

**NATIONAL GEOLOGICAL AND GEOPHYSICAL
DATA PRESERVATION PROGRAM
Fiscal Year 2008**

FINAL TECHNICAL REPORT

**From Acquisition to Access: Developing a Long-Range
Data Preservation and Archiving Plan for WGNHS Research Collections**

Program Announcement No. 08HQPA0015

Project Period: September 1, 2008 – August 31, 2009

NOVEMBER 20, 2009

Principal Investigator: **Dr. Thomas J. Evans**
Wisconsin Geological and Natural History Survey
3817 Mineral Point Road
Madison, Wisconsin 53705
tevans@wisc.edu

Submitted to: **Dr. Tamara L Dickinson**
NGGDPP Project Coordinator
911 National Center
12201 Sunrise Valley Drive
Reston, VA 20192
tdickinson@usgs.gov

**From Acquisition to Access: Developing a Long-Range
Data Preservation and Archiving Plan for WGNHS Research Collections**

Program Announcement No. 08HQPA0015

Abstract

The Wisconsin Geological and Natural History Survey (WGNHS) received a grant for FY 2008 from the United States Geological Survey (USGS) National Geological and Geophysical Data Preservation Program (NGGDPP). A Revised Work Plan and Budget for FY 2008 was submitted on June 20, 2008 for a twelve-month grant period beginning September 1, 2008 and ending August 31, 2009.

The Revised Work Plan and Budget for FY 2008 identified the following tasks for the twelve-month project period:

- (1) Review and maintain the Wisconsin contributions now in the on-line inventory for the WGNHS physical research collections,
- (2) Prepare point-specific metadata records to be added to the NGGDPP National Data Catalog for rock core data currently in WGNHS research collections, and
- (3) Develop a comprehensive Long-Range Data Preservation and Archiving Plan for Wisconsin consistent with the requirements of the NGDDPP.

Accomplishments for FY 2008 include:

- (1) two additional direct physical collections—thin-sections (P15 90) and sediment samples (P1591)—were added to the Wisconsin entries in the Online Collection Inventory;
- (2) metadata for 1,861 rock cores were prepared and uploaded to the National Digital Catalog; and
- (3) a Long-Range Data Preservation Plan for WGNHS Research Collections was developed from the preparation of the WGNHS Data Catalog, the WGNHS Collections Policy, and the WGNHS Scanning Policy.

Accomplishments in FY 2008

1. The WGNHS added two direct physical collections to the Online Collection Inventory in FY 2008: thin-sections (estimated at 12,700 records) [collection ID #1590] and sediment samples (approximately 14,900 records) [collection ID #1591]. Both of these collections will be revised as more detailed study continues in FY 2009 and FY 2010.

2. The WGNHS prepared NDC-compliant metadata for 1,861 rock cores [collection ID #516]. These data items are described in detail within WGNHS Geobase, the enterprise database of point-specific geological and geophysical data for Wisconsin. These records were successfully transferred as digital files in XML to the National Digital Catalog in September 2009 according to the schema described in the Revised Plan and Budget for FY 2008 and shown below:

National Catalog Metadata Element*	National Catalog Concept	WGNHS Geobase Attributes (Given as Table.Field)
SiteID	A unique and sustained identifier within a collection.	Location.Well_ID provides a unique identifier for each site in Geobase.
CollectionID	The unique identifier for a collection as reported in the National Catalog.	Typedrecd.Suffix provides a code for the type of sample in the WGNHS collection. This code will be transformed to the unique collection identifier in the National Catalog.
Latitude Longitude	Location of a sample site expressed in geographic coordinates.	Location.Latitude Location.Longitude
SiteDescription	A brief description of a site.	WellOwner.Site_Name provides a short, textual name for a site.
DataDescription	A short, human-readable description of a material or type of data.	Typedrecd.Suffix provides a code for the type of sample in the WGNHS collection. This code will be transformed to a standard prescribed value, such as “rock cuttings”, “rock core”, “hand sample”, or “thin section”.
VerticalExtent**	The vertical extent of a sample.	Core.Core_Top, Core.Core_Bottom, Lithology.Lith_Top, and Lithology.Lith_Bottom all provide an interval for various samples. These fields will be aggregated to produce an overall vertical extent for a sample.

* Attachment F of Program Announcement No. 08HQPA0015

** Metadata Profile for the National Digital Catalog, version 1.0

[<http://datapreservation.usgs.gov/docs/NGGDPPMetadataProfile.pdf>,]

3. We prepared the Long-Range Data Preservation Plan for WGNHS Research Collections (LRDPP). The document is an evolving document built from the WGNHS Data Catalog, the WGNHS Collections Policy, and the WGNHS Scanning Policy, with goals/timelines/strategy

assessments prepared for specific data preservation activities related to Survey collections. These documents were developed in FY 2008 as the fundamental elements of our LRDPP.

a. WGNHS Data Catalog

The Data Catalog is a Microsoft Access® application tied to a database containing an inventory of all data collections (>150) at the Survey. An excerpt from the training manual developed for this application is included below. In response to staff input at training sessions for using the Data Catalog, we are currently developing a web-based application and improving the search function for the “DataCat.” Excerpt from training manual (page 1):

Introduction

The WGNHS Data Catalog (DataCat) is an inventory of all Collections at the Survey. It is organized into a series of related SQL Server database tables and accessed through Microsoft Access.

What are Collections? Collections are groups of related data and interpretations, and include both Projects and Resources. Projects are defined units of work that are focused on a specific question or issue and usually have a defined lifespan. Examples include a geologic mapping project or a groundwater flow model. The Bedrock Geology of Marathon County and Groundwater Recharge in Southeastern Wisconsin are specific Projects. Resources are materials collected or maintained outside of a particular Project and are typically for shared reference or research use by staff members. For example, the core samples housed in Mt.Horeb and the Geobase database are Resources.

Information in the WGNHS Data Catalog is organized by the concepts of Collections, Elements, and Media. Collections may contain one or more Elements, and Elements may contain one or more Media. However, all Collections must contain at least one Element and one Media. This will become clearer as you progress through this document and work with the various Collections.

Purpose and Goals

We collect, interpret, disseminate, and archive natural resource information.

This statement from the Mission of WGNHS is at the heart of our interest in data preservation. Knowing what data we have now and are likely to collect in the future is a basic part of doing our work. The WGNHS Data Catalog is a tool the Survey and its staff will use to maintain a complete and accurate inventory of its various data Collections.

Provide content, media and location information on all Survey resources in one place;
 Provide knowledge of available resources to survey staff and the public;
 Encourage communication among staff members with regard to individual and group projects;
 Assist the Survey to properly evaluate, preserve and maintain its resources;
 Assist in the Survey’s contributions to the USGS National Digital Catalog; and
 Permit the Survey to better utilize its resources in outreach and educational projects.

WORDS THAT APPEAR IN THE DATA DICTIONARY (pp. 9-12) ARE UNDERLINED WHEN FIRST MENTIONED IN THIS DOCUMENT.

b. WGNHS Collections Policy

In addition to the WGNHS Data Catalog or inventory of all Survey data collections, the WGNHS Long-Range Data Preservation Plan includes the Collections Policy, which is a 22-page document describing our approach to collections: their definition, location, use, and care.

From Acquisition to Access...

COLLECTIONS POLICY WISCONSIN GEOLOGICAL AND NATURAL HISTORY SURVEY JULY 1, 2009

The Mission of the Wisconsin Geological and Natural History Survey

The Survey conducts earth-science surveys, field studies, and research; provides objective scientific information about the geology, mineral resources, water resources, soil, and biology of Wisconsin; collects, interprets, disseminates, and archives natural resource information; communicates the results of its activities through publications, technical talks, and responses to inquiries from the public. WGNHS activities support informed decision making by government, industry, business, and individual citizens of Wisconsin.

Table of Contents

Overview	1	<i>Legacy Materials and Field Notebooks</i>	
Scientific Information		WGNHS Projects	
Historic Information		<i>Pre-Project</i>	
Non-WGNHS Materials and Partnerships		<i>During Project</i>	
WGNHS Data Catalog		<i>Post-Project</i>	
General Collections Issues	3	<i>Ongoing</i>	
Use of Collections		<i>Project Storage</i>	
<i>By Staff</i>		Physical Data Sets	
<i>By Others</i>		<i>Core</i>	
Data Preservation		<i>Cuttings</i>	
<i>Staff Expectations</i>		<i>Hand samples</i>	
<i>Staff Archivist</i>		<i>Sediment samples</i>	
<i>Storage Practices</i>		<i>Thin sections</i>	
Orienting New Staff		Digital Data Sets	
Departing Staff		<i>Scanning Policies</i>	
Appraisal of Collections:		<i>Project data files</i>	
A Tutorial		<i>Project map files</i>	
De-Accessioning Collections		<i>Archives</i>	
		<i>Inventory</i>	
		Publications	
		<i>Inventory</i>	
Specific Collections Issues	10	<i>Archives</i>	
Background Information		Water-well Records	
Reference Collection		<i>Well construction Reports</i>	
<i>WGNHS and Pre-1897 Publications</i>		<i>Geologic Logs</i>	
<i>UW Digital Collection</i>			
<i>Publications of</i>		Bibliography of Online Resources	20
<i>Other Organizations</i>		Science Organizations	
<i>Theses and Manuscripts</i>		Library and Archival Organizations	

c. WGNHS Scanning Policy

A critical element of the WGNHS LRDPP is the digital capture of information—referred to here as the Scanning Policy. This policy is currently being reviewed and currently stands as a separate document, but we anticipate incorporating it into the WGNHS Collections Policy either as an Appendix or as an integrated section. It is being used to guide decisions on “what to scan” and “at what resolution should data be digitally captured.” The first page of the 3-page scanning policy follows:

SCANNING POLICY
WISCONSIN GEOLOGICAL AND NATURAL HISTORY SURVEY
September 2009

Background

The Survey is fully committed to preserving and providing appropriate access to information that is critical to understanding the geoscience heritage of the State of Wisconsin, and will take all reasonable measures to ensure that information is collected, maintained and preserved in accordance with its Collections Policy. Digital scanning is one method to preserve and facilitate access to analog documents in WGNHS collections. This Scanning Policy expands upon the Collections Policy by stating the purpose and goals for scanning projects and by setting minimum standards for evaluation of existing scanned files and for preparation and capture of new documents.

Scanning of documents within WGNHS collections provides increased access to project files for further research, an archive of the information in historic collections, and a format for continued reproduction and distribution of publications through Map Sales.

General Information

Scanning at the Survey takes place both as focused scanning projects and within the context of broader research projects. Scanning projects should be approached with the cooperation and assistance of other Survey staff members. Without proper oversight and expertise, scanning projects can be a misuse of WGNHS resources and scanned images may be inaccessible or unusable for their intended purpose. GIS staff have experience in scanning and particular expertise in the requirements for reproduction from digital copies.

Proposals for focused scanning projects should consider anticipated costs, intended purpose, and the value of the project relative to the broader WGNHS mission. Ensure that scanned images will meet appropriate quality standards for their intended use and that the files will be properly organized and maintained. Consider whether scanning will allow for de-accessioning of the original materials.

The type of source document will determine preparation, scanning, and file storage requirements. Not all survey collections can or should be scanned. Some materials, such as samples and core, are not physically compatible with the technology. More recently prepared maps and reports have been produced digitally and scanned versions would be, at best, redundant. Some materials have been scanned and the quality of the scans varies. Before scanning documents, determine if a digital copy exists and the quality of the digital version so that Survey resources are used as efficiently as possible. Scan resolution and subsequent image quality will vary, depending on the purpose of the file.

Document Management

Document Tracking

An inventory of scanned documents will be maintained within the publications database, including information about the analog source document as well as attributes of the scanned document, such as resolution, color depth, and amount of post-processing. This inventory will facilitate access to scanned documents, help track progress of preservation activities, and reduce duplication of effort.

Image Resolution and Processing

Existing scanned document files must be reviewed and classified to determine if they are of high enough resolution for their intended use. Re-scanning files not at the highest resolution should only be done if there is an immediate demand for production and distribution of the image.