Final Technical Report

Prepared for the
National Geological and Geophysical Data Preservation Program

Generating Metadata Records and Digitizing Documents in the Tennessee Geological Survey's Coal Exploration Core Hole Logs, Coal Mine Maps, Mineral Resources, and Field Notes & Maps Collections from September 1, 2014 to August 31, 2015

United States Geological Survey
Award No. G14AP00148

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

November 25, 2015

Abstract

For FY2014 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 work consisted of four projects:

(1) To create the metadata records and convert to a digital format the remaining 500 documents in the Coal Exploration Core Hole Logs Collection. The result of this project was that 507 core hole logs were digitally captured with associated metadata.

(2) To create the metadata records and convert to a digital format 500 documents in the Coal Mine Maps Collection. The outcome of the project was that 557 coal maps were digitally captured with associated metadata.

(3) To create the metadata records and convert to a digital format 1000 documents in the Mineral Resources Collection. The metadata file currently consists of 690 records and represents nearly all of the documents in the Mineral Resources Collection located in the Knoxville Field Office. More than 1000 document scans are associated with the 690 records.

(4) To create the metadata records and convert to a digital format 100 field books and ancillary documents in the Field Notes and Maps Collection. The outcome of the project was that the metadata for 324 field books and ancillary documents were compiled and 294 field books were scanned.

All of the digital files are safely stored, duplicated, and available for distribution as Portable Document Format files. All of the metadata files have been uploaded onto the United States Geological Survey's ScienceBase website, but have yet to be harvested.

Introduction

The importance of cataloging Tennessee Geological Survey (TGS) collections and digitizing the documents in them continues to increase as potential users expect to have the ability to remotely query collection records and to acquire a document in a useful electronic format. Although a slow and somewhat tedious process, which has enlisted staff geologists that are trained to catalog geologic data in the field, the value added to the mission of TGS cannot be overstated. 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 continues to make steps towards reaching that goal. The following is a brief summary of the work completed for each of the four data preservation projects that received USGS support in FY 2014.

Project 1: Coal Exploration Core Hole Logs Collection Project

Prepared by Barry W. Miller
Tennessee Geological Survey

Work Completed

The proposed objective of this project was to create the metadata records and convert to a digital format 500 documents from the Coal Exploration Core Hole Logs Collection located in the Tennessee Geological Survey's Knoxville Field Office. This year's project focused on core hole logs located in the northwestern portion of the Tennessee coal field. Numerous core hole logs have been obtained in the past for this area from neighboring geologic surveys, coal companies, and consulting firms. An effort to review and organize this material has not been undertaken because this part of the Tennessee coal field is currently not a target for coal mining and exploration. The Tennessee Geological Survey is using this project as an impetus to catalog and locate these various core hole logs. During this project it was determined that many of the records did not have any location information attached to them. TGS's Coal Mine Map Collection was utilized to determine the coordinates and surface elevations for most of the core hole logs.

Page-sized and up to A4-sized documents were scanned in Portable Document File (PDF) format at 300 dpi resolution in either grayscale or RGB color format on a Canon ImageRunner. All larger documents were scanned at 400 dpi resolution in either grayscale or RGB color, uncompressed tagged image file format (.tiff) on a large format scanner. When necessary, free software called GIMP (GNU Image Manipulation Program), was used to crop, rotate, and digitally enhance the scanned documents. Adobe Acrobat software was used to create a second digital file in PDF format for all files originally scanned in .tiff format. Acrobat software was also used to combine separate scanned files into a single document file when required. Multiple copies of the digital files are stored on hard drives at separate locations with plans to move them to a secure server hosted by the Tennessee Department of Environment and Conservation.

The result of this project was that 507 core hole logs were digitally captured with associated metadata. This outcome slightly exceeded the target of 500 documents originally proposed. 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) awaiting harvesting 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) Our in-house Coal Exploration Core Hole Logs Collection Excel database 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 in the collection;

(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.

Tennessee Geological Survey Staff Participation

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.

Project 2: Coal Mine Map Collection Project

Prepared by Barry W. Miller
Tennessee Geological Survey

Work Completed

The proposed objective of this project was to create the metadata records and convert to a digital format 500 coal maps from the Coal Mine Map Collection located in the Tennessee Geological Survey's Knoxville Field Office. This year's project concentrated on the 15 minute quadrangle coal compilation map portion of the Coal Mine Map Collection. These maps were produced in the 1950's and consist of a compilation of coal geology information for mineable coal seams on a 15 minute quadrangle base outline at a scale of 1;24,000. These maps were drawn on "onion skin" tracing paper with coal information consisting of underground and surface coal mine locations and outlines; locations of coal exploration drill holes with seam data; and locations of measured coal sections with seam information. Most of the maps measured approximately 40 inches by 50 inches and many were in delicate condition due to the age of the maps.

The coal maps were scanned at 400 dpi resolution in either grayscale or RGB color, uncompressed tagged image file format (.tiff) on a large format scanner. Due to the fragile nature of many of the 60 year old coal maps, a large sized protective Mylar sleeve was obtained to protect the maps during the scanning process. When necessary, free software called GIMP (GNU Image Manipulation Program), was used to crop, rotate, and digitally enhance the scanned maps. The digital scan files were organized and identified according to the 15 minute quadrangle number designation and coal seam name. Multiple copies of the digital files are stored on hard drives at separate locations with plans to move them to a secure server hosted by the Tennessee Department of Environment and Conservation.

The outcome of the project was that 557 coal maps were digitally captured with associated metadata. This surpassed the objective of 500 coal maps originally proposed. The metadata file for the Coal Mine Map 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) awaiting harvesting by USGS staff.

In addition to converting coal maps in the Coal Mine Map Collection into a digital format and creating metadata for the maps, the following results were realized by working on this objective:

(1) We continued to assess and better organize the material in the Coal Mine Map Collection;

(2) The digital scan files can be reviewed for coal information without disturbing the original fragile coal maps. These digital scan files can be also be utilized in GIS software to create shapefiles, databases, and maps that display the characteristics of each coal seam;

(3) Our in-house Coal Mine Map Collection Excel database 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 in the collection;

(4) 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.

Tennessee Geological Survey Staff Participation

Geologist Barry Miller in TGS's Knoxville Field Office was responsible for scanning, image processing, and creating the metadata files for the Coal Mine Map Collection.

Project 3: Mineral Resources Collection Project

Prepared by Peter J. Lemiszki
Tennessee Geological Survey

Work Completed

The proposed objective of this project was to create the metadata records and convert to a digital format 1000 documents in the Mineral Resources Collection located in the Tennessee Geological Survey's Knoxville Field Office. This goal was achieved, but entailed a greater effort than originally planned. The initial proposal only considered the two sets of Mineral Resource files in the Knoxville Field Office, which were organized and cross filed based on two criteria: mineral commodity and geographic location (county). However, the Knoxville Field Office acquired two more sets of mineral resource files from our Nashville office in 2014. These mineral resource files were obtained years ago from the Tennessee Valley Authority and were also organized and cross filed based on mineral commodity and geographic location (county). As a result, the Mineral Resource files that were examined as part of this year's project consisted of four separate sets of files.

We chose to incorporate all four sets of files into this year's project because quite often there were duplicates of the same document in two or more sets of files. Therefore, it became clear that the most efficient procedure to review, organize, and combine the four sets of files was to examine all of the documents at the same time. The primary reason for this approach being that the attributes of a particular document would be fresh on our mind and we could more easily identify duplicate documents, recall its location, and check to be sure that it is an exact copy. This ultimately prevented us from creating multiple records of the same document in the metadatabase, as well as mistakenly scanning a document more than once. This also afforded us

the opportunity to determine which of the duplicate copies of a document were of the best quality and in the best condition to be scanned and preserved.

There were two important outcomes achieved by completing this data preservation objective. First, creating the metadata records for the documents in the Mineral Resources Collection eliminated any further need to cross file them based on commodity and geographic location. The previous file scheme made sense in the days prior to computer databases. Computer databases provide the capability to search for key terms, such as a county name or mineral resource commodity, in order to effortlessly determine what documents are available. A second related outcome is that scanning all of the documents also eliminated the need for maintaining multiple copies of a particular document to ensure its preservation. As a result, this effort has dramatically reduced the physical file space required to preserve the documents in our Mineral Resources Collection.

The metadata file for TGS's Mineral Resources Collection currently consists of 690 records and represents nearly all of the documents in the Mineral Resources Collection located in the Knoxville Field Office. It is important to note that many individual records actually consist of multiple documents so that the total number of scans completed in FY2014 is over 1000. During our comprehensive review of the documents in the collection we recognized that it made more sense to package closely related correspondence, memoranda, and sometimes reports into a single record. We also combined those documents into a single digital file for their distribution.

The metadata file for the Mineral Resources 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 waiting harvesting by USGS staff. In addition to providing the metadata records for the Mineral Resources Collection, the following outcomes were realized as a result of completing this objective:

(1) Our in-house Mineral Resources Collection Excel database 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 in the collection;

(2) 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.

Page-sized and up to A4-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. All larger documents were scanned at 400 dpi resolution in either grayscale or RGB color, uncompressed tagged image file format (.tiff) on a large format scanner. When necessary, free software called GIMP (GNU Image Manipulation Program), was used to crop, rotate, and digitally enhance the scanned documents. Adobe Acrobat software was used to create a second digital file in PDF format for all files originally scanned in .tiff format. Acrobat software was also used to combine separate scanned files into a single document file when required. Multiple copies of the digital files are stored on hard drives at separate locations with plans to move them to a secure server hosted by the Tennessee Department of Environment and Conservation.

Tennessee Geological Survey Staff Participation

Chief Geologist Dr. Peter Lemiszki in TGS's Knoxville Field Office was responsible for reviewing, sorting, and organizing the documents in the Mineral Resources Collection. He completed the metadata file and supervised secretarial staff to assist him with document scanning. He was also responsible for completing the Final Technical Report and uploading all of this year's NGGDPP project datasets onto the USGS ScienceBase website.

Project 4: Field Books (and Related Documents) Collection Project

Prepared by
Dr. Michael A. Gibson
Department of Agriculture, Geosciences, and Natural Resources
University of Tennessee at Martin

Work Completed

At the onset of the Tennessee Geological Survey (TGS) Field Book Digitizing subproject, it was initially estimated that the TGS housed approximately 100 field books spanning nearly 100 years of work in Tennessee by the TGS and contracted geologists. There were four primary goals of the TGS Field Book Digitization Project. The outcome for each of the goals is listed below:

(1) ***Digitally scan ~100 field books and loose field notes into PDF format at specified resolution.*** – Digital scanning of all field books was completed with the final total being 294 field books and loose note packets. Scanning was accomplished using a Digital Library Systems Group Knowledge Imaging Center (KIC) machine with scanning at the prescribed color 300dpi resolution and saved as PDF's. Blank pages were not scanned to keep file size down. Each field book was scanned with a frame mask that identified the field book author, title, and book number, UT Martin label as the scanning institution, along with an English and metric graphic scale for sizing. Each field book scan was cropped and rotated for easy reading and for space reduction.

(2) ***Digitally scan large format maps into PDF format at specified resolution.*** – There are 39 large format items within the collection that were not scanned; however, the information files were entered into the metadatabase. The large-format scanner necessary for this task is available at the TGS office in Nashville, but requires specialized covering be used so as to not damage the maps. Most of these maps are in fragile shape and deteriorating. Running them through the TGS feed scanner would damage many of the maps and potentially damage the scanner itself. TGS has not obtained the necessary coverings needed to complete the scanning of the large format items. Once the necessary protective covers are obtained by TGS, it will take approximately 2 days to scan and update the information for the 39 maps in the metadatabase.

(3) ***Enter pertinent data for each field book into metadata file***. – Each field book has been entered into a metadata spreadsheet. All available pertinent data for each field book and loose note packet were entered into the database resulting in 294 entries. Entry data includes field book names, authors, dates, abstract of contents, coordinates of studies, and additional information about each book useful for later historical study (e.g., type of book, condition of book, etc.). It should be noted that the field books exhibit a wide range of quality in terms of content and completeness of information within the field books, which often made identification of the prescribed metadata difficult to obtain. This unevenness in quality is a result of the age of the field studies, which are largely before the development of geologic field camps for training geology graduates in standardized mapping procedures and notation. Quality of the writing and informational content of the field books varies from very poor to very good, and vary from author to author. Many of the field books do not contain date information, locations of some study areas, field book authors, map names, etc. Several field books contain conflicting information. Many of the field books contain random notes and "doodling" with little identifying information, summaries from published literature read by the author, and nearly all books contained multiple projects and locations, with many locations only indicated by whose property was being visited (no map information). Field books also were used to record laboratory analyses, but often the author did not include the necessary information to relate these numerical data to specific sample locations, etc. Location coordinate information was necessarily irregular in outline due to the hop-scotch nature of geographic locations within each book. Most of the field books did not lend themselves to giving single set coordinates of individual or contiguously grouped quadrangles. Many projects were laid-out according to irregularly shaped physiographic regions, or multiple regions spanning the entire state. In such instances, we generally used the coordinates for the largest regular quadrangle size that encompassed all of the maps within the project to make the coordinates fit the coordinate field within the metadatabase.

(4) ***Enter pertinent data for each large format map (> 11 x 7 inch) into metadata file*** – All available pertinent data for each large format map were entered into the metadatabase resulting in 39 entries. Entry data includes map names, authors, dates, abstract of contents, coordinates of map, and additional information about each map useful for later historical study (e.g., condition of map, etc.).

Tennessee Geological Survey Staff Participation

Chief Geologist Dr. Peter Lemiszki in TGS's Knoxville Field Office was responsible for providing technical guidance to Dr. Gibson and his students. He also reviewed and made final revisions to the Field Books Collection metadatabase prior to uploading the file onto the USGS ScienceBase website.