Award Number G14AP00140

National Geological and Geophysical Data Preservation Program

Preservation of Geologic Data and Collections in Illinois: Compilation, Documentation and Planning

Final Technical Report

Fiscal Year 2014

9/1/2014-8/31/2015

Ronald Klass
Scientific Database Coordinator
Tel. 217-244-5585
Fax 217-333-2830
rklass@illinois.edu

Zakaria Lasemi
Senior Geologist and Head, Bedrock Geology and Industrial Minerals
Tel. 217-244-6944
Fax 217-333-2830
zlasemi@illinois.edu

Illinois State Geological Survey
Prairie Research Institute
University of Illinois
615 East Peabody Drive
Champaign, IL 61820

November 2, 2015

Abstract

This report summarizes the work performed for the Illinois State Geological Survey (ISGS) FY14 NGGDPP project:

- USBM Illinois Mineral Development Atlas (IMDA) mining borehole datasheets and corresponding location maps for the Lead-Zinc mining district were scanned. The maps were georeferenced and used to precisely locate each borehole. Metadata for each borehole was entered into a database, and the boreholes were reconciled with our existing data. NGGDPP metadata records were created and uploaded to the National Catalog.

- Fluorspar mining district historical documents from the ISGS files were scanned and metadata entered into a database. NGGDPP metadata records were created and uploaded to the National Catalog.

- Lead-Zinc mining district historical documents from the ISGS files were scanned and metadata entered into a database. NGGDPP metadata records were created and uploaded to the National Catalog.

We believe the collection metadata generated and uploaded for this project represents an appropriate and successful use of the resources provided under the project contract.

Collections Processed in FY14

The following collections were processed for FY14. Student hourlies were used to perform most tasks with the guidance and assistance of the PI, including pre-processing/organization, scanning, georeferencing, and data entry.

USBM Illinois Mineral Development Atlas (IMDA)

The IMDA records are likely the most complete and best organized records available for the Lead-Zinc District in northwestern Illinois, covering the period from approximately 1949-1970. They consist of two parts: large-scale (1”=200’) 36”x30” sheet section maps showing mining digs and borehole locations, and a set of 8-1/2” x 11” datasheets containing corresponding logs and mineral analyses (assays and/or visual estimates at various depths). The maps had been separated from the datasheets in different storage areas for many years, and only recently were reunited with the datasheets when discovered during work for a previous NGGDPP project.

The datasheets are masters from USBM and very clear and sharp, but are on very thin onionskin with a carbon impression of the text on the reverse side that apparently was used to make copies, probably through spirit duplication. We also have a set of copies on thicker paper without carbon, but they are faded and fuzzy compared to the originals. The very thin onionskin as well as carbon buildup and smudging/streaking issues precluded the use of an auto-feed scanner so each datasheet was individually hand-scanned on a flatbed scanner to a PDF file. We scanned datasheets for 4,401 boreholes, consisting of downhole logs with formation tops. The majority of logs also include downhole Lead, Zinc, and Iron concentrations (as sulfides), either visually estimated or assayed.

Key information about each borehole, including section, serial number, hole number, elevation, total depth, date, landowner, driller, and water depth was entered from the scanned PDFs into a project database. The data entry form was custom designed to be extremely efficient, taking full advantage of the format and sequencing of the datasheets. The next datasheet to enter was automatically displayed, with many fields of the data entry form pre-populated to minimize keystrokes and data entry errors while at the same time significantly increasing data entry speed over a traditional form. Queries were developed to cross-check the final data for errors.

We also scanned 246 corresponding large-scale section maps depicting precise locations for 4,371 boreholes from the datasheets in a roller scanner as 300 dpi TIFF files; this was more than adequate resolution as these large-scale maps do not contain fine detail. Some of the paper maps had had extensive handling and were torn; archival repair tape was used to repair damage.

While each map corresponded to a section, township, and range in the PLSS system, the original IMDA developers assigned each section in the Atlas a unique “section reference” number (simply 1 through 652) that is used to indicate the section, rather than the unwieldy and error-prone PLSS set of 5 individual items (section, township, township direction, range, range direction).

Any borehole in the Atlas is fully qualified with just two numbers: section reference number and serial number, and in fact the datasheet logs are organized this way. The Atlas includes a section index map to convert between systems, and each section map includes the PLSS section along with the section reference number in the map border.

Using GIS, the maps were georeferenced using section corners. To facilitate borehole and other feature digitization, the 246 map images were programmatically trimmed at the section lines and combined into a single mosaicked GIS layer using automated methods developed by the PI. This eliminated having to constantly re-order and/or turn on/off map layers in GIS when the borders of one map overlapped part of the image of the adjoining map as well as having to handle such a large number of layers.

Each borehole location was digitized and annotated in a GIS database with the section reference number, serial number and hole number. This allowed creation of precise coordinates for each borehole, and allowed each borehole to be tied to its log record by section and serial number, with the hole number as a cross check. Besides boreholes, the maps also depicted major mines as well as numerous mining digs which were also digitized for possible future use.

Probably the most challenging part of the project was reconciling the IMDA boreholes with our existing Lead-Zinc data. We knew roughly half of the boreholes were in our primary wells and borings database, either from prior survey projects/work, or from a prior NGGDPP project which entered paper log records from our Industrial Minerals Section.

Using parameters available in both sets of data such as location, elevation, total depth, and hole number, 2,674 IMDA boreholes were matched to existing data. Some boreholes could be matched visually in GIS by location and hole number (with other data cross-checked), and others by using various database queries. Specialized queries were developed, including distance computation, “fuzzy” matches for certain fields such as hole number (often entered in multiple slightly different formats in our records), and “fuzzy” matches to PLSS and other data to find matches even in the case of a data entry error in either database.

Some boreholes had insufficient shared data, too much data in common, were too closely spaced to easily match, or had data entry errors. These required careful review of all data including logs where available. While a student performed much of the reconciliation in several phases using GIS, queries and related data provided by the PI, the PI performed the more difficult remaining reconciliation phases. Additional benefits of the reconciliation process included correction of data entry errors in both databases, and population of an occasional missing data field in our wells and borings database.

Due to the state line, the northernmost row of sections in the Illinois Lead-Zinc district consists of irregular sections with a relatively short N-S length. During reconciliation, it was discovered that our main wells and borings database had coordinate locations for many boreholes in the second northernmost row of sections located too far south, usually in the adjacent section south of the correct section. This was due to the fact that these boreholes have footage locations measured from the state line, in effect skipping over the short section. The automated system the wells and borings database uses to create coordinates from PLSS locations is not designed to handle footages measured from a different section, and thus produced improper coordinates.

A total of 1,697 new boreholes were added, making a total of 4,371 IMDA borehole records; the IMDA section reference and serial number are included with these records as well as our API number for those holes in our primary wells and borings database.

After reconciliation with our wells and borings database, we used the IMDA logs to enter formation tops and downhole Lead, Zinc, and Iron concentrations for each of the new boreholes as well as any existing boreholes for which we did not have this data. As with the borehole header information, custom forms were designed to efficiently display the proper log file, to facilitate efficient data entry, and to check for potential data-entry errors.

The number of IMDA logs does not precisely match the number of IMDA mapped boreholes because the maps and datasheets were updated periodically, and in a few cases we did not have the latest map or the map was missing. Obvious errors in the original data were corrected (for example, serial number and hole number transposed). In some cases where the map was missing, we were able to match a log to an existing borehole record in that section using other parameters to obtain a good location; otherwise, they are not included.

A total of 4,968 metadata records were uploaded to the NGGDPP National Catalog. These include new records for 1,697 IMDA boreholes and 246 maps, and updated records for 2,674 IMDA-identified boreholes previously uploaded from previous projects. In total, 4,617 IMDA records were uploaded. The remaining records were previously uploaded from other sources and/or more recent time periods and were included to complete the set.

Fluorspar and Lead-Zinc Mining Districts Historical Documents

A large amount of information and data covering both of these important and unique mining districts was accumulated and developed by the Illinois State Geological Survey (ISGS) generally over the time period 1940s-1970s; also included are a few items from earlier periods. This paper-based data is in basement storage file cabinets, mostly in folders and also some binders of varying sizes. The documents are quite varied, and include for example unpublished manuscripts, prospect reports and detailed information on many mines and mining companies, and confidential reports from the 1940s that were commissioned to aid U.S. government agency wartime efforts. In addition there are field notes, logs, and local newspaper clippings and other accounts of new prospects and discoveries.

In a few cases the material corresponds to data behind a published report. One folder includes original photographs from a report published in 1914. The photographs show significantly better detail than the halftone “screened” photographs in the printed report.

There are also some data and/or various notes that likely require interpretation and research to determine whether they might be useful or important. We did not attempt to filter or edit the material scanned, as the time required to make such a determination (if it even could be made with available resources) would be extensive and take much longer than the processing time for any particular document. Further, it is difficult (perhaps impossible) to accurately predict what may or may not be useful in the future, and presumably some determination was made when the documents were originally filed.

The project proposal as written was to process the documents for the Fluorspar mining district in southern Illinois. However, as the project progressed, we realized we had enough resources and time to also process similar documents for the Lead-Zinc mining district of northwestern Illinois. The Lead-Zinc district documents also complement the USBM Illinois Mineral Development Atlas (IMDA) project.

Generally, each document is one of multiple related documents contained within a file folder, a large filing envelope, or a binder. We scanned each document, filing the digital version in a manner parallel to the paper versions: first by folder or binder name (general subject covered), then by document name (specific subject covered). Most documents were scanned as PDF files at 300dpi or 600dpi depending on detail level; detailed photos and maps were generally scanned as TIFF files at 600dpi.

The bulk of the documents were generated in the 1940s and thus are about 75 years old. Many documents are fragile and deteriorating (sometimes folded), often requiring careful handling. In many cases related pages were of varying sizes (for example, typed pages interspersed with log pages and hand-drawn maps of varying sizes). For these reasons as well as conservation the use of an auto-feed scanner was not practical nor advisable for most documents; in general each page was individually hand-scanned on a flat-bed scanner. Large documents such as maps were carefully unfolded and scanned in a roller scanner. Archival repair tape was used to repair damage.

Data about each document was entered into a project database, including folder, name, document type, author, date or date range, and page count. Each document was reviewed and one or more locations were assigned to it. For a few extensive documents that referenced numerous locations, a page number was entered for each location.

Summary of project scanning:

Mining District	Pages Scanned	Folders created	Documents created	Locations
Fluorspar	7,079	114	590	1,187
Lead-Zinc	3,106	85	444	773
Total	**10,185**	**199**	**1,034**	**1,960**

In the proposal, it was envisioned that each folder would be made into a single PDF document. In practice this resulted in unwieldly and large documents covering multiple subjects. It was decided to only group very similar documents within a folder into a single PDF document. This resulted in substantially more documents than estimated for the Fluorspar district; however as expected the folder count does track with the original document estimate of approximately 100. There were also significantly more Fluorspar district locations than estimated, mostly due to a few documents with a large number of locations.

Metadata records corresponding to 1,187 Fluorspar District locations/documents, and 773 Lead-Zinc District locations/documents were created and uploaded to the NGGDPP National Catalog.

Updates to Other Metadata

In addition to the current project year collections metadata, metadata for collections with significant changes were updated and uploaded to the NGGDPP National Catalog.

Comparison of Proposal Goals with Project Accomplishments

The proposal goals were to preserve the large amounts of unique and irreplaceable paper data in these collections, and to improve the discoverability, accessibility, and usability of such data through use of metadata and data entry of key information.

USBM Illinois Mineral Development Atlas (IMDA)

This project was completed as planned with proposal goals met. All paper data was scanned and key metadata entered and uploaded. The data was reconciled with our primary wells and borings database.

Fluorspar and Lead-Zinc Mining Districts Historical Documents

This project was completed as planned with proposal goals met. In addition the Lead-Zinc Mining District Historical Documents collection which was not included in the original proposal was added to the project and completed. All paper data was scanned and key metadata entered and uploaded.

Summary of FY14 metadata record estimates versus actual:

Collection Name	Estimated Records	Documents	Records Uploaded
USBM Illinois Mineral Development Atlas	4,000	4,617	4,617
Fluorspar Mining District Historical Documents	500	590	1,187
Lead-Zinc Mining District Historical Documents	--*	444	773

*Not included in proposal.

Conclusion

We believe the collection metadata generated and uploaded for this project represents an appropriate and successful use of the resources provided under the project contract.