National Geological and Geophysical Data Preservation Program
Final Technical Report for FY 2014

Digital Infrastructure and Data Rescue of Historic Louisiana Geological Publications

Award Number: G14AP00145

Principal Investigator: Patrick O'Neill, Research Associate III
Louisiana Geological Survey

Louisiana State University
Baton Rouge, LA 70803
Phone: 225/578-8590
Fax: 225/578-3662
Email: poneil2@lsu.edu

Project Term: September 1, 2014 to August 31, 2015

Submittal Date: November 30, 2015

Table of Contents

Abstract 2

Introduction 2

Key tasks 3

Results 4

Digital Archive effort 4

Digital Infrastructure effort 5

Data Preservation 5

Abstract

The Louisiana Geological Survey (LGS) participated in the FY 2015 phase of the National Geological and Geophysical Preservation Program (NGGDPP) in order to digitize its publications dating back to 1869. Among these publications are geological bulletins and pamphlets, water resources data, technical reports, and field guides.

The physical inventory of these publications in hard copy form has dwindled over time. At the same time, there has been an increase in demand for them in digital form. The earliest reports were published by the State of Louisiana's Department of Conservation, which predated the establishment of the LGS in 1934. Original copies of these documents are considered rare, and can only be found in libraries or in the possession of private collectors.

The LGS Publication Catalog served as the inventory for this phase of the archival project. The Catalog is organized into many categories. This is largely due to the fact that LGS is an 81 year-old state agency which has undergone many changes in organization and priorities over time.

Documents were prepared for scanning by selecting those in the best conditions and physically dissembling each one. Much of the preparation process involved removing staples and/ or trimming adhesive off of the edge of each document, which was then scanned at 300 pixels per inch (dpi) and stored as a portable document file (pdf). The nine publications categorized as Early Reports were handled differently due to fragility and the impracticality of dissembling rare books. Metadata records of each file were then uploaded into the National Digital Catalog.

Introduction

Funding for this phase of the NGGDPP award was granted in August 2014.

The earliest LGS publications were printed between 1869 and 1909, and the remaining copies are in fragile condition. There is a limited number still in existence. Rare books such as these required delicate handling, and were scanned with more sophisticated equipment by the Digital Services Staff at Louisiana State University's Middleton Library. The rest of them were scanned by LGS' digital infrastructure team. This Final Technical Report describes project details for the FY 2014.

Background: The LGS publications storage rooms contain hundreds of geoscience bulletins, pamphlets, field guidebooks, technical reports, and maps and atlases dating back as far as 1869. This material encompasses the complete catalog of LGS publications and its predecessor agencies, with the exception of large format documents such as maps and folios. Many of these publications are out of print and are becoming increasingly difficult for the research community to acquire. Much of the remaining stock is water damaged or moldy, and soon to be discarded. Many of the Early Reports are over 100 years old and in fragile condition.

Objective: The goal of this phase of NGGDPP is to systematically and digitally archive a large collection of valuable geological and geophysical publications and data that have until now been available in hard copy form only. Making these historic illustrated publications searchable with national Geophysical Data Center (NGDC) compliant metadata will improve public access to LGS publications.

Team: Patrick O'Neill, John Snead, and Reed Bourgeois were the LGS staff members involved in this phase of the NGGDPP.

Key tasks

Project Description: Digitizing of LGS documents began in FY 2010, when it participated in a NGGDPP project that took inventory and assessed documents for digital conversion. This phase of the NGGDPP simply augments the effort begun in 2010 by adding the catalog of published bulletins, pamphlets, field guides, and technical reports to its digitizing efforts.

The project also cataloged and scanned the individual text pages of the geologic publications including thousands of photographs and page-size illustrations contained within. Those publications were saved as portable document files (pdf) for digital archiving and access via web site. Metadata records were uploaded to the national catalog.

Methodology: The inventory is defined as all items in the LGS Publications Catalog minus large format items such as maps and folios, which were digitized in the first phase of NGGDPP. Each document had to be assessed for condition and dissembled before scanning, with the exception of the nine Early Reports.

Digitization and archival is accomplished in the following steps:

1 – Choose a copy of the publication that is in the most ideal condition, taking into account any discoloration, water damage, and condition of the binding.

2 – Dissemble publication by removing staples or fasteners, then carefully tear adhesive used along edges of each page.

3 – Trim frayed edges of each page to uniform size using a straight-edged paper cutter.

4 – Feed publication through the document scanner, and save the temporary digitized versions as portable document files (pdf). Covers are scanned in color, while the bodies of most publications were scanned in grayscale mode.

5 – Edit and proofread each document, using software to arrange covers, crop pages to uniform size, rotate pages to proper view, and save a completed, permanent digital version of the document as a pdf.

6 – Create a metadata record for each digital document file, and digitally archive each file, metadata record, and publication database on multiple hard drives and server.

The nine rare and fragile Early Reports were handled differently. For the sake of preserving these rare books, dissembling was not a practical option, and the fragile condition of these books does not allow for feeding through a standard automated document scanner, so a more delicate process known as non-destructive scanning was employed. Personnel at LSU's Middleton Library digitized these publications on a sophisticated book scanner manually, two pages at a time, and edited to separate the pages and crop off the "black gutter" area near the spine. For consistency, the same resolution settings -- grayscale mode at 300 pixels per inch (dpi) -- were used.

Results

Digital Archive Effort

The LGS Digital Infrastructure Team successfully scanned, archived, and created metadata records for over 180 historic Louisiana publications, many of which are now very scarce in paper form. Surprisingly, some of the bulletins published in the 1930's and 1940's were in greater supply than those published much more recently. This was in part due to the fact that the research community was much more dependent of paper documents in those years, but also reflects the priorities and budgetary resources at the time of their printing. In recent years, printing costs have increased, but so has the demand for quick access to digital data. Moreover, many of the paper documents have been stored in areas without climate control, making them susceptible to mold, water damage, and general decay. Thus digital archiving is the logical solution to preserving historic documents. A large number of paper items – books, pamphlets, and folios – will now be discarded, having been deemed unmarketable, and now available in digital form.

Among the 195 documents scanned and archived were:

111 Books over 100 pages
50 Booklets under 100 pages
34 Open-File reports (unbound photocopies)

Among the types noted were:

9 Early Reports
44 Geological Bulletins

47 Water Resources publications
36 Publications relative to oil, gas and minerals
34 Miscellaneous topics (Open-File reports)
9 Coastal Geology Technical Reports
7 Field trip guide books
5 Anthropological studies
4 Educational booklets

Digital Infrastructure effort

Of the 317 items in the LGS Publications Catalog, 195 were deemed suitable as subjects for this phase of the NGGDPP. Two were excluded because of a copyright issue with Geological Bulletin 37, and Geological Pamphlet 9, for which a copy could not be found.

Metadata records for these documents were developed by creating a spreadsheet database with the following fields, so as to comply with the NGGDPP format: *collectionID, Title, Alternate Title, Abstract, datatype, supplementalInformation, coordinates, alternateGeometry, onlineResource, browseGraphic, Date, datasetReferenceDate, and verticalExtent*. However, several of the fields -- *coordinates, alternateGeometry, onlineResource, browseGraphic, datasetReferenceDate, and verticalExtent* -- contain no data because these parameters were not applicable to the types of publications being archived. Data for the applicable fields was entered, and a sample metadata record sample follows.

collectionID: (left blank, to be assigned by USGS)

Title: Geology of Iberia Parish

Alternate Title: GB-1

Abstract: Includes discussions of physiography, with special reference to Pleistocene terraces and to drainage; stratigraphy; origin of the salt domes and caprock; and mineral resources. Includes annotated bibliography. 187 pages Authors: H. V. Howe and C. K. Moresi

datatype: digital publication pdf

supplementalInformation: http://www.lgs.lsu.edu/deploy/publications/publicationcatalog.php

Date: 1931

After final review and consultation with Natalie Latysh, a digital database file was uploaded into the National Digital Catalog.

Data preservation

Project data management and preservation: The digital inventory created by this project is in the form of 195 portable document files (pdf's). This data is not directly available online, but can be uploaded to a web site in which files can be shared with interested parties. The digital files generated by this project are stored on several hard drives with additional copies on two LGS servers. Standard protocols for

backup and security are in place. In time, the data will also reside in LSU's planned Institutional Repository and enjoy full internet access with NSF-compatible, professional data management and preservation protocols.

Long-term strategy for data preservation: The long-term strategy is the integration of LGS digital publications into a preservation scheme compatible with that of Louisiana State University (LSU). The University's scheme is presently undergoing evaluation to comply with the National Science Foundation's (NSF) data management mandate. The LSU Center for Computation and Technology has established a web-based Institutional Repository that is considered the best way to meet the goals of campus libraries and research institutes and to satisfy requirements for data management and preservation. When fully implemented, the LGS digital data will become part of the LSU Institutional Repository. Until then, the University Archives, managed by the LSU Libraries Special Collections unit, is the official repository for all permanent non-current records of enduring historical and administrative value from all units of the University.

According to LSU Policy Statement 74,

"Procedures for implementing this policy will be developed by the University Archivist during the course of a two year University Records Survey. One product of the Survey will be a Records Management Procedures Manual, which will outline the procedures to be followed by University personnel."
http://sites01.lsu.edu/wp/policiesprocedures/policies-procedures/74/