Final Technical Report for NGGDPP 2008-2009

# Increased Usability through Data Preservation at the South Carolina Geological Survey

USGS Award Number 08HQGR0105

Prepared by

C. Scott Howard
South Carolina Department of Natural Resources
Geological Survey
5 Geology Rd
Columbia, SC 29212
(803) 896-7712 – office
(803) 896-7695 – fax
howards@dnr.sc.gov – email

Final Technical Report for NGGDPP 2008-2009

# Increased Usability through Data Preservation at the South Carolina Geological Survey

## Abstract

The FY08 NGGDPP project is separated into five tasks, which are summarized in Table 1. The inventory of SCGS Core repository was completed with 1,489 cores. Metadata records of 239 cores were submitted to NDC. The discrepancy is related to finding location information for the cores. A methodology for digitizing geophysical logs into LAS files was developed, and 88 logs were digitized. Metadata records of all 88 digitized logs also were submitted to NDC. The results of these tasks are the initial steps by the SCGS to develop a comprehensive digital data inventory of all holding s and the transfer of old records into an accessible digital format.

Final Technical Report for NGGDPP 2008-2009

"Increased Usability through Data Preservation at the South Carolina Geological Survey"

Award Number 08HQGR0105

C. Scott Howard
South Carolina Department of Natural Resources
Geological Survey

## Introduction

This report documents the results of the South Carolina Geological Survey's (SCGS) third year of participation in the National Geologic and Geophysical Data Preservation program. The priorities for the FY08 plan included completing the inventory of the SCGS Core Repository collection, entering the inventory results into a database, establishing standards for data capture in the Core Repository, digitizing geophysical well logs, and providing metadata of the inventory and digitization projects to the National Data Catalog (NDC). All these objectives were achieved with varying levels of success.

## Task Results

For the purpose of reporting, the FY08 project was divided into five tasks. Table 1 summarizes the objectives and results of each task.

**Table 1 Summary of NGGDPP Tasks**

| Task | Task Status | Notes |
|---|---|---|
| 1) Database Development | MS Access database constructed to contain results of Core Repository Inventory. | Early use of the database has resulted in some modifications to design. |

| Task | Task Status | Notes |
|---|---|---|
| 2) Complete Inventory and Data Entry | Completed, 1,489 cores inventoried, minimum footage = 230,453. All records entered into database. | Extraneous cores are still being found. Majority of records entered into database are only partially complete. |
| 3) Data Capture Standards | Completed. Included travel to Delaware Geological Survey, March 2009. | Discussions with geologists, data entry personnel, and IT personnel have helped define the rudimentary needs of a data model. |
| 4) Digitize Geophysical Logs | Digitized 88 geophysical logs. | Developed technique for creating LAS files and digital images of logs. |
| 5) Metadata | Two sets of metadata developed and submitted to NDC. | Metadata records created for cores in Repository and for digitized geophysical logs. |

**Task 1** – A Microsoft Access database was developed to maintain the results of the core repository inventory. The design of the database was accomplished by examining existing records in SCGS files, comparing data models from other sources, and incorporating comments from SCGS geologists. After the first version of the database was completed, it was tested and modified before a final version was implemented. In some respects, the initial database design was done in a vacuum by being completed prior to examining core material. To accommodate this situation, the database was constantly tested, tweaked, and improved during the early stages of the inventory of core samples.

A paper data-entry form was created concomitant with database development. The paper form followed the database design and provides a permanent archival record of each core in the repository. Additionally, these hard-copy forms are easier to use in the repository and were printed on Tyvek paper to ensure the longevity of the record.

**Task 2** – The inventory of material in the Core Repository consisted of two phases: 1) the physical inventory, and 2) population of the MS Access database with inventory results. A part-

time geologist was hired (University of South Carolina undergraduate geology major) to perform most of the inventory work. Additional part-time student help was added during the late spring and summer of 2009. Before the inventory began, it was necessary to do two things: 1) clean and organize the facilities and 2) map the core repository facilities. Cleaning was necessary because the shelves and aisles have accumulated a lot of clutter, primarily because of a lack of designated storage space. This clutter prevented access to some areas and core material. A temporary solution was to move the clutter, do the inventory, then move the clutter back so the inventory could proceed. A map of the core repository facilities had two purposes: First, it added a data attribute to each core record so that the core could be located quickly. Second, it allowed us to make some numerical estimates of shelf space used and amount of space available.

Recording inventory information directly onto a computer and the database was considered; but the logistics of juggling hardware, while doing the inventory, made this option impractical. Instead, a data-recording sheet was printed on Tyvek paper to make the record more permanent. These durable data forms give us a way to preserve original records and to ensure that future acquisitions of cores are documented properly when accepted.

Populating the Core Repository database was a straightforward endeavor. The inventory found 1,489 cores in the repository. The main problem resulted from incomplete information for many of the cores in the repository. To date, locations to only 239 cores have been found. This result was unexpected as it had been assumed from the beginning that the paper records accompanying cores, although in disarray, were detailed.

**Task 3** – Standards for collecting and recording information about the cores were developed through internal discussions with SCGS geologists, GIS, and database developers and in

consultation with the Delaware Geological Survey (DGS). The DGS has a mature digital data program, having a robust water-well database and an extensive core and sample library. In addition, a procedure for developing metadata records was finalized. We considered the Excel template to be the most practical means of transferring metadata to the NDC.

**Task 4** – Geophysical logs from SCDNR's Hydrology Section were digitized. This task supported Hydrology's efforts to develop digital records of their log library in preparation of developing an online database. Our contribution was to develop a simpler method for generating LAS files. Previously, the Hydrology Section relied on outdated software and a no longer supported digitizing board to make digital logs of old paper records. Using our experience in generating digital map products, we offered Hydrology a simpler, on-screen method of digitization. We experimented with several methods of digitizing before we settled on a program that was part of our RockWare suite. We also developed a methodology to translate the LAS files into clean digital images (PDF or jpeg). A total of 88 various geophysical logs were digitized.

**Task 5** – Metadata records were created for all records with location coordinates. This included all 88 geophysical logs and 239 cores in the Repository. Since the submission of this metadata, location information for more cores has been found from maps in files. At this time, it is unknown how many other cores can be located and added to the database. The format for submitting metadata records to the NDC was the Excel template. After consulting with DGS and internal discussions, the Excel template was considered the most flexible means of submitting metadata.

## Conclusions

Overall, the results of FY08's NGGDPP project are positive. Some tasks were successful and completed; and other tasks were successful, but not entirely finished.

- The inventory of the Core Repository was completed; however, only 239 records were submitted to NDC from a total of 1,489 cores. This discrepancy indicates that an inadequate job was done when cores were first received into the Repository. We hope to rectify this problem by implementing better procedures for accepting new cores. We will continue to seek locations of cores and add them to the Core Repository database.
- Digitizing geophysical logs was successful. We developed a methodology for making LAS files and digital illustrations from the LAS files. With these two successes, we can continue to increase our digital collection of geophysical logs.
- Metadata records of locatable cores and digitized geophysical logs were successfully developed and delivered to NDC. As other cores are located and new geophysical logs digitized, additional information will be added to NDC.