Award Number G11AP20176

National Geological and Geophysical Data Preservation Program

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

Final Technical Report

Fiscal Year 2011

7/1/2011-6/30/2012

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

Zakaria Lasemi
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

September 24, 2012

Abstract

This report summarizes the work performed for the Illinois State Geological Survey FY11 NGGDPP project:

- Metadata records were created for 3 additional collections: Groundwater Potential Reports, Field Notes, and Thin Sections. Paper records were also scanned for the Typed Field Notes collection.

- Additional metadata records and scans were created for the Historic Clay Mineralogy Reports Collection.

- Existing metadata records for other non-static collections were updated.

- All new or modified metadata records were 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 FY11

Groundwater Potential Reports Collection

The Groundwater Potential Reports collection contains over 100 years of valuable geologic background, data, and historical information for specific sites and local areas throughout Illinois. Pertinent previously prepared reports are generally included in responses to requests for information on groundwater resources for a given area. This information is helpful to landowners and drillers evaluating groundwater resource potential or drilling factors such as depth, type of well and well diameter, aquifer material and water quality. Considerable time and cost savings can be generated when information for a specific area has already been compiled and is readily available.

Pertinent information along with locations for each Groundwater Potential Report was entered into a database. In addition, we entered data for Groundwater Potential Contacts which can contain additional information about a specific area.

Programmatically converting PLSS locations to coordinates for a significant subset of the data was challenging as we needed to determine the appropriate meridian for a PLSS location referencing a Quad (either 7.5 or 15 minute). GIS polygon processing and database queries were used to assign the appropriate meridian for PLSS to coordinate conversion.

We created metadata for about 24,000 Groundwater Potential Report and Contact locations, allowing reports for a given locality to be readily discovered.

Typed Field Notes Collection

The ISGS field notes collection contains information that survey scientists have recorded for over 120 years. Field notes provide an account of geologic information for specific sites without the need to visit the sites personally. They provide background and a starting point for research, often for sites that are no longer accessible. This information is invaluable for geological mapping, mineral resource evaluation, paleontological research, and environmental assessment.

Our field notes are divided into two collections: the “original” field notes collection, original and in some cases retyped versions of field notes organized by author; and the “typed” field notes collection, consisting of typed versions of original field notes organized by location. For this project year we concentrated on the typed field notes; we plan to address the original field notes next year. The entire typed field notes collection of 161 volumes covering Illinois and nearby states was scanned. In addition, 8 volumes of Industrial Minerals field notes discovered during the project were scanned.

The PLSS location, geologist(s), and year corresponding to each field note was entered into a custom-designed database and tagged to a specific page. Most field notes reference very specific locations, often with 3 or more section quarters of precision denoted as points on an 8x8 grid subdivision of a section. We maintained this precision

in the data by entering a grid code along with additional codes further refining the point location, resulting in a maximum possible precision of between 4 and 5 quarters.

Endpoints of marked lines and traverses were entered as 2 related points, and where applicable, specific areas were denoted as 2 corners of a rectangle. Less specific locations (2 or fewer quarters) were entered using standard quarter notation.

The data entry user interface was designed to be as efficient, streamlined, and automated as possible, with additional design changes based on user input to maximize data input efficiency. A primary goal was to minimize the number of key presses and hand/mouse movements to enter the data. Checks were included to make it impossible to enter invalid or incomplete data. Various cross-checks were performed on the data by the PI to detect and minimize data entry errors.

Converting the wide variety of location information to coordinates or GIS features presented a number of challenges. Sophisticated pre- and post- processing software had to be developed to handle the extra precision codes, as well as to properly capture and convert non-point locations to lines and polygons.

In order to differentiate location precision or extent when viewing a GIS representation of field notes locations, it was decided a location would be represented by a point only if was denoted to at least 3 quarters. Less precise locations were converted to polygons representing the corresponding part of the section (or entire section). Lines and rectangles were captured by converting their precise endpoints or corners to coordinates, then using those coordinates to build GIS lines or rectangular polygons. Quadrangles were captured by matching name and type to existing Quadrangle GIS polygon data.

For NGGDPP reporting purposes, the centroids of polygons and midpoints of lines were computed to provide the requisite point data. The origin of each point (point, line, or polygon) as well as the precision is denoted in the “alternateGeometry” section of the NGGDPP XML data package.

Software first developed for the FY08 project was used to interrogate Bureau of Land Management web services in order to obtain coordinates for PLSS locations outside of Illinois. The web service does not handle subdivided sections so all out of state locations are at best no more precise than the section centroid (with the effective precision noted). Although we captured location information for non-PLSS states Kentucky and Tennessee, time did not allow researching and developing methods to convert the specialized coordinate systems used in these states and instead, county centroids were used to provide coordinates.

Metadata records generally containing precise coordinates for about 97,000 locations were created, allowing field notes for a given locality to be readily discovered and viewed.

Thin Sections

Thin section petrography used in conjunction with well log analyses, seismic interpretation, regional geology, and numerous other related studies is an invaluable tool for interpreting the composition, origin, and properties of rock. Thin sections are invaluable for depositional environmental assessment and detailed characterization of lithology, petrophysical properties and geological framework.

Collections of thin sections made by retired scientists have accumulated in various offices and need to be reviewed, categorized, and digitally catalogued to make them discoverable and accessible to researchers, industry, and academic institutions.

Work on collecting and organizing our thin sections has been ongoing within the survey for some time and this project has helped that cause, though more work needs to be done.

We uploaded 206 metadata records covering a total of 1,812 samples. We plan to continue work on thin sections as time allows, and will update the National Catalog as appropriate.

Historic Clay Mineralogy

Early in the project year, since we still had trained students and access to a key retired geologist, we created additional records for the Historic Clay Mineralogy project from FY10. An additional 82 metadata records and corresponding scans were created for this collection.

Updates to Existing Metadata

In addition to the collections previously described, the following collections were updated to include new data and data revisions (static collections were not updated). With these updates, all of our collections with metadata in the National Catalog are current as of August 2012.

- Rock Cores
- Rock Cuttings

Location Issues

Virtually all of our historical locations are in PLSS format. This format is prone to typos and transcription errors in many more ways than a simple x,y coordinate format, and a number of metadata locations in PLSS fail to convert to a coordinate because they do not reference a valid PLSS location. This can be used as a data entry integrity check, and in fact a few data entry typos are caught this way. However the majority of the issues are with the original data.

In general most data without a good location are essentially “invisible”. The PI and project staff reviewed and where feasible corrected invalid PLSS locations in order to have a precise coordinate location for as many items as possible. Sometimes the issue was obvious, for example wrong county referenced when working near a county boundary; transposed numeric digits; wrong township or range direction (i.e., “E” instead of “W”). In other cases the reason is not obvious and the location cannot be easily corrected, in which case a less precise coordinate is substituted by the conversion software (usually the township or the county centroid), and noted in the “alternateGeometry” section of the NGGDPP XML data package.

Comparison of Proposal Goals with Project Accomplishments

Groundwater Potential Reports

Goals: Create metadata to index up to 5,000 Groundwater Potential Reports. Optionally scan the reports if time allows, with additional scanning in future years.

Results: A total of 18,455 metadata/location index records for 14,195 Groundwater Potential Reports were created (some reports reference multiple locations). In addition, we decided to create metadata for 5,463 related Groundwater Contact Reports for a total of 23,918 metadata items.

Upon review of cost/benefit, and because it would require significant resources over multiple years, it was decided not to begin scanning the massive amount of corresponding paper files. The files are well-organized and can be easily referenced to obtain a desired report.

Field Notes

Goals: Scan and create metadata for as many field notes volumes as possible. It was estimated that processing the entire collection will take multiple years.

Results: We completed scanning and metadata creation for all 161 volumes in the "typed" portion of the field notes collections. In addition we scanned and created metadata for 8 volumes of Industrial Minerals field notes, resulting in a total of 169 volumes; 68,137 scanned pages; and 97,395 locations. Next year we plan to begin working with the "original" portion of the collection.

Thin Sections

Goals: Collect and organize thin sections throughout the survey. Create metadata for up to an estimated 3,000 thin sections.

Results: We created metadata records representing 1,812 thin sections at 206 locations. We plan to continue work on thin sections in the future.

Summary of FY11 metadata record goals versus actual:

Metadata Records		
Collection Name	**Estimated**	**Created and Uploaded**
Groundwater Potential Reports	Up to 5,000	23,918
Field Notes	TBD	97,395
Thin Sections	Up to 3,000	206 (1,812 samples)

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.

Acknowledgements

Several staff members at the Illinois State Geological Survey assisted this project year, including Jared Freiburg, Anne Huber, Myrna Killey (retired), Donald Mikulic, Bob Mumm, and Richard Rice.