

# NGGDPP Final Technical Report

**Award Number:** G13AP00087

**Project Title:** Digital Conversion of Paper Metadata Records for 5000 Wells that have Cores and Cuttings stored at BEG Houston Research Center: Ultimate Inclusion into National Digital Catalog

**Name of the State Geological Survey:** University of Texas Bureau of Economic Geology

**Principal Investigator:** Beverly Blakeney DeJarnett  
Research Science Associate  
University of Texas at Austin  
Bureau of Economic Geology  
Houston Research Center  
11611 West Little York Road  
Houston, TX 77041  
Ph: 713-983-9420  
e-mail: [bev.dejarnett@beg.utexas.edu](mailto:bev.dejarnett@beg.utexas.edu)

**Authorized Institution:** University of Texas at Austin  
Bureau of Economic Geology  
University Station, Box X  
Austin, TX 78713-8924  
Ph: 512-471-1534 Fax: 512-471-0140

**Total Project Period:** 9/1/2013 through 9/1/2014

**Date of Report:** 4/1/015

## **Digital Conversion of Paper Metadata Records for 5000 Wells that have Cores and Cuttings stored at BEG Houston Research Center: Ultimate Inclusion into National Digital Catalog**

### **Abstract**

The University Of Texas Bureau Of Economic Geology (BEG) serves as the geological survey for the state of Texas. In that capacity, BEG serves as the repository and curator for over 2 million boxes of geological samples (cores and cuttings) and an abundant amount of geoscience data (geophysical logs, thin sections, geochemical analyses, etc.). Most material is accessible and searchable online, and it is publically available to the geological community for research and education.

BEG has received the vast majority of its rock materials through donations of collections formerly acquired and held privately by industry. As part of one of these donations, BEG received thousands of cores and cuttings with either no associated digital metadata, or digital metadata that was woefully incomplete.

All of the information regarding the geologic samples (cores and cuttings) from this donation is currently on approximately 4000 pages of letter-sized paper and on an additional 25,000 4" x 6" index cards. This compilation of information for the physical samples does not exist anywhere else besides BEG because the company donated it to BEG with the rock material. To make the geologic samples discoverable and accessible to the public, BEG converted this data to digital files by scanning the index cards and by manually entering API numbers for each scanned image

to name each file. BEG attempted to use commercially available optical character recognition (OCR) software (ABBY Fine) in order to quickly capture the metadata on each card and organize it into spreadsheets, but unfortunately the OCR software was not capable of recognizing handwritten text in an accurate way. (The OCR software was not purchased with NGGDPP funds.) The digital file for each well with samples was entered into BEG's online sample and log SQL database, and it was provided to the National Digital Data Catalog as an update to metadata that BEG has provided to the National Digital Catalog as part of past NGGDPP projects.

### **Project Summary**

Metadata included on each paper index card included, but was not limited to, the following: location (country name; state; county; township, section and range; and latitude/longitude where possible); Operator Name; Well Name and Number; Field Name; API number (unique well identifier); top depth of interval sampled; bottom depth of interval sampled; type of physical sample; whether or not the samples had been analyzed by micropaleontologists, and a unique sample identifier.

Two high school seniors scanned 9,407 paper index cards in the summer of 2014. While scanning, the student entered the API number for each well as a file name and organized the files into folders based on state and then counties in each state. The PI (DeJarnett) was responsible for introducing and describing the project to the students, and the PI was also present as the actual scanning part of the project began to ensure quality control of both the scanned images

and data entry. The PI also periodically checked the data entry and scanned images for errors and/or omissions throughout the life of the project.

The technical clerks also utilized commercially available OCR software (ABBY Fine) in an attempt to quickly capture the metadata that was included on each index card. However, the cards contained a mixture of typewritten and handwritten text, and the OCR software was not able to recognize the handwritten text accurately. Therefore, the clerks organized the scanned images into folders (state and then county), and then individual files for each well with the file name being the API number for each well. These files were then sent to Aaron Averett at BEG Austin for inclusion into BEG's database (<http://igor.beg.utexas.edu/crc2/>) and ultimately to the National Digital Catalog in the requested format.

### **Projected Goals and Project Deliverables**

The two technical clerks scanned almost twice as many paper copies of well information (9407 index cards as opposed to the original proposed estimate of 5000 scans (with each index card and scan representing a well with physical samples in BEG's collection). Excellent progress was made; however, much remains to be done to finish the rest of the paper material. We used Optical resolution software (OCR – ABBY Fine) in an attempt to categorize the metadata found on each index card into a spreadsheet format; however, the current OCR software was still not able to adequately read handwritten text. Therefore, the clerks manually entered the API number for each scanned image (representing 1 well per image). These were then filed digitally under county and state. The API numbers were then used to import the scanned images into BEG's database <http://igor.beg.utexas.edu/crc2/> .

This information is now available to users of the USGS National Digital Catalog, and 9,407 pieces of metadata (the scanned images for each well) were provided to USGS in the format requested.

Projected Goal: estimated 5000 scanned images

Actual # of scanned images: 9,407

Choctaw Co., Alabama  
Pruet & Hughes  
# 1 Gibson  
S. Gilbertown  
Sec. 11 - 10N - 3W

~~8-11-67~~

8-8-67 11,000 - ~~11,400~~

Bus

12380

**L2361**

"A"

*Rinsed Samples Only*

ENTERED INTO EDP 6/80

API UNIQUE # 20008

8-29-68

Figure 1. Scanned example of a typical index card from the original inventory of cuttings and cores received from Amoco's domestic collection. The inventory consists of thousands of paper index cards with metadata for all wells in the original Amoco collection of physical samples that were donated to University of Texas Bureau of Economic Geology.