

**FINAL TECHNICAL REPORT**

**Award Number:  
G09AP00095**

**Title:  
FDEP/Florida Geological Survey Metadata Project (Part III)**

**Author:  
Clint Kromhout, Principal Investigator  
903 West Tennessee Street  
Tallahassee, FL. 32304  
Phone: (850) 488-4191  
Fax: (850) 488-8086  
[Clint.Kromhout@dep.state.fl.us](mailto:Clint.Kromhout@dep.state.fl.us)**

**Term of Award:  
September 15, 2009 to September 14, 2010**

**Submittal Date of Final Technical Report:  
December 15, 2010**

## Proposal Abstract

The purpose of this project is to continue the work by the Florida Department of Environmental Protection/Florida Geological Survey (FDEP/FGS) in the effort to enhance metadata records as started in the previous phase of the National Geological and Geophysical Data Preservation Program (NGGDPP) granting period. The end goal is to submit to the National Catalog portal metadata relating to the FDEP/FGS sample collections that have undergone a strict quality assurance and quality control (QA/QC) process and are therefore reliable and accurate data sets. The previous project was comprised of two tasks 1) a GIS-based location refinement of up to 10,000 well entries and 2) metadata creation and refinement in XML format for those wells. In order to complete metadata for the FDEP/FGS collections, well record validation and refinement of locational accuracy of well cuttings, cores, and core chips are required. The FDEP/FGS sample collection is made up of three sample types (collections) which have up to five different recorded locations which may, or may not, all agree. Combined, these three collections represent approximately 19,000 wells. The current phase of the NGGDPP grant is projected to refine up to 10,000 records. We wish to continue this process in order to complete the metadata clean up for the entire collection of 19,000 wells.

The FDEP/FGS has been working for a considerable number of years to develop a digital database of its geological collections. Impediments to progress include insufficient staffing, funding, and bureaucratic obstacles. As a result, the preservation of and accessibility to the FDEP/FGS data has suffered. End users of the data are faced with quality assurance and quality control (QA/QC) issues within the metadata. Rejuvenated interest in the last four years has led to some success in database development; however, metadata population challenges exist due to locational uncertainties and limited staffing to migrate relevant data from paper records to the database following appropriate QA/QC procedures. The granting of the requested funds will greatly benefit both the Nation and State by not only improving and preserving the data in a nationally consistent format, but by opening the collection to serve a wider audience through the NGGDPP National Catalog portal.

## Technical Report

### *Overview*

In the 2009 project year (fiscal year 2009/2010), the Florida Geological Survey was awarded funds by the United States Geological Survey (USGS) through the National Geological and Geophysical Data Preservation Program (NGGDPP) to continue (phase three) providing the USGS with metadata from our well cuttings and cores collections. This work was initiated in phase two, 2008 (fiscal year 2008/2009), and was funded by the same program. Funding from the first phase in 2007 (fiscal year 2007/2008) allowed FDEP/FGS to inventory our sample collections, determine their current standing, assess future needs and conduct a number of improvements. These improvements helped to ready FDEP/FGS for undertaking the second and third phases which involved validation and refinement of individual records within the collections.

The first phase of the NGGDPP program led to the establishment of two goals for the FDEP/FGS's collections. The first goal was to address physical preservation of the collection data that is in danger of loss through the decline of historic data sources such as fading ink and deteriorating boxes in the physical sample collections. The second goal was to clean up paper and electronic records for the sample collections and create corresponding metadata for an internal database. Work on this second goal comprised a majority of the effort expended during the second (2008) and third (2009) phases of funding. This need was expanded to our own internal database to include submission to the National Catalog. Developing plans of action for these goals as well as the work done in the three phases of the NGGDPP granting program have helped to conceptualize a draft of a long range data-preservation plan.

During phase two much internal learning and growth took place which led to the redevelopment of the GIS data refinement process used to verify a digital database of recorded locations for the cuttings and core collections against historical hardcopy data sources. Phase two's efforts yielded a total of 3,316 verified locations and 5,799 (includes locations previously verified by other projects) records uploaded to the National Catalog (Appendix A).

A process redevelopment evolved in the form of a new master reference dataset which combines the well location information from the best paper data sources (Blackbooks and Graybooks) with recent project-specific data sets. Creation of the new master reference dataset was initiated at the end of phase two in 2008 and continued on into phase three in 2009. A detailed discussion regarding the paper data resources and creation of the new master reference dataset along with a data dictionary is available in phase two's final technical report.

### *Progress and Momentum Shifts*

In August 2009, before the start of phase three, project team-member responsibilities were purposefully shifted to maximize professional experience opportunities for the GIS Analyst. The GIS Analyst was already heavily involved in the project, charged with compiling and creating the new master reference dataset and shapefile. To facilitate the managerial transition, the principal investigator set up budget tracking and provided training for the GIS Analyst, emphasizing the need to keep the QA/QC team on schedule. Creation of the file was not expected to require significant time. The principal investigator still remained involved with the project, but only in a peripheral role with regards to core project planning and direction.

After experiencing management-related challenges, the new master reference dataset and shapefile were completed. Noting significant delays in project forward momentum, project management was shifted back to the principal investigator.

As anticipated the new master reference dataset and shapefile drastically sped up the well-location verification process; however the turnover of part-time staff also slowed progress. Initially, the QA/QC team was verifying each county twice, but challenges to the project management, budget and timeline required revision of the QA/QC process to include only single-pass location verification.

As an additional QA/QC measure, the first 730 well locations verified in phase two (Year Two) were re-verified during phase three. These 730 wells were initially verified during training of project staff. Only one percent of those wells were found to contain errors - a satisfactory outcome as it indicated the original method of verification, although comparably slow and laborious, was very accurate. For consistency, metadata for all 730 locations was re-uploaded to the National Catalog to append the previously uploaded records.

For comparison, during phase two 3,316 locations were verified at a rate of approximately 276 verifications per month. Phase two used the older method for verifying locations which was clearly slower. During phase three, despite experiencing management-related challenges and staff turnover, 5,659 locations were verified and the average verification rate was approximately 470 well records per month. In the most productive months of the project, verification rates exceeded 1,000 per month. The aforementioned delays resulted in an unanticipated shortfall of achieving the project goal of up to 10,000 verifications.

Metadata for 5,659 verified locations were uploaded to the National Catalog. It is apparent that verifying each county twice would not have led us to reach our goal in a single project year; however, verifying each county once would have yielded over 10,000 locations verified and uploaded. In total for phase two through three, metadata for 11,458 total records have been uploaded to the National Catalog. Just over 7,700 records remain to be verified and uploaded. The FGS, having refined the process, hopes to gain future funding from the NCGDPP to complete the location verification and we are confident our overall goals can be met.

*Appendix A*

*Areas of verified well locations and total numbers of wells uploaded to the National Catalog.*

