FOSSIL ENERGY

DOE/BC/14847-3
(OSTI ID: 752072)

NATIONAL GEOSCIENCES DATA REPOSITORY SYSTEM PHASE III: IMPLEMENTATION AND OPERATION OF THE REPOSITORY

Final Report
September 1999

Date Published: March 2000

Work Performed Under Contract No. DE-FG22-97BC14847

American Geological Institute
Alexandria, Virginia

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

DISCLAIMER

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

DOE/BC/14847-3
Distribution Category UC-122

National Geoscience Data Repository System

Phase III: Implementation and Operation of the Repository

March 2000

Work Performed Under Contract No DE-FG22-97BC14847

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

Robert E. Lemmon, Technology Manager
National Petroleum Technology Office
P.O. Box 3628
Tulsa, OK 74101

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

Table of Contents

Introduction and Background

Goals and Objectives

All scientific and technical investigations require access to basic fundamental data. The capture and long term preservation of data are required to address a wide range of scientific issues. The National Research Council released a report entitled *Preserving Scientific Data on Our Physical Universe (1995)*, that took a broad look at the challenges of scientific data preservation and management in Federal Agencies. The report concluded "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.

When fully implemented, the NGDRS will serve as an important and valuable source of information for the entire geoscience community for a variety of applications, including environmental protection, water resource management, global change studies, and basic and applied research. The repository system will also contain 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.

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

Data Rescue and National Energy Security

In the course of their exploration and development activities over the past several decades, major U.S. oil and gas companies have acquired enormous amounts of domestic geological and geophysical data. These companies have the most comprehensive geoscience datasets and sample inventories of information related to the Earth's crust of any organizations in the world. As major oil and gas companies 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 is to reduce U.S. vulnerability to oil supply disruptions by increasing domestic oil and gas production and reducing our dependence on unstable sources of foreign oil. In announcing the Initiative, President Clinton expressed the need to undertake "a determined effort to find and produce more domestic energy, with an industry-led solution that especially recognizes the role of independent drillers and producers." According to many independent oil and gas producers, the establishment of a National Geoscience Data Repository System would do more to aid the future search and development of domestic petroleum resources than most current or proposed programs or facilities. Such a data repository system could provide critical and hard-to-obtain information that would prevent premature abandonment of producing fields, and assist domestic producers in their evaluation of geologic trends, development of new plays, and assessment of remaining resources in existing fields.

Crisis in the Oil and Gas Industry

Oil and natural gas will remain critical components of energy supply in every nation for the foreseeable future. The United States, long a dominant producer of oil for domestic consumption, 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 geologic al 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.

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

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.

Phase III: Implementing the Repository System

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. Efforts continue to add repositories to the metadata catalog.

1.0 Clearinghouse

1.1 The Clearinghouse Review Committee

The Clearinghouse Review Committee manages all project aspects, but particularly those involving establishing the goals and performance metrics. 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. To date and Ad Hoc subcommittee chaired by the Chair of the NGDRS Steering Committee has servers as the Clearinghouse Review Committee.

1.1.1 NGDRS Steering Committee Meeting

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. Three NGDRS Steering Committee meetings were held during Phase III. Each meeting was called when the management team identified a number of major issues requiring community-wide resolution.

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.

1.1.2 Project Management

Three formal project management review meetings were held during Phase III. Each meeting responded to specific issues that were arising at the time, as well as provided strategic planning as the project moved forward.

The first project review meeting was held on August 27, 1997 at the offices of The Information Store in Houston. M. Milling, Jan F. van Sant, R. Graebner, Christopher Keane, G. Breed, O. Teoh and A. Eloy attended the meeting. The following items were discussed and reviewed:

- The difficulties faced by the NGDRS project, resulting from the current downturn in the industry level of activity, led to the suggestion to reconvene the NGDRS Steering Committee during October 1997 to reassess the priorities and opportunities for data preservation.

- Possible new targets for GeoTrek links were identified, particularly well-established and mature repositories (such as EROS) that manage data from satellites, gravity surveys, geochemical surveys, bathymetry, and other data types.

- An agreement was reached regarding the content of the Internet site that will house GeoTrek. The following general principles were adopted:
  - Site should have the look-and-feel of a **"yellow pages"** advertising or announcing data from public as well as private service company vendors.
  - Site will emphasize **breadth** (versus depth) of information.
  - Site will strive for **simplicity**. This should be reflected in the user interface, user access, and uniform (possibly standard) metadata for the different data types.
  - The site will be clearly identifiable as an **AGI site.**
  - **GeoTrek** was identified as the software browser for the NGDRS web-based catalog.

The second meeting was held on January 22, 1998 to review the beta testing and the development and staging of additional features and databases. As a result of this meeting, a number of issues were identified to be addressed before the public launch of the metadata repository.

- Improve the memory efficiency of the GeoTrek browser.
- Complete user documentation.
- Finalize the data advertising and ordering process through the NGDRS.
- Improve the user interface in response to suggestions from the beta-testers.
- Partition the front-end web material between technical and business issues.
- Format the web materials to "look and feel" like the AGI home page.

The third meeting was a two-day project review meeting was held in Houston on February 22-23, 1998. The current system, including the front-end web pages, was demonstrated. Three major issues were identified during the course of the meeting:

- Completion of the front-end web pages and related documentation

- Completion of data ordering procedures for the commercial data providers

- Inclusion of the Oklahoma Geological Survey data.

1.1.3 Pre-planning of Geoscience Data Preservation Workshop – National Academy of Science

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 have focused on defining the scope, goals, outcomes, and timing of the workshop. The American Geological Institute is encouraging 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. The expected outcome will be a report of the National Academy of Science based on the results of the workshop detailing a recommended national geoscience data preservation strategy, including an assessment of the need and priorities for preservation.

The Board on Earth Sciences and Resources is raising support for this study from various agencies and private organizations, including the US Geological Survey, US Department of Energy, National Science Foundation, POSC, and AGI. Additional supporting organizations are expected to be confirmed during the third quarter of 1999.

1.2 Core, Cuttings, and Thin Section Transfer

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 (Table 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.

Table 1. Industry Core Repository Unit Cost Analysis

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.

1.2.1 Unocal/Spirit Energy Cores and Cuttings

Unocal reported that they are nearing completion of cataloging and indexing their non-proprietary cores and cuttings. Unocal has expressed interest in the transfer of cores and cuttings to the NGDRS and the listing of their non-proprietary holdings in the NGDRS metadata catalog. However, given the current situation in the oil industry, Unocal has delayed the physical transfer of the cores and cuttings, but is willing to make their non-proprietary holdings available to the public through the NGDRS metadata catalog. Once Unocal has completed indexing and an arrangement made between Unocal, AGI, and C&M Storage, the holdings will be made publicly available on the metadata catalog.

The NGDRS is also assisting in the transfer of Unocal's Utah cores and cuttings from Shulemburg, 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, needs to be listed in GeoTrek™. Discussions were initiated with Utah about listing the data, but have been temporarily delayed given the change in leadership at the Utah Geological Survey.

1.2.2 Chevron Cores and Cuttings

Chevron has agreed to index their core and cutting holdings at the C&M Storage facility. The non-proprietary cores and cuttings records will be released to the NGDRS for inclusion in the metadata catalog. Similar to Unocal, once an agreement between AGI, Chevron, and C&M Storage is reached for access to the non-proprietary data, the Chevron holdings will be made publicly accessible.

1.2.3 Discussions with DOSECC

Discussions with DOSECC began in the second quarter of 1999. DOSECC is a consortia of 48 universities and research laboratories who 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 invited AGI to their annual workshop in Hilo, Hawaii to present to their membership about the NGDRS program and the areas which AGI felt the two organizations could cooperate.

DOSECC currently has one major operation underway, drilling 5 kilometers of core from the flank of Mauna Kea in an effort to determine the developmental history of the Hawaiian Islands. This core represents the first project where long term preservation issues are facing DOSECC. DOSECC also recognizes that all future projects with 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.

Following the DOSECC meeting in Hawaii, additional discussions were held to further investigate collaboration. In an effort to find a facility to store their equipment in the Salt Lake City, Utah area, Dennis Neilson encountered a facility which has potential as a data repository on a decommissioned Army base west of Salt Lake. Neilson is gathering additional information and AGI plans to investigate this option further.

1.2.4 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 Table 2. 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.

Table 2. Summary results of 1998 data capacity survey

State	Core & Cuttings	Well Logs	Digital Seismic	Current Repository Size (ft^2)	Available Repository Space (ft^2)	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

1.2.5 National Geoscience Data Repository and Research Center

The Phase II studies indicated that the volume of data available for transfer into the NGDRS is far greater than the existing 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, 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: Over the last few years 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 are reviewing the environmental reports provided by the Stapleton Development Corporation and the City and County of Denver.

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 have 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 has 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 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. AGI has likewise stopped pursing the acquisition of this facility at this time.

1.2.6 Discussions with National Science Foundation

Marcus Milling, David Applegate, and Christopher Keane 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 repository. 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.

1.3 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 proposes 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.

1.3.1 Mobil Digital Seismic and Well Log data

Mobil E&P is currently negotiating for 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, will represent an introduction of an enormous data holding previously unavailable to the wider geoscience community.

Work continues on arranging the transfer of a substantial portion of Mobil's domestic digital seismic data holdings to the NGDRS. Mobil is currently in the processes of determining which tapes are eligible for transfer. Issues being investigated include the tape's transfer eligibility including whether the data still has proprietary value, whether Mobil 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.

1.3.2 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 has indefinitely delayed the transfer.

1.4 Digital Data Media Conversion

AGI signed a letter of agreement with Oil Data of the Hays Business Services Group, Houston, Texas to assist in identifying digital seismic data for transfer to the NGDRS. Oil Data is contacting individual companies and data vendors on AGI's behalf to solicit seismic data contributions to the NGDRS. They will assess and inventory the potential seismic data contributions and provide a summary of results for AGI's review and prioritization.

Additionally, Oil Data will also provide recommendations to AGI related to the following digital data conversion parameters:

- Transcription options
- Data format standards
- Archive media selection
- Archive media indexing criteria
- Multi-line per original input treatment criteria
- Support data treatment considerations

1.5 Electronic Database Conversions

The transfer of electronic databases and their conversion for use within the NGDRS has been redefined as the effort to integrate and add new databases to the metadata catalog, as described in section 2.3

1.6 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 thus far 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.

1.6.1 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.

1.6.2 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.

1.6.3 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.

1.6.4 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 is willing to transfer 350 side-cores from the Gulf of Mexico and Atlantic Offshore to the NGDRS.

2.0 Creating the Metadata Catalog

2.1 Project Plan and Management

Oversight and management of the creation of the GeoTrek metadata catalog was handled as part of the effort to establish the clearinghouse organization. The metadata catalog was quickly realized as a central technology to assist the development of the clearinghouse and to be the linkage mechanism between the participating repositories. Details of management activities with respect to the creation of the metadata repository are outlines in section 1.1.

2.2 Setup Computer Operations

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.

2.2.1 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 is currently being 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 will be available when the back-up installation is completed. 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. A porting effort of the GeoTrek server-side software to the Linux platform for use as the back-up system is underway.

2.2.2 Software

The GeoTrek server software is currently stable and has been deployed in production environments. The browser software is also very stable, especially on Windows 95, Windows NT 4.0, Solaris 2.4, and Linux 2.0.33.

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, has been 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. It has undergone minor software improvements. The beta testing effort was of high value and resulted in identification of several critical issues before the software was released.

Additionally, experiments with new data types, such as publication citations will be explored on the AGI-based GeoTrek site. Also, AGI is looking at developing turn-key, low-cost solutions for installation of the GeoTrek standard system at State Geological Survey's to enable their active participation in the NGDRS metadata catalog effort.

2.3 Create 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.

2.3.1 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 indexed and added to the NGDRS metadata catalog. Additionally, the BEG has installed GeoTrek for local use on their systems, providing a new target user community and their feedback. The BEG, along with the NGDRS management committee has identified a number of issues and pilot projects which need to be completed as part of the integration process.

The installation of GeoTrek at the BEG-Austin was completed on December 12, 1997. The 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 is 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 implementation is 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. Meanwhile, C. Keane and A. Eloy are examining options to improve Macintosh-based performance, in particular, use of Sun's HotJava browser. However, no suitable solution has been identified for the Macintosh and a "wait and see" approach appears to be the only course of action concerning the Macintosh.

The BEG received funding from AGI to develop a core inventory control application to work hand-in-hand with GeoTrek. This program will allow AGI to market the core inventory management system together with GeoTrek in an effort to increase participation by core repositories across the country using the GeoTrek standard system. The first release version of the software was completed in July 1999.

The BEG requested funding in early 1998 to complete the geocoding of all well locations in Texas with latitude/longitude. Currently, wells within each county are not uniquely locatable. Therefore, in GeoTrek, all wells are shown at the geographic center of their county. The geocoding will provide 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. As of July 31, 1999, 8,505 wells in the BEG database have been geocoded.

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.

2.3.2 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 has yet to release their metadata, citing a desire to further populate their data set on the NGDRS before release.

2.3.3 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.

2.3.4 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 king of samples are available, well status, etc.

2.3.5 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.

2.3.6 Kansas Geological Survey (KGS)

Susan Bolton, from the Geoscience Data Resource Management group, organized a meeting on September 25,1 997 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 installing GeoTrek, probably when the conversion of their databases is completed.

2.3.7 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.

2.3.8 Oregon Department of Geology and Mineral Industries

Discussions have been initiated 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 expect to have the project completed and a copy of the data transferred for integration into the NGDRS by the end of 1998.

2.3.9 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 currently 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.

2.4 Setup Business and Marketing Operations

Most of the business operation issues were handled by the clearinghouse organization development. However, the public face of the business and marketing operations for the metadata catalog are through the World-Wide Web.

2.4.1 Web Pages

The development of the front-end pages of NGDRS was completed in February 1998. These web pages will provide the first exposure for users to the NGDRS. Two sets of materials are being released on the web site. First, business-issue related materials are presented in their own section. This includes such things as project information, history of the NGDRS, and a list of frequently asked questions with answers concerning joining and accessing the NGDRS. The other part of these pages relates to technical issues of the GeoTrek browser. This includes information on system requirements, troubleshooting tips, contact points for assistance, and frequently asked questions.

It was also determined that once a geographically dispersed set of data is available, a new page be created to facilitate user access to data from different parts of the country. This page, to be accessed just before entering GeoTrek, will present a map with hyperlinks into the geographic areas for which catalog data is available. This map will initially point to our target areas — GOM, Texas, Kansas and Oklahoma.

3.0 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

3.1 Project Plan and Management

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

3.1.1 Metadata Catalog Steering Committee

The committee headed by Robert Merrill (Spirit Energy) and made up of Madelyn Bell (Mobil), John Deery (Amoco), Jacob Loga (PGS), and Jan F. van Sant (AGIF) started their activities. The committee's first meeting was held on January 23, 1998. A. Eloy provided an update on the NGDRS and other related subjects. The committee expressed strong interest in working with the BEG to get their data sets in order with latitude/longitude and to work at attracting additional data vendors to joining the NGDRS.

3.2 Routine Computer Operations

Routine operations have encompassed all aspects of maintaining the GeoTrek system. A major aspect of the routine operations is ensuring that capacity is meeting demands, which to date has been satisfied by the originally installed hardware and software.

3.3 Systems and Network Management

Systems and network management has ensured a minimum of downtime since the NGDRS metadata catalog came online in May, 1998. Most of this effort is focused on the equipment and the Internet connection in Houston, TX. However, additional effort has been taken to ensure that system redundancy is established with the recently initiated development of a back-up and experimental server at AGI headquarters in Alexandria, VA. Additionally, GeoTrek performance and usage is monitored both locally in Houston, as well as cross-internet performance monitored from AGI in Alexandria, VA.

3.4 Data Administration

To date the efforts of data administration have focused on the addition of databases to the metadata catalog, as outlined in section 2.3. It is expected that routine data updates will begin early in 1999 as the GeoTrek system becomes increasingly distributed and initial datasets require updating.

3.5 User Access Administration

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 are 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.

Another major aspect of User administration though is the monitoring of usage and the understanding of user patterns. The statistics for the use of the metadata catalog and the associated web pages concerning the NGDRS are as follows:

1998 GeoTrek Usage Statistics

	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	YTD
NGDRS Website Hits	874	2934	3919	4040	5857	5860	4405	3022	30,911
Unique Visitors to NGDRS	46	165	76	61	58	344	333	248	1331
Unique GeoTrek Users	15	33	42	31	31	10	3	2	167
Number of GeoTrek Searches	479	406	468	282	313	46	4	2	2000
Total GeoTrek databases transactions	3100	3755	4971	3885	1892	261	47	19	17,930

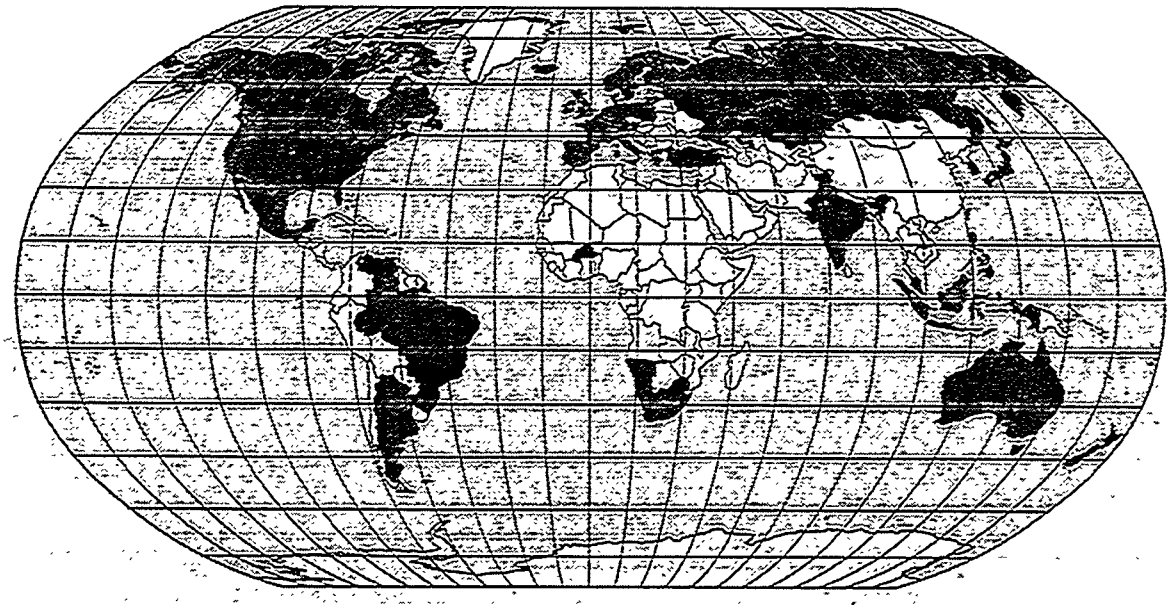
1999 GeoTrek Usage Statistics

	Jan	Feb	Mar	Apr	May	Jun	Jul	YTD
NGDRS Website Hits	6844	5904	6398	4905	5030	3902	4978	37,947
Unique Visitors to NGDRS	344	334	456	364	401	346	346	2591
Unique GeoTrek Users	23	27	24	11	15	5	26	131
Number of GeoTrek Searches	221	290	96	182	135	39	104	1067
Total GeoTrek databases transactions	789	1196	605	391	456	215	33	3685

Total GeoTrek Usage (May 14, 1998 – July 31, 1999)

	Project to Date
NGDRS Website Hits	68,858
Unique Visitors to NGDRS	3922
Unique GeoTrek Users	298
Number of GeoTrek Searches	3067
Total GeoTrek databases transactions	21,615

GeoTrek Accesses By Country



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

- 121 U.S. Companies
- 143 U.S. universities
- 44 Countries

3.6 Help Desk and Technical Support

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 emailed to the various staff at the Information Store, Inc., while other questions about the NGDRS and overall system function are emailed and handled by AGI staff. The need for assistance has been relatively low, partially credited by the beta-testing period and partially by the generally good level of technical expertise by the user community.

3.7 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.

3.7.1 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.

3.7.2 Marketing through Presentations

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

Table 4. Marketing Presentations

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 14,1 998

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

3.7.3 Print Media Advertising

Both Hart Publications and the PTTC have included articles about the metadata catalog in their respective publications. Additionally on May 17, 1998 in Salt Lake City, Marcus Milling and Christopher Keane attended the AAPG Core Preservation Committee meeting to discuss the status of the NGDRS. A consensus to encourage the AAPG *Explorer* to run an article about the metadata catalog and the NGDRS was reached from this meeting.

3.7.4 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 and 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.

By working with the GCMD personnel, the NGDRS will be more usable to users as the scope and number of data sets increase. Additionally, the availability of linkages into the NGDRS metadata catalog will hopefully increase traffic.

4.0 Budget

The original budget for Phase III was $1,867,169, with the Department of Energy providing $1,493,735 of it with industry contributions providing the balance of $373,434. A total of $472,814 in contributions were booked to Phase III, including $145,814 in labor by the Texas Bureau of Economic Geology for geocoding and installation of the GeoTrek metadata catalog system. Additional contributions included $97,000 from Micro-Stat, $150,000 from Locklin Oil, and $80,000 from James Thorne for data contributions, including the data value and costs for indexing and transfer of data.