

Final Technical Report

Prepared for the  
National Geological and Geophysical Data Preservation Program

**Generating Metadata for  
Part of the Tennessee Division of Geology's Coal Mining Collections  
and Digitizing Paper Records for  
Part of the Coal Mining and Zinc Mining Collections  
from July 1, 2011 to June 30, 2012**

United States Geological Survey  
Award No. G11AP20172

Principal Investigator  
Ronald P. Zurawski

Tennessee Department of Environment and Conservation  
Division of Geology  
13th Floor, L & C Tower  
401 Church Street  
Nashville, TN 37243-0445  
Telephone: 615-532-1502  
Fax: 615-532-1517  
E-mail: [Ronald.Zurawski@tn.gov](mailto:Ronald.Zurawski@tn.gov)

September 28, 2012

## **Abstract**

For FY2011 the National Geological and Geophysical Data Preservation Program awarded the Tennessee Division of Geology funds to continue the process of producing metadata records for the National Digital Catalog and to start the conversion of its paper records into digital format. The proposed work was divided into three projects: (1) to digitally capture 1000 items in the Zinc Mining Maps and Reports Collection (P1326); (2) to digitally capture and create metadata records for the 500 items in the Coal Exploration Core Hole Logs Collection (P1334); and (3) to digitally capture and create metadata records for all 325 items in the Coal Geology Maps Collection (P1345). All of the objectives were successfully completed resulting in a total of 506 Coal Exploration Core Hole Log records and 346 Coal Geology Map records submitted for uploading into the National Digital Catalog. The documents associated with these records have been scanned and the digital files are safely stored, duplicated and available as Portable Document or Tagged Image Format files. In addition, a total of 1057 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 as Portable Document or Tagged Image Format files. We have also updated the Zinc Mining Maps and Reports Collection metadata to indicate which documents are now available in digital format and submitted the revised metadatabase. As a result of this effort customers will find it easier to determine that this geologic information is available and it will benefit the Tennessee Division of Geology by saving time responding to requests for the documents and information contained in these collections.

## **Introduction**

The importance of accurately cataloging Tennessee Division of Geology (TDG) 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) TDG is making steps towards reaching that goal.

Nearly all of TDG'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 TDG 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 condition of the items in these collections.

### **Work Completed**

The two broad objectives for our NGGDPP FY 2011 project were to compile metadata in a format suitable for input into the NDC and to convert paper/mylar documents into a digital format. The first objective involved compiling the metadata for the entire Coal Geology Maps Collection and for 500 items in the Coal Exploration Core Hole Logs Collection. This objective was achieved and a total of 952 records were submitted for uploading into the NDC; 506 Coal Exploration Core Hole Log records and 346 Coal Geology Map records.

In addition to providing the metadata records for the NDC, the following outcomes were realized as a result of completing this objective:

- (1) a review of the material in the Coal Geology Maps Collection and 500 logs in the Coal Exploration Core Hole Logs Collection;
- (2) an Excel database of the items in these collections that contains additional fields for each record to assist TDG staff with the sorting and retrieval of pertinent information when a request is made;
- (3) a GIS-based (ArcMap) project to better visualize the geographic extent and location of items in these collections; and
- (4) the specific storage location of all of the items in these collections was provided to TDG staff.

The second objective successfully completed this year launched TDG into the formal data preservation process. This objective consisted of three separate components:

- (1) to digitally capture 1000 documents in the Zinc Mining Maps and Reports Collection (a total of 1057 documents were completed);
- (2) to digitally capture 500 documents in the Coal Exploration Core Hole Logs Collection (a total of 506 documents were completed); and
- (3) to digitally capture all of the documents in the Coal Geology Maps Collection (a total of 346 documents were completed).

Although the scanning and digitizing procedure is a relatively straightforward process on paper, the effort required to complete this second objective was quite substantial for a number of reasons. First, the 400 dpi, grayscale or color, RGB, uncompressed tagged image file format used for the majority of the scanned documents (larger than 8½ by 11 inches) resulted in large file sizes. As a result, the transfer of files between computers was a slow process when the network was used because it required that the file first be transferred to the server and then from the server to the appropriate desktop computer. We therefore moved towards using external hard

drives to initially store the raw scanned files and to transfer them between computers. Second, our desktop computer processing speed, amount of RAM, and available internal hard-disk space affected our ability to efficiently conduct image processing on the scanned images. We were able to overcome this issue by temporarily transferring RAM from another computer and by using a newer computer that was available to us. In addition, we downloaded free software called GIMP (GNU Image Manipulation Program), which worked well with the large file sizes (Figure 1). Third, we did not initially realize the importance of properly naming the digital files so that the names would be concise, consistent, and clear enough to identify the original document. As a result, some file names used for documents scanned at the beginning of the project had to be modified later, which then required making changes in the pertinent databases. Finally, some documents listed as a single record in the metadata for that collection required numerous steps in order to complete the scanning process. For example, a report in the Zinc Mining Maps and Reports Collection may consist of pages of text, fold-outs, and plates. The report had to be disassembled, the text and maybe the fold-outs scanned on one machine and the large plates scanned on a separate machine.

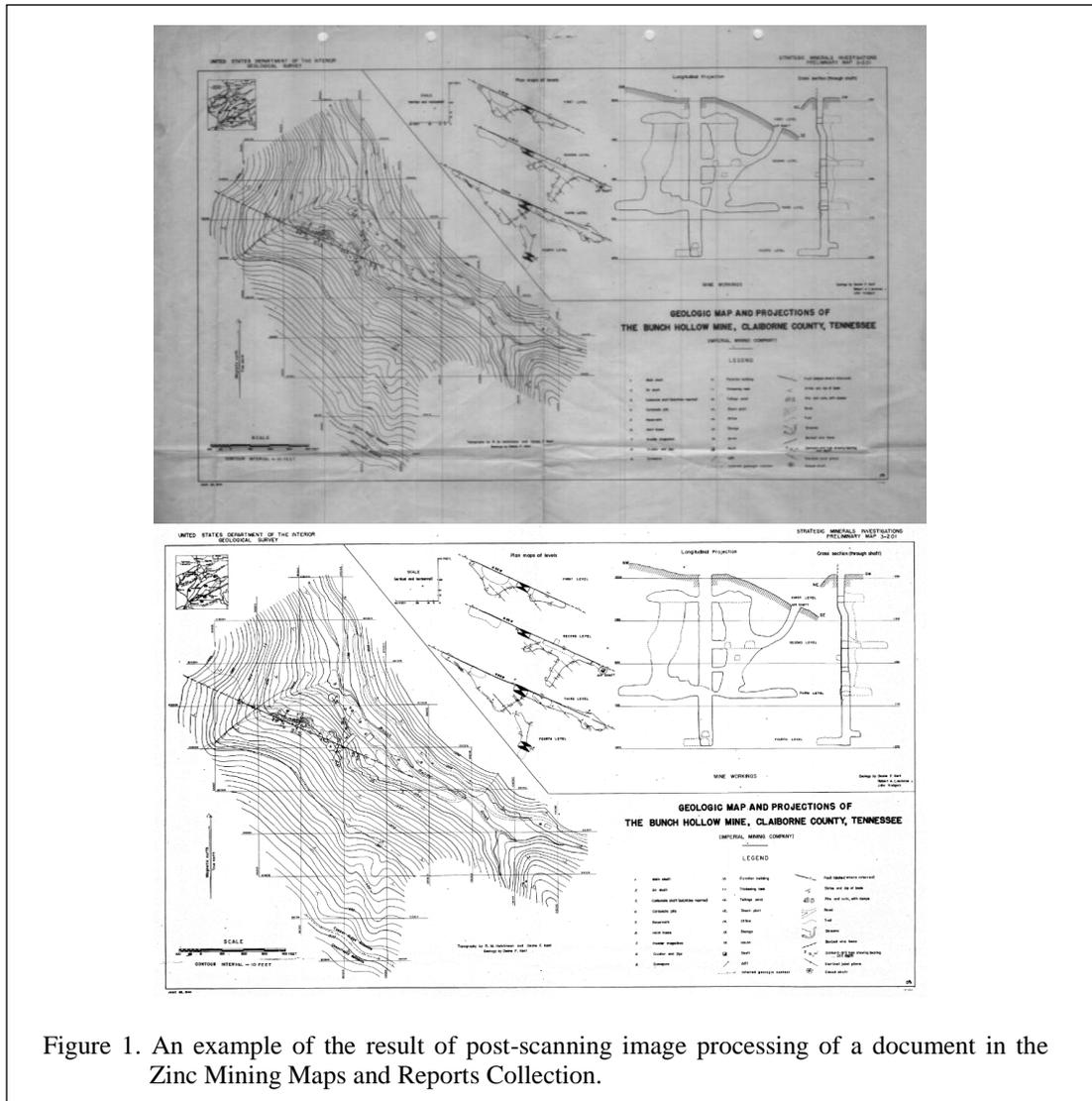


Figure 1. An example of the result of post-scanning image processing of a document in the Zinc Mining Maps and Reports Collection.

The separate scanned files were then transferred to the only computer containing Adobe Acrobat software and reassembled digitally (after any necessary image processing) to reconstruct the original report in Portable Document Format. All of that effort contributed towards the completion of only one scanned item in the collection. So to conclude, because of the aforementioned and other reasons, TDG has benefitted from the experience gained this year related to the digital preservation of paper records.

The metadata files for the Coal Geology Maps Collection and the Coal Exploration Core Hole Logs Collection completed this year were created in Microsoft Excel and saved in .csv file format using the pipe (|) record delimiter character as a field separator. Although all of the mandatory metadata elements were provided, we had to incorporate the use of a bounding box to define the extent of the information contained in some of the documents. ESRI's ArcMap software was used to determine the geographic coordinates (lat/long) for the centroid of all bounding boxes. The centroid coordinates were entered into the metadata Coordinates field and the associated bounding box coordinates were entered into the AlternateGeometry field for the non-point specific items. In addition, the metadata file for the Zinc Mining Maps and Reports Collection was updated in order to indicate which items in the collection are available in digital format. All of the .csv files for the three collections have been uploaded onto the USGS ScienceBase (NGGDPP website) awaiting final loading by NGGDPP staff into the NDC.

### **Tennessee Division of Geology Staff Participation**

A number of TDG personnel were involved in completing the work for this year's NGGDPP project. Chief Geologist Dr. Peter Lemiszki in TDG's Knoxville Field Office was responsible for scanning, image processing, and updating the metadata file for 1057 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 TDG's Knoxville Field Office was responsible for scanning, image processing, and creating the metadata files for the Coal Geology Maps and Coal Exploration Core Logs collections. Secretary Becky Hawkins in the Knoxville Field Office assisted Pete and Barry with the scanning of page-size to tabloid-size documents from the above named collections.

### **Publications**

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