

OPEN ACCESS GEOLOGY: USING THE INSTITUTIONAL REPOSITORY TO HOST STATE GEOLOGICAL SURVEY PUBLICATIONS

Lisa Johnston and Harvey Thorleifson

Science and Engineering Library, University of Minnesota, 108 Walter Library, 117 Pleasant St. SE, Minneapolis, MN 55455, ljohnsto@umn.edu

Minnesota Geological Survey, University of Minnesota, 2642 University Ave. W, St. Paul, MN 55114

The Minnesota Geological Survey (MGS) hosts information systems containing data sets that are valuable historic and scientific resources for the state. Several options are being pursued to curate, preserve, describe, and disseminate these data to broader audiences, including web services, national data archives, and institutional repositories. One of the challenges has been to find a means to make available scanned versions of geological survey publications. Geoscience research literature is increasingly published electronically and made digitally available for immediate online access. For state geological survey publications, a library-run institutional repository (IR) can be an excellent solution to host digitized and born-digital content. In the past few years, MGS has scanned all of its publications published since 1872 through a number of library and state preservation grants. The comprehensive project included over 40,000 pages of reports, guidebooks, and bulletins, and over 600 maps, along with GIS data files from recent decades. This talk will describe how the MGS content was archived in the university's institutional repository and the issues and challenges we faced such as format decisions, workflow issues, and modes of user access.

<http://gsa.confex.com/gsa/2010AM/finalprogram/index.html>

INTRODUCTION

The Minnesota Geological Survey¹ (MGS) was established in 1872 under the direction of Newton Horace Winchell (see Figure 1). Winchell started teaching as the University of Minnesota's first professor of Geology and Mineralogy and subsequently the MGS became a permanent unit of the University of Minnesota's Geology and Geosciences Department when it later was established in 1874.

Since 2007, the library has partnered with the MGS to scan and digitally preserve their complete publication record, including all final reports, published maps, and field guides. With the support of library technical service staff and key MGS metadata and GIS experts, we successfully completed the scanning and upload this summer and all MGS final publications are available electronically online for public download in our institutional repository², the University of Minnesota Digital Conservancy (UDC) (see Figure 2). This paper will describe this project in detail and focus on our approach, workflow, complications and issues that led us toward successfully implementing a digital open access plan for the state survey's information systems. Finally, we will describe our next steps for archiving other MGS collections, such as Open File Reports, future born digital works, GIS files, and other data sets.



Figure 1: The Minnesota Geological Survey (MGS) was established in 1872 by an act of State Legislature directing the University of Minnesota to investigate the geology of Minnesota and make that information available to its citizens.

UNIVERSITY OF MINNESOTA

One Stop | Directories | Search U of M

[about](#)
[policies](#)
[f.a.q.](#)
[contact](#)
[help](#)
[best practices](#)
[guide to submitting your work](#)
[theses and dissertations](#)

login to:
[submit your work](#)
(authorized users)
[new user?](#)

 **digital conservancy**
UNIVERSITY OF MINNESOTA
University of Minnesota Digital Conservancy >

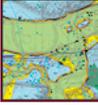
search the **digital conservancy**

[advanced search](#)

browse:
[collections](#) :: [authors](#) :: [titles](#) :: [subjects](#)

featured **collections**

 **Aurora Sporealis**
Congratulations to the University of Minnesota Department of Plant Pathology on its 100th birthday! This collection contains the alumni news magazine, *Aurora Sporealis*, from 1926 forward. For more information on the department's history and centennial celebration, please visit the [website](#).

 **Minnesota Geological Survey**
The Survey is the University outreach center for the science and technology of earth resources in Minnesota. Conducting basic and applied earth science research, the Survey conveys that information to the public through publications and service activities, and promotes earth science education. Minnesota is its exclusive focus. The Miscellaneous Map Series, 1966-2006, represents a variety of Minnesota's geological features and regions, including simple bouguer maps, lineament maps, and maps of surficial and bedrock geology.

the digital conservancy provides:

- :: Free, open access to university digital works
- :: Higher search engine rankings - increased visibility
- :: Compliance and accountability for your publicly-funded research
- :: Expert consultation on copyright, digital formats, and [authors' rights](#)

[University of Minnesota Web Archive](#)
(1996 to present)

Try [AqEcon Search](#), our Agricultural and Applied Economics subject repository

Figure 2: The University of Minnesota's institutional repository, dubbed the University Digital Conservancy (UDC), went online in 2007². Build on DSpace repository software, using a basic variation of Dublin core metadata schema, it allows for individual self-archiving capability, in open-access format. For a campus-based publisher, like the Minnesota Geological Survey, this presented an excellent solution to move toward electronic publishing.

ARCHIVING PROCESS

The MGS project encountered a number of issues and resulted in the formulation of several policies throughout the process of archiving content in the university's institutional repository. The issues and challenges we faced included: the scanning process, format decisions, workflow issues, born digital files and modes of user access.

Scanning Process

Our library's in-house digital collections unit provided the scanning service for the monographic series and text-based map supplements. This was possibly with the support of an internal, library

grant program. As a historic university-based collection, the MGS publications were a priority for digitization and upload in the newly formed UDC. Once in the UDC, the text was indexed and made full-text searchable. The monographic series available in the UDC to date include:

- ANNUAL REPORTS
- BULLETINS
- EDUCATIONAL SERIES
- GEOLOGY OF MINNESOTA: A CENTENNIAL VOLUME
- GEOLOGY OF MINNESOTA—THE FINAL REPORT. Vol 1-7 Folio
- GUIDEBOOK SERIES
- INFORMATION CIRCULARS
- MINNESOTA AT A GLANCE
- MISCELLANEOUS PUBLICATIONS
- NEWSLETTERS
- OPEN-FILE REPORTS (not complete)
- REPORTS OF INVESTIGATIONS
- SPECIAL PUBLICATION SERIES
- SUMMARY REPORTS

Simultaneously, a grant from the state provided the support to scan around 600 maps for their own digital library collection, Minnesota Reflections, where most of the MGS maps are duplicated online. The library augmented these maps in our collection with the original GIS and metadata files when available (mostly post-1990 map publications). The map series held in the UDC are:

- AEROMAGNETIC MAP SERIES
- COUNTY ATLAS SERIES
- GEOLOGIC MAP INDEX
- GEOLOGIC MAP OF MINNESOTA [Scale 1:250,000]
- GEOLOGIC MAP SERIES
- MISCELLANEOUS MAP SERIES
- REGIONAL HYDROGEOLOGIC ASSESSMENTS

- STATE MAP SERIES

The MGS was enthusiastic to get their content digitized, and they prepared the print copies from the Survey's library collection. Creating an up-to-date inventory of all MGS publications was not an easy process, and many publication lists had to be cleaned and double-checked by a student worker hired by the MGS. Most of our scanning metadata was generated from the MGS "publication order form" as it was the most up to date source for author, title, and related bibliographic information.

Format

MGS publications scanning resulted in over 40,000 pages of reports, guidebooks, and bulletins, and over 600 maps. The primary format was PDF with special treatment for the images in the reports, plates, and the maps, which were scanned at a higher resolution than the text-based pages. This process, essential for quality images and zoom capabilities, resulted in large PDF file sizes. For example, several of the bulletin issues were over 50MB in file size. Modern Internet connections are capable of downloading such a file quickly, however, some browser PDF preview functions, such as Firefox 6, do not function with files larger than about 25MB, resulting in a "Broken Link" error message. Therefore we decided to chunk larger files into smaller parts (see Figure 3). Alternate formats, such as page flip views, are planned for the future.

Workflow Issues

In addition to the authors, we had the assistance of several people working on the uploading process, from the library and the MGS staff. Uploading files into a digital archive is primarily a manual practice. Some of the work was batch loaded into the repository, but many of the series, such as the Miscellaneous Maps, had associated GIS files that required special treatment and must be loaded one-by-one. Therefore, to avoid conflicting practices, we created a policy for uploading MGS publications into the UDC to streamline our workflow. Our policy followed that all MGS records in the UDC must:

- Be self contained (ie. do not link to other web pages, or non-archived files)
- Include at least one file (ie. citations or placeholders such as "Report 9 was skipped" were not included in the UDC to avoid "dead ends" for users. Rather, the map index was archived to preserve this type of information.)
- Contain files in preferred archival formats (files in open, standardized file formats)
- UDC records should not include any information that could change over time (ex. Price info)

Issue Date: 2002
Publisher: Minnesota Geological Survey
Series/Report no.: M-123
Description: Scale 1:24,000.
Permanent URL: <http://purl.umn.edu/715>
Appears in Collections: [Miscellaneous Map Series](#)

Files in This Item:

File	Description	Size	Format	
m123_Extras.zip	Supplementary GIS files	2194Kb	Zip	View/Open
m123.pdf	PDF file of the map plate	1302Kb	PDF	View/Open
readme.txt	readme	1Kb	Text	View/Open

[Show full item record](#)

Figure 4: This screenshot illustrates the file structure used to upload MGS map publications with associated GIS files into the University Digital Conservancy.

Preferred File Formats

The primary file format was PDF for the maps and text. However special consideration was needed for the GIS files of the born-digital maps. Without an open non-proprietary standard format available, the archiving process of the MGS GIS files included zipping them in their original ERSI ArcGIS formats (eg. .e00). The zipped files provides a user-centered view of the record holding, rather than containing upwards of 20 or more files to choose from, each map record holds three objects:

- PDF of the map. This file format should meet the needs of most users and is maintainable for long-term preservation and migration.
- A zipped file containing the supplementary GIS files and FGDC metadata. The software used to zip the files was carefully chosen to not require additional software to run.
- A text “readme” file containing instructions on how to open and use the GIS files, software requirements and other metadata of how the map was created. The MGS geologists, not librarians, contributed all of this info, including the GIS metadata.

Born Digital Publications

Going forward, the MGS will be publishing their born-digital maps and publications directly into the institutional repository. The library provided training sessions for MGS staff on how to

upload the files into the UDC. Now the Survey can take advantage of the publishing platform directly from their office and within their existing workflow without relying on the library staff. This practice has recently resulted in the publication of a new County Atlas and the creation of two additional MGS digital collections: Geology of Minnesota Parks and Non-Series Publications.

Access

Access to the digital MGS publications and maps is as easy as a web search (see Figure 5). The UDC indexes the full-text of the documents and the maps include the scanned supplement documents and augmented metadata, such as keywords. The searchable MGS collection has been a valuable addition to the reference desk in the Science & Engineering Library at Minnesota. And the statistics for access to the digital files have topped the UDC download rates for all publications.

Of course browsing is very important, and therefore, the library created a dump of the UDC archived collection and marked it up in HTML so the Survey can host links directly from their site along with a collection-specific keyword search box. This interface allows their website to be the primary website for public access to the MGS digital publications³.

The screenshot shows a Google search interface. At the top, there are navigation links: Web, Images, Videos, Maps, News, Shopping, Gmail, and more. The Google logo is on the left. The search query is "Guidebook for the Precambrian Geology of East-Central Minnesota", with approximately 1,360 results found in 0.24 seconds. The left sidebar includes "Everything" with a "More" dropdown, "New York, NY" with a "Change location" dropdown, and "Show search tools". The main results list includes:

- [PDF] [FIELD TRIP GUIDEBOOK FOR THE PRECAMBRIAN GEOLOGY OF EAS](#)
File Format: PDF/Adobe Acrobat - [Quick View](#)
THE **PRECAMBRIAN GEOLOGY OF EAST-CENTRAL MINNESOTA**. MINNESOTA GEOLOGICAL SURVEY. UNIVERSITY OF MINNESOTA. ST. PAUL, MINNESOTA 55108. **GUIDEBOOK SERIES NO. 12 ...**
conservancy.umn.edu/bitstream/58968/1/MGS_GB_12.pdf
[University of Minnesota Digital Conservancy: Guidebook 12. Field ...](#)
Field Trip **Guidebook for the Precambrian Geology of East-Central Minnesota ...**
conservancy.umn.edu/handle/58968 - [Cached](#)
- [PDF] [PRINCIPAL SOURCES OF GEOLOGIC DATA](#)
File Format: PDF/Adobe Acrobat - [Quick View](#)
Geologic map of Minnesota--**east-central Minnesota**, bed- consin ...
conservancy.umn.edu/bitstream/59786/1/mgs-384.pdf
[+ Show more results from umn.edu](#)
- [ILSG - Index of Proceedings Volumes and Field Trip Guidebooks](#)
Field Trip: **Guidebook to the Precambrian geology** of northeastern and ... mineral resources of **east-central Minnesota** (MGS Field Trip **Guidebook** Series no. ...
www.lakesuperiorgeology.org/publications/proceedings.html - [Cached](#) - [Similar](#)
- [Institute On Lake Superior Geology Index](#)
Oct 14, 2008 ... **Guidebook to the Precambrian geology** of northeastern and north-central ... of **east-central Minnesota** (MGS Field Trip **Guidebook** Series no. 9) ...
www.lib.mtu.edu/mtuarchives/ilsgindex.aspx - [Cached](#)

Figure 5: An example web search for a MGS publication yields the PDF and the UDC record for the digitized publication in the first two search results.

DATA: THE NEW SPECIAL COLLECION

Next we have turned our preservation eye on the MGS data sets: the new special collection for libraries. The MGS hosts information systems containing data sets that are valuable historic and scientific resources for the state. Several options are being pursued to curate, preserve, describe, and disseminate these data to broader audiences, including web services, national data archives, and institutional repositories. To begin, we've archived three MGS datasets: Gravity and Aeromagnetic Data of Minnesota, the MGS's Rock Properties database, and the database of Karst Features of Minnesota (see Figure 5).

The datasets were archived in an open database format (.dbf) so that users may download a copy and run on their favorite platform. Although archival, this access method is not the ideal, compared to a custom interactive web-interface to the data. However, our first goal was preservation ensure that an archival copy exists as we move to more useable platform. The library is changing their backend of our repository to a fedora-based platform where we hope to be able to better host a variety of data format types and the MGS is building a custom GIS web service. Other datasets we plan to make available include:

- Geological observations
- Till texture and lithology
- Geochemical data
- Borehole geophysics index
- Water well data
- Geotechnical data
- Hydrogeological data

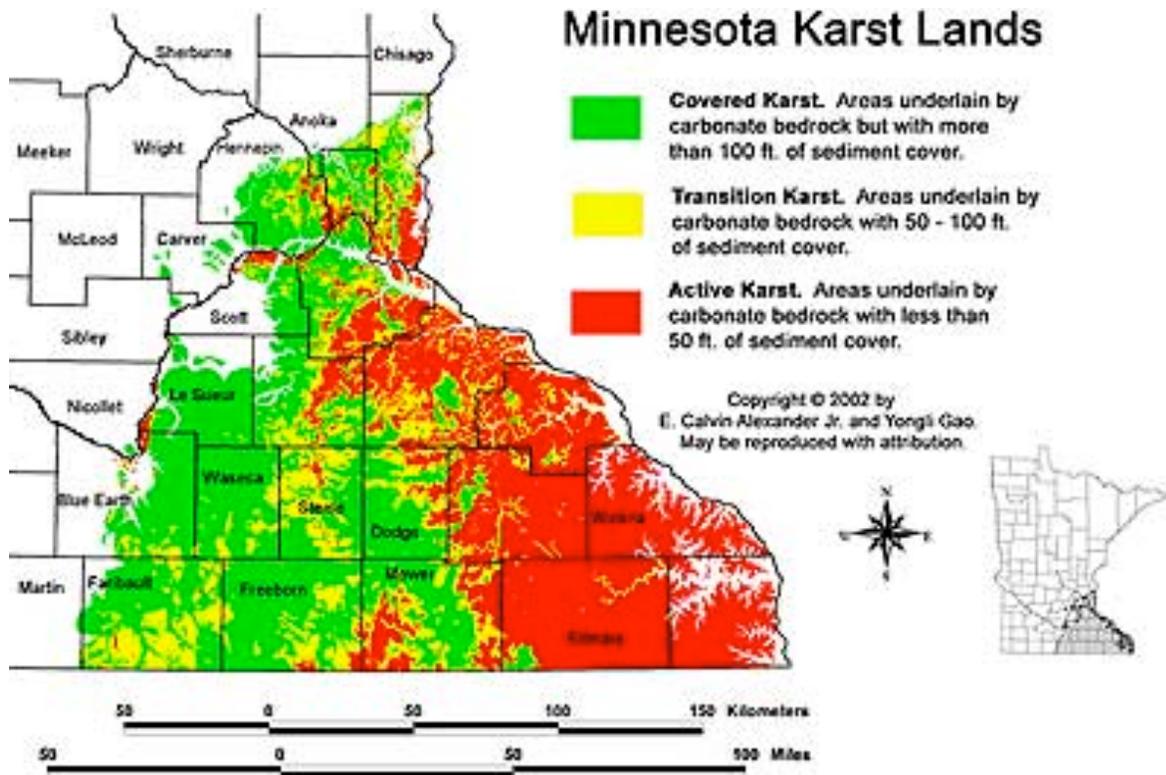


Figure 5: Created from the Karst database included in the MGS digital collection, this image represents the type of information that can be created from archival data sets. Image created by Calvin Alexander, University of Minnesota.

FUTURE NEXT STEPS

Now that we have everything scanned and archived, there are some important next steps to consider. As we mentioned above, the monographic content will be made available in alternate formats more suitable for web viewing. Aiding this effort is our plan to archive the MGS materials in the HathiTrust digital library⁴ as part of the University of Minnesota's participation in the Google Book project. This will not only provide a page-flip view of the book, but will allow all libraries to harvest the content from OCLC's WorldCat to ingest into their library catalogs for direct access. Also, the UDC is an OAISTER-compliant repository, thus the maps and data files can be pushed to WorldCat as well. Finally, with our GIS web services moving into place, federated repositories such as OneGeology.org and other Minnesota GIS services, can link to our archived GIS files for interactive access to the data.



Another area for archival consideration is the Survey's physical data collections: samples that may be in need preservation, cataloging and wider dissemination. There are many ways that the University Library can partner with a state survey to succeed in this area. For example, the library was brought in to consult on the MGS's metadata and preservation plan for applying to the USGS's National Geological and Geophysical Data Preservation Program (NGGDPP). As a part of this proposal the MGS's physical collections that we will work toward archiving and providing enhanced access include:

- Hand samples
- Thin sections
- Sediment samples
- Geochemical samples
- Cuttings
- Drill cores
- Fossils (collection managed by the University of Minnesota's Bell Museum)

Project files and orphan data sets are another potential area for archiving. But how do we begin to shift through all the file cabinets, both physically and digitally. Progress has already been made here, with over 300 print field notebooks and other historical administrative files moved to

the University Archives in the Library. In the short term the documents will receive descriptive records and be made more usable through finding aids. The future plan is to scan and add them to the digital archive as a UDC collection. Also, for future projects the library might work with the MGS on building a data management plan at the beginning of the project, to describe in detail how the supplementary information will be transitioned after the project is complete and providing descriptive metadata directly from ArcGIS in accessible XML file formats.

Finally, the MGS has been scanning their historic photo collection in-house and there is an opportunity to archive these currently inaccessible files into the Library's new media repository, UMedia Archive⁵. This Fedora based repository hold university affiliate created images and video with the goal of open access and digital perpetuity. This would not only give the mgs an excellent searchable database of their image archive but would expose their unique digitized collection to the world.

CONCLUSION

Partnership with the University Library and a State Geological Survey is a great fit. With the library's dedication to long-term preservation and wide dissemination of information, and the State Survey's open access publishing model and scientific expertise to provide the public with quality information and data, each bring the necessary components for a robust digital library of geological information. The MGS publications project is a proof-of-concept for future projects for data, GIS, and other media forms with the goal of preservation, access and the widest-possible dissemination of content.

Acknowledgements

Thanks to those who were responsible for successfully implementing the archiving process: Nicole Bretall, Philip Herold, Rich Lively, Beth Petsan, and Kadie Seitz.

REFERENCES

- 1 Minnesota Geological Survey. (2010). Minnesota Geological Survey Accessed December 29, 2010 from <http://www.mnngs.umn.edu>.
2. University of Minnesota Digital Conservancy. (2010). Minnesota Geological Survey. Accessed December 29, 2010 from <http://conservancy.umn.edu/handle/708>.
3. Minnesota Geological Survey. (2010). MGS Publications and Data. Accessed December 29, 2010 from <http://www.mnngs.umn.edu/mgsdocs.html>
4. HathiTrust Digital Library. (2010). HathiTrust Home. Accessed December 29, 2010 from <http://www.hathitrust.org>.
5. University of Minnesota UMedia Archive. (2010). Media Archive Beta: Connecting Our Collections with Yours. Accessed December 29, 2010 from <http://umedia.lib.umn.edu/>