FINAL TECHNICAL REPORT

Indexing and Inventorying the Lake Superior Legacy Collection

Program Announcement No. G2AS20015
Grant / Cooperative Agreement No. G12AP20148

Project Period: September 1, 2012 – August 31, 2013

November 27, 2013

Abstract:
Roland D. Irving, Wisconsin's first true geologist, established the Lake Superior Division of the USGS in 1882. As head of the division, Irving began a detailed survey of the Upper Midwest during which he pioneered the application of microscopic petrography. The Division's fieldwork continued long after Irving's untimely death in 1888, and in the course of 40 years of investigation and research the survey geologists published nine monographs and several bulletins. The raw data used to produce these influential works makes up the Lake Superior Legacy Collection and currently consists of approximately 400 field notebooks, 62 maps, 9,800 hand samples, 15,500 thin sections, and 37 ledgers cataloging hand specimens, locations, thin sections, macroscopic and microscopic lithologic descriptions, and chemical analyses. For this project, staff and students at the Wisconsin Geological and Natural History Survey made a comprehensive inventory of this collection. We identified hand samples and thin sections with geographic locations, created metadata for them and added this metadata to our existing collections in the National Digital Catalog. We identified a set of at-risk paper records, namely notebooks and ledgers of lithological descriptions, and, in partnership with the University of Wisconsin Digital Collections, had the paper records scanned and placed online. We also created a comprehensive database, a digital infrastructure that describes the collection, integrates paper records with physical samples, and gives researchers extensive access to these important historical materials.

Principal Investigator: **Dr. M. Carol McCartney**
Wisconsin Geological and Natural History Survey
3817 Mineral Point Road
Madison, Wisconsin 53705
mmccartney@wisc.edu

Submitted to: **Betty Adrian**
Grants Program Manager, NGGDPP
United States Geological Survey
Denver Federal Center
Box 25046, MS 975
Denver, Colorado 80225
badrian@usgs.gov

Introduction:

The Wisconsin Geological and Natural History Survey (WGNHS, the Survey) received a grant for FY 2012 from the United States Geological Survey's (USGS) National Geological and Geophysical Data Preservation Program (NGGDPP) to inventory and index the Lake Superior Legacy Collection. A proposal for a work plan and budget was submitted in May 2012 and revised August 23, 2012, for the 12-month period beginning September 1, 2012, and ending August 31, 2013.

For the proposal, we outlined the following goals. The remainder of this document discusses our accomplishments towards each goal.

1. Inventory the Lake Superior Legacy Collection
    a. Systematically examine and document the Survey's holdings of the collection
    b. Identify materials with clear geographic locations
    c. Determine the suitability and feasibility of incorporating materials with location information into GeoBase, WGNHS's database
    d. Determine what metadata describing the collection would be suitable for uploading to the National Digital Catalog (NDC)

2. Create metadata
    a. Create a database of rock specimens and thin sections with geographic coordinates, descriptions, unique identifiers and related information
    b. Add this data to our GeoBase database and our Data Catalog
    c. Prepare NDC-compliant metadata in XML
    d. Upload these files to the National Digital Catalog

3. Create a digital infrastructure that integrates data from the collection's various paper records
    a. Review field notebooks, ledgers and related analog materials
    b. Integrate information from the paper records into a single database
    c. Migrate the records we have determined as suitable into GeoBase
    d. Extract metadata suitable for uploading to the NDC, and modify it to make it NDC-compliant

4. Rescue a subset of at-risk paper documents
    a. Select a subset of paper documents for scanning
    b. Partner with the University of Wisconsin Digital Collections (UWDC) to scan the documents and create metadata for them
    c. Partner with the UWDC to provide online access to the documents
    d. Obtain high-resolution TIFFs from the UWDC with the goal of making them available on the WGNHS web site as well for additional access.

Assessment of FY 2012 accomplishments relative to work plan

1. Inventory the Lake Superior Legacy Collection

A. Systematically examine and document the Survey's holdings of this collection

We have made a complete inventory of the items from the Lake Superior Legacy Collection in our possession. By comparing our samples to the written records and comparing various pieces of the written records to one another, we have reached a good understanding of what was in the original collection and what remains. In addition, a geologist at WGNHS familiar with historical terminology has checked a set of samples against their descriptions in the field notebooks and lithological descriptions to assure that the physical samples we have match the paper records (that they are, indeed, a part of this collection). This was the first and most critical of our quality assurance checks for this collection. As can be imagined with an old collection such as this, there are large gaps in nearly every part. We are missing field notebooks, hand samples, thin sections, and lithological descriptions. There is a book listing photograph negatives, but we have found none of these and none of the prints that were made from them. We even discovered, more than halfway through the project, a set of maps of which we were completely unaware, mostly of parts of Michigan's Upper Peninsula, showing sample numbers and rock types. A complete inventory of what was originally in the collection (as far as we could ascertain) and what is in our possession is attached to the end of this report as an appendix.

B. Identify materials that have a clear geographic location

The table below lists the physical samples in our possession by level of location information.

Sample type	**Number in our possession**	**Number with Section-Township-Range location**	**Number with state location only**	**Number without location**
Hand sample	9,798	6,643	2,968	187
Thin section	15,540	9,064	5,011	1,465

Location information for these samples was obtained from the collection's paper records, mainly the field notebooks and specimen catalog. We performed continuous checks for consistency among the documents and reconciled inconsistencies during quality assurance reviews by the archivist.

Our inability to provide complete locations for all samples is due to the following factors:

- Incomplete locations in the notebooks—the large majority of these notebooks were written in Wisconsin, Minnesota, and especially Michigan's Upper Peninsula. However, samples were taken in places as far away as Colorado, Massachusetts, Nevada, and provinces of Canada. More than 2,000 samples were taken in Ontario alone. Ontario does not have a Public Land Survey System (PLSS), nor do the original 13 colonies. Even when geologists worked in surveyed areas such as Colorado or South Dakota, they generally did not note PLSS locations.

- Missing notebooks—we are missing 77 of the field notebooks (see appendix) and 47 were written by C.K. Leith in a personalized version of an obsolete shorthand. Approximately 50 of the notebooks are still in the USGS warehouse in Herndon, Virginia, and we are hoping to borrow these to fill in data gaps in the next phase of this project.

- Inconsistencies in the paper records—a number of samples are listed in the specimen catalog as being lost. Many were lost in 1884 when the UW–Madison Science Hall burned down, while others apparently were lost either during fieldwork or in shipping. We know some were dispersed into teaching collections at other campuses. Meanwhile, some of the numbers listed as lost have actually appeared in the physical collection. Additionally, it is possible that some lost samples were re-collected during the summer following the Science Hall fire. Several of the notebooks from 1885 mention that they revisited sites from the previous summer but do not state specifically that samples were replaced.

- Incomplete information in the specimen catalog—the first sample number in the first notebook is 4012. Data for all previous samples is recorded in the first volume of the specimen catalog (see appendix). However, often the location information was not recorded or is vague.

- Work in progress—locations for the remaining 1,465 thin sections will be completed to the best of our ability in our FY 2013 project.

C. Determine the suitability and feasibility of incorporating materials with location information into GeoBase, WGNHS's database

The inventory revealed that we had a large number of the thin sections, but only a small fraction of the hand samples. We had, therefore, paper records providing extensive data on lost samples. We considered putting this information into GeoBase, but finally decided to add metadata only for physical samples in our possession. We believe, too, that adding information for samples with a vague location, such as state only, would not be useful to historians or current researchers. The work here at WGNHS is focused on the geology of Wisconsin, and the bulk of the Lake Superior Division's samples come from Wisconsin, Ontario, Minnesota, and especially Michigan's Upper Peninsula. As mentioned before, Ontario has no PLSS, so location information for samples collected there was vague. In the end, we added to GeoBase metadata that met the following criteria:

- Sample was collected in Wisconsin, Minnesota, or Michigan
- Sample has at least a township and range location, preferably a section-quarter section as well
- We possess the hand sample, the thin section, or both

D. Determine what metadata describing the collection would be suitable for uploading to the National Digital Catalog

The same criteria used for determining which samples will be recorded in GeoBase will also determine what metadata will be uploaded to the NDC. That is, metadata will be uploaded for physical samples in our possession for which we have at least township and range location information in Wisconsin,

Minnesota, or Michigan. As the metadata was compiled we performed continuous checks for consistencies in the total number of samples and for unique sample identifiers. We also checked locational information in the collection's paper records against USGS topographic maps, online plats from the original Public Land Surveys, and historical publications about mines and mining.

2. Create Metadata

A. Create a database of rock specimens and thin sections with geographic coordinates, descriptions, unique identifiers and related information.

Recording metadata for the physical samples was carried out simultaneously with the inventory (goal #1). Student employees entered locational data into a database describing where each sample is currently housed at WGNHS and where it was originally collected. Providing locations for the large number of thin sections for which we had no hand sample proved particularly challenging. In the paper records, all metadata including location information is attached to the hand sample number, not to the thin section number. Both sets of samples are numbered consecutively, with no real correlation between the two numbers. So, for a large portion of our thin sections, the students painstakingly combed through the paper records to find each thin section's related hand sample number. Even after this process, there were gaps, so one of the students examined the slides to read the hand sample number written on it. Once the thin section and hand sample numbers were correlated, students used the specimen catalog and the field notebooks to attach locations and rock types to the samples. They then recorded the information in spreadsheets. Each spreadsheet was reviewed by the archivist for completeness and consistency.

After the above information was recorded, the spreadsheets were combined in a Microsoft Access database. Metadata for samples in our possession was then integrated with metadata for the paper records into a single master table. This step included continuous quality assurance checks for consistency of unique sample identifiers and total number of samples. A detailed description of the database can be found in our assessment of goal #3, part B.

B. Add this information into GeoBase and our Data Catalog (our project database)

Metadata for hand samples and thin sections in our possession has been extracted from the Lake Superior Legacy Collection's database and imported into GeoBase, our geological database. A general entry has been made for the collection in the Data Catalog, our project database, with a specific entry for each element of the collection (thin section catalog, field notebooks, hand samples, etc.).

C. Prepare NDC-compliant metadata in XML

Pete Schoephoester, one of our GIS specialists, converted the metadata to XML prior to uploading it to the National Digital Catalog. When the text location was translated into coordinates, only valid coordinates could be assigned by the algorithm, providing another level of quality assurance for the entire database.

D. Upload these files to the National Digital Catalog

Pete Schoephoester has uploaded metadata for 6,058 hand samples to our existing collection P517 and for 8,056 thins sections to our existing collection P1590. (Both collections have now more than doubled; previously, our hand sample collection had 4,562 records, and our thin section collection had 2,540.)

Originally, we proposed to upload metadata for 7,000 hand samples and 10,000 thin sections. We have come close to this goal, but have fallen short. The primary reasons for this shortage of metadata records are:

- The difficulties we have had attaching locations to some samples (see goal #1)
- The surprising number of samples taken from outside the main area of study (Michigan's Upper Peninsula, Minnesota, and Wisconsin) or out of the U.S. For example, more than 2,000 samples were collected in Ontario and nearly 600 in New York and Massachusetts.
- Problems with metadata conversion: Location information for the NDC was converted from a text description (section-township-range) to latitude-longitude coordinates, and we encountered some difficulties during the conversion. The biggest problem we have detected is the recording of more than one number in the section, township, and range fields. (Two numbers were recorded whenever a sample was taken on a section, township, or range line.) This type of entry caused nearly 600 thin section records and more than 1,000 hand sample records to be kicked out of the converted metadata. This quality assurance step prevented the assignment of coordinates that were not valid. We will correct these entries and upload the metadata during the FY 2013 project.

3. Create a digital infrastructure that integrates data from the collection's various paper records

A. Review field notebooks, ledgers, and related analog materials

A review of the paper records showed that most were at risk, and that while some contained information that could easily be captured in a database, others were of enough historical and geological interest to warrant making scans of the original document and making it publicly available (see goal #4). The latter included mainly the field notebooks themselves and the ledgers of lithological descriptions (see appendix). While paper records of the chemical analyses did not lend themselves to easy capture in a database, we determined that they were not ideal candidates for scanning because they covered only a small number of samples, analytical methods have changed significantly, and some samples were not from the Lake Superior Legacy Collection.

Information that could be captured easily in a database without scanning the original documents is contained in the specimen catalog (see appendix). Though the catalog's early volumes contain a great deal of information, such as a sample's exact location, who took the sample, and what type of rock or mineral it was, this information is not consistently recorded, and it vanishes altogether from the later books. Also, the information contained in the thin section catalog amounted to a simple correlation of hand sample number to thin section number, so these, too, were deemed less suitable for scanning. Instead, we opted to

extract from these sources as much information as we could about the hand samples and thin sections in our possession. (See goal #2.)

B. Integrate information from the various paper records into a single database

To support the scanned images of the field notebooks and lithological descriptions, and to integrate that information with the metadata related to physical samples, the Survey archivist and several student employees created spreadsheets of metadata for these paper records. From the notebooks, a separate entry for each sample described was created and included the fields: notebook number, author, year, section, township, range, state, specimen number, rock type, informal location (e.g., "Vicinity of Lake Gogebic"), and notes. A separate entry was made for each sample so that researchers seeking information on a particular sample can go directly to the point in the scanned version of the paper records at which it is described. Fields for the lithological descriptions metadata include hand sample number, thin section number, and book and page number. After students entered data, the archivist did a quality assurance review of each of the spreadsheets, and as with the physical samples, verified locations against topographic maps and online plats from the original Public Land Surveys.

The spreadsheets containing the metadata for these paper records were imported into the Microsoft Access database described in goal #2. As the metadata were compiled, we did continuous quality assurance checks for consistency.

In addition to metadata for our physical samples, the database now contains metadata for the 100 notebooks available online (see goal #4), which are notebooks numbered 1–45 (minus what we are missing), and all of the notebooks authored by Charles Van Hise (scanned for a previous project).

The metadata for the Lake Superior Legacy Collection is contained in three tables:

1. A master table with a row for each physical specimen showing its hand sample number, thin section number, location at WGNHS, where in the field the sample was taken, rock type, notebook and page number where the sample is described, informal location, notes, and whether or not we have a chemical analysis for the specimen.
2. A table recording the specimen numbers and the book and page number in which that sample is described in the books of lithologic descriptions. The specimens described in these books represent a small subset relative to the collection as a whole, and the master table already shows a field notebook and page number for each specimen, so this metadata could easily be overlooked if included in the master table. Also, the lithologic descriptions comprise a unique resource, so we decided that keeping this information in a separate table would provide the simplest and most straightforward access to this piece of the collection.
3. A table giving the author, year, and number of each field notebook.

C. Migrate the records we have determined as suitable into GeoBase

The majority of the Lake Superior Legacy Collection's hand samples are no longer in our possession, and some of the thin sections have been lost as well. Our approach to capturing metadata for the notebooks meant that we would have metadata for a large number of physical samples that are not in our possession.

We evaluated the option of adding to GeoBase metadata for such samples. Some of the paper records are quite extensive, and the lithological descriptions in particular provide information that would be valuable even without having the physical sample. However, in the end, we decided against doing this during this phase of the project. Metadata from notebooks covering lost samples will be captured in the database dedicated to the Lake Superior Legacy Collection, and our database will expand and continue to incorporate metadata from notebooks as they are scanned. We will revisit this issue with the intent of creating a best practice for adding metadata to GeoBase that describes lost samples.

D. Extract the metadata suitable for uploading to the NDC, and modify it to make it NDC-compliant

The discussion about which records to upload to the National Digital Catalog followed a similar path to our decisions about what to include in GeoBase. In the end, we decided to upload only metadata for physical samples in our possession in FY 2012, but to explore the possibility of creating a new collection for the paper records in FY 2013.

4. Rescue a subset of at-risk paper documents

A. Select a subset of at-risk documents for scanning (3,000 pages)

Our ultimate goal is to scan and place online all of the Lake Superior Legacy Collection's field notebooks that are dedicated to geology. However, due to our limited resources, we have focused on a subset for this project. Our original intention was to scan the 56 notebooks produced by the Lake Superior Division while Roland Irving was its director, which covered the period from 1882, when the division was formed, to 1888, the year Irving unexpectedly died (Notebooks authored during this period by Charles Van Hise had been scanned for another project.) . However, after inventorying the collection, we saw that another group of books, nine ledgers containing lithological descriptions, also fit our criteria for scanning. Our subset pushed us well over the page limit dictated by our budget. In the end, we slated notebooks 1–45 and the first two books of lithological descriptions for scanning.

B. Partner with the University of Wisconsin Digital Collection (UWDC) to scan the documents and create metadata for them

We delivered the selected books to be scanned to the UW Digital Collections. They spent the year scanning the documents and creating their metadata. The metadata scheme used by the UWDC is fairly basic, but users will have more extensive access to the collection using our digital infrastructure (described in goals #2 and #3) in conjunction with the online images of the notebooks and lithological descriptions at the UWDC's web site.

C. Partner with the UWDC to provide online access to the scanned documents

The UW Digital Collections is now hosting the scanned images of the field notebooks and lithological descriptions on their web site. Researchers have ready access to the information without the original documents being placed at risk for physical damage. The UWDC will maintain an archive of high

resolution scans to back up the more web friendly, lower resolution images displayed on their web site. We have also paid a great deal of attention to our front page in the UWDC so that users have easy access to the Lake Superior Legacy Collection.

The notebooks and lithological descriptions can be viewed at:
http://uwdc.library.wisc.edu/collections/EcoNatRes/WGNHS

D. Obtain high resolution TIFFs from the UWDC with the goal of placing them on our web site for additional access.

The UWDC has delivered high resolution TIFFs of all the scanning done this year. The Survey web site is undergoing a complete overhaul, and we don't yet have a time frame for hosting this material, but it remains an important goal for us.

Appendix: Inventory of the Lake Superior Legacy Collection

Collection element	**Original extent**	**Portion in our possession**	**Brief description**	**Possible reasons for gaps**
Hand samples	Approx. 80,600	9,798	Rock samples taken from the field	Unused sample numbers, samples lost in field, in a fire, or in transportation, samples loaned to other institutions.
Thin sections	Approx. 18,000	15,440	Microscope slides made from hand samples	Same as above; also, slides are easily damaged. Some may have been discarded.
Field notebooks*	486	409	Geologic and topographic notebooks, including drill hole logs	Approx. 50 notebooks were not shipped from Herndon with the rest of the collection in 1981. The rest, unknown.
Thin section catalog	3 ledger books	3 ledger books	Catalog correlating hand sample numbers with thin section numbers	
Chemical analyses	Unknown	2 ledger books	Chemical analyses of samples—not all are Lake Superior Division samples	
Photograph negative catalog	Unknown	1 ledger book	List of photograph negatives including location taken and corresponding notebook and page number	Unknown. The photographs have not been located.
Specimen catalog	24 ledger books	24 ledger books	Listing of specimens from 1 to 80,816, some with location information, thin section number, notebook number and page	
Lithological descriptions	9 ledger books	6 ledger books	Microscopic and macroscopic descriptions of samples written by	Unknown

			Irving and Van Hise, organized by rock type	
Maps	Unknown	62	Plat maps showing sample numbers and rock types, mostly of parts of Michigan's Upper Peninsula	Unknown

*Some specimen catalog entries refer to notebook numbers preceded by a "C." C1, C2, etc. We do not have these, and do not know what information they contain apart from sample descriptions.