

Final Technical Report

Prepared for the
National Geological and Geophysical Data Preservation Program

**Generating Metadata for
Part of the Tennessee Geological Survey's Coal Mining Collection
and Digitizing Paper Records for
Part of the Coal Mining and Zinc Mining Collections
from July 1, 2012 to June 30, 2013**

United States Geological Survey
Award No. G12AP20132

Principal Investigator
Ronald P. Zurawski

Tennessee Department of Environment and Conservation
Tennessee Geological Survey
William R. Snodgrass TN Tower
312 Rosa L. Parks Ave., 12th Floor
Nashville, TN 37243
Telephone: 615-532-1502
Fax: 615-532-0199
E-mail: Ronald.Zurawski@tn.gov

September 25, 2013

Abstract

For FY2012 the National Geological and Geophysical Data Preservation Program awarded the Tennessee Geological Survey funds to continue the process of producing metadata records for the National Digital Catalog and to continue the conversion of its paper records into digital format. The proposed work was divided into two projects: (1) to create metadata records and digitally capture 1000 items in the Coal Exploration Core Hole Logs Collection (P1334); and (2) to digitally capture 1000 items in the Zinc Mining Maps and Reports Collection (P1326). Both of the objectives were successfully completed within the grant award period. A total of 1026 Coal Exploration Core Hole Log metadata records were compiled and the documents associated with these records have been converted to digital format scans that are safely stored, duplicated, and available for distribution as Portable Document or Tagged Image Format files. In addition, a total of 1020 maps, cross-sections and drill logs contained in the Zinc Mining Maps and Reports Collection have been scanned and the digital files are safely stored, duplicated, and available for distribution available as Portable Document or Tagged Image Format files. We also updated the Coal Exploration Core Hole Logs and Zinc Mining Maps and Reports collection metadata to indicate which documents are now available in digital format. The revised files have been uploaded and the records harvested onto the United States Geological Survey's ScienceBase website. As a result of this effort customers will find it easier to determine that this geologic information is available and it has benefitted the Tennessee Geological Survey by saving time responding to requests for the documents and information contained in these collections.

Introduction

The importance of accurately cataloging Tennessee Geological Survey (TGS) collections and creating digital versions of the documents in them continues to increase as potential users expect to have the ability to remotely query collection records and to obtain them electronically through various digital media. It is a goal worth striving for because the outcome helps ensure the preservation of historic documents, as well as decrease the staff time and the expense required to fulfill requests for information contained in these collections. Through the financial assistance provided by the U.S. Geological Survey (USGS) National Geologic and Geophysical Data Preservation Program (NGGDPP) TGS is making steps towards reaching that goal.

Nearly all of TGS's Coal Mining collections and Zinc Mining Maps and Reports Collection are unpublished and in paper/mylar form. As a result, companies, government agencies, and the general public are unaware of the abundant site specific geologic data available in our Zinc and Coal Mining collections. At the present time, the TGS has not reached the point where it can provide the on-line presence needed for its collections. Therefore preparing metadata records for uploading into the National Digital Catalog (NDC) is the first step towards providing the public with a method to search for, locate, and evaluate the type of information that is available in them.

The Coal and Zinc Mining collections are a valuable source of information that can be used to improve estimates of the state's remaining coal and zinc resources and for companies to evaluate potential areas for future exploration. Some of the items in these collections are between 50 to 100 years old and rapidly deteriorating. We are unable to estimate the relative cost in today's dollars that would be incurred to reproduce the geologic information contained in these collections, but it would be substantial. For that reason, it is imperative that we make every effort to preserve the information contained in the items in these collections.

Work Completed

The two broad objectives for our NGGDPP FY2012 project were to compile metadata in a format suitable for input into the NDC and to convert documents into a digital format.

The first objective focused on the Coal Exploration Core Hole Logs Collection. The metadata for a total of 1026 records was completed, which slightly surpassed the target of 1000 supported by this year's award. The metadata file for the Coal Exploration Core Hole Logs Collection was created in Microsoft Excel and saved in .csv file format using the pipe (|) record delimiter character as a field separator for seamless uploading. The file was uploaded onto the USGS ScienceBase (NGGDPP website) and the records have been harvested by USGS staff.

In addition to providing the metadata records for the Coal Exploration Core Hole Logs Collection, the following outcomes were realized as a result of completing this objective:

- (1) We continued to review and better organize the material in the Coal Exploration Core Hole Logs Collection;
- (2) An in-house Excel database started last year was populated with the new records. The importance of this database is that it includes additional information not included in the NDC to assist TGS staff with the searching and sorting of records, as well as the storage location of each document cataloged to date;
- (3) The existing Excel database also feeds directly into a GIS (ArcMap) project to better visualize the geographic location and query the documents in the collection.

The second objective successfully completed this year was to continue the conversion of paper, mylar, and linen documents into a digital format. In order to fulfill this objective we were required to:

- (1) digitally capture 1000 documents in the Coal Exploration Core Hole Logs Collection (a total of 1026 documents were completed) and;
- (2) digitally capture 1000 documents in the Zinc Mining Maps and Reports Collection (a total of 1020 documents were completed);

All documents larger than 11x17 inches were scanned at 400 dpi resolution in either grayscale or RGB color, uncompressed tagged image file format (.tiff) on a large format scanner. Page-sized and some tabloid sized documents were scanned in Portable Document File (PDF) format at 300 dpi resolution in either grayscale, black and white, or RGB color format on a Canon ImageRunner. Adobe Acrobat software was used to create a second digital file in PDF format for all files originally scanned in .tiff format. When necessary, free software called GIMP (GNU Image Manipulation Program), was used to crop, rotate, and digitally enhance the scanned documents.

The TGS staff involved in the project built upon lessons learned from the previous year's document scanning and digitizing effort. Even though we were using the same equipment, prior experience resulted in a smoother workflow from document retrieval, to scanning, to digital enhancement of the documents in the collections. This was also true with regard to file naming,

duplicating, storing, and all the accounting that goes along with verifying the completed tasks as a particular document moves through the process. With the advantage that experience provides, one may question why the final document scan total for the Zinc Mining and Reports Collection turned out to be less than last year (1020 vs 1057 items). This occurred in part because some of the very large plate-sized documents required special handling, the use of a scanning sleeve, and more time for the computers to scan, save, open, modify and copy the digital files.

In addition to converting documents in the Coal Exploration Core Hole Logs Collection and Zinc Mining Maps and Reports Collection into a digital format, the following results were realized by working on this objective:

- (1) An additional 73 records were added to the Zinc Mining Maps and Reports Collection. Most of the new records represent driller's logs. These logs were either previously represented by a single record or not included in the dataset because we did not have their location coordinates. As our familiarity with this collection continues to improve we are able to understand and decipher relationships between documents that were once a poorly identified.
- (2) We updated the metadata records to indicate which documents are now available in digital format in each of the collections.
- (3) We uploaded the .csv format files for the updated collections on the USGS ScienceBase website, revised the metadata describing the collections, and requested that the old versions of the collections be removed from the website.

Tennessee Geological Survey Staff Participation

A number of TGS personnel were involved in completing the work for this year's NGGDPP project. Chief Geologist Dr. Peter Lemiszki in TGS's Knoxville Field Office was responsible for scanning, image processing, and updating the metadata file for 1020 documents in the Zinc Mining Maps and Reports Collection. He was also responsible for completing the Final Technical Report and submitting the datasets onto the USGS ScienceBase website. Geologist Barry Miller in TGS's Knoxville Field Office was responsible for scanning, image processing, and creating the metadata files for the Coal Exploration Core Logs Collection. Secretary Becky Hawkins in the Knoxville Field Office assisted Pete and Barry with the scanning of page-size to tabloid-size documents from both collections.

Publications

Currently there are no plans to publish this information in any other form than what is available on the USGS ScienceBase website.