NATIONAL GEOSCIENCE DATA REPOSITORY SYSTEM

PHASE II: PLANNING AND PILOT STUDY

PROGRESS REPORT

**3rd Quarter
July - September, 1995**

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

November 1995

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

MASTER
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

NATIONAL GEOSCIENCE DATA REPOSITORY SYSTEM

Activities Report — July-September 1995

DE-FG22-95BC14833

Task 1. Planning and Specification

1.1 & 1.2 Organization, Operations and Management Plan

The Business and Management Task Force prepared cost analysis summaries for transfer, indexing, and cataloging core/cuttings samples and digital seismic records. A straw-man business model plan for an AGI metadata Internet service has been developed. Both work-ups are being presented and reviewed with company sponsors.

Documentation was submitted in July for the Gas Research Institute (GRI) to join in sponsorship of the project and was approved by GRI. PGS joined in support of the project in August. Current company sponsors are: Conoco, Chevron, Pennzoil, Unocal, Mobil, GRI, and PGS. Contacts have been made with an additional six companies that are considering joining in support of the project.

1.3 Data Audit and Prioritization

Metadata requirements definitions based on POSC standards (1.3.2) and on criteria for data retention and purging (1.3.5) continued. Research on generalized metadata requirements continued via use of the internet and contacts within the National Laboratory system (LLNL, LANL, etc.). Workgroup recommendations from 1.1.4 are still a problem area. Survey questionnaires for sponsor company organizations and vendors dealing in data storage activities were created to address data retention and data purging. Questionnaires were mailed out to the seven sponsor companies and 10 vendors dealing in data-warehouse services.

1.4 Data Preparation, Transcription, Imaging, and Preservation

Work on sections 1.4.3 resulted in moving the delivery date for the Ampex tape drive system to September 15, due to the state of the MobilView software activity. Review of data, data sources, and transcription methodologies continued. Questionnaires on data storage and transcription methods were developed and will be sent to sponsors and the appropriate service companies in August. Questions were also posted on the Internet in various bulletin boards, but these received little response. Information gathering and collation continued on 1.4.1 and 1.4.2. Similarly, data collection and collation continued on sections 1.4.4 through 1.4.6 using other repository standards (such as CIDAL, DISKOS, and POSC), data warehousing publications, and via Internet queries. Questionnaires for service companies and major petroleum companies were created and sent out in August. Requests to the NTTC and NML were made for research into subtasks 1.4.2 to 1.4.6.

1.5 Data Catalog and Indexing

An expanded outline was developed for this section. Work continued on cataloging and indexing standards, and on the POSC-compliant model for mapping of BEG database attributes to Epicentre data model attributes using EXPRESS-M language. Epicenter Model ASCII files compliant with POSC Utility Programs Loader Formats were written for entering BEG Cores and Cuttings data in a database complying with the Epicentre Relational Implementation. These model files are written in Epicentre/Express notation using the Epicentre Entity Data Dictionary.

The Epicenter Model ASCII files were used to build a POSC Compliant Epicentre Model of BEG Cores and Cuttings Data for input of this data into the MobilView System. Our first effort added Core and Cuttings data to a POSC compliant UniSQL database. At the present time, it is planned to obtain an Oracle account with access to an Oracle database and enter the BEG Cores and Cuttings data into a POSC compliant Oracle database.

The NTTC and NML were requested to provide information pertinent to subtasks 1.5.2 to 1.5.4.

1.6 Data Access, Distribution, and Delivery

Surveys for approximately 30 potential end-user organizations that would want to access the NGDRS system were sent out. The potential end-user addresses included academic institutions, state geology and natural resource organizations, independent oil companies, minerals companies, and service companies. The questionnaires are to determine the current status of the end-user community and to determine the type and cost of the infrastructure that would be necessary to allow a broad range of potential users. Eight responses have been received to date from two minerals companies, several state geologic groups, universities, and one coal company. Interest appears to be high, including from the minerals companies. Initial write-ups were completed for each subtask; much data remains to be compiled on this and other subtasks, and the write-ups will change considerably as data compilation continues.

Task 2. Directory of Geoscience Data Centers

The data center survey questionnaire was completed after review by end-users and data center managers. Compilation of the Database of North American Geoscience Data Centers is continuing. To date 327 sites have been identified and most will be incorporated into the database.

Task 3. Pilot Projects

3.1 Digital Seismic Data Transfer

A seismic dataset has been donated by Mobil Corporation. This is the Grand Island 90 dataset, which includes a 3-D seismic survey. Meetings were held with Mobil in Dallas to set out formats, media, QC, and delivery schedules for the data. Conversion and reformatting of the seismic data is underway and it will be transferred to the subcontractor upon completion. Problems concerning the storage of the original paper data associated with the dataset have been resolved. Seismic datasets for Haddock and Halibut were loaded into the MobilView system in August.

3.2 Core and Cuttings Processing

The core set has been digitally imaged and the results are being edited. An unanticipated problem has cropped up concerning the format of the TIFF files of the images. Currently the files cannot be read by a UNIX system, which MobilView is based on. The files will be reformatted before they can be loaded into the demo database. Resolution of the problem is underway. A final report on procedures, time and cost elements was sent by the service company. The core was returned to the BEG in Austin, along with a second copy of the digital images and the laminated photographs that were created in the scanning process. Save for the reformatting of the digital data and the mapping of the data into a POSC-compliant database, Pilot 3.2 is nearing completion. Thin-section digitization has been requested by AGI and is being pursued. Write-up on this section will be done in collaboration with the BEG.

3.3 Catalog and Index Database

Activity increased greatly with the acquisition of MobilView by AGI in mid-July. New personnel were added and a clone environment was being created in Dallas for the transfer of MobilView software to a physical environment controlled by AGI. Two complete datasets are currently being tied to the software; these are the Haddock and Halibut datasets, donated by industry and already POSC-compliant. Data loading and data mapping exercises with POSC assistance are underway.

Activity concerning MobilView continued at a high level. A requirements definition document for the cataloging and indexing system was completed as a first pass. The document contains four sections on cataloging and indexing objectives, scope and responsibilities, along with a functional requirements section. Data loading and data mapping exercises and discussions with POSC assistance continued.

A Sun Sparc 10 clone system was set up and configured for the development environment. A week was spent trying to get the POSC compatibility layer to work. The result was not good. A method to write directly to the relational projection is being pursued while POSC looks at the compatibility layer. As mentioned above, the digital core images and the Haddock dataset were loaded; the Halibut dataset needed new disks and is currently being loaded. Much was learned about navigating the

model. There may be a need to acquire additional tape drives to cover the tape for sizes, but this is not an issue at this time. The project appears to be on schedule for a mid-October demonstration of the system. The geographic browser/catalog database has been named GeoTrek for purposes of system demonstration.

A homepage for the NGDRS project was completed and PPP for dial-in was installed and configured. Numerous tapes with data for the GeoTrek system were transcribed and retrievals tested.

3.4 Paper Records Data Capture

The integrated paper dataset has been identified and certain transfer procedures are underway (see comments in Pilot 3.1 above). Companies for data conversion have been identified and contacted to review their costs and methodologies. Data on paper data conversion continues to be gathered even though work on this section has been pushed back until the fall.

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.