

National Geosciences Data Repository System

Phase III: Implementation and Operation of the Repository

Final Report

Performed Under Grant No. DE-FG26-99BC-15115



**Submitted by the
American Geological Institute
to the
Office of Fossil Energy, U.S. Department of Energy**

October 2003

*American Geological Institute
4220 King Street, Alexandria VA 22302
Voice: (703) 379-2480 Fax: (703) 379-7563*

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, express 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 or any agency thereof.

ABSTRACT

The NGDRS has facilitated 85% of cores, cuttings, and other data identified available for transfer to the public sector. Over 12 million linear feet of cores and cuttings, in addition to large numbers of paleontological samples and are now available for public use. To date, with industry contributions for program operations and data transfers, the NGDRS project has realized a 6.5 to 1 return on investment to Department of Energy funds.

Large-scale transfers of seismic data have been evaluated, but based on the recommendation of the NGDRS steering committee, cores have been given priority because of the vast scale of the seismic data problem relative to the available funding. The rapidly changing industry conditions have required that the primary core and cuttings preservation strategy evolve as well. Additionally, the NGDRS clearinghouse is evaluating the viability of transferring seismic data covering the western shelf of the Florida Gulf Coast.

AGI remains actively involved in working to realize the vision of the National Research Council's report of geoscience data preservation. GeoTrek has been ported to Linux and MySQL, ensuring a purely open-source version of the software. This effort is key in ensuring long-term viability of the software so that it can continue basic operation regardless of specific funding levels. Work has been on a major revision of GeoTrek, using the open-source MapServer project and its related MapScript language. This effort will address a number of key technology issues that appear to be rising for 2003, including the discontinuation of the use of Java in future Microsoft operating systems. The recent donation of BPAmoco's Houston core facility to the Texas Bureau of Economic Geology has provided substantial short-term relief of the space constraints for public repository space.

TABLE OF CONTENTS

ABSTRACT.....	1
TABLE OF CONTENTS.....	2
TABLE OF FIGURES.....	3
INTRODUCTION.....	4
INTRODUCTION.....	4
EXECUTIVE SUMMARY.....	9
EXPERIMENTAL APPROACHES.....	10
RESULTS AND DISCUSSION.....	11
Summary of Phase I: Feasibility & Assessment Study.....	11
Summary of Phase II: Planning, Directory, & Pilot Studies.....	12
Summary of Phase III: Implementing the Repository System.....	14
Phase III – Operating the Repository System.....	15
NGDRS Steering Committee Meetings.....	17
Evaluating Available Repository Space.....	22
Establishing New Public Data Repository Space.....	23
Discussions with National Science Foundation.....	24
Discussions with DOSECC.....	25
Academic Data Preservation and Archiving.....	25
Unocal/Spirit Energy Cores and Cuttings.....	25
Chevron Cores and Cuttings.....	26
Altura Midland Core Facility.....	26
Phillips Seismic Tapes.....	26
Marathon Oil Cores and Cuttings.....	27
Texaco/Chevron Midland Proposal.....	27
Legacy Log Library Well Logs.....	27
Digital Data Transfer.....	27
Mobil Digital Seismic and Well Log data.....	27
Vastar/Arco Appalachian Seismic Data.....	28
ChevronTexaco West Coast United States Reflection Seismic Data.....	28
Hardcopy Data Media Conversion.....	28
Lockin Oil Company.....	29
Bryan Winberly Private Collection.....	29
James Thorne Private Collection.....	29
McMoRan Oil and Gas Company.....	29
Data Transfer Summary.....	29
Assessing the Return on Investment of the NGDRS.....	31
The Metadata Catalog.....	32
Initial Development and Deployment.....	32
Hardware.....	32
Software.....	33
Redesign of GeoTrek.....	33
Creating Metadata Catalogs.....	34
Bureau of Economic Geology.....	34
PGS and Fairfield - Databases Conversion.....	36

Minerals Management Service.....	36
Eastern Gulf Region PTTC	36
A2D - Conversion and Installation.....	37
Kansas Geological Survey (KGS).....	37
Oklahoma Geological Survey	38
Oregon Department of Geology and Mineral Industries.....	38
West Virginia Geological Survey	38
Utilizing the Metadata Repository.....	38
Business and Marketing Support.....	40
Internet-based marketing of GeoTrek	40
Marketing through Presentations	40
International Geoscience Data Repository Meeting.....	41
Print Media.....	41
Interaction with NASA’s Global Change Master Directory	41

TABLE OF FIGURES

Figure 1. Phase I Identified Available Data.....	11
Figure 2. Industry Core Repository Unit Cost Analysis.....	21
Figure 3. Summary results of 1998 data capacity survey	22
Figure 4. NGDRS Cores and Cuttings Transfers.....	29
Figure 5. NGDRS Well Log Transfers	30
Figure 6. NGDRS Paleontological Transfers.....	30
Figure 7. NGDRS Seismic Data Transfers	30
Figure 8. Public Data Integrated into the NGDRS	30
Figure 9. Rate of Data Transfer by Data Type.....	30
Figure 10. Data Transferred and Value under Auspices of the NGDRS	31
Figure 11. Assessment of the NGDRS Financial Performance.	32
Figure 12. GeoTrek Usage Statistics during Phase III.....	39
Figure 13. Marketing Presentation of the NGDRS Project.....	40

INTRODUCTION

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 "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 serves 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 contains critical data that 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 other 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 from 29 states to the university along with its warehouse and a \$1.3 million endowment to the university. All of these data entered the public domain for the first time.

The NGDRS project is highlighted in the 1997 *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."

Since 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. 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." Additionally, the goals and current effort of the NGDRS are further justified by the 1997 Report of the Energy Research and Development Panel, of The President's Committee of Advisors on Science and Technology (PCAST). This report recognizes that this effort, through the American Geological Institute and the geoscience societies, to preserve important scientific data and complementary efforts to archive core specimens and other geoscience data will contribute significantly to increased understanding from and use of a very large base of well-drilling experience.

Additionally, in response to the suggestion of the NGDRS steering committee, AGI, and the US Department of Energy, the National Research Council commissioned the study "Geoscience Data and Collections: National Resources in Peril" in 2000 and released a final report on the state of geoscience data preservation in April 2002. Based largely on a review of the survey data collected in Phase I of the NGDRS, the NRC report discovered substantial additional data resources at risk following the downturn in petroleum in 2000. The report made a call for establishing a system of regional public repositories to accommodate holdings that could not be placed at existing state facilities, as well as for direct increases in support from federal and private stakeholders to ensure the long-term preservation of the materials.

These endorsements add to the list of professional, industrial and state organizations that have gone on record in support of geoscience data preservation, including the 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 directly support the concept of 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 have downsized their domestic 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' files, warehouses, and repositories is in jeopardy of being lost due to the general decline in a business need and 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 will 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 was 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.

With the new world we face following the terrorist attacks of September 11, 2001, national security has again risen as a national priority. Energy security should follow closely on any broader initiatives at establishing an increased security for the country as a whole. Following the war in Iraq, much uncertainty remains regarding the direction and future of the dynamics of global energy supply, with a distinctive possibility for requiring the United States to focus on improving reliance on domestic hydrocarbon sources if global stability decreases. Through the efforts of establishing both method and mindset for geoscience data preservation, participants in the upstream petroleum sector can realize new opportunities to assess domestic sources for both new exploration and perhaps even more importantly, secondary recovery efforts in abandon 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, in the 1990's it 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, the cost for imported oil have amounted to more than 60 percent of our trade deficit. Between

1980 and 1995, the United States has 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. In the same year, the U.S. imported 3.2 billion barrels of crude oil and refined products for domestic consumption. 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, the petroleum industry has cut its funding for R&D by nearly half. At the same time, the federal government has substantially reduced its funding for the DOE Fossil Energy Program, particularly the upstream petroleum research accounts. 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 their less profitable 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 oil embargo and price shocks of the early 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, through the course of time were transferred to small companies and independent producers. This 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 and reservoir characterization. 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 assess the heterogeneity of complex reservoirs.

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 need to be 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 preserve and improve access to 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

- Supporting the needs of university-based research
- Planning for environmental protection
- Managing and evaluating water resources
- Reducing risks from earthquakes and other geological hazards
- Screening sites for municipal, toxic, and nuclear waste disposal
- Designing highways, bridges, dams, and utility lifelines
- Exploring and developing energy and mineral resources

There are many potential users of the system: universities, consultants, service companies, engineering companies, data brokers, government agencies, petroleum companies, lending institutions and the public at-large. In short, the NGDRS would benefit all elements of the geoscience profession, both private and public sectors, including the major oil and gas companies that would contribute much of the data.

EXECUTIVE SUMMARY

With increased oil prices in 2000 and early 2001, the NGDRS has seen a distinctive increase in activity and interest. This increase in activity is the result of increased activities in the petroleum sector, including new funding to examine infrastructure issues facing many of the companies over the long-term. Over the past several years, the petroleum industry has been focused on short-term issues and cost-savings. However, with increased activities and continued industry consolidation, longer time horizons have reemerged.

Despite a lack of available public repository space, the NGDRS has pressed ahead in coordinating transfers both to existing facilities and to virtualize some transfers, whereby previously private data is made public. This has resulted in the NGDRS attaining 85% of the targeted cores and cuttings transfers, with over 14 million linear feet of cores and cuttings being transferred to in the public sector. Additionally, large-scale transfers of seismic data have been evaluated, but based on the recommendation of the NGDRS steering committee, cores have been given priority because of the vast scale of the seismic data problem relative to the available funding. The rapidly changing industry conditions have required that the primary core and cuttings preservation strategy evolve as well.

Major efforts have been underway to facilitate the transfer of over 50,000 line-miles of 2-D reflection seismic data from ChevronTexaco to the public domain. AGI has been actively pursuing the effort to identify appropriate recipients, data conversion contractors, and funding sources to enable a successful transaction.

AGI remained actively involved in publicizing the National Research Council's report of Geoscience Data Preservation through meetings and presentations. The GeoTrek metadata catalog system continued to expand. Additionally, the metadata catalog is now served directly from AGI headquarters, and the software has been ported to Linux and all datasets transferred from Oracle to MySQL databases. Revision efforts are being retargeted given the identification of new open source technologies which will provide for a flexible, user-friendly GeoTrek, Version 3. This includes adopting MapServer and MapScript for future versions, and examining user-interface issues given the downward pressure on the use of Java in web browsers with the introduction of Internet Explorer 6 and Windows XP.

EXPERIMENTAL APPROACHES

The National Geoscience Data Repository System, Phase III is an operational project focused on coordinating and facilitating transfers of at-risk geoscience data from the private sector to the public domain. As such, the project does not have a consistent “experimental approach.” During the operation of Phase III, no efforts undertaken required experimental approaches to arrive at specific conclusions.

RESULTS AND DISCUSSION

The National Geoscience Data Repository System project has been a three phase effort, including an evaluation, a planning, and an implementation phase. Each phase built upon the findings and activities of the previous efforts. This report provides brief summaries of Phase I and Phase II, as well as the implementation component of Phase III. The primary results and discussion are focused on the second part of Phase III, namely the operation of the system.

Summary of Phase I: Feasibility & Assessment Study

Goals and Objectives

In 1994, the American Geological Institute completed Phase I of the NGDRS project to determine the feasibility of establishing the National Geoscience Data Repository System. The Phase I Report 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 I 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

Figure 1. Phase I Identified Available Data.

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 proposed rock core and cuttings contributions 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 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 environmental restrictions that place land off limits to 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 existing public domain geoscience data could also be made available through the NGDRS.

The Phase I 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 potential users of the repository system.

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.

Summary of Phase II: Planning, Directory, & Pilot Studies

The positive findings of the Phase I study led to the initiation of Phase II, which addressed specific organizational and operational requirements for establishing and implementing the NGDRS. Phase II began in the first quarter 1995 and was completed in April 1997. It was 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. Phase II consisted of three major components:

Planning and Specification: The planning and specification component of the Phase II study identified the requirements and operational business plan for implementing the NGDRS. The organization structure and appropriate oversight committees required for efficient operation of the NGDRS were defined.

National Directory of Geoscience Data Repositories: The directory was produced to determine what data are available in existing repositories before implementation of the NGDRS, where they are located, and how the data can be acquired. The directory was published in April, 1997 by the American Geological Institute and made available on the world-wide web in June 1998 at <http://www.agiweb.org/datadirectory>

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 pilot projects contributed critical information for the design and development of an efficient and cost-effective repository operations process for the implementation of the NGDRS. The results from the pilot projects yielded a working model for the NGDRS and specific plans for implementation of individual system

functions.

The NGDRS Model

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.
- Control costs of the repository system by sharing information and repository facilities, thereby taking advantage of economies of scale.
- Develop a centrally managed metadata catalog and a decentralized network of geographically dispersed data repositories.
- 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 industry standard data models to eliminate unwarranted data-formatting operations.

Implementing Clearinghouse, Data Transfer, and Cataloging Functions

One objective of Phase II was to establish a series of action plans to determine the cost and benefits, technical viability, and procedural requirements 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, data services vendors, data standards organizations, and project managers of similar repository projects worldwide.
- Workgroups, made up of data management specialists, 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 procedural aspects of data collection, preparation, loading, indexing/cataloging, and accessing.

Creating the Repository System

Phase II transformed the vision and strategies identified in Phase I into specific objectives for creating a repository system. The principal vehicle for doing so was the implementation of a pilot metadata catalog that permitted potential users and contributors to have a clearer understanding of the system's future capabilities. The pilot implementations supported:

- installation of more than twenty distributed data servers
- development of the GeoTrek data browser prototypes
- testing viability of the Internet as the wide-area network for the NGDRS.

In parallel with the technology-oriented activities, the project team continued their efforts to develop the 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 in Phase III.

Utilizing the Repository System

User requirements were tested via appropriate pilot projects, and demonstrations testing the viability of the GeoTrek software browser were conducted. The practical experience acquired during the installation and running of the pilot metadata catalog 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.

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 an NGDRS pilot project and direct contacts with private and public-sector repositories. 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.

Summary of Phase III: Implementing the Repository System

The first component of Phase III focuses on the implementation of the National Geoscience Data Repository System, particularly establishing the metadata catalog and initiation of data transfers under the auspices of the NGDRS. Data transfer issues are part of establishing the Clearinghouse Review Committee, in which a number of issues and actions have occurred related to overall NGDRS strategic planning, as well as specific data transfer targets. The metadata catalog was tested and installed and has been publicly accessible since May 1998.

Establishing the clearinghouse

A central role of the NGDRS was establishing and operating a clearinghouse function for establishing transfer models, coordinating actual data transfers, and setting priorities in data preservation. The clearinghouse operated largely as a dual function of day-to-day efforts by AGI staff under the guidance and review of the NGDRS Steering Committee. The Clearinghouse function formalized the model established by the Shell transfer to the Texas Bureau of Economic Geology of including facility and cash endowment with data transfers. The Clearinghouse function also established the virtual transfer of data by release data to the public domain access but retaining the data in private facilities as keeping the data in-place was substantially less expensive than moving the materials.

One major directive from the Steering Committee was for the NGDRS effort to focus on transfer of materials in the Gulf of Mexico region as these represent the greatest data holdings and the highest potential for short-term reuse by independent petroleum companies and as research analogs.

Another major initiative of the Clearinghouse was to evaluate additional space for public repositories, with the most intensive effort at examining transforming a hanger at the former Stapleton Airport in Denver into a central repository for otherwise implacable material. That effort

did not succeed as the combined effects of depressed oil prices hurt industry's funding ability and ongoing environmental concerns at the targeted site raised concerns of long-term liability and bonding issues with the City government.

Data Transfers

A number of major data transfers occurred during both parts of the Phase III effort. The details of these data transfers are documented in the body of this report. Data transfers occurred from the beginning with the Shell donation to the Texas Bureau of Economic Geology and ranged in scale from full facilities to small private collections. The precise business model of the data transfers did vary based on specific circumstances, but effort was made in all cases for the contributed data to be endowed by the donating organization.

Establishing the Metadata Catalog

The metadata catalog, GeoTrek, was developed during the first part of Phase III and released for public access in May 1998. The initial system design used a centralized Oracle database server in Houston, TX to host the metadata that participating organizations contributed data to, and was interfaced through a heavy-weight Java 1.1 applet. During the first 24 months of operation a number of user issues were identified, as well as barriers to adoption by additional public data repositories. A major effort of the second part of Phase III was to address these issues, as will be detailed in the body of the report.

Phase III – Operating the Repository System

Steering Committee and the Clearinghouse

The NGDRS Steering Committee is composed of representatives from industry, government, and academia. The committee provides overarching guidance on the direction and strategic operation of the NGDRS. Additionally, the Steering Committee represents a means of determining the potential candidate companies for transfer of data into the NGDRS given the changing economic conditions. Formal Steering Committee meetings were generally held in October of each year, from 1997 to 2002. These meetings represented opportunities for review of ongoing activities, evaluation of the current state of geoscience data preservation, and potential transfer targets and business models.

The Steering Committee, in cooperation with the AGI staff represents the Clearinghouse Review Committee manages all project aspects. Major issues were evaluated during the annual meetings of the Steering Committee, however, ad hoc meetings and subcommittees were established on an as-needed basis to address particular issues of concern, just as specific transfer targets or issues related to transfer business plan implementation. Additionally, the clearinghouse organization is the focus of data transfer coordination, including setting data priorities, identification of data contributors, and setting terms for the actual transfer of data. The members of the NGDRS Steering Committee are:

Wayne Ahr, Texas A&M University
Edith Allison, US Department of Energy, Liaison
Lee Allison, Kansas Geological Survey
David Archer, Petrochemical Open Software Consortium

Richard Bajura, National Research Center for Coal and Energy
Michael Baranovic, Shell Oil Company
Madelyn Bell, ExxonMobil
Ron Broadhead, New Mexico Bureau of Mines and Minerals
Marvin Carlson, Nebraska Geological Survey
Susan Cisco, Texas Railroad Commission
Paul Cutler, National Academy of Sciences
Peter D'Onfro, ConocoPhillips
John Dave, Unocal
Anthony DeSouza, National Academy of Sciences
Tammy Dickenson, National Academy of Sciences
Robert Finley, Illinois Geological Survey
Del Fortner, Bureau of Land Management
Richard Fritz, American Association of Petroleum Geologists
Robert Graebner, Texas Bureau of Economic Geology
Linda Gunderson, US Geological Survey, Liaison
James Handschy, ConocoPhillips
Allen Hittelmann, NOAA National Geophysical Data Center
William Hottman, Haliburton
Michael Hunt, Minerals Management Service
Tom Jordan, University of Southern California
Christopher Keane, American Geological Institute
William Kempner, ChevronTexaco
Rick Lahann, ConocoPhillips
William Lawson, US Department of Energy, Liaison
Ray Levy, EGI
Susan Longacre, ChevronTexaco
Walter Lynn, PGS
Charles Mankin, Oklahoma Geological Survey
Christopher Maples, Indiana University
Robert Merrill, Samson
Marcus Milling, American Geological Institute
Dennis Neilson, DOSECC
David Nicklin, Petroleum Development Associates
Michael Padgett, EEX
Don Paul, ChevronTexaco
Jon Price, Nevada Geological Survey
Doug Ratcliff, Texas Bureau of Economic Geology
Skip Rhodes, El Paso Energy
Peter Scholle, New Mexico Bureau of Mines and Minerals
Dan Smith, American Association of Petroleum Geologists
Robert Sneider, Sneider Exploration
Walter Snyder, National Science Foundation
Steven Stanley, Johns Hopkins University
John Stienmetz, Indiana Geological Survey
Jack Thomas, American Association of Petroleum Geologists

Ray Thomasson, Thomasson Exploration Associates
Scott Tinker, Texas Bureau of Economic Geology
Greg Wahlman, BP
Tony Walton, University of Kansas
Larry Woodfork, West Virginia Geological Survey (retired)

NGDRS Steering Committee Meetings

The first Steering Committee meeting was held on October 5, 1997 at Amoco's Houston offices. The meeting focused on identification of potential repository facilities and status of the GeoTrek Metadata Catalog. The committee recognized the need for a new repository facility because existing state and regional repositories are near capacity and can not accept the vast quantities of data available for contribution to the NGDRS. Two facilities for a central geoscience data repository were discussed — the Amoco Houston Core Repository located on Little York Road in northwest Houston and Hanger Four at Denver's former Stapleton Airport.

The GeoTrek Metadata Catalog was also demonstrated and reviewed. The Committee was supportive of the GeoTrek development effort. It recommended establishing two activities committees — one to address the needs for a repository facility and related contributions of cores and cuttings and a second to address opportunities for including additional service company databases on GeoTrek. John Deery, Amoco, chaired the Repository Core Subcommittee and Robert Merrill, Spirit of 76, chaired the GeoTrek Metadata Catalog Subcommittee. Each company was given an opportunity to discuss and review their data management issues and how they might be addressed through the NGDRS program.

The second NGDRS steering committee meeting was at the Unocal/Spirit Energy offices in Sugarland, Texas on October 1, 1998. Twenty-seven members of the steering committee attended the meeting, including representatives from 16 companies, the USGS, MMS, BLM, DOE, Texas BEG, and Texas Railroad Commission.

The meeting focused on identifying and addressing issues concerning transfer of data into the NGDRS, and particularly the proposal for establishing a central core facility at the former Stapleton Airport in Denver. Formal topics presented during the meeting were:

1. Overview and status report on the NGDRS project
2. Update on the NGDRS metadata catalog, GeoTrek™.
3. Overview of the proposed National Geoscience Data Repository and Research Center
4. Discussion of the letter of intent signed by AGI and Stapleton Development Corporation
5. Discussion of the appraisal of value of core and other geoscience data

From these items and the general discussion, a number of identified action items were identified.

1. The committee must focus on obtaining commitments for 2 million boxes of core in order to develop the endowment required to ensure that the central facility becomes a reality
2. Need to clarify the position of IHS (PI/Dwights) and Tobin for releasing the API number, Lat/Long and well numbers for GeoTrek use.

3. Companies should work with their vendors to encourage them to provide lat/long on all wells as part of the log header.
4. Participation of government agencies in the metadata catalog could bring into question whether they are competing with data providers. Susan Cisco of the Texas Railroad Commission plans to look into this issue.
5. A pro forma agreement and information packet must be developed and distributed to all of the companies to enable them to make data contribution commitments to the NGDRS.
6. Companies need a better understanding of the IRS's position on the deductibility of geoscience data to the NGDRS.
7. A standard listing contract and terms for vendors to list data on GeoTrek needs to be more widely circulated.
8. More commercial vendors are needed to participate on GeoTrek to establish a centralized, more efficient data shopping mechanism for data users. Company sponsors feel they can assist in encouraging vendors to join in support of the metadata catalog.
9. AGI needs to develop a GeoTrek user information packet for distribution to companies for their staff and other potential users of the system.

The third NGDRS steering committee meeting was held at the Chevron offices in Houston, Texas on March 2, 1999. Most of the major companies, as well as the USGS, MMS, BLM, DOE, Texas BEG, and Texas Railroad Commission were represented.

The meeting focused on the viability of establishing a central core facility at the former Stapleton Airport in Denver, and the overall priorities for the NGDRS given the downturn in crude oil prices.

Three central issues were discussed:

1. The viability of establishing an endowment to support the Stapleton initiative.
2. Defining viable near-term alternatives to a central facility.
3. Setting the NGDRS priorities in the new oil price climate.

Don Paul, Chevron Vice President, initiated the discussion about the viability of raising a \$10-12 million endowment for establishing a central facility at Stapleton. Given the current oil prices and the aggressive search for less expensive core storage options, a number of companies have lowered their core storage costs sufficiently that it is substantially more economical to continue to store core compared to either transfer to an NGDRS facility or disposal. A number of other companies agreed that the required \$4 per box contribution to an endowment was too high for the current climate, though they also recognized the proposed contributions reflect the real costs.

Marcus Milling also outlined the current status of negotiations with the Stapleton Development Corporation. At the time, AGI had established that the initial environmental liability assessment for the site was insufficient, and that the City of Denver would need to indemnify AGI for any current or future environmental liability. Additionally, the Continental Airlines lease on the facility was expiring, endangering the maintenance of the building if AGI did not move expeditiously to settlement.

The committee recognized that the need for additional storage capacity continues to exist, however, the capital required to establish the facility does not appear to be available at the current time. Given that, a recommendation was put forward to facilitate a near-term solution that would parallel and support the goals and objectives of the NGDRS mission. Bob Merrill (Spirit Energy) and Jan Van Sant (AGI Foundation) were asked to visit with C&M Storage in Schulemburg, Texas concerning use of their facility for core storage. It was proposed that AGI establish a joint operating agreement with C&M. Coupled with this initiative was a pledge by Unocal/Spirit Energy and Chevron to open substantial portions of their core holdings at C&M to the public domain and into the GeoTrek catalog until conditions improve to institute a transfer of core to the NGDRS.

As part of this effort, Unocal and Chevron have completed a substantial portion of the indexing of their holdings for inclusion into the NGDRS GeoTrek Metadata catalog. Likewise, other companies participating in the NGDRS with core held at C&M would need to catalog their holdings as well.

At the root of this initiative was a call by the steering committee for the NGDRS to broaden to promote the public access of geoscience data, and where needed, for preservation, the transfer of data into the public domain.

The 2000 and 2001 Steering Committee meetings were focused on discussing the ongoing National Research Council's report on Geoscience Data Preservation and the future directions needed to address the ongoing space shortage. The 2001 meeting was held October 12, 2001 at the ChevronTexaco facility in Bellaire, Texas and focused on new opportunities for data transfer in the area of 2D seismic. Additionally the first joint meeting between the NGDRS Steering Committee and the AAPG Committee on Core and Cutting Preservation was held during the AAPG Annual meeting in Denver in 2001. This broad community meeting provided an opportunity to expand the discussion and scope of the geoscience preservation effort and to recognize that publicity and verification of geoscience data's importance was critical in moving forward.

In 2002, two Steering Committee meetings were held. The first was a joint meet between the Steering Committee and the AAPG Committee on Core and Cutting Preservation at the AAPG Annual Meeting in Houston, Texas. A preliminary report from the NRC study panel was made and ongoing community initiatives were identified. In addition, discussion was held regarding the transfer of the BP facility in Houston to the Texas Bureau of Economic Geology.

The final meeting of the Steering Committee occurred on October 2, 2002 at the Houston Geoscience Research Center in Houston, Texas. This meeting was largely to analyze the success of the transfer of the former Amoco facility to the Texas Bureau of Economic Geology and how the long-term business model would work. In addition, discussions continued regarding efforts in seismic data transfers.

National Research Council Geoscience Data Preservation Panel and Report

Discussions began in April 1999 between the American Geological Institute and the Board on Earth Sciences and Resources staff of the National Academy of Science concerning holding a workshop on the issue of geoscience data preservation. The discussions focused on defining the scope, goals, outcomes, and timing of the workshop. The American Geological Institute encouraged the National

Academy of Science to conduct the workshop in the later part of 1999, with a primary focus on determining the priorities, optimal methods, and development of a national geoscience data preservation strategy. At that time, the NRC began soliciting for support of the effort and decided that a more comprehensive study was warranted.

The Board on Earth Sciences and Resources staff of the National Research Council (NRC) secured sufficient funding in the Fall 2000 to commence the study on the issue of geoscience data preservation. The study formally began in March 2001, with an initial meeting of the panelists on April 6, 2001. AGI worked closely with the NRC in developing its project scope:

With budget cuts and the downsizing of the U.S. oil industry and some federal agencies, combined with the lack of space in private and public museums, the preservation of geoscience data (e.g., cores, cuttings, maps, paper reports, digital data) is becoming a critical issue for federal agencies, academic researchers, museums, institutes and industry. This study will (1) develop a strategy for determining what geoscience, paleontological, petrophysical and engineering data to preserve; (2) examine options for long-term archival of these data; (3) examine 3-5 accession and repository case studies as examples of successes and failures; and (4) distinguish the roles of the public and private sectors in data preservation. The overall goal of the study is to develop a comprehensive strategy for managing geoscience data in the United States.

The NRC panel consisted of the following members:

- Christopher Maples, University of Indiana (**Panel Chair**)
- Beth Driver, National Imagery and Mapping Agency
- Robert Schafer, Kinross Gold
- Kevin Biddle, ExxonMobil
- Robert Sneider, Sneider Exploration
- Sally Zinke, Society for Exploration Geophysicists
- Thomas R. Janecek, Florida State University
- John Steinmetz, Indiana Geological Survey
- Linda R. Musser, Penn State University
- Warren Allmon, Paleontological Research Institute
- Donald D. Clark, City of Long Beach

The NRC panel report was released in April 2002, entitled *Geoscience Data and Collections – National Resource at Peril*. The report builds on AGI's previous analyses of the issues facing the geosciences and society regarding data preservation. In particular, the NRC committee also recognized a severe shortage of adequate repository space and the importance of quality indexing and the need for the national metadata catalog. In particular, the NRC report recommends the funding and operation of at least three regional geoscience data repositories, built onto existing repositories if possible. Likewise, the report recognized that the major long-term costs in data

preservation is the transport of cores, cuttings, and paper logs between facilities, and applying appropriate indexing to the data holdings. We recognize that the NGDRS program to date has been addressing most of the needs noted in the NRC report, and our steering committee has been evaluating the appropriate priorities of the effort relative the report's recommendations.

The Board on Earth Sciences and Resources has raised support for this study from various agencies and private organizations, including the US Department of Energy, US Geological Survey, National Science Foundation, Smithsonian, AAPG, POSC, and AGI.

Physical Data Transfer Efforts

One of the primary objectives of the NGDRS is the physical transfer of at-risk data from the private sector to public sector repositories. Given their high storage and curation costs, physical data such as core and cuttings are particularly at-risk for disposal. Therefore, under the guidance of the steering committee, the NGDRS has paid particular attention to the transfer of cores and cuttings. The transfer of the Shell Midland facility to the University of Texas in 1995 was the first major core transfer accomplished by the NGDRS. The identification of vast quantities of available data for transfer with limited existing repository space has required the NGDRS to identify and create new geoscience data repository space.

Based on an industry survey it is estimated that between 3.5 and 4 million boxes (~ 12 million feet) of core is available for transfer to the NGDRS in the next two to three years (Figure 1). Curation of this volume of material would require 250,000 square feet of high capacity storage space. Most existing state and regional repositories are filled near capacity and are accepting only material on a limited basis. To accommodate transfer of identified private sector cores and cuttings will require new repository capacity.

Company	Units (Boxes)	Operating Costs	Unit Cost
Amoco	850,000	\$785,000	\$0.93
UNOCAL	380,000	574,000	1.51
Shell	50,000	53,000	1.06
Phillips	144,000	190,000	1.32
Marathon	130,000	163,000	1.25
Mobil	180,000	195,000	1.08
Chevron	1,170,000	1,170,000	1.00
Texaco	60,000	72,000	1.20
BP	30,000	30,500	1.02
Exxon	350,000	500,000	1.43
Conoco	218,000	215,000	0.99
Total	3,562,000	\$4,426,000	\$1.24

From 1995 industry study by Amoco.

Figure 2. Industry Core Repository Unit Cost Analysis

Evaluating Available Repository Space

In October 1998 the American Geological Institute sent a survey form to all state geologic surveys in the United States in an effort to assess the true geoscience data repository capacity. Twenty-three surveys responded, including all of the states recognized as having the major public repositories. The results of the survey are shown in Figure 3. In general, most states hold cores and cuttings, while few curate digital seismic data. A total of 325,000 square feet of existing repository space was identified by the twenty-three state surveys responding to the questionnaire, and of that over 89% of the capacity is filled. Most states have standing policies to only accept material from their state, and in many cases, they are selective of this material as well. A few states, including Utah, Wisconsin, Illinois, and Oklahoma indicated that facility expansions are being considered. However, the quality of this expansion is variable, with some considering using unimproved metal containers and pole-barns in place of building a sustainable curation infrastructure.

The existing available quality repository space represents little more than 1% of the required capacity to handle the identified geoscience data available for transfer from the private sector to the public domain.

State	Core & Cuttings	Well Logs	Digital Seismic	Current Repository Size (ft²)	Available Repository Space (ft²)	Digital Catalog
North Dakota	Y	Y	N	18,000	7,200	Y
Oregon	Y	Y	N	600	120	N
Louisiana	Y	Y	N	1,600	0	Y
Florida	Y	Y	N	10,000	250	N
New Mexico	Y	Y	N	25,000	2,500	Y
Georgia	Y	Y	Y	10,000	5,000	N
Oklahoma	Y	Y	N	20,000	2,000	Y
New Jersey	Y	Y	Y	5,000	500	N
North Carolina	Y	Y	N	3,000	1,500	Y
Kansas	Y	Y	N	26,521	5,304	Y
Pennsylvania	Y	Y	N	5,300	2,120	N
Michigan	Y	Y	N	10,000	2,000	Y
Missouri	Y	Y	N	16,000	480	N
Mississippi	Y	Y	N	6,400	-1,280	Y
Utah	Y	Y	Y	14,000	5,600	Y
Wisconsin	Y	Y	N	29,500	0	N
Texas	Y	Y	Y	96,000	1,920	Y
California	N	N	N	0	0	N
Massachusetts	N	N	N	0	0	N
Nevada	Y	Y	Y	5,000	250	Y
South Dakota	Y	Y	N	6,000	1,200	N
Illinois	Y	Y	Y	17,500	350	N
Wyoming	N	N	N	0	0	N
Total	20	20	6	325,421	37,014	11

Figure 3. Summary results of 1998 data capacity survey

Establishing New Public Data Repository Space

The Phase II studies, confirmed in the 1998 AGI survey and 2001 NRC survey indicate that the volume of at-risk data available for transfer is far greater than the capacity of existing state and regional geoscience data repositories. Because of this situation, the NGDRS steering committee recognized the need to develop substantial new repository space. Following this lead in 1998, facilities were identified for potential transfer to the NGDRS as a central Geoscience Data Repository in Houston, Tulsa, and Denver. Based on negotiations concerning availability and suitability of these facilities, AGI focused on developing a facility in Denver at the former Stapleton Airport.

On August 29, 1998, Dr. Marcus Milling, Executive Director of the American Geological Institute and Mr. Richard Anderson, President of the Stapleton Development Corporation signed a Letter of Intent (LOI) to purchase the former Continental Aircraft Hanger and office complex located at 8250 Smith Road. The LOI established a framework to allow AGI to conduct the “due diligence” of the building and site to determine if the facility was suitable to serve as the NGDRRC. The LOI provided for 120 days due diligence to evaluate the facility and a 60-day period to negotiate a purchase agreement.

The due diligence period was instituted in September to pursue the requirements to gain a Certificate of Occupancy, assessing environmental liabilities, determine projected costs of building out the facility, and evaluate operating costs of the facility.

The initial results of the due diligence indicates that some modifications and improvements to the facility will be necessary, however AGI may be granted waivers in a number of areas:

Roofing: In the mid-1990’s Denver has received a number of very costly hailstorms. The proposed facility for the National Geoscience Data Repository and Research Center was heavily damaged in 1997. To determine the type, quality and identify areas of repair and the potential costs, AGI contracted with WeatherSure Systems. The report concludes that the East Hanger Section and the Office Section require repairs, but not immediate replacement, however the Mid and West Hanger sections require replacement before occupancy.

Environmental Assessment: Due to its location and prior function, assessing the potential environmental liabilities has been one of the most critical aspects the due diligence. Members of the AGI’s Executive Committee with extensive environmental experience reviewed the environmental reports provided by the Stapleton Development Corporation and the City and County of Denver and determined that without substantial liability releases from the City of Denver, that the long-term bonding and liability issues of the property make it uneconomical as a data repository.

Capital and Operations Costs: The capital costs for necessary improvements to the facility, including structural repairs and build-out, such as the installation of core racks are being assessed. Preliminary estimates project it will cost between \$.25-3 million to prepare the facility for operations.

Support for the development of the facility has been extensive. The companies represented on the NGDRS Steering Committee all expressed their interest in seeing a National Geoscience Data Repository and Research Center established. Even more importantly, local support in the Denver area, particularly from the local independent oil and gas producers, professional trade organizations, key state and local community leaders has been strong. Additionally, Philip Bradford, President of the Colorado Advanced Technology Institute pledged \$200,000 to cost-share the development of the facility.

AGI completed its due diligence of the Stapleton property in February 1999. This effort determined that substantial environmental concerns about the facility remained. Given the continued pressure on the petroleum industry from low oil prices, the steering committee also suggested that raising sufficient support for the needed endowment is not currently feasible.

Upon reviewing these circumstances, the American Geological Institute formally discontinued discussions with Stapleton Development Corporation on March 31, 1999 concerning the acquisition of the facility for use as a geoscience data repository and research center.

AGI also met with BP Amoco in 1999 concerning the possible transfer of their core research facility in Houston. However, BP Amoco expressed interest in only selling the property to a company to then operate the facility for BP Amoco on a lease basis. Though AGI stopped directly pursuing this option, as the mergers and cost pressures in the petroleum industry increased, BP became interested in striking a deal for converting their Houston data facility to the public sector.

In 2002, BP made a contribution of the former Amoco core repository on West Little York Road in Houston, Texas to the Texas Bureau of Economic Geology. With the transfer of the facility, the Texas BEG has also been provided additional land, rights to the proceeds from adjacent land sales, an initial endowment investment, along with 400,000 boxes of non-proprietary cores and cuttings previously held at that facility. With the addition of this substantial volume of new space, the severe space shortage has been temporarily relieved within the Gulf Coast region, and in particular, within Texas. BP is in similar negotiations regarding other former Amoco facilities which may further alleviate the immediate space issues for core storage.

Discussions with National Science Foundation

Marcus Milling, David Applegate, and Christopher Keane, all of the American Geological Institute met with Robert Corell of the National Science Foundation to discuss any interest NSF may have in supporting the development of a central geoscience data repositories. Robert Corell detailed the requirements which the National Science Foundation would need in order to consider support for a central geoscience data repository. In particular, NSF would like to see a consortium of Universities be developed, which would be the base of developing NSF support. NSF does recognize the need for geoscience data curation, particularly as a number of major NSF programs generate cores and cuttings and are required by grant conditions to preserve the data. Corell recommended that AGI open discussions with the Drilling, Observation and Sampling of Earth's Continental Crust, Incorporated (DOSECC), an NSF supported consortium of 48 Universities and

National Laboratories. DOSECC has and is continuing to produce new cores and cuttings, all in need of curation.

Discussions with DOSECC

Discussions with DOSECC began in May, 1999. DOSECC is a consortium of 48 universities and research laboratories that are engaged in research on onshore crustal studies and drilling techniques. Given DOSECC's interest in onshore cores, AGI made contact with their Executive Director, Dennis Neilson.

DOSECC currently has two major operations underway, drilling 5000 meters of core from the flank of Mauna Kea and deploying a mobile floating drill rig for coring of lake bottoms, such as the Great Salt Lake. DOSECC recognizes the long term scientific core preservation issues and recognizes that all projects face similar circumstances in being unable to find data repositories willing to accept the core for curation. This situation represents a potential point of collaboration.

With their focus on core and equipment, DOSECC has found itself with an immediate need for storage space. DOSECC, in communication with AGI, contacted the agent for the former Toole Army Depot west of Salt Lake City to inquire about potential storage space. At this point the property prices are too high to be viable for acquisition by either DOSECC or the NGDRS. However, DOSECC is leasing a smaller lot within the same property now for storing cores and equipment in sea containers.

DOSECC was a major discussion point of the NRC report of geoscience data preservation, representing a clear example of the need for NSF to establish formal guidelines and support for long-term curation of research samples. At this time, NSF is considering supporting the new Houston Geoscience Research Center as a center for storing NSF research funded samples, such as those of DOSECC and other drilling and non-cold region sampling programs.

Academic Data Preservation and Archiving

The NGDRS program was represented at an NSF-sponsored workshop held in Bloomington, Indiana in January 2003. The National Science Foundation convened the meeting to discuss the state of data preservation in academia and for NSF-funded research. The workshop was attended by approximately 30 people, mostly from academic departments and university museums. The workshop determined that data archiving and preservation is an unfunded mandate that is very poorly executed within NSF-funded programs. The primary recommendation was for NSF to support a national repository where NSF-funded programs could deposit their data collections once support ended and the data was deemed of future value to science. From this, the Texas Bureau of Economic Geology with encouragement from the NGDRS program submitted a proposal to NSF to perform this function at their new Houston Geoscience Research Center

Unocal/Spirit Energy Cores and Cuttings

Unocal's onshore cores and cuttings are now cataloged in the NGDRS metadata catalog. The metadata is house at the American Geological Institute in its database servers. This transfer covered

2082 core records, representing 1,109,016 linear feet of core from across the nation. Quality control by AGI allowed the inclusion of 1198 core records into the metadata catalog. Additional work on the data has determined the geolocating of the additional 884 core records is not possible given the existing metadata. However, the data is included in the system for those queries not dependent on geographic location. Unocal continues to store their core and cutting holdings at C&M Storage in Schulemberg, Texas. Users of the GeoTrek metadata catalog can arrange for access to listed cores by contacting C&M Storage directly.

The NGDRS is assisting in the transfer of Unocal's Utah cores and cuttings from Shulenburg, Texas to the new core repository at the Utah Geological Survey. Unocal has made as a condition of this transfer, that all of Utah's data holdings, including the Unocal data, need to be listed in GeoTrek. In March 2002, the Utah Geological Survey sent a copy of its metadata to the American Geological Institute for review for integration into GeoTrek. The data was found suitable and coordination of the actual transfer is being discussed between Unocal and the State of Utah.

Chevron Cores and Cuttings

Chevron transferred its metadata catalog of over 180,000 core and cuttings records to the NGDRS. These records represent 934,157 feet of cores, over 10M feet of cuttings, 14M washed paleo sample bags, 41,942 paleontology slides, and 56,621 oil samples.

Similar to the arrangement by Unocal, Chevron is maintaining the cores and cuttings at the C&M Storage facility. However, all of the cores and cuttings in the database are now released to the NGDRS for inclusion in the metadata catalog. The data is not extensively geocoded, so translations from TRS coordinates to latitude and longitude needs to be performed. Full integration of the data into the metadata catalog, including geolocation of the records is an ongoing effort, however the database is fully searchable within GeoTrek except for the geocoded component.

Altura Midland Core Facility

Altura has transferred ownership of some 85,000 boxes of core and cuttings to the Bureau of Economic Geology at the University of Texas at Austin in 2000. The construction of a new repository in Midland was completed and physical movement of core and cuttings boxes began in October 2000. The metadata records for the Altura core was processed and integrated into the BEG's metadata catalog under direction of the NGDRS. The consolidated BEG catalog has been integrated into GeoTrek.

Phillips Seismic Tapes

Phillips Petroleum has transferred selected seismic holdings for the Santa Barbara Channel in California to the NGDRS. AGI completed a pilot project to evaluate the feasibility and costs for digitizing and transcribing the analog data to current media and format. The data was stored on 1-inch analog tapes, for which there are few known working readers. A selected number of tapes, representative of the Santa Barbara channel were transcribed and processed to check for validity. The processed seismic lines demonstrated excellent quality and provide a new set of data for the

geoscience community to use. Copies of the tapes are available on request for private sector and academic researchers from the Texas Bureau of Economic Geology.

Marathon Oil Cores and Cuttings

Marathon Oil approached AGI concerning the contribution of their cores and cuttings to the NGDRS from the Littleton, CO facility. That facility has been slated for closure, and their Permian Basin cores and cuttings were also in danger of being discarded. After initial discussions with Marathon representatives, the company decided to contract with C&M Storage to hold and maintain their data holdings in Schulenberg, TX. Discussions are ongoing regarding the incorporation of non-proprietary holdings of Marathon into GeoTrek using similar arrangements as those with Chevron and Unocal. Approximately 100,000 boxes of core are at issue in these discussions.

Texaco/Chevron Midland Proposal

Texaco's Midland operations had approached AGI concerning the development of a Permian Basin-wide core facility, including identification of non-proprietary cores and cuttings for release to the public. Following a meeting in October 2000, both Texaco and Chevron indicated a strong desire to move forward expeditiously with this process. However, with the merger between Texaco and Chevron no further activity has occurred on this effort. At this time, AGI does not expect any activity regarding this issue in the near future.

Legacy Log Library Well Logs

The Legacy Log Library has contributed the well log records for the Texas Railroad Commission Districts 1-6 to the NGDRS and are now being integrated into the well log records of the Texas Bureau of Economic Geology. As these logs are integrated into the Bureau metadata catalog, they are available via GeoTrek.

Digital Data Transfer

Large quantities of digital geoscience data are also held by private industry. Like cores and cuttings, this digital geoscience data represents decades of data collections previously not available to the wider geoscience community. Additionally, storage costs for the tapes this data is stored on continues to increase, as does the concern that media degradation may make much of the data unrecoverable in the near future. To address these concerns, the NGDRS has attempted to coordinate the transfer and transcription of contributed digital data to new media and to be placed at public repositories around the country. Estimates made during Phase II indicate that millions of 9-track tapes, mostly reflection seismic surveys, are available for transfer.

Mobil Digital Seismic and Well Log data

Mobil E&P proposed the transfer of over 200,000 digital seismic survey tapes, representing nearly 4 million line-miles of seismic coverage, to the NGDRS. In November 1998, Mobil approached AGI to transfer not only seismic tapes, but also over 14,000 digital well logs. The transfer of these

digital data into the NGDRS, and thus the public domain, would have represented an introduction of an enormous data holding previously unavailable to the wider geoscience community.

Though substantial effort was made to coordinate the transfer of a substantial portion of Mobil's domestic digital seismic data holdings to the NGDRS, the merger with Exxon coupled with fundamental funding issues at executing such a transfer indefinitely delayed these discussions. Additional issues that will need to be determined by ExxonMobil for future action on these holdings is determining which tapes are eligible for transfer, including whether the data still has proprietary value, whether ExxonMobil has total ownership so that they can initiate the transfer, and if not, can they arrange with co-owners for permission to transfer the data.

Vastar/Arco Appalachian Seismic Data

Vastar verbally agreed to contribute over 6000 digital seismic reflection field tapes to the NGDRS. The collection represents some 700 seismic lines extending along the Appalachian Thrust Belt from Southern New York State to Northern Alabama. The 9-track seismic tapes were to be transcribed to new high-density tape media prior to transfer to the NGDRS. However, a change in management at Vastar and the subsequent acquisition of Arco by BP has indefinitely delayed the transfer. The NGDRS staff understands that parts of this data set were given to the University of Alabama and the University of Tennessee, but the bulk of the data was sold to a seismic vendor.

ChevronTexaco West Coast United States Reflection Seismic Data

AGI and the USGS were approached by ChevronTexaco concerning the transfer of 50,000 line-miles of reflection seismic data from the offshore of the US West coast. This data is at risk for disposal in the next couple years and it is believed to be of significant research value. Christopher Keane of AGI made a presentation on this data and the NGDRS efforts to preserve it at the EarthScope meeting in Denver on March 3, 2003. From that meeting a major effort was launched to identify the science problems that could be most immediately addressed using the data – a technique to facilitate funding from NSF.

AGI assessed the costs and the operations model for transferring the data from ChevronTexaco into the public domain. AGI determined that included transport, baking, and transcription to modern media, the 50,000 line-miles will cost approximately \$400,000 to transfer into the public domain. AGI had built commitments for funding the data transcription and transfer from federal agencies and industry partners when operational staff at ChevronTexaco decided that they would like to pursue the effort with the US Geological Survey as lead. The US Geological Survey has used the funding and business model developed by AGI to solicit transcription bids for this effort. Currently the US Geological Survey does not have funding secured to conduct this project, and funding partners AGI had secured have declined to support the US Geological Survey's activities.

Hardcopy Data Media Conversion

Hardcopy, or paper-based data records remain prevalent throughout the geosciences, particularly for datasets of historical importance and value. Hardcopy represents a unique challenge for the NGDRS in that it requires identification of receiving repository space as well as intensive indexing

to make the material accessible. Several hardcopy data transfer occurred in Phase III facilitated by AGI, involving the transfer of hardcopy data to the Texas Bureau of Economic Geology. The long-term intent for the hardcopy is to digitize the materials and then dispose of the original hardcopy. However, given the greater data transfer priority to cores, cuttings, and digital seismic and well log data, contributed hardcopy data is being transferred to repositories for file storage.

Lockin Oil Company

Locklin Oil Company of Tyler, Texas contributed 9,000 well logs and 40,000 PI scout tickets to the Texas Bureau of Economic Geology in April 1998, through the coordination of the NGDRS. Estimated replacement cost of these data is \$150,000.

Bryan Winberly Private Collection

In January 1998, Bryan Winberly, Midland, Texas contributed six file-boxes and two map-boxes containing reports, maps, well logs, and scout information concerning the Gulf Coast area. The value of these data has not been assessed.

James Thorne Private Collection

James Thorne of Pleasanton, Texas transferred a total of 20,000 well logs to the Texas Bureau of Economic Geology through the NGDRS. These data were primarily from the Texas Railroad Commission District #1 area. Estimated value of these contributions, which contained many logs that had not been previously released, is \$80,000.

McMoRan Oil and Gas Company

AGI is current negotiating with McMoRan Oil and Gas for transfer of 12,000 well logs from the states of Mississippi, Oklahoma, Colorado, New Mexico, Montana, North Dakota, Michigan, and Kansas. Additionally, McMoRan transferred 350 side-cores from the Gulf of Mexico and Atlantic Offshore to the NGDRS.

Data Transfer Summary

Source	Liner Ft. of Cores & cuttings
Unocal	1,109,016
Chevron (Cores)	934,157
Chevron (Cuttings)	10,038,898
Shell	1,350,000
BP/Amoco	1,200,000
Altura	255,000
Total To Date	14,887,071

Figure 4. NGDRS Cores and Cuttings Transfers

Source	Well Logs
Legacy Well Log Library	180,000

Figure 5. NGDRS Well Log Transfers

Source	Section Equivalent (ft)
Chevron	43,200,000

Figure 6. NGDRS Paleontological Transfers

Source	Line-miles
Phillips	2000

Figure 7. NGDRS Seismic Data Transfers

Source	Boxes/Logs
Texas BEG cores	100,656
Texas BEG well logs	87,772
Texas RRC logs	552,524
Alabama cores	1,091
Oklahoma cores	4,604
MMS logs	44,455
US Geological Survey cores	370,000
Total	1,161,102

Figure 8. Public Data Integrated into the NGDRS

The NGDRS project has successfully managed the transfer of nearly three-quarters of the volume of identified at-risk cores and cuttings. The NGDRS Steering Committee gave priority to cores and cutting data given their particular economic risk, and thus to date, that data type has been the primary transfer target. A test case regarding seismic transfers was performed, but given the vast volumes of data that needs to be transferred and converted, current funding levels preclude major initiatives into that area.

Data Type	Phase I Target	Phase III Transfers	Percent Completion
Cores & Cuttings	17.5 M liner feet	14.9 M liner feet	85%
Seismic Data	100 M line-miles	961 line-miles	<1%
Paleo Data	Not quantified	43.2 M section ft.	>100%

Figure 9. Rate of Data Transfer by Data Type

Assessing the Return on Investment of the NGDRS

The American Geological Institute has performed a basic assessment of the financial dynamics of geoscience data, including the value of contributions, either in data or financial support, by the various stakeholders in the NGDRS. This effort is greatly facilitated by the recent survey of minimum total available geoscience data conducted by the National Research Council Committee of the Preservation of Geoscience Data and Collections.

Data Type	Min. Total Volume	At-Risk Data Volume	Data Transferred	Total National Value of Data	Value of Transferred Data	Percent of At-Risk Data Trans.	Percentage of Total Data Trans.
Cores & Cuttings	55M ft	17.5 M ft	14.9M ft	\$73M	\$19.9M	85%	27%
Seismic Data	357M line-miles	100M line-miles	961 line-miles	\$3.57B	\$960,000	<1%	<1%
Well Logs	46M logs	7.1M logs	530,000 logs	\$184M	\$2.1M	7.5%	1%

Figure 10. Data Transferred and Value under Auspices of the NGDRS

The National Geoscience Data Repository System received \$3.8 million in financial support from the US Department of Energy, \$2.87 million in financial support from industry, and \$22.3 million in base-line valued data contributions. Additionally, financial and facility contributions to the Texas Bureau of Economic Geology of \$14 million in 1996 and 2002 transfers from Shell and BP also represents a derivative of the NGDRS project. The total cost-share investment by industry of \$25.2 million represents over a 6.5 to 1 return on investment by the US Department of Energy, even without attempting to place future use value on the data or attributing the University of Texas and University of Oklahoma transactions. When including these facility transactions, the total derivative NGDRS contribution by Industry becomes \$58 million, with a return on Department of Energy investment of over 15:1.

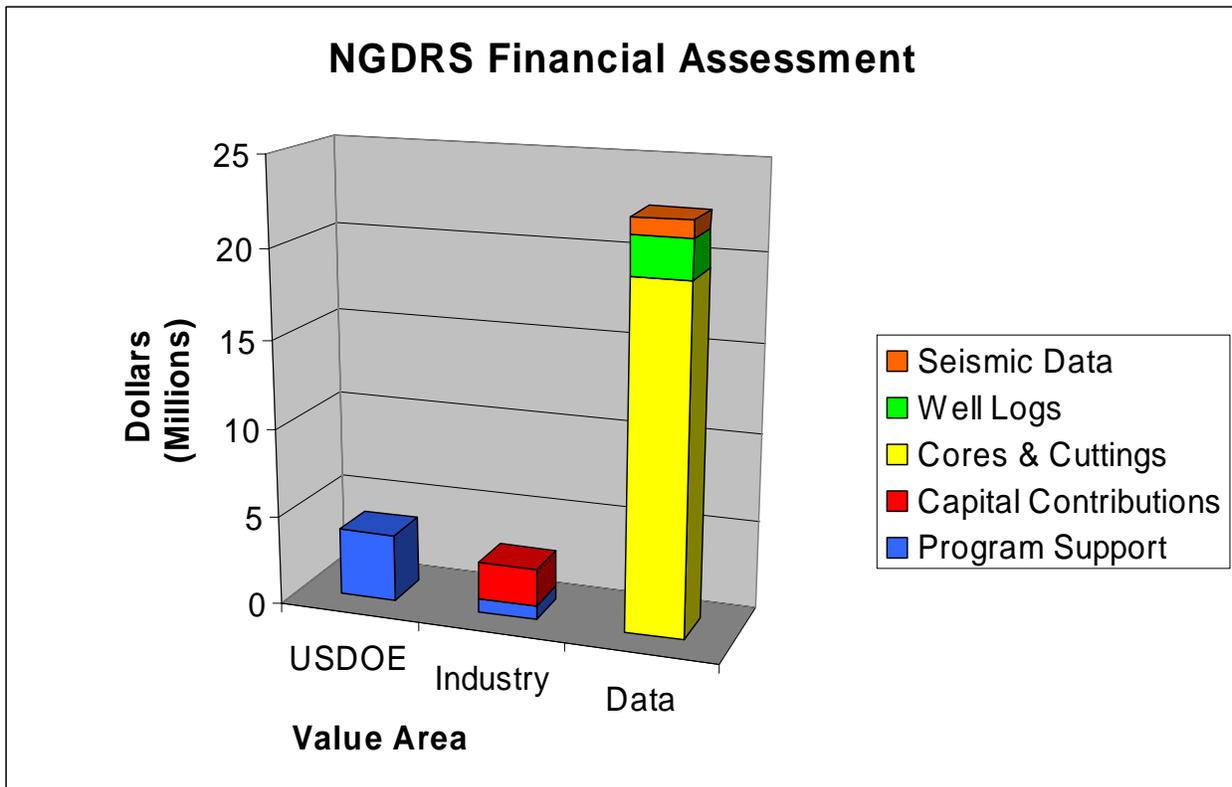


Figure 11. Assessment of the NGDRS Financial Performance.

The Metadata Catalog

Initial Development and Deployment

The basic infrastructure for establishing the metadata catalog included the installation of hardware and software development and testing. Primary hardware installation occurred in September 1997 and most of the final software development and testing occurred in late 1997 and early 1998.

Hardware

The necessary rental hardware for the central metadata catalog system was installed in the offices of the project's Houston-based consultants. The hardware consisted of Sun Ultra II and 3000 servers, DLT tape unit for backups and storage array for data storage.

A back-up system was configured at AGI headquarters in Alexandria, Virginia. The main web pages for the NGDRS are hosted at this site, and an installation of GeoTrek, including a mirror of the databases held in Houston were available through the back-up system. This system will be available in cases where access to the NGDRS metadata catalog in Houston is unavailable from the rest of the Internet. A leased-line connection was established to AGI headquarters, and two Intel-based Linux servers were installed.

Software

The GeoTrek server software is stable and has been deployed in production environments. The browser software is also stable, especially on Windows 98, Windows NT 4.0, Windows XP, Solaris 2.4, and Linux 2.4.

The software underwent extensive testing by a variety of users in late 1997 and early 1998. A wide range of users, both in discipline interest and computer skill levels, were enlisted to test the NGDRS metadata catalog system. The beta testers were:

Person	Organization
Edith Allison	DOE, Washington, DC
Madelyn Bell	Mobil, Houston
Mary Blount	MMS, New Orleans
Todd M Boyd	NASA Maryland
Ronald Brignac	MMS, New Orleans
George Dellagiarino	MMS, Virginia
John Deery	Amoco, Houston
Shawn Devlin	Viking Resources-Kansas
Eric Hatleberg	POSC, Houston
Allen Hittelman	NOAA, Denver
Christopher Keane	AGI, Virginia
Shinji Kanai	JNOC, Japan
Ben Lin	Unocal, Houston
Jacob Loga	PGS, Houston
Jenny Meader	POSC, Houston
Robert K. Merrill	Spirit Energy, Houston
Chandra Nautiyal	DOE, Tulsa
Esteban Quijano	SECTEC, Mexico
Gary Stone	Five States Energy
Sherilyn Williams-Stroud	USGS, Denver

The metadata catalog was released to the public on May 14, 1998. During 1998 and 1999 it underwent minor software improvements. The beta testing effort was of high value and resulted in identification of several critical issues before the software was released.

Redesign of GeoTrek

GeoTrek, the metadata catalog for the NGDRS was redesigned in 2001 based on extensive user input related to usability, both by clients and data repositories. This redesign revolved around 2 areas: backend remote database integration and an improved user interface.

The reliance on Oracle database systems and the Solaris operating system became a major deployment hurdle for public repositories as the licensing and support fees for these systems were not compatible with their budget for the initiative. Additionally, the long-term goal was that repositories host their own databases and GeoTrek would query each database independently. The

reliability and scalability of the Internet environment in 1997 precluded adoption of this option during initial development, but by 2001 the infrastructure was sufficiently mature to pursue this effort.

To address the cost issue of support GeoTrek, the entirety of the GeoTrek code was ported to Linux and interfaced with a number of Open Source database systems in August 2001. This effort was targeted to ensure the greatest viability of the metadata catalog regardless of funding levels. This has allowed GeoTrek to now be run on lower cost servers and against a wider range of databases, including several without annual service licenses. It is expected that these arrangements will assist in bringing additional public data holders into the NGDRS fold. As a result, GeoTrek can now support numerous backend databases, including Oracle, MySQL, PostgreSQL, Sybase, and MS SQLServer.

Additionally, the system now is capable of querying remote database systems over the Internet, allowing data holders to control the availability and extent of data accessible through the metadata catalog. Additionally, it allows transparent integration of additional information, such as core disposition, core photos, analytical data, etc., to be readily accessible by the end-user.

A new user interface system has been identified in the Open Source community that will neatly tie into the GeoTrek metadata catalog. An effort was made to examine how to convert the GeoTrek front end to use MapServer technology. However, the decrease in user interaction capability was reviewed as a major drawback, and preference was maintained for the original Java-based interface.

A major obstacle facing GeoTrek in the future is the decline in support for Java “out of the box” on Microsoft platforms. Given the reliance of GeoTrek’s front-end to Java, as well as MapServer’s most interactive front-end is based on Java, concern is raised about the need to explore other strategies for user interfaces. In general, AGI expects the lack of default Java support to be a major hurdle a number of users do not have dedicated IT staffs to address these issues.

Creating Metadata Catalogs

The real value of the metadata catalog is the underlying information. A major effort was made to populate the GeoTrek system with metadata from both participating repositories and cooperating commercial vendors, in an effort to demonstrate the viability of the system to enable the marketing of NGDRS participation to a wider range of repositories and users. During the process of identifying candidates for joining the NGDRS metadata catalog in the first part of Phase III, DOE (Mike Ray) recommended that coverage be focused on the Gulf Coast region. From this point, additional data sets from a wide variety of localities would then be added, expanding the geographic and topical scope of the system.

Bureau of Economic Geology

The Bureau of Economic Geology has been the primary test case for integration of participating repositories into the NGDRS and the GeoTrek metadata catalog. The BEG’s core and well log data is being continuously indexed and added to the NGDRS metadata catalog. The BEG, along with

the NGDRS management committee has been key in identifying operational and organization issues with the system, leading to ongoing improvements in the system.

An installation of GeoTrek at the BEG-Austin was completed on December 12, 1997. The then-current BEG Core Repository data and Geophysical Log data were loaded into an Oracle database and installed GeoTrek was installed as a browser for this data. Minor compatibility issues were encountered with the Oracle Web Server software but were solved during the installation process.

The system was running under Unix (Solaris operating system) using Oracle Web (HTML) Server (v3.0). The BEG tested the system in December 1997 and January 1998 with various combinations of platforms and browsers. As part of the testing, they allowed users from the Core Research Centers in Austin and Midland to access the BEG database through the system. The BEG compiled the feedback they received from their users during this testing period.

A review of the Bureau of Economic Geology (BEG) staff experiences using the GeoTrek system was performed in February 1998. As part of the testing process, no tutorial or training sessions were provided to the BEG staff. This was considered a key test of the intuitiveness of the GeoTrek system. The feedback focused largely on technical hardware and software issues. In general, the databases system worked as expected. However, the BEG is still dominantly a Macintosh-based organization, and the Macintosh Java virtual machine implementations are not stable enough to run the GeoTrek browser consistently. However, access via PC's and Unix boxes worked as expected. The BEG is currently evaluating their desktop computing environment and are considering moving to PCs.

In 1999, the BEG determined the ongoing costs of the system were beyond their financial means and their local system was decommissioned with BEG data being hosted in Houston, TX at a service provider. During 1999 and 2000, the BEG redesigned their data management systems and adopted the use of MS SQLServer. Based on this effort and the integration of their data on the new servers, support for SQLServer became a key item for the GeoTrek backend.

The BEG received funding from AGI to develop a core inventory control application to work hand-in-hand with GeoTrek. This allows dynamic inventory integration into the end-user experience of GeoTrek and allows end users greater understanding of the disposition of their targeted cores and cuttings.

The BEG requested funding in early 1998 to complete the geocoding of all well locations in Texas with latitude/longitude. Wells within each county are not uniquely locatable. Therefore, in GeoTrek, all wells are shown at the geographic center of their county. The geocoding provides a latitude/longitude for each well. Under this contract with AGI through the NGDRS program, the Bureau of Economic Geology is geocoding the well log and core location information. The BEG has contributed \$128,564 of like-in-kind support for the NGDRS by covering a share of the labor expenses for this initiative. Using information from the Texas Railroad Commission, the actual locations of the wells from which these data were acquired are being geocoding. The updated locations will be added to the GeoTrek installations both in the NGDRS and at the BEG.

An additional database being considered for use in the BEG installation of GeoTrek is a geographically indexed version of the BEG publications catalog. Most of the BEG's publications are geographically oriented and can be fixed to a location or region. This would allow users to geographically browse an area and see what publications are available.

PGS and Fairfield - Databases Conversion

Loading of PGS Gulf of Mexico seismic surveys database was completed in August 1997. Several additional meetings were held to review data quality and other technical details. Both clients (PGS, Fairfield) requested that their data sets not be publicly displayed until after the 1997 SEG annual meeting. Following the meeting, Fairfield's metadata was made public. However, PGS delayed release of their metadata, citing a desire to further populate their data set on the NGDRS before release. During discussion on the revision of GeoTrek in 1999, it was decided that commercial data listings would be discontinued in 2001.

Minerals Management Service

All MMS Gulf of Mexico (GOM) well-wellbore data sets, consisting of approximately 35,000 wellbore entries, were loaded in the NGDRS metadata catalog. The data set was also "tiled" to allow for the loading of the data displayed on the working area of GeoTrek. The tiling of the data shows a summary of the wellbores available for the each GOM block. GeoTrek queries can be applied for further filtering the data set. This allows users to develop, for example, a tiled map for the GOM containing all wells owned by company XYZ.

The field descriptions for the MMS databases have been edited to improve their intuitiveness. This activity was completed before the public release of the metadata catalog, and additional revisions may be made depending upon feedback from the system's users.

Eastern Gulf Region PTTC

The original data set deals with data from the Mississippi Interior Salt Basin and is available at the web site <<http://egrpttc.geo.ua.edu/pttc/reports/smacko~1/contents.html>>. The set is part of a report aimed at providing the information obtained from a geological study of the Mississippi Interior Salt Basin. The work focused on inventorying the data files and records of the major information repositories in the Northeastern Gulf of Mexico and making these inventories easily accessible in an electronic format. The study is part of the Advanced Geologic Basin Analysis Program promoted by the DOE. The program designed to provide an avenue for studying and evaluating sedimentary basins and to improve the efficiency of the discovery of the nation's remaining undiscovered oil resources.

The following actions were required to integrate the database:

- Export of the original files from HTML to ASCII format.
- Merging of the original 15 data sets that were split by county into one single data set.
- Identification of the data schema, unique keys and data issues. The following is a summary of our findings:

- The unique data set key is the Permit Number, not the API number. The same API number appears repeated on different permits.
- There are approximately 1,200 permit numbers (rows in the table).
- About 3% of the 1,200 entries did not have Lat/Long data.
- The data set is a fairly rich metadata data set containing information on what well logs were taken, what kind of samples are available, well status, etc.

A2D - Conversion and Installation

A2D, a Houston-based commercial digital well log vendor installed their well data on the GeoTrek site. During January 1998, the Information Store worked with Ray Kivimaki, A2D's technical consultant, to discuss the data elements they wish to display and procedures for data installation and refreshing their well log dataset. The A2D data was successfully loaded during February 1998. The initial datasets cover the Gulf of Mexico - both state waters and continental shelf. Procedures for ordering were developed and were reviewed and tested by A2D staff. Like with the seismic data, listing of commercial vendors was suspended in 2001.

Kansas Geological Survey (KGS)

Susan Bolton, from the Geoscience Data Resource Management group, organized a meeting on September 25, 1997 at KGS headquarters in Lawrence. The meeting agenda included several items: an overview of the NGDRS project, a demo of the GeoTrek browser, a discussion of the current status of KGS projects, and scoping a preliminary plan of action for KGS's participation in the NGDRS. Some general comments follow:

- The Kansas Geological Survey director for Information Services felt that the NGDRS and GeoTrek are of value to KGS.
- The following is a preliminary, prioritized list of databases identified as the potential targets for installation in the NGDRS metadata catalog: electrical logs, well cuttings, plugged wells, cores, scout tickets, production data, list of operators, cultural data, and other geophysical (magnetic and gravity).
- Most of the KGS data is identified using the Township/Range/Section (TRS) coordinate system which need to be converted to Lat/Long.
- Most of their clients are majors or independents typically not located in the state of Kansas, more commonly in Houston or Calgary. Remote access to Kansas data for all operators was acknowledged to be very important.
- Beyond the Kansas Geological Survey, an officer of the Kansas Geological Society believes their log database (about 2,000 wells) is a good candidate for the NGDRS. The Society is a for-profit organization.
- Most of KGS databases do not follow a standard data model. In the future they would like to take advantage of a standard E&P data model and they are leaning towards the PPDM data model.
- KGS indicated that they are *definitely* interested in participating in GeoTrek, probably when the conversion of their databases is completed.

Oklahoma Geological Survey

The Oklahoma Geological Survey core data has been successfully converted and installed on the NGDRS metadata catalog. The BLM's TRS2LL software initially believed to handle the conversion between T/R/S to latitude/longitude did not handle Oklahoma. However, Michael Schmidt (Deputy Director of Oil and Gas Conversation Division, Oklahoma Corporation Commission) indicated that a new service was being provided by the University of Oklahoma called "Spatial Calculator." This service was used, for a fee, to calculate the latitude and longitude based on the legal descriptions provided in the data.

Oregon Department of Geology and Mineral Industries

Discussions occurred between the Oil and Gas group of the Oregon Department of Geology and Mineral Industries and AGI for listing of their core and well log holdings on the NGDRS metadata catalog. They hold cores and records for 400 oil and gas wells and 50 geothermal wells.

Oregon's metadata is in digital format and should be readily integrated into the metadata catalog. However, the data is currently stored in MapInfo and they are working on exporting it to a simple table format. They are currently geocoding the data and completed the project in 1998. However, they decided that they would prefer to charge users for access to the metadata, thus they declined to participate in the metadata catalog.

West Virginia Geological Survey

Initial contact was made concerning placing the West Virginia Geological Survey's Oil & Gas metadata on the NGDRS metadata catalog. The WVGS has a database of 123,715 wells; for which they charge a \$10/hour data access fee. Listing in the NGDRS metadata catalog is counter to this policy and does not seem reasonable at this time. However, it has been noted that the data for West Virginia is in digital form and easily integrated into the metadata catalog when it may become more appropriate to pursue this avenue.

Utilizing the Metadata Repository

On May 14, 1998, the metadata catalog for the NGDRS became operational and open to the public. At launch, the following databases were made available:

- Fairfield Seismic
- A2D Well Logs
- MMS Well Logs
- Alabama Eastern Gulf PTTC Well Logs
- BEG Well Logs
- BEG Cores
- Oklahoma Geological Survey Cores
- MMS Block and Lease Boundaries

Maintaining the operations of the NGDRS metadata catalog operated under the guidance of the management committee and the Steering Committee. Overall, the metadata catalog saw steadily increasing usage of the system, no major user support requirements, and generally smooth systems and user management.

Routine operations have encompassed all aspects of maintaining the GeoTrek system. The major operational parameters which were of concern include connectivity uptime and system backups. Both of these issues were reasonably addressed through the course of the grant-funded operation.

Systems and network management has ensured a minimum of downtime since the NGDRS metadata catalog came online in May, 1998. Most of the initial effort was focused on the equipment and the Internet connection in Houston, TX. However, after the preliminary addition of redundancy at AGI's Alexandria, VA facility, and then the complete migration of the primary operations to Alexandria, VA, the system did not experience any downtime related to the migration.

A key design element in the metadata catalog that has greatly reduced the level of labor required to maintain the NGDRS metadata catalog is the automated user access system. Users were allowed to create their own accounts over the Internet and then may immediately begin accessing the metadata catalog. By not requiring administrative interaction, the addition of users has been a low-labor effort. However, during the first 18 months of operation, it was determined that the registration process was a barrier to a number of perspective users, unwilling to invest time in registering while apparently uncertain of the value of the product. So in 1999, we dropped the user registration processes and instead focused on basic traffic numbers to evaluate usage of the system. The statistics for the use of the metadata catalog and the associated web pages concerning the NGDRS are as follows:

	1998	1999	2000	2001	2002	Through 10/1/2003	Total
NGDRS Website Hits	30,911	61,152	48,656	64,746	99,169	187,518	492,152
Unique Visitors to NGDRS	1,331	4,336	6,218	7,900	6,660	3638	30,083

Figure 12. GeoTrek Usage Statistics during Phase III

A wide variety of education, government, and private organizations have accessed the NGDRS web pages and metadata catalog. Users of the system include:

- 174 U.S. Companies
- 143 U.S. Universities
- 44 Countries

Hand-in-hand with the user access administration, the help desk function is critical to ensuring the success of the NGDRS metadata catalog. The extensive beta-testing period of GeoTrek was designed to minimize the help desk and technical support load. Currently help requests are dealt with on a case-by-case basis, and generally are less than 1 per month. This low level of support requests is partially credited by the beta-testing period and partially by the generally good level of technical expertise by the user community.

Business and Marketing Support

The public face of the NGDRS is the metadata catalog. A focused effort has been ongoing throughout Phase III to market GeoTrek as both an interface to existing NGDRS repositories, but as the mechanism for new repositories to join the NGDRS.

Internet-based marketing of GeoTrek

Notification of the NGDRS Metadata catalog becoming available was made to the USENET community, particularly the sci.geo.* newgroups. A notice was also sent to several geoscience-related listservs, and links and indexing were requested from all of the major search and content sites such as Yahoo! and AltaVista.

Marketing through Presentations

The following is a list of events that have been identified as part of the on-going marketing campaign:

Meeting	Location	Date
Repositories Meeting	Calgary	June 16-17, 1997
DOE Contractors Workshop	Houston	June 20, 1997
PTTC Board Meeting	DC	July 13-15, 1997
NGDRS Steering Committee Meeting	Houston	Oct 14, 1997
SEG Annual Meeting	Dallas	Nov 2-7, 1997
European Association of Geoscientists and Engineers	Germany	May 1998
AAPG Annual Meeting	Salt Lake City	May 17-20, 1998
NGDRS Steering Committee Meeting	Houston	Oct 1 1998
Department of Interior	DC	September 2, 1998
GeoInfo VI	DC	September 19, 1998
British Geological Survey – Geosci. Data Warehousing	England	October 14, 1998
National Global Change Research Program	Maryland	November 4-5, 1998
AGU Annual Meeting	San Francisco	December 6-7, 1998
DOSECC Workshop	Hawaii	June 10-12, 1999
GSA Annual Meeting	Reno	November 2000
AGI Spring Meeting	Boston	June 1, 2001
GSA Annual Meeting	Boston	November 5, 2001
NDR4	Stavanger	March 6, 2002
NSF Workshop	Bloomington	January 6, 2003

Figure 13. Marketing Presentation of the NGDRS Project

International Geoscience Data Repository Meeting

The National Geoscience Data Repository System was represented during the fourth meeting of the National Geoscience Data Repositories, held in Stavanger, Norway. This meeting provided an opportunity for representatives of the various National Data Repositories to meet and discuss issues facing them, strategies for data management and funding, and to assess differences in data preservation policies around the world. A summary report of the meeting was published in the June 2002 issue of *Geotimes*.

The fundamental issue facing all National Data Repositories is the issue of valid and high quality metadata. Metadata appears to be a universal problem with most data providers not documenting their data acquisitions adequately, leading to the disposal of a large amount of data throughout the world. However, most countries appear to have systems in place to encourage metadata improvement, either through legislative threat or incentives by the NDR providing enhanced data management to the acquiring entities.

Universally, National Data Repositories are focusing on well logs and seismic data, in particular the digitization of well logs to reduce paper costs is the single largest effort occurring throughout the world. Cores and cuttings are also of high interest, and generally seen as the detailed data which the well logs will lead researchers to. Most NDR programs are funded either directly or indirectly from royalties and fees collected from petroleum production, while providing time-limited confidentiality of selected data and free access to public data for educational purposes.

The NGDRS program is working with numerous federal agencies to organize the 5th International Meeting of National Data Repositories to be held in 2004 in Washington, DC.

Print Media

Various print publications have run full and partial pieces on the NGDRS, including *Hart Publications*, the *PTTC*, *AAPG Explorer*, and *Geotimes*. A number of these publications also expanded their coverage on the issue of geoscience preservation based on the launching of the National Academy study on data preservation, which also provided additional exposure for the NGDRS.

Interaction with NASA's Global Change Master Directory

In January 1998, a meeting was held at the operational center for NASA's Global Change Master Directory in Greenbelt, Maryland. The GCMD is a NASA effort to provide a first-tier metadata catalog of available datasets related to earth science issues. The GCMD is part of the National Spatial Data Infrastructure, and provides an important service in assisting users to identify appropriate databases based on subject keywords and geographic location.

A dialogue was opened with the GCMD about how the NGDRS and GCMD could cooperate, in a mutually beneficial manner and ensure that no duplication of effort was being made. From a mission, data source, and user level, it is clear that the NGDRS and GCMD are complimentary to each other and do not represent a duplication of effort. Additionally, Lola Olsen, project director

for the GCMD, volunteered to have her staff create the Database Information Files (DIF) for the NGDRS component databases. A preliminary DIF has been written for the MMS well database as an example. These DIF's will be incorporated into the GCMD system, helping publicize the NGDRS's existence. Additionally, the NGDRS inclusion into the GCMD represents the first databases of analog resources to be included, opening an entirely new avenue for expansion of the GCMD.

It was agreed that open communications should continue, and that an overall efficiency in cataloging the available analog geoscience data is best handled by the NGDRS. The GCMD will continue to write appropriate DIFs for NGDRS-cataloged databases as they become available, providing subject and keyword access to the repository databases to the wider geoscience community.

In follow-up to the meeting, Christopher Keane met with Todd Byrd, the geology specialist on the Global Change Master Directory, during the 1998 AAPG annual meeting in Salt Lake City. The GCMD plans to write complete DIFs for the data holdings of the NGDRS metadata catalog. This capability will enable users to search on subject-based key words for the appropriate data sets within the NGDRS. This also provides an interface from the National Spatial Data Infrastructure (NSDI) to the NGDRS metadata catalog.