#####

NATIONAL GEOSCIENCE DATA REPOSITORY SYSTEM

PHASE III: IMPLEMENTATION AND OPERATION OF THE REPOSITORY

PROGRESS REPORT DE-FG22-97BC-14847

**1st Quarter
February - March, 1997**

**Submitted by the
AMERICAN GEOLOGICAL INSTITUTE
to the
Office of Fossil Energy, U.S. Department of Energy**

April 1997

#####

DISCLAIMER

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

National Geoscience Data Repository System

Phase III Activities Report - February-March, 1997

DE-FG22-97BC-14847

Overview

The pace of activities in several of the tasks was accelerated. Activities on Project 1.0, the conversion and installation of electronic databases, were kicked-off with meetings at the BEG in Austin and with PGS in Houston. Initial project plans were drafted for all major sub-projects that were jointly agreed by AGI and The Information Store. Eight major sub-projects were identified: three dealing with data transfer and preservation, four dealing with the conversion of electronic databases (BEG, PGS, Kansas Geological Survey, MOBIL Seismic Sales), and one dealing with the deployment of the hardware, software and networking infrastructure for the Metadata Repository. Details follow.

Progress by Specific Project

1.0 ESTABLISHING THE CLEARINGHOUSE

Task 1.1 - Project Plan and Management

Project plan sheets for eight sub-projects were open to permit tracking and reporting on the activities of the NGDRS Phase III - Implementation:

PROJECT NAME	PROJECT TYPE	PROJECT
1. Houston Core Facility Project	Transfer of Core and Cuttings	Underway
2. Well Log Data Tape Conversion Project Not Defined	Transfer of Digital Data	
3. Seismic Data Tape Conversion Project Not Defined	Transfer of Digital Data	
4. BEG Project	Electronic database transfer	Underway
5. PGS Project	Electronic database transfer	Underway
6. Kansas Geological Survey Project	Electronic database transfer	Underway
7. Mobil Seismic Sales Project	Electronic database transfer	Not Defined
8. Metadata Repository Infrastructure	Creation of the Metadata Repository	Underway

The format includes the list of Steps, Completion Date by step (Planned and Actual), M-days Required by task and person, and M-days Actual by task and person. It is possible we may need to break Project no. 8 (Infrastructure) into three or four projects to more easily track and report on its activities.

From this new project planning exercise, it became apparent that two decisions need to be made by the NGDRS Management Team (Team) as soon as possible:

- A. At the moment projects no. 2 and 3 (above) are not clearly defined, meaning the Team needs to decide if these projects will (or will not) take place using the FY96 budget funds.
- B. From the defined scope of projects 4, 5 and 6, it seems that the budget is tight for delivering four electronic database projects. We need to considering dropping one project, (probably project 7), to be able to deliver the other three. Alternatively we can use funds from projects 2 and/or 3, if they will not be started with FY96 funds.

The progress for each project is discussed below.

Task 1.2 - Houston Core/Cuttings Facility (HCF)

A project plan sheet was prepared updating some of the information created back in February/97.

Further progress and decisions on this project, are now dependent on the positive feedback, by the Petroleum companies, on the volumes of cores and cuttings to be donated. Once it is determined that the HCF will be created in Houston, we are prepared to bring in the appropriate people to expedite the physical movement of cores and oversee the establishment of the repository. An experienced consultant in the procedures dealing with geosciences data cataloging and transfer - Thomas D. Burns (Tom) - was brought in March (for ten days) to help us refine the project plan for this activity. Tom accompanied us in the BEG visit for general discussions on this subject.

Task 1.3 - Digital Data Media Conversion

Project plan sheets were drafted for two sub-projects under this task. One deals with the conversion and transcription of tapes containing well log data (Project no. 2) and the other deals with conversion and transcription of magnetic tapes containing seismic data (Project no. 3). Project plan sheets are included in Appendix B.

At the moment the Team has not identified either project as part of FY96 implementation plans. No further work is planned for either project until we have direction from the Team.

Task 1.4.1 - BEG - Databases Conversion

This project was officially kicked-off on March 26 with a working meeting at the BEG in Austin. The meeting, was organized by Doug Ratcliff and Morgan Ives (at the BEG), and by A.Eloy from The Store. The meeting attendees included:

NAME	TITLE	DEPARTMENT	E-MAIL ADDRESS
Ed Garner		Well Log Data Center	garner@be2v.be2.utexas.edu
Daniel Ortuno		Well Log Data Center	ortunod@be2v.be2.utexas.edu
George Bush	Curator	Core Research Lab	bushg@begv.beg.utexas.edu
James Donnelley	Quality Assurance Specialist	Core Research Lab	donnelleyj@begv.beg.utexas.edu
Sammy Jacobo		Core Research Lab	crcfront@begv.beg.utexas.edu
Rodney Heathcott		Computing	heathcottr@begv.beg.utexas.edu
Morgan Ives		Computing	ivesm@begv.beg.utexas.edu
Doug Ratcliff	Associate Director	BEG Management	ratcliff@begv.beg.utexas.edu
Jay A. Raney	Deputy Associate Director	BEG Management	raneyj@begv.beg.utexas.edu

The agenda for the meeting included:

- Description of the National Geosciences Data Repository System (NGDRS) and the plans for the BEG. This was done via a slide presentation by A.Eloy followed by Q&A.
- Demonstration of the PetroTrek System browser using the PC demo package.
- General discussion of BEG users' needs, the gap between BEG needs and the functionality provided by PetroTrek and BEG user expectations for the project.
- General discussion of the current BEG systems (core and well log header) and plans for migrating to a PetroTrek-based system.

- Discussion of a major issue affecting the migration: the lack of coordinate well location for the core database. Initial discussion of the options to acquire the coordinate information from other databases or using a manual input process.
- Definition of the next steps.

General Observations:

- The BEG is currently moving away from their existing computing platform [VAX/VMS, Ingres] to a [Sun/Solaris, Oracle] computing platform. This is very favorable to the NGDRS project since we use exactly the same computing platform [Sun/Solaris, Oracle].
- They want to coordinate their efforts and requirements with the goals and objectives of the NGDRS. This is understandable and we may encounter it again when dealing with other non-for-profit organizations. It presents both challenges (such as, managing the portion of the budget allocated to the BEG) and opportunities (such as, BEG is willing to provide funds for items that are definitely out of the scope of the NGDRS).
- BEG staff is genuinely interested in PetroTrek (GeoTrek). Many of the "mundane" needs that they have today from their databases (such as "give me all the wells we have in county XYZ") can be directly met by Version 1 of PetroTrek.
- They have some specialized needs, currently not in PetroTrek, that we need to specify in more detail, and understand if they fall or not within the NGDRS budget. Specifically they need:
 - To track physical movements of boxes of cores and cuttings between Austin, Midland and the patrons. Volumes and transactions for these physical movements will have to be defined.
 - To input modifications or additions to the their databases (cores and well-log headers). This is a routine process that requires the dedication of two BEG data administration persons. The volume of database changes is not small; it was identified as in thousands/week.
 - To create some specialized reports.

Next Steps:

- 1) Install the larger data set provided by BEG for both databases consisting of approximately 100,000 core items and 250,000 well-log headers. We will use this initial installation to diagnose more closely the issue around the well coordinate/location and possible other issues relating to the data.
- 2) Analyze the set of forms and processes currently being used by the BEG, particularly in reference to the updates of the databases and the physical movements of core boxes. A set of forms is being organized by BEG personnel and was promised to be sent to The Store shortly.
- 3) Determine the hardware, software and network requirements specific to the BEG facilities in Austin and Midland.
- 4) Hold a second meeting in Austin during the week of April/21. Bring along Iain Morison or Rod Hanks to discuss, in more detail, technical issues. This date is conditioned on early arrival of the materials promised in step 2 above.

Task 1.4.2 - PGS - Databases Conversion

This project was kicked-off with a meeting held between PGS' management and technical staff, and Oscar Teoh and Glenn Breed from The Store. It was held on March/25 with John Babec (Sales Mgr., PGS Data Management Inc.) and it was aimed at addressing the following items:

- To answer or clarify the questions addressed in the G. Breed's memo of 3/21/97.
- To have a demonstration of the GeoBank system, created by IBM and marketed by PGS.

- To have an in depth discussion of the *business* requirements for selling/buying seismic data via the web.
- To have an initial discussion of the *technical* requirements for selling/buying the data seismic via the web. Included in the discussion was a preliminary sizing of the likely data set to be installed under the NGDRS.

From this meeting, some initial estimates were put together for the potential PGS data set to be installed. The figures below represent upper bounds. Either Richard Koseluk or Jacob Loga will need to make the final decision on what becomes available on the NGDRS server.

- PGS total seismic data (for the GOM) is in the order of 2 Terabytes. This is not the metadata but the actual seismic data. These 2 terabytes equates to 250 million to 2 billion traces, depending on the number of bytes/trace.
- Estimated number of blocks within GOM is 5,000. PGS estimates reaching that within a year or so
- There are altogether about 20-30 surveys (tops 50) in the GOM.
- Estimated number of CDP bin centers (i.e., X-Y locations) is about 250 million. This means we will have to support, for PGS alone (assuming they agree to release this information) about 250 Million coordinate pairs, or 2 GB.

In conclusion, the initial estimates indicate the PGS data set to be installed as part of the NGDRS will require from 4-6 gigabytes of disk space.

Next steps will be to hold meetings, in early April, with PGS Data Management executives to discuss further the *business* aspects of the NGDRS implementation.

Task 1.4.3 - Kansas Geological Survey (KGS) - Databases Conversion

Human resources for this project will be allocated in April. An e-mail memo was sent to Marcus Milling requesting that an initial contact be established with KGS Management in order to officially kick-off this project.

Task 1.4.4 - MOBIL Seismic Sales Organization - Databases Conversion

Because of FY96 budget constraints we need the Team to define if funds will be allocated to this project. Once funds are identified, human resources will be allocated to the project.

Task 1.5 - Hardcopy Data Media Conversion

This project has no funds allocated in the FY96 budget. No activity is planned.

2.0 CREATING THE METADATA REPOSITORY

It is possible we will need to break up this project plan into three or four projects to track and report easily and accurately on its activities.

Activity for this project continued. The bulk of the effort was in the procurement of the hardware, software and network configuration items required by the NGDRS Metadata Repository. In summary, the following was achieved:

Sun Microsystems

Sun Microsystems will be our provider for the major hardware system components, the so-called server systems. It is in these systems that the databases, from the diverse projects, are installed and run. Servers are the backbone of the PetroTrek system and will need to be sufficiently powerful to handle 25-50 users accessing simultaneously the NGDRS databases. Users will use their client (PC) infrastructure and Internet to reach our hardware servers. Sun has the largest share of server market; most petroleum (upstream) companies use Sun as their basic Unix platform, both for performance and economics reasons. The BEG is currently moving entirely to the Sun/Solaris platform for similar considerations.

The major Sun hardware components for the NGDRS system will include:

- (1) Two servers, model Sun Ultra 2, will be used to run the production environment. One machine will support users that need access to public domain data (such as the BEG or KGS) and the second will be dedicated to support NGDRS commercial subscribers (such as PGS). Each server will be configured with about 20 gigabytes of disk space.
- (2) One Sun Ultra 3000 server will be dedicated to development and staging of the databases to be installed. This server will also backup the databases on the Sun Ultra 2 servers. This server will be configured with about 40 gigabytes of disk space.
- (3) One Sparc Disk Storage Array (Model 112) with an additional 20 gigabytes of disk space.

Initial discussions were held with Michael J. Heagney, Petroleum Industry Manager for Sun Microsystems, regarding the configuration components and possible price discounts. Michael has directed us to the Sun Microsystems Catalyst program. This is a program directed at software developers (such as The Store). Discounts for this program are substantial (40%) and we have filled-in the necessary paperwork to become part of the program. We should hear back from the Catalyst program management by the middle of April. Meanwhile we will be talking with Sun hardware/software engineers to further tune the configuration component to our needs.

Oracle Corporation

Oracle will be our provider for database management systems (DBMS), a key software component of the NGDRS project. The DBMS is the component that supports the PetroTrek software within the server hardware. It is the software that translates the user requests, formulated using PetroTrek, into accesses to the databases. Oracle has the largest market share of all DBMS providers and by a large margin. Most, if not all, upstream petroleum companies, use Oracle as their standard relational DBMS.

Initial discussions were organized by Ezat Zarasvand at Oracle and included Douglas Benson, John Hood, George Colonna and Claudio Grossling. It was their initial impression that the Sun Ultra 2 machines may only require Oracle DBMS run-time system, while the Ultra 3000 will require a full license. We found out that Oracle has a similar program to the Catalyst program of Sun Microsystems and Ezat Zarasvand will direct us to their program. At this time we do not know what discounts their program provides. During April we will also continue talking with Oracle software engineers to further tune Oracle specifications to the NGDRS environment.

Tyco Electric

Will be our provider for the ethernet cabling local area network (LAN) that will be needed to support the NGDRS project. To start the process, we prepared for Tyco a physical layout of all our facilities in Houston — old and new area. The layout was used by Tyco to specify and quote prices for cabling job.

Office move

The installation of the cables by Tyco is scheduled for April/8. Once completed we can hook-up the other ethernet cabling components and will move the technical staff to the new area. The necessary furniture will be acquired to allow the movement of the technical staff. The room of M. Graebner should be available during the week of April 14 to be occupied by Jan F. Van Sant.

Organization

We interviewed two candidates for the job of systems manager. Eric Kellen and Wilfred Teoh.

Eric is a former POSC employee, and well known to all of us. He is currently a networking consultant and contractor with EDS Corporation. Because of his current work load and career choice he can not give us the full-time attention required by the NGDRS project. We have, however, retained him as a contractor to help us in the design of the networking components.

Wilfred Teoh is currently providing systems management support for the Houston Community College system. This job is very similar to the one we require for NGDRS. He is quite experienced with PC and Unix environments. He will be joining us on April 11, for a trial period of six weeks, one day per week.

3.0 UTILIZING THE METADATA REPOSITORY

This project saw no activity during March. Activities in this project are only expected to start after August or September 1997.