NATIONAL GEOSCIENCE

DATA REPOSITORY SYSTEM

PHASE III: IMPLEMENTATION AND OPERATION OF THE REPOSITORY

SEMIANNUAL PROGESS REPORT
DE-FG26-99BC-15115

2nd Half FY01
April 2001-October 2001

Submitted by the

AMERICAN GEOLOGICAL INSTITUTE

to the

Office of Fossil Energy, U.S. Department of Energy

October 2001

American Geological Institute, 4220 King Street, Alexandria, VA 22302-1502
(703) 379-2480 Fax: (703) 379-7563

DISCLAIMER

ABSTRACT

The NGDRS has attained 72% of its targeted goal for cores and cuttings transfers, with over 12 million linear feet of cores and cuttings, in addition to large numbers of paleontological samples and are now available for public use. 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. Additionally, the NGDRS clearinghouse is evaluting the viability of transferring seismic data covering the western shelf of the Florida Gulf Coast.

AGI remained actively involved in assisting the National Research Council with background materials and presentations for their panel convened to study the data preservation issue. A final report of the panel is expected in early 2002. GeoTrek has been ported to Linux and MySQL, ensuring a purely open-source version of the software. This effort is key in ensuring logn-term viability of the software so that is can continue basic operation regardless of specific funding levels. Work has commenced 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 2002, including the discontinuation of the use of Java in future Microsoft operating systems. Discussions have been held regarding establishing potential new public data repositories, with hope for final determination in 2002.

Table of Contents

Table of Contents....3
List of Graphical Material....3
Introduction....4
Executive Summary....5
Experimental Approaches....6
Results and Discussion....7
National Academy of Science Geoscience Data Preservation Panel....7
NGDRS Steering Committee....8
Discussions with DOSECC....8
Data Transfer Status....8
Total Priority Data Transfers Relative to Phase I Targets....9
Unocal/Spirit Energy Cores and Cuttings....9
Unocal/Spirit Energy Utah Core Transfer....9
Chevron Cores and Cuttings....10
Altura Midland Core Facility....10
Phillips Seismic Tapes....10
Marathon Oil Cores and Cuttings....11
Texaco/Chevron Midland Proposal....11
Operating the Metadata Catalog....12
Redesign of GeoTrek....12
Enhanced targeting of public databases....13
Conclusions....14
References....15

List of Graphical Material

Figure 1. Map of transferred seismic lines from Phillips Petroleum....11

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) project was initiated in the face of the fact that billions of dollars worth of domestic geoscience data is 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. The goal of the project is to act before valuable data are permanently displaced.

The NGDRS will serve as an important and valuable source of information for the entire geoscience community and the nation at large for a variety of applications, including environmental protection, water resource management, global change studies, and basic and applied research. It 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 mineral resources.

A model for transferring data from the private to public sector is provided by the 1995 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 some 39 states to the university along with its warehouse and a $1.3 million endowment to cover annual operating expenses. All of these data entered the public domain for the first time.

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 72% of its targeted cores and cuttings transfers, with over 12 million linear feet of cores and cuttings now in the public domain. 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.

AGI remained actively involved in assisting the National Research Council with background materials and presentations for their panel convened to study the data preservation issue. The GeoTrek metadata catalog system continues to operate. 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. Current revision efforts are being retargeted given the identification of new open source technologies with 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 second half of the FY01, no efforts undertaken required experimental approaches to arrive at specific conclusions.

Results and Discussion

National Academy of Science Geoscience Data Preservation Panel

The Board on Earth Sciences and Resources staff of the National Academy of Science secured sufficient funding in the Fall 2000 to commence a 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 membership of the NRC committee is:

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 expected outcome will be a report of the National Academy of Science detailing a recommended national geoscience data preservation strategy, including an assessment of the need and priorities for preservation. The NRC staff has targeted completion of the study by the end of 2001 with a report released in 2002.

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, POSC, and AGI.

To date, AGI has made presentations and/or liaison contacts with the panel during each of their meetings and fact-finding missions, including the April 2001 meeting inWashington, DC, a June 2001 meeting in Denver, and the August 2001 meeting in Houston, TX. In addition, AGI continues to provide information to panel members as requested concerning AGI's previous studies and activities.

NGDRS Steering Committee

A steering committee last met on November 30, 2000 in Houston Texas, chaired Robert Merrill of Samson. The committee discussed the current status of the NGDRS, recent data transfers, improvements in GeoTrek, and future directions for the program. Representatives from major companies, federal and state agencies, including DOE, and several smaller companies were present. A full report of the committee meeting was attached to the first half of FY01 report.

Discussions with DOSECC

Discussions with DOSECC began in the second quarter of 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. Depending on the success of commercializing Toole, DOSECC and AGI are keeping the option of a facility in Utah open.

Data Transfer Status

NGDRS Cores and Cuttings Transfers

Source	**Liner Ft. of Cores & cuttings**
Unocal	1,109,016
Chevron (Cores)	934,157
Chevron (Cuttings)	10,038,898
Shell	450,000
Altura	85,000
Total To Date	**12,617,071**

NGDRS Paleontological Transfers

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

NGDRS Seismic Data Transfers

Source	Line-miles
Phillips	2000

Public Data Integrated into the NGDRS

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

Total Priority Data Transfers Relative to Phase I Targets

To date for the NGDRS project, nearly three-quarters of the volume of identified at-risk cores and cuttings have been transferred into the public domain. The NGDRS steering committee established 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	12.6 M liner feet	72%
Seismic Data	100 M line-miles	961 line-miles	<1%
Paleo Data	Not quantified	43.2 M section ft.	>100%

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.

Unocal/Spirit Energy Utah Core Transfer

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, the Utah Geological Survey sent a copy of its metadata to the American Geological Institute for review for integration into GeoTrek. The review is ongoing and integration of the Unocal Utah cores is expected in 2001.

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 metadata records of the Chevron cores and cuttings are undergoing quality control at this time. The data is not extensively geocoded, so translations from TRS coordinates to latitude and longitude need to be performed. Full integration of the data into the metadata catalog, including geolocation of the records will occur in 2001.

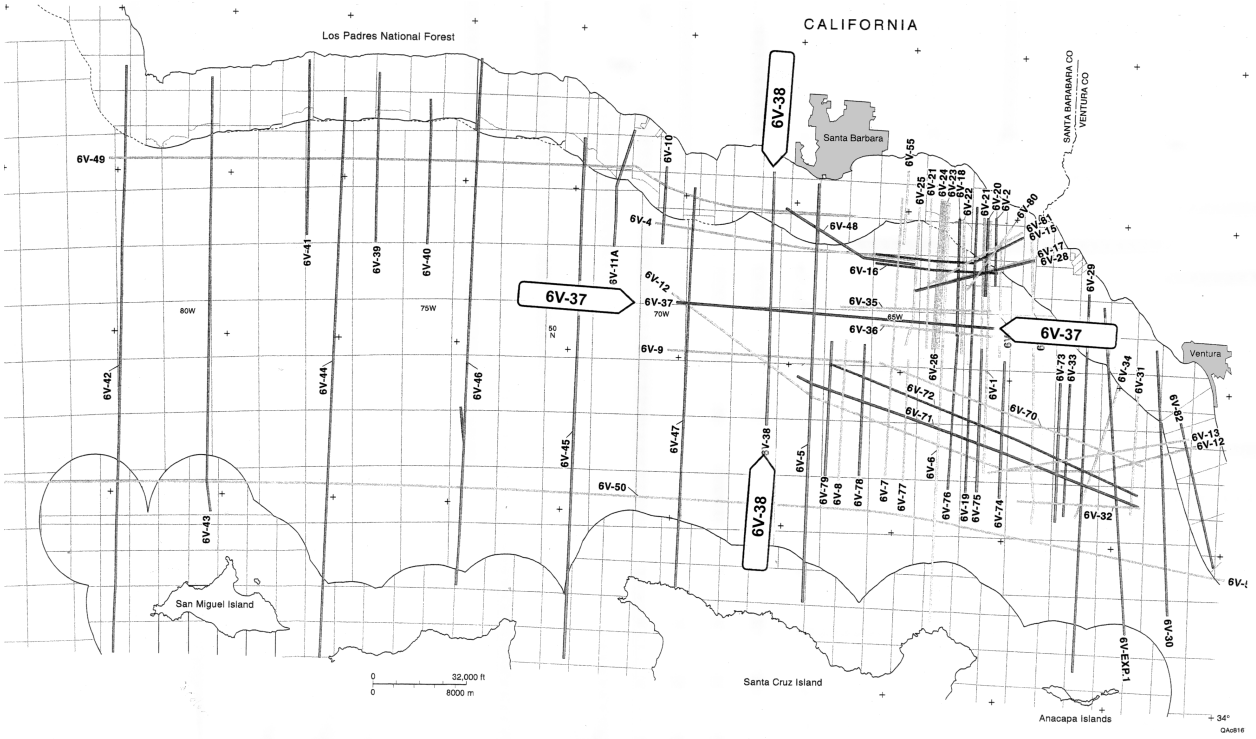
Altura Midland Core Facility

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 will be reintegrated into GeoTrek.

Phillips Seismic Tapes

Phillps 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. A map of the transferred seismic lines is below.

Figure 1. Map of transferred seismic lines from Phillips Petroleum



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 underway, Texaco affected no further activity on this effort. At this time, AGI does not expect any activity regarding this issue in the near future.

Future Repository Opportunities

An additional prospect of future new public repository space and data contribution is occurring in Texas. Progress continues in negotiating the arrangement with hopes that an agreement will be reached in 2002. This facility would represent a major addition to the public repository space and data holdings in the United States.

Operating the Metadata Catalog

The operation of the metadata catalog continued during the first half of 2001. The following databases are currently available on the metadata catalog:

- 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
- Texas Railroad Commission Well Logs
- Unocal Onshore Cores and Cuttings
- Chevron Onshore Cores, Cuttings, and Paleontological samples

The current access statistics are provided, as well as Project to Date (PTD).

	1998	1999	2000	Jan-Oct 2001	Total
NGDRS Website Hits	30,911	61,152	48,656	30,637	171,356
Unique Visitors to NGDRS	1,331	4,336	6,218	5,826	17,711

Redesign of GeoTrek

GeoTrek, the metadata catalog for the NGDRS went through an entire redesign. This redesign revolved around 2 areas: backend remote database integration and an improved user interface.

The back-end remote database change was completed and implemented with the integration of the Unocal cores. 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.

Additionally, in August 2001, the entirety of the GeoTrek code was ported to Linux and interfaced with a number of Open Source database systems. This effort was targeted to ensure the greatest viability of the metadata catalog regardless of funding levels. This has allow 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.

A new user interface system has been identified in the Open Source community that will neatly tie into the GeoTrek metadata catalog. A concerted effort will be engaged to integrate this widely used interface, which has the benefit of being well-tested on a number of browser technologies, into the end-users GeoTrek experience.

A major obstacle facing GeoTrek over the next year is the recent development of Microsoft no longer supporting Java "out-of-the-box" with its web browsers and operating systems. Given the reliance of GeoTrek's front-end to Java, as well as MapServer's most interactive front-end 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 within the next 12 months as a number of users do not have dedicated IT staffs to address these issues.

Enhanced targeting of public databases

In 2001, AGI will be accelerating its efforts to integrate existing public databases into GeoTrek, mainly through interactions with the State Geological Surveys. Among the primary targets for full integration into the system include:

- U.S.G.S. Core and Cuttings, Denver, Co - Acquired
- Texas Bureau of Economic Geology – Live Integration
- Kansas Geological Survey Cores and Cuttings, Lawrence, KS - Acquired
- Utah Geological Survey Cores, Salt Lake City, UT - Acquired
- Kentucky Geological Survey Cores and Cuttings, Lexington, KY
- Kentucky Geological Survey Coal Sample Cores, Lexington, KY
- Illinois Geological Survey, Bloomington, IN
- Indiana Geological Survey, Champaign, IL
- West Virginia Economic and Geological Survey, Morgantown, WV

Conclusions

The NGDRS has seen a distinctive increase in activity and interest, particularly since October 2000. This new spike 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 press 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 72% of its targeted cores and cuttings transfers, with over 12M linear feet of cores and cuttings now in the public domain. Additionally, large-scale transfers of seismic data have been evaluated, but bases 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.

AGI remained actively involved in assisting the National Research Council with background materials and presentations for their panel convened to study the data preservation issue. The GeoTrek metadata catalog system continues to operate. Porting of GeoTrek to Linux and MySQL was completed in August 2001 and now efforts are underway to examine the use of MapServer and other open source software to further enhance the metadata catalog.

References

None applicable.