computer, underneath applications like word processors and spreadsheets). UNIX is designed to be used by many people at the same time (it is multi-user) and has the TCP/IP protocols built-in. It is the most common operating system for servers on the Internet.

URL

Uniform Resource Locator - the standard way to give the address of any resource on the Internet that is part of the World Wide Web. The most common way to use a URL is with a WWW browser program.

Wide Area Network (WAN)

Any internet or network that covers an area larger than a single building or campus.

Web

Same as World Wide Web.

Web Browser

A client program (software) that is used to look at various kinds of Internet resources, for example Netscape Navigator or Microsoft Explorer. The browser gives the user means of viewing the contents of nodes (or "pages") and of navigating from one node to another.

World Wide Web (WWW)

An Internet client-server hypertext distributed information retrieval system which originated from the CERN High-Energy Physics laboratories in Geneva, Switzerland. On the WWW everything (documents, menus, indices) is represented to the user as a hypertext object in HTML format. Hypertext links refer to other documents by their URLs. These can refer to local or remote resources accessible via FTP, Gopher, Telnet or news, as well as those available via the HTTP protocol used to transfer hypertext documents.

X-Terminal

An intelligent terminal which operates using the X-Windows protocol directly connected to Ethernet. The X-Windows protocol was developed at MIT and it allows for creating device-independent windowing operations on bitmap display devices (such as computer screens).

X.25 Protocol

A standard for packet-switched network, approved by ISO. X.25 defines standard physical layer, datalink layer and network layers. It was developed to describe how data passes into and out of public data communications networks. X.25 networks are in use throughout the world.

APPENDIX A: PHASE II INFORMATION EXCHANGES

Table A1. Phase II Information Exchanges With Industry, Government, and Academic Groups

ORGANIZATION			INFORMATION EXCHANGE EVENT
NAME	LOCATION	MISSION	
American Petroleum Institute (API)	Washington, DC	National trade association serving the major integrated petroleum companies	Exploration Committee Endorsed NGDRS plan and concept
Bureau of Economic Geology, University of Texas	Austin, TX	Geosciences data repository for core and cutting	Core and cuttings data model for transfer, indexing, storage, and access via browser
Bureau of Land Management (BLM)	Washington, DC	To sustain the health, diversity and productivity of the public lands	NGDRS project presentation and joint BLM/MMS proposal for POSC compliant systems
California State Natural Resources Commission	Sacramento, CA	Enforcement of regulatory and policy issues involving state lands	Assessed opportunities for establishing West Coast repository facility
California State University	Bakersfield, CA	Preserve and provide access to geoscience data from the San Joaquin Basin	Reviewed and assessed facility operations procedures
California State University	Monterey Bay, CA	Prospect for new West Coast data facility	Assessed opportunities for establishing West Coast repository facility
Delaware Geological Survey	Newark, DE	Geological research, exploration, and services.	Reviewed offshore Atlantic data holdings and procedures for access
3-M Corporation	Minneapolis, MN	Develop various business products	Assessed available media for high-density electronic storage of geoscience data
Geomap, Inc.	Plano, TX	Supplier of geologic data and mapping services	Assessed storage and access capabilities
Louisiana Geological Survey	Baton Rouge, LA	Establish a Petroleum Information Center and the Louisiana Energy Access System	Inspected design of proposed new data repository facility

Minerals Management Service (MMS)	Washington, DC	Manage mineral resources of the outer continental shelf	Reviewed status of data inventory for offshore areas
North Dakota Geological Survey	Bismarck, ND	Geological research, exploration, and services.	Assessed data holdings and access procedures
Oak Ridge National Laboratory	Oak Ridge, TN	DOE multi-program energy laboratory managed by Lockheed Martin Energy Research Corporation,	Reviewed opportunities for establishing a digital seismic data repository
Department of the Interior Outer Continental Shelf Policy Committee	Washington, DC	Review policy and regulations controlling offshore oil, gas, and mineral development	Endorsed NGDRS plan and concept
Oklahoma Geological Survey	Norman, OK	Geological research and services.	Assessed data holdings and acquisition procedures
Petroleum Institute Corporation	Denver, CO	Provider of information services to the energy industry.	Reviewed data holdings and access procedures
Petroleum Technology Transfer Council (PTTC)	Washington, DC	A national clearinghouse for Oil & Natural Gas technology	Focused Technology Workshop, at the University of Alabama, Tuscaloosa, Alabama, March 96
PGS	Houston, TX	Seismic data acquisition and processing services	Specifications for seismic data for metadata supercatalog
Geosciences Data, Inc.	Dallas, TX	Commercial data storage and management	Reviewed data holdings, indexing, and access procedures
Texas Railroad Commission (TRRC)	Austin, TX	Regulatory policy for state oil and gas operations	Reviewed plans for digitizing railroad commission files
TGS	Houston, TX	Seismic vendor	Request seismic data for metadata supercatalog
Unocal Corporation	Brea, CA	Management of UNOCAL E&P data	Reviewed data holdings and opportunities for transfer of facility to public sector
USGS Federal Center	Denver, CO	Management of Rocky Mountain area data	Reviewed data holdings and procedures for assessment

Table A.2. Phase II Information Exchanges With Data Storage and Management Organizations

ORGANIZATION			TOPIC OF INFORMATION EXCHANGE
NAME	LOCATION	PROJECT	
Norwegian Petroleum Directorate (NPD)	Stavanger, Norway	DISKOS. National databank for all geoscience data supporting the North Sea Petroleum Industry	Metadata definition for index and catalog, data entry quality control for seismic data
Department of Trade and Industry (DTI)	London, United Kingdom	Common Data Access (CDA). North Sea databank for all geoscience data supporting petroleum operations	Metadata definition for index and catalog, data entry quality for well log data.
URGENT Consortium	Europe (various company locations)	URGENT (User Reference Group For Exploitation of New Technology)	Standards implementation (POSC) and definition of E&P business processes.
CANNOGIS	Calgary, Canada	Cannogis. Databank of geosciences data for the western provinces.	Metadata definition for index and catalog, data entry quality for well log and production data
POSC	Houston, TX	Several projects including the Industrial Pilot Project (IPP), POSC Caesar	Metadata definition for index and catalog (Epicentre footprint), implementation of geophysical data set, data loading procedures.
PeruPetro	Lima, Peru	Countrywide data repository for geosciences and engineering data.	Metadata definition for index and catalog, data entry quality for hardcopy and well log data. Definition of architecture for a country repository.
Sonatrach (PECC)	Algiers, Algeria	Countrywide data repository for geosciences and engineering data.	Metadata definition for index and catalog, data entry quality control for seismic data, data center and data loading operations
Alaska Energy InfoBank Consortium	Anchorage, Alaska	Joint industry data repository for geosciences and engineering data.	Metadata definition for index and catalog.

Geological Survey of the Netherlands	The Hague, Netherlands	Country database	Metadata definition for index and catalog
Geological Survey of Denmark	Copenhagen, Denmark	Country database	Metadata definition for index and catalog
New Zealand Ministry of Commerce, Energy and Resources Division	Auckland, New Zealand	Country database	Metadata definition for index and catalog
Indian Directorate General of Hydrocarbons	New Delhi, India	Country database	Metadata definition for index and catalog

APPENDIX B: METADATA EXAMPLES FOR A VARIETY OF DATA TYPES

Note: All examples below do not include all the required metadata items; they each represent one particular logical record of information requiring possibly multiple physical records (or tables).

Table B.1. 3-D Seismic Data - Metadata Example

	Metadata Item	Example - Values	Comments
1	Geographic Location	Offshore Louisiana, Grand Isle South Addition	
2	Block(s) Occupied	Block 90, portions of 91, 92, 93 and portions of Ewing Bank blocks 347, 348	
3	X-Y Coordinates: (Latitude and Longitude Corners) NE	X=2420043, Y=-17209 Lat=28°36'46", Long=90°01'28"	Coordinate data merged with traces
4	X-Y Coordinates: (Latitude and Longitude Corners) SW	X=2398399, Y=-46491; Lat=28°31'58", Long=90°05'35"	Coordinate data merged with traces
5	Size of Dataset	Approx. 4 blocks, 36 sq. miles	
6	Landmark (data file) Size	750MB	
7	Vintage of Dataset	1984	
8	Acquired by	M/V T.W. Nelson (S-7)	
9	Processed by	Mobil Exploration & Producing	
10	Processing History	On side label, no tape	
11	Ownership of Dataset	Mobil E&P - 100%	
12	Ellipsoids/Projections Used	Louisiana State Plane Zone 2, Clarke 1866 Spheroid.	
13	Observers Logs	Microfiche: Lines 1001 - 1265 on fiche #s 14496-86	
14	Type of Physical Media	Field & processed tapes 1/2"9 track	
15	Field Tape format	SEG Y	
16	Processed tapes format	SISMO	
17	Landmark disks format	SEG Y	Mobil internal format
18	Navigation tapes format	UKOOA	
19	Stack tapes - Filtered/Unfiltered	None available	
20	Migrated tapes - Filtered/Unfiltered	Filtered available	
21	Relative Amplitude Tape	Not available	On the Landmark project database only
22	Scaling Function Used	800ms AGC	
23	Shotpoint & Bin center maps	Bin center and Crossline location maps	
24	Binset Grid, Binset Name	100546, HHH-02-A Binset	

Table B2. Offshore Well Data - Metadata Example

	Metadata Item	Example - Values	Comments
1	Name	South Timbalier 260-1	
2	API Code Number	177180001100	
3	Type	Simplex	
4	No. of wellbores	Two (2)	Separate records
5	Purpose	Development	
6	Status	Plugged and abandoned (P&A)	
7	Short description	"Short description"	
8	Service Class	Production testing	
9	State	Not applicable	
10	Ellipsoids/Projections Used	Louisiana State Plane Zone 2, Clarke 1866 Spheroid.	
11	Lease	South Timbalier 260	
12	License number for Offshore:	OCS 00159	
13	Block Number	260	
14	Field Name	South Timbalier	
15	Slot Offset for Development wells	None	
16	Well Ownership	Mobil 100%	
17	Date Spudded	67/07/07	
18	Date Completed	67/09/13	
19	Observers logs	Available	Separate records
20	Drilling Company	Zapata	
21	Rig Name	Heron	
22	Water depth (feet)	217	
23	Kelly bushing (feet)	98	
24	Slot Offset for Development wells	None	
25	Drill bit	Hughes milled-tooth	
26	Coring bit	Christensen diamond	
27	Drilling Fluids	Oil-based, barite	
28	Initial pressure	1500 psi	
29	Initial Production	3,500 bbls/day	
30	Drill Stem Test	Available	Separate records

Table B3. Wellbore Data - Metadata Example

	Metadata Item	Example - Values	Comments
1	Wellbore Number	2	Sequence no.
2	Kind	Re-spud	
3	Part of Well Number	177180001100	
4	Status	Plugged and abandoned (P&A)	API Code Number
5	Purpose	New pool test	Original
6	Surface point vertex (X)	2406863	
7	Surface point vertex (Y)	29812	
8	Surface point type	Perforation	
9	Surface point coordinate: latitude	28.74244499	
10	Surface point coordinate: longitude	-90.06395721	
11	Bottom point vertex (X)	2406899	
12	Bottom point vertex (Y)	29845	
13	Deviation survey	Available	Separate records
14	Date Spudded	67/07/07	
15	Date completed	67/09/13	
16	Observers logs	Available	Separate records
17	Cored interval top (feet)	2366	
18	Cored interval bottom (feet)	4223	
19	Logged interval top (feet)	2500	
20	Logged interval bottom (feet)	4350	

Table B.4 Well Logging (Trace) Data - Metadata Example

	Metadata Item	Example - Values	Comments
1	Logged interval top (feet)	2500	
2	Logged interval bottom (feet)	4350	
3	Part of Well Number	177180001100	
4	Wellbore Number	2	
5	Sonde manufacturer, serial no.	"Sonde manufacturer, serial no."	
6	Logs physical media	Field & processed tapes 1/2" 9 track	
7	Formats	Field tapes LIS	
8	Processing History	On log, not on tape.	
9	Induction log	Available	Separate record
10	Dual induction log	Not available	Separate record
11	Laterolog log	Available	Separate record
12	Microlog log	Not available	Separate record
13	Micro-laterolog log	Not available	Separate record
14	Spontaneous potential (SP)	Available	Separate record
15	Gamma Ray (GR)	Available	Separate record
16	Density log	Available	Separate record
17	Neutron log	Not available	Separate record
18	Sonic	Available	Separate record
19	Dipmeter	Available	Separate record
20	Directional survey	Available	Separate record
21	Caliper log	Available	Separate record

Table B5. Core Data - Metadata Example

	Metadata Item	Example - Values	Comments
1	Cored Interval Top (feet)	2366	
2	Cored Interval Bottom (feet)	4223	
3	Bar Code Number(s)	123552366-12354223	
4	Part of Well Number	177180001100	
5	Wellbore Number	2	
6	No. of Boxes for Cored Interval	110 boxes	
7	Core Barrel	"Manufacturer and serial number"	
8	Coring Start Date	67/07/10	
9	Coring Completed Date	67/09/11	
10	Geologist Logs	Available	Separate record
11	Analysis	Porosity, lithology	
12	Analysis History	Not available	
13	Analysis Physical Media	1/2" 9 track tapes	
14	Media Format	ASCII	
15	Core Images	Available	Not across all the interval
16	Core Image Format	3"x4"	Polaroid
17	Porosity Measurements	Available	Separate records
18	Permeability Measurements	Not available	Separate records
19	Lithology descriptions	Available	Separate records
20	Formation Age	Tertiary	
21	Donation Number	12355	
22	Donation Company	Mobil	
23	Date Received	88/12/06	
24	Date Data Labeled	89/01/06	
25	Re-box Date	89/01/06	
26	Date Sampled Sorted	89/01/20	
27	Slab Date	68/03/06	
28	Sample Half Location	MOBIL, Dallas	
29	Use restrictions	None	
30	Use restrictions comments	Open to all	
31	Special projects core	No	
32	Core chips	None	
33	Secondary gas project core ?	No	
34	Tight gas project core ?	No	
35	Core loaned or donated ?	Donated	
36	Samples related to Thesis	No	
37	Quality assurance Guidelines	Bureau of Economic Geology	
38	Repository Location	Austin, Texas	
39	Warehouse Location	Midland, Texas	
40	Shelf Location	15N-12E, 3rd Row	

Table B6. Hardcopy Data - Metadata Example

	Metadata Item	Example - Values	Comments
1	Data Type	Seismic section	Other: logs, maps, scout tickets
2	Standard Data Code	"S1"	To be established
3	Geographic Location	X=2420043, Y=-17209 Lat=28*36'46", Long=90*01'28"	
4	State	LA	
5	County	Not applicable	
6	Legal Description	"Legal description"	
7	License number for Offshore	OCS 00159	
8	Block Number	260	
9	Field Name	South Timbalier	
10	Vintage of Dataset	1985	
11	Ownership	Mobil 100%	
12	Ellipsoids/Projections Used	Louisiana State Plane Zone 2, Clarke 1866 Spheroid.	
13	Observers Logs	Available	Separate records
14	Type of Physical Media	Paper	Other: Film, Fiche, Sepia
15	Formats	Mobil standard	
16	History	Processed, interpreted by AJJ, February 1995	
17	Donation Company	Mobil	
18	Date Labeled	92/06/03	
19	Date Sorted and Inventoried	92/06/08	
20	Date Received	92/07/01	
21	Re-labeled Date	92/07/10	
22	Date Cataloged	92/08/01	
23	Item on Loan or Donated	Donated	
24	Bar Code Number	1667788	
25	Number of Items or Pages	2	
26	Repository Location	New Orleans	
27	Warehouse	15th Street	
28	Location	Area 35, Cabinet 1, Shelf 6	
29	Warehouse Contact Information	JJR	
30	Use Restricted	No	
31	Use Restricted Comment	None	
32	Check out Date	93/10/10	
33	Estimated Return Date	93/11/10	
34	Actual Return Date	93/12/10	
35	Quality Assurance Guidelines	Keep in dark area	

APPENDIX C: NATIONAL DIRECTORY OF GEOSCIENCE DATA REPOSITORIES QUESTIONNAIRE

Please complete this form and return it to AGI in order to have your organization included in the National Directory of Geoscience Data Repositories and Data Providers. You may also complete the form on-line at <http://www.istore.com/>.

Primary Contact Person:

Data Access Fee:

Data Center Name:

- ☐ Per Search
- ☐ Membership Dues
- ☐ Other

Parent Organization:

Catalog and Index Type:

Street/P.O. Box Address:

- ☐ Paper
- ☐ Digital
- ☐ CD ROM
- ☐ Fiche
- ☐ Other: _____

City, State, Zip Code:

Data Retrieval:

Telephone:

- ☐ On-line
- On-line search retrieval URL: _____

FAX:

E-Mail:

Data Delivery Media:

URL (http, ftp, etc):

- ☐ Paper
- ☐ CD ROM
- ☐ On-line
- ☐ Mail out
- ☐ Other: _____

Data Center Staff Size: _____

Type of Repositories:

Approximate Quantity of Holdings:

- ☐ Public:
 - ☐ Federal ☐ State/Local
 - ☐ Academic ☐ Other Public
- ☐ Private:
 - ☐ Membership ☐ Assn.
 - ☐ Commercial ☐ Proprietary

_____ line miles of seismic 2-D
_____ square miles of seismic 3-D
_____ number of well logs
_____ feet of core
_____ scout tickets/completion cards

Geographic Areas

Media

Type of Data

Geophysical Data

Seismic Refraction

2-D Seismic Reflection

3-D Seismic Reflection

Potential Fields

Velocity, VSP

Drill Hole Data

Well Logs

Lithology Logs

Cores

Cuttings

Paleo Reports/Samples

Drilling/Completion Data

Scout Data/Strat Tops

Drill Stem & Other Test Data

Production Data

Field/Pool Information

Field Studies

Reserve Studies

Other:

Maps & Location Info

Base Maps

Digital Well Spots

Lease Ownership Maps

Topographic Maps

Geologic Maps (Surface)

Subsurface Maps

Aerial/Satellite Imagery

Other

Geochemical Data

Paleontological Collections

Rock & Mineral Specimens

Thin Sections

Other: _____

Eastern Onshore
Atlantic Offshore
Gulf Onshore
Gulf Offshore
Permian Basin
Mid-Continent
Rockies-N. Plains
Colorado Plat-B&R
Pacific Onshore
Pacific Offshore
Alaska
Inter

ational
Paper
Digital
Film
Microfiche

[illegible]

Key for Coding Table

No holdings:

Minor holdings:

Major holdings:

Mark as applicable

☒

Services Provided:

- | | |
|---|---|
| <input type="checkbox"/> Core/Sample services | <input type="checkbox"/> Petrography |
| <input type="checkbox"/> Core Analysis | <input type="checkbox"/> Seismic Processing |
| <input type="checkbox"/> Thin Section | <input type="checkbox"/> Staff Research |
| <input type="checkbox"/> Geochemistry | <input type="checkbox"/> Interpretation |
| <input type="checkbox"/> Scanning/Digitizing | <input type="checkbox"/> Data Packages |
| <input type="checkbox"/> Paleo Picks | <input type="checkbox"/> Bibliographic Searches |
| <input type="checkbox"/> Tape Transcription | <input type="checkbox"/> Other Services: |

Comments on holdings and general services:

Completed and submitted by:

Name

Title

() _____
Phone

() _____
Fax

Please return to: American Geological Institute
ATTN: Nicholas H. Claudy
4220 King Street
Alexandria, VA 22302-1502
Telephone: 703/379-2480
Fax: 703/379-7563

APPENDIX D: SUMMARY OF DATA FROM THE NATIONAL DIRECTORY OF GEOSCIENCE DATA REPOSITORIES

Type of Data	Geographic Areas											Media			
	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital
Geophysical Data															
ALABAMA															
Department of Geology (University of Alabama)			•											•	
ALASKA															
Alaska Geologic Materials Center											•			•	
Alaska Outer Continental Shelf Region											•			•	•
Bureau of Land Managment -Alaska											•			•	
ARIZONA															
Heinrichs GEOEXploration Co. and Associates							•								
CALIFORNIA															
Geology-Mining-Petroleum Room														•	
COLORADO															
Gas Research Institute Information Center	•	•	•	•	•	•	•	•	•	•	•			•	•
Petroleum Information	•	•	•	•	•	•	•	•	•	•	•			•	•
CONNECTICUT															
Geological and Natural History Survey of Connecticut			•											•	•
DELAWARE															
Delaware Geological Survey	•	•												•	•
FLORIDA															
Florida Geological Survey	•			•										•	
ILLINOIS															
Illinois State Geological Survey Division					•									•	•
INDIANA															
Energy Resources Section (Indiana Geological Survey)					•										
IOWA															
Geological Survey Bureau					•									•	•

Geographic Areas

Media

Type of Data

Geophysical Data

LOUISIANA

Cambe Geological Services, Inc.

Louisiana Geological Survey

MAINE

Natural Resources Information
and Mapping Center

MISSISSIPPI

Jackson Log Library

USM GeoData

NEVADA

Nevada Bureau of Mines
and Geology

NEW JERSEY

New Jersey Geological
Survey

NORTH CAROLINA

Coastal Plain Office and
Repository (North Carolina
Geological Survey)

OHIO

Ohio Division of Geological
Survey

TENNESSEE

Tennessee Division of
Geology

TEXAS

C&M Storage, Inc.

Geco-Prakla, NSA, Houston

Midland Energy Library

Ocean Drilling Program

(Texas A&M University)

Pan American Center for
Earth and Environ Studies
(Univ of Texas at El Paso)

**Southern States Professional
Log Association, Inc.**

	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital	Film	Microfiche
				
				
	.													.			
				.										.			
			.	.										.			
												.		.			
	.	.													.		
	.	.															
					.									.	.		
	.													.		.	

		
				.										.			
		
					

Geographic Areas

Media

Type of Data

Geophysical Data

UTAH

Energy and Geoscience

Institute (University of Utah)

WASHINGTON

Division of Geology and Earth

Resources

Drill Hole Data

ALABAMA

Alabama State Oil and Gas

Board

Department of Alabama

(University of Alabama)

ALASKA

Alaska Geologic Materials

Center

Alaska Oil and Gas

Conservation Commission

Alaska Outer Continental

Shelf Region

Bureau of Land Management

- Alaska

Division of Oil and Gas

Earth Science Collections

(University of Alaska

Museum)

ARIZONA

Arizona Geological Survey

Oil and Gas Conservation

Commission

ARKANSAS

Arkansas Geological.

Comission

U.S. Army Corps of

Engineers

Geographic Areas

Media

Type of Data

Drill Hole Data

CALIFORNIA

California Well Sample

Repository

Geology-Mining-Petroleum
Room

Teale Data Center

COLORADO

Colorado Geological Survey

Core Research Center (USGS)

Denver Earth Resources

Library

Field Records Library (USGS)

Gas Research Institute

Information Center

Petroleum Information

CONNECTICUT

Connecticut Department of
Environmental Protection

Core Repository

Geological and Natural

History Survey of

Connecticut

DELAWARE

Delaware Geological Survey

FLORIDA

Antarctic Marine Geology

Research Facility (Florida State University)

Florida Geological Survey

GEORGIA

Georgia Geologic Survey

HAWAII

Core Analysis Laboratory

Honolulu Board of Water

Supply

[illegible]

Geographic Areas

Media

Type of Data

Drill Hole Data

ILLINOIS

Illinois State Geological
Survey Division
Illinois Oil and Gas
Association

INDIANA

Energy Resources Section
(Indiana Geological Survey)

IOWA

Geological Survey Bureau

KANSAS

Geoscience Data Resources
(Kansas Geological Survey)
Independent Oil and Gas
Service
Wichita Well Sample Library

KENTUCKY

NREPC Office of Information
Services

LOUISIANA

Ark-La-Tex Log Library, Inc.
Basin Research Institute
Resource Center
Cambe Geological Services
Louisiana Geological Survey

MAINE

Natural Resources Information
and Mapping Center

MARYLAND

Maryland Geological Survey

MINNESOTA

Minnesota Division Core
Library

Water Well Information Unit

MISSISSIPPI

Geological Data Center, Inc.
Jackson Log Library

[illegible]

Geographic Areas

Media

Type of Data

Drill Hole Data

Mississippi Office of Geology

USM GeoData

MISSOURI

McCracken Core Library

MONTANA

Northern Rockies Geologic

Data Center

Department of Geology

(Rocky Mountain College)

NEBRASKA

Nebraska Conservation and

Survey Division

NEVADA

Nevada Bureau of Mines and

Geology

Nevada Department of

Conservation

Yucca Mountain Site Charac-

terization Project

NEW JERSEY

New Jersey Geological Survey

NEW MEXICO

New Mexico Library of

Subsurface Data

Roswell Energy Library, Inc.

NEW YORK

Collections Unit (New York

State Geological Survey)

NORTH CAROLINA

Coastal Plain Office and

Repository (North Carolina

Geological Survey)

NORTH DAKOTA

Wilson M. Laird Core and

Sample Library

	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital	Film	Microfiche
Mississippi Office of Geology		•	•											•		•	•
USM GeoData		•	•											•			
MISSOURI																	
McCracken Core Library													•				
MONTANA																	
Northern Rockies Geologic Data Center							•							•		•	•
Department of Geology (Rocky Mountain College)							•									•	•
NEBRASKA																	
Nebraska Conservation and Survey Division														•	•		
NEVADA																	
Nevada Bureau of Mines and Geology													•	•		•	
Nevada Department of Conservation							•							•	•		
Yucca Mountain Site Characterization Project							•							•	•	•	•
NEW JERSEY																	
New Jersey Geological Survey	•	•												•	•		
NEW MEXICO																	
New Mexico Library of Subsurface Data					•		•							•	•		
Roswell Energy Library, Inc.					•		•										
NEW YORK																	
Collections Unit (New York State Geological Survey)														•	•		•
NORTH CAROLINA																	
Coastal Plain Office and Repository (North Carolina Geological Survey)														•			
NORTH DAKOTA																	
Wilson M. Laird Core and Sample Library							•										

Drill Hole Data

OHIO

Ohio Geologic Sample Repository

Subsurface Data Center
(University of Toledo)

OKLAHOMA

L.S. Youngblood Energy
Library

OGS NRIS Facility

Oklahoma City Geological
Society Library

Riley Electric Log, Inc.

Sooner Well Log Service

OREGON

Nature of the Northwest Info
Center (OR Dept of Geology
& Mineral Industries)

PENNSYLVANIA

Pennsylvania Geological
Survey

SOUTH CAROLINA

South Carolina Geological
Survey

SOUTH DAKOTA

South Dakota Geological
Survey

TENNESSEE

Tennessee Division of
Geology

TEXAS

Balcones Energy Library, Inc.

C&M Storage, Inc.

Cambe Geological Services, Inc.

Core Research Center (Bureau
of Economic Geology)

Media

	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital	Film	Microfiche
					
				
					.									.			.
					.									.			
					.									.			
					.									.			.

					.									.			
								.						.		.	
		
	.													.	.		
					.									.	.		
	.													.			

			
			.											.			

Geographic Areas

Media

Type of Data

Drill Hole Data

Energy Information, Inc.
Herold Geological Research
Center
Herold Geological Research
Center (Abilene)
Midland Energy Library
Midland County Public
Library
Ocean Drilling Program
(Texas A&M University)
Oil Information Library of
Fort Worth
Oil Information Library of
Wichita Falls
Pan American Center for
Earth and Environ Studies
(Univ of Texas at El Paso)
Southern States Professional
Log Association, Inc.

UTAH

Energy and Geoscience
Institute (University of Utah)
Sample Library (Utah
Geological Survey)

VERMONT

Vermont Geological Survey

VIRGINIA

Division of Mineral Resources
U.S. Army Corps of
Engineers

WASHINGTON

Division of Geology and
Earth Resources

WEST VIRGINIA

Oil and Gas Section (West
Virginia Geological and
Economic Survey)

	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital	Film	Microfiche
Energy Information, Inc.				•										•	•		
Herold Geological Research Center	•	•	•	•	•	•	•	•	•	•				•	•	•	•
Herold Geological Research Center (Abilene)				•	•												
Midland Energy Library				•	•	•								•	•	•	•
Midland County Public Library				•													
Ocean Drilling Program (Texas A&M University)	•	•							•					•	•		
Oil Information Library of Fort Worth		•		•	•									•	•	•	•
Oil Information Library of Wichita Falls				•	•									•		•	•
Pan American Center for Earth and Environ Studies (Univ of Texas at El Paso)				•	•	•	•	•	•					•	•		
Southern States Professional Log Association, Inc.			•														
UTAH																	
Energy and Geoscience Institute (University of Utah)						•	•	•			•			•	•		
Sample Library (Utah Geological Survey)						•											
VERMONT																	
Vermont Geological Survey	•													•			
VIRGINIA																	
Division of Mineral Resources	•													•			
U.S. Army Corps of Engineers	•	•												•			
WASHINGTON																	
Division of Geology and Earth Resources							•	•						•	•		
WEST VIRGINIA																	
Oil and Gas Section (West Virginia Geological and Economic Survey)	•													•	•	•	•

Media

Drill Hole Data

Wisconsin Geological and
Natural History SurveyHerold Geological Research
CenterInternational Archive of
Economic Geology

Wyoming State Geological
Survey

Wyoming Oil and Gas
Conservation Commission

Alabama State Oil and Gas
Board

Department of Geology
(University of Alabama)

Alaska Oil and Gas
Conservation Commission

Alaska Outer Continental Shelf Region

Bureau of Land Management
- Alaska

Division of Oil and Gas

Earth Science Collections
(University of Alaska)

Arizona Geological Survey
Heinrichs GEOEXploration

Co. and Associates

Oil and Gas Conservation Commission

105

Geographic Areas

Media

Type of Data

Field/Pool Information

CALIFORNIA

Teale Data Center

COLORADO

Field Records Library (USGS)

Gas Research Institute

Information Center

Petroleum Information

FLORIDA

Florida Geological Survey

ILLINOIS

Illinois State Geological

Survey Division

Illinois Oil and Gas

Association

INDIANA

Energy Resources Section

(Indiana Geological Survey)

KANSAS

Geoscience Data Resources

(Kansas Geological Survey)

KENTUCKY

Office of General Information

(Kentucky Geological Survey)

LOUISIANA

Louisiana Geological Survey

MINNESOTA

Minerals Division Core

Library

MISSISSIPPI

Jackson Log Library

MONTANA

Geology Department

(University of Montana)

NEBRASKA

Nebraska Conservation and
Survey Division[illegible]

Geographic Areas

Media

Type of Data

Field/Pool Information

WEST VIRGINIA

Oil and Gas Section (West
Virginia Geological and
Economic Survey)

PTTC Appalachian Resource
Center

WYOMING

Herold Geological Research
CenterWyoming State Geological
Survey

Maps & Location Information

ALABAMA

Alabama State Oil and Gas
Board

Department of Geology
(University of Alabama)

ALASKA

Alaska Geologic Materials
Center

Alaska Oil and Gas

Conservation Commission

Alaska Outer Continental Shelf Region

Division of Oil and Gas

Earth Science Collections

(University of Alaska
Museum)

Technical Data Unit (USGS)

ARIZONA

Arizona Geological Survey
Henrichs GEOEXploration
Co. and Associates

Eastern Onshore
Atlantic Offshore
Gulf Onshore
Gulf Offshore
Permian Basin
Mid-Continent
Rockies-N. Plains
Colorado Plat-B&R
Pacific Onshore
Pacific Offshore
Alaska
Internat'l

International
State Specific
Paper

Paper
Digital
Film
Microfiche

[illegible]

Maps & Location Information

Media

[illegible]

Geographic Areas

Media

Type of Data

Maps & Location Information

ILLINOIS

Illinois State Geological
Survey Division

Illinois Oil and Gas
Association

INDIANA

Energy Resources Section
(Indiana Geological Survey)

IOWA

Geological Survey Bureau

KANSAS

Geoscience Data Resources
(Kansas Geological Survey)

Wichita Well Sample Library

KENTUCKY

Office of Geologic
Information (Kentucky
Geological Survey)

University of Kentucky

LOUISIANA

Ark-La-Tech Log Library, Inc.

Cambe Geological Services

Louisiana Geological Survey

MAINE

Natural Resources Information
and Mapping Center

MARYLAND

Maryland Geological Survey

MINNESOTA

Minerals Division Core Library

Water Well Information Unit

MISSISSIPPI

Jackson Log Library

Mississippi Office of Geology

USM GeoData

[illegible]

Type of Data

Maps & Location Information

MISSOURI

NASA Regional Planetary
Image Facility

NEBRASKA

Nebraska Conservation and
Survey Division

NEW JERSEY

New Jersey Geological Survey

NEW MEXICO

New Mexico Library of
Subsurface Data

Roswell Energy Library, Inc.

NEW YORK

Collections Unit (New York
State Geological Survey)

NORTH CAROLINA

Coastal Plain Office and
Repository (North Carolina
Geological Survey)

OHIO

Ohio Division of Geological
Survey

Ohio Geologic Sample
Repository

Subsurface Data Center
(University of Toledo)

OKLAHOMA

L.S. Youngblood Energy
Library

OGS NRIS Facility

Oklahoma City Geological
Society Library

OREGON

Nature of the Northwest Info
Center (OR Dept of Geology
& Mineral Industries)

Geographic Areas

Media

Eastern Onshore
Atlantic Offshore
Gulf Onshore
Gulf Offshore
Permian Basin
Mid-Continent
Rockies-N. Plains
Colorado Plat-B&R
Pacific Onshore
Pacific Offshore
Alaska
International
State Specific

Paper
Digital
Film
Microfiche

	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital	Film	Microfiche
MISSOURI																	
NASA Regional Planetary Image Facility														•	•	•	
NEBRASKA																	
Nebraska Conservation and Survey Division													•	•	•		
NEW JERSEY																	
New Jersey Geological Survey	•													•	•		
NEW MEXICO																	
New Mexico Library of Subsurface Data					•	•								•			
Roswell Energy Library, Inc.					•												
NEW YORK																	
Collections Unit (New York State Geological Survey)	•													•	•		
NORTH CAROLINA																	
Coastal Plain Office and Repository (North Carolina Geological Survey)	•													•			
OHIO																	
Ohio Division of Geological Survey					•									•	•	•	
Ohio Geologic Sample Repository					•									•	•		
Subsurface Data Center (University of Toledo)	•	•		•	•	•	•			•				•		•	
OKLAHOMA																	
L.S. Youngblood Energy Library	•	•		•	•	•	•	•		•	•			•	•		
OGS NRIS Facility					•									•			
Oklahoma City Geological Society Library					•									•	•		
OREGON																	
Nature of the Northwest Info Center (OR Dept of Geology & Mineral Industries)								•	•					•		•	

Type of Data

Maps & Location Information

PENNSYLVANIA

Mine Map Repository (U.S.
Department of the Interior)

Pennsylvania Geological
Survey

SOUTH CAROLINA

South Carolina Geological
Survey

SOUTH DAKOTA

EROS Data Center (USGS)
South Dakota Geological
Survey

TENNESSEE

Tennessee Division of Geology

TEXAS

Balcones Energy Library, Inc.

C&M Storage, Inc.

Cambe Geological Services,
Inc.

Energy Information, Inc.

Geco-Prakla, NSA, Houston

Geophysical Log Facility

(Bureau of Economic
Geology)

Herold Geological Research
Center

Herold Geological Research
Center (Abilene)

Midland County Public
Library

Midland Energy Library

Ocean Drilling Program
(Texas A&M University)

Oil Information Library of
Fort Worth

Geographic Areas

Media

Eastern Onshore
Atlantic Offshore
Gulf Onshore
Gulf Offshore
Permian Basin
Mid-Continent
Rockies-N. Plains
Colorado Plat-B&R
Pacific Onshore
Pacific Offshore
Alaska
International
State Specific

Paper
Digital
Film
Microfilm

	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital	Film	Microfilm
Mine Map Repository (U.S. Department of the Interior)																	
Pennsylvania Geological Survey																	
South Carolina Geological Survey																	
EROS Data Center (USGS)																	
South Dakota Geological Survey																	
Tennessee Division of Geology																	
Balcones Energy Library, Inc.																	
C&M Storage, Inc.																	
Cambe Geological Services, Inc.																	
Energy Information, Inc.																	
Geco-Prakla, NSA, Houston																	
Geophysical Log Facility (Bureau of Economic Geology)																	
Herold Geological Research Center																	
Herold Geological Research Center (Abilene)																	
Midland County Public Library																	
Midland Energy Library																	
Ocean Drilling Program (Texas A&M University)																	
Oil Information Library of Fort Worth																	

Geographic Areas

Media

Type of Data

Maps & Location Information

Oil Information Library of

Wichita Falls

Pan American Center for

Earth and Environ Studies

(Univ of Texas at El Paso)

Southern States Professional

Log Association, Inc.

UTAH

Energy and Geoscience

Institute (University of Utah)

Department of Geology

(Southern Utah University)

Harold B. Lee Library (Brigham

Young University)

VERMONT

Vermont Geological Survey

VIRGINIA

Division of Mineral Resources

WASHINGTON

Division of Geology and Earth

Resources

WEST VIRGINIA

Oil and Gas Section (West

Virginia Geological and

Economic Survey)

WISCONSIN

Wisconsin Geological and

Natural History Survey

WYOMING

Herold Geological Research

Center

International Archive of

Economic Geology

Wyoming State Geological

Survey

Media

Other

Alabama State Oil and Gas
Board

Department of Geology
(University of Alabama)

Alaska Geologic Materials
Center

Alaska Outer Continental Shelf Region

Earth Science Collections
(University of Alaska
Museum)

Technical Data Unit (USGS)

Arizona Geological Survey
Oil and Gas Conservation
Commission

Arkansas Geological
CommissionCalifornia Well Sample
Repository

Colorado Geological Survey
Core Research Center (USGS)
Field Records Library (USGS)
Gas Research Institute
Information Center

Geological and Natural History
Survey of Connecticut

Delaware Geological Survey

114

Media

Other

FLORIDA

Antarctic Marine Geology
Research Facility (Florida
State University)

Florida Geological Survey

HAWAII

Honolulu Board of Water
Supply

ILLINOIS

Illinois State Geological Survey Division

INDIANA

Energy Resources Section
(Indiana Geological Survey)

IOWA

Geological Survey Bureau

KENTUCKY

Office of Geologic Information
(Kentucky Geological Survey)

MAINE

Natural Resources Information
and Mapping Center

MARYLAND

Maryland Geological Survey

MINNESOTA

Minerals Division Core
Library

Water Well Information Unit

MISSISSIPPI

Mississippi Office of Geology
USM GeoData

MONTANA

Geology Department
(University of Montana)

NEBRASKA

Nebraska Conservation and
Survey Division

[illegible]

Geographic Areas

Media

Type of Data

Other

NEW JERSEY

New Jersey Geological Survey

NEW MEXICO

New Mexico Library of
Subsurface Data

NEW YORK

Collections Unit (New York
State Geological Survey)

NORTH CAROLINA

Coastal Plain Office and
Repository (North Carolina
Geological Survey)

NORTH DAKOTA

Wilson M. Laird Core and
Sample Library

OHIO

Ohio Division of Geological
Survey

Ohio Geologic Sample
Repository

Subsurface Data Center
(University of Toledo)

OKLAHOMA

National Coal Resources Data
System

OCS NRIS Facility

Oklahoma City Geological
Society Library

OREGON

Nature of the Northwest Info
Center (OR Dept of Geology
& Mineral Industries)

PENNSYLVANIA

Pennsylvania Geological
Survey

	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital	Film	Microfiche
NEW JERSEY																	
New Jersey Geological Survey																	
NEW MEXICO																	
New Mexico Library of																	
Subsurface Data				•		•								•			
NEW YORK																	
Collections Unit (New York																	
State Geological Survey)	•													•	•		
NORTH CAROLINA																	
Coastal Plain Office and																	
Repository (North Carolina																	
Geological Survey)	•																
NORTH DAKOTA																	
Wilson M. Laird Core and																	
Sample Library						•											
OHIO																	
Ohio Division of Geological																	
Survey					•									•			
Ohio Geologic Sample																	
Repository					•									•			
Subsurface Data Center																	
(University of Toledo)	•	•	•	•	•	•	•	•	•	•	•	•	•				
OKLAHOMA																	
National Coal Resources Data					•												
System																	
OCS NRIS Facility					•												
Oklahoma City Geological																	
Society Library					•												
OREGON																	
Nature of the Northwest Info																	
Center (OR Dept of Geology																	
& Mineral Industries)								•									
PENNSYLVANIA																	
Pennsylvania Geological																	
Survey	•													•	•		

Other

RHODE ISLAND

Marine Geological Samples
Library (University of
Rhode Island)

SOUTH CAROLINA

South Carolina Geological
Survey

SOUTH DAKOTA

South Dakota Geological
Survey

TENNESSEE

Tennessee Division of Geology

TEXAS

Core Research Center (Bureau
of Economic Geology)

Midland Energy Library

Ocean Drilling Program
(Texas A&M University)

**Southern States Professional
Log Association, Inc.**

UTAH

Energy and Geoscience
Institute (University of Utah)

VIRGINIA

Division of Mineral Resources

WASHINGTON

Division of Geology and Earth Resources

WYOMING

International Archive of
Economic Geology
Wyoming State Geological
Survey

Media

	Eastern Onshore	Atlantic Offshore	Gulf Onshore	Gulf Offshore	Permian Basin	Mid-Continent	Rockies-N. Plains	Colorado Plat-B&R	Pacific Onshore	Pacific Offshore	Alaska	International	State Specific	Paper	Digital	Film	Microfiche
							
	.													.			
					.									.	.		
	.																
			.														
				.										.			
		
			.														
							
	.													.			
								
										.				.			
					.									.			

APPENDIX E: AGI NATIONAL GEOSCIENCE DATA REPOSITORY SYSTEM STEERING COMMITTEE

Roger B. Brown **Chair**
Conoco, Inc.
Houston, TX

Thomas D. Coffman
Thomas D. Coffman Inc.
Austin, TX

W. K. Dirks
Shell Western E&P
Houston, TX

Randy S. Ewasko
Mobil Exploration & Production
Dallas, TX

William L. Fisher
University of Texas at Austin
Austin, TX

Michael C. Forrest
Maxus Energy Corporation
Dallas, TX

James A. Gibbs
Five States Energy Company
Dallas, TX

Robert J. Graebner
R.J. Graebner, Inc.
Dallas, TX

Robert D. Hatcher Jr.
University of Tennessee
Knoxville, TN

James E. Hooks
Chevron U.S.A. Production Co.
Houston, TX

T. M. LaHouse
Vastar Resources
Houston, TX

Jacob Loga
PGS Data Management
Houston, TX

Robert K. Merrill
UNOCAL
Sugar Land, TX

John H. Miers
Amoco Production Company
Houston, TX

Michael J. Padgett
Pennzoil Company
Houston, TX

Bobby J. Patrick
Seismic Exchange, Inc.
New Orleans, LA

Jonathan G. Price
Nevada Bureau of Mines & Geology
Reno, NV