

DOE/BC/14833-72

NATIONAL GEOSCIENCE

DATA REPOSITORY SYSTEM

PHASE II: PLANNING AND PILOT STUDY

PROGRESS REPORT

1st Quarter
February - April, 1995

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

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

THE ATTACHED DOCUMENTS HAVE BEEN
ENTERED INTO THE DTC AND DISTRIBUTED
ON 6-21-95

THIS IS THE COPY FOR THE AWARD FILE
DOCUMENT CONTROL CENTER

14833
dh
MASTER

May 1995

RECEIVED
DOE/PETC
65 JUN 12 AM 9:31
ASSISTANCE DIV.

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

DISCLAIMER

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

BACKGROUND AND OBJECTIVES

The American Geological Institute (AGI) recently completed the first phase of a multiphase program to study and implement a National Geoscience Data Repository System (NGDRS) to capture and preserve valuable geoscientific data. The study was initiated in response to the fact that tens of billions of dollars worth of domestic geological and geophysical 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.

The NGDRS would 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, reducing risks from earthquakes and other geologic hazards, and basic and applied research. The repository system would also contain critical data that would enable domestic energy and minerals companies to expand their exploration and production programs in the United States for increased recovery of domestic oil, gas, and mineral resources.

Phase I: Feasibility and Assessment Study

The first phase of the project was designed to assess the feasibility of establishing the NGDRS. It focused on two major issues. First, it documented the types and quantity of data available for contribution to the NGDRS. Second, it documented the data needs and priorities of potential users of the system. There would be no point in proceeding with the project without large contributions of data that would be of great interest to the potential user community.

The results of AGI's National Geoscience Data Repository System Study (1994) were extremely positive. Major oil companies, large independent petroleum producers, and minerals companies have indicated they would consider contributing vast amounts of data to a NGDRS. Many state geological surveys, federal agencies, and a number of existing public repositories indicated they would include their data in the NGDRS.

The total amount of seismic data identified in the Phase I study is conservatively estimated to represent more than 15 million line miles, which constitute a substantial fraction (perhaps 25 percent) of all seismic data collected in the United States since 1950. Likewise, the rock core and cuttings identified in the Phase I study are estimated to represent a significant fraction (perhaps 10 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 after the NGDRS was established.

Proposed industry contributions represent billions of dollars worth 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 restrictions that place certain areas off limits to resource exploration and development activities. It is proposed that all material placed in the NGDRS would enter the public domain and become available to all users.

State and federal agencies have indicated that large amounts of geoscientific data already in their possession would be made available through the NGDRS.

There is 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 independent petroleum producers and other potential users of the repository system.

Phase II: Planning and Pilot Study

The second phase of the NGDRS project is being proposed in response to the positive findings of the Phase I study. The Phase II study would address specific organizational and operational requirements for establishing and implementing the NGDRS. The joint industry, academic, government agency Steering Committee established in the Phase I feasibility study would provide oversight and guidance for the Phase II study. The proposed Phase II study has four major components:

- 1. Planning and Specification.** The planning and specification component of the Phase II study would identify the requirements and develop an organizational and operational plan for implementing the NGDRS.
- 2. Directory of Geoscience Data Centers.** A comprehensive directory of geoscience data centers providing information on data types and access will be critical to the implementation and operation of a decentralized NGDRS.
- 3. Pilot Projects.** Pilot projects developing the metrics for the transfer of various types of data from the private sector to the public sector are critical to the development of an efficient and cost-effective NGDRS.
- 4. Steering Committee Operations.** The 25-person, industry-led Steering Committee established to oversee the Phase I Data Repository system Feasibility and Assessment Study will continue to provide oversight for the Phase II Planning and Pilot Study. The Committee will have scheduled biannual review meetings and otherwise meet as required to consider technical and policy issues.

Phase III: Implementation and Operation

The goal of Phase III would be to provide for a coordinated, routine orderly transfer of significant volumes of geoscience data from the private sector to the NGDRS. The business plan and network support base defined in the Phase II study would assure efficient operation of the NGDRS. Major data contributors would be matched to appropriate regional data repositories to facilitate data transfer. Procedures would be established for easy and efficient access to newly acquired data by data users.

PROJECT ACTIVITIES

The project was formally initiated in February 1995. To date activities have focused on the three following areas: technical architecture, technical acquisitions for pilot projects, and technical evaluations.

Technical Architecture

Technical architecture efforts have concentrated in defining a preliminary architecture for the NGDRS system of repositories. Four major design principles are being used to define the repository model: flexibility, integration, ease of access to data, and continuous improvement or streamlining of the system. The correct balance of these principles will be applied during Phase II to provide the most appropriate repository system design.

Specifically the initial architectural efforts included:

- 1) Research and discussion on similar efforts to develop comprehensive data/information management environments for geoscience technical data, in particular the QC Data/Alberta (Canada), CIDAI-UK, NPD-DISKOS (Norway), Sonatrach (Algeria), and PeruPetro (Peru) repositories.
- 2) Diagramming a preliminary architecture that embodies the principles mentioned above and identifying the major repository components — data entry, data validation, data preparation, database management, data warehousing, data archival, data browsing, data ordering, data retrieval, and data distribution.
- 3) Diagramming likely subsystems required to operate on the geoscience technical data. This includes a repository usage subsystem, a data ownership subsystem, a data brokerage subsystem, and so forth.
- 4) A draft description of the components and subsystems mentioned above.

Technical Acquisitions

Activities were undertaken and assistance given in acquiring certain technical components for various aspects concerning initiation of the Pilot Projects in the NGDRS program. These acquisitions are necessary in order to begin and plan the individual pilots.

Specifically, these acquisitions included:

- 1) An integrated seismic data set of sufficient size to use a testbed in the Digital Seismic Data Transfer and Cataloging and Indexing Database pilots, and to use in evaluating a data browsing/data archiving/data retrieval demonstration system.

- 2) Commitment of a robotic tape drive and data archival system for the demo center.
- 3) Commitment of POSC personnel and software technology for standardizing the data in the Pilot Programs.
- 4) A data management and data browsing system (MobilView) from Mobil Corporation. This is a significant acquisition that will add tremendous leverage to the project, but will also require a re-evaluation of project focus and resource allocation.

Technical Evaluations

Technical evaluations of various components and companies that would supply services to test these components were initiated for the Data Audit and Prioritization, Catalog and Index, and Paper Records Data Capture sections of Phase II. Pilot Project coordination and pilot preliminary design were also initiated.

Specifically, the following activities were undertaken:

- 1) Companies in Calgary, Denver, and Houston were researched to evaluate their capabilities, methodologies, costs, and compatibility for digital data conversion using scanning, digitizing, and photo-imaging techniques.
- 2) Research on metadata requirements and data retention and data purging criteria was initiated.
- 3) Solicitations for a Core and Cuttings data set were undertaken and potential donors identified. Procedures for core acceptance, transfer, and processing were also initiated.
- 4) A generic diagram to indicate timing and dependency relationships between tasks assigned to The Information Store has been developed. This diagram will be used to structure The Store activities and to coordinate other subcontractor activities and to redesign future activities as the project evolves, such as comprehending the effects of the acquisition of MobilView on the project and in redirecting or refocusing resources and time as necessary.