

MA Data Preservation Program FY13: Conversion of Geological Documents To Digital Format

National Geophysical and Geological Data Preservation Program
Award No. G13AP00103

Final Technical Report

September 15, 2013 to September 14, 2014



Principal Investigator/Author
Stephen B. Mabee, Ph.D., PG
State Geologist
Department of Geosciences
University of Massachusetts
611 North Pleasant Street
Amherst, MA 01003
413-545-4814 (Phone)
413-545-1200 (Fax)
sbmabee@geo.umass.edu

Geologist/Project Manager
Nick Venti, Ph.D.
Massachusetts Geological Survey
Department of Geosciences
University of Massachusetts
611 North Pleasant Street
Amherst, MA 01003
413-545-2538 (Phone)
413-545-1200 (Fax)
nventi@geo.umass.edu

July 10, 2014

Abstract

Metadata for 50 publications was created and uploaded to the National Digital Catalog. The metadata was created in CSV format in accordance with the procedures and schema outlined in the Metadata Profile for the National Catalog (March 17, 2008; version 1.0) and the NGGDPP Metadata Preparation Guide (01/2013).

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

http://datapreservation.usgs.gov/docs/NGGDPP_MetadataPreparation_CatalogGuide_012013.pdf

The metadata can be found at:

<https://www.sciencebase.gov/catalog/item/53bc2cb9e4b084059e8bffe5>

These 50 publications are part of a larger collection of paper documents that comprise the University of Massachusetts, Department of Geosciences Contribution Series. This series consists of field trip guidebooks, field reports, theses and dissertations prepared primarily by faculty and students in the Department of Geosciences at the University of Massachusetts in Amherst. These documents were printed by the University of Massachusetts and were made available to the public upon request. Many of the documents are now out of print or have been damaged (moisture) due to poor storage environments and recent flooding. Accordingly, the Massachusetts Geological Survey, which is housed in the Department of Geosciences, took the responsibility of scanning a subset of this collection so that the documents can be available to the public online.

The entire Contribution Series collection contains about 74 publications ranging in time from 1969 to 2000. For this project, 50 have been scanned. The remaining 24 publications will be scanned as part of NGGDPP FY15, for which funding was just received. The online digital versions of these publications can be found at:

<http://www.geo.umass.edu/research/list.html#field>

Mandatory and optional fields included in the metadata file are: collectionID, Title (gives the Contribution Series Number), Alternate Title (gives the title of the publication), Abstract, which gives the full reference for the publication, DataType, SupplementalInformation, which includes the URL where the data can be found, Coordinates, the centroid (or thereabouts) of the study area or map, and DatasetReferenceDate. All metadata records were inspected visually twice for accuracy and typographical errors. Any errors were corrected.

Introduction

The following technical report describes the work completed on Award No. G13AP00103. The award period was from September 15, 2013 to September 14, 2014. This award is for the creation or updating of digital infrastructure. The metadata provided is for 50 published field trip guides, field reports, theses and dissertations that are part of the University of Massachusetts, Department of Geosciences Contribution Series. This collection of geologic reports was originally printed by the University of Massachusetts for the purpose of public distribution. Many of the reports are out of print or have been damaged (flooded) due to poor storage conditions. The Massachusetts Geological Survey took responsibility to scan these documents so that they can be available to the public for download. All metadata for this project were uploaded to the ScienceBase.gov as instructed on July 9, 2014. The metadata file can be found at: <https://www.sciencebase.gov/catalog/item/53bc2cb9e4b084059e8bffe5>

The scanned documents can be found at:

<http://www.geo.umass.edu/research/list.html#field>

Work Completed Between September 15, 2013 and September 14, 2014

Good quality hard copies of the 50 geologic documents were retrieved from various sources. If the volume is a rare book that cannot be taken apart, each individual page was scanned on a flat bed scanner using Adobe Acrobat Pro XI and Adobe Creative Suite 6 software. Once scanned, the pages were compiled into a single PDF. Margins were cropped as necessary. Pages containing photographs were scanned and processed separately to improve quality. Large maps and plates were scanned on a large format scanner and added to the PDF. After each scan, the individual pages were inspected visually for quality control and a page count was conducted to make sure no pages were missing. Random pages were zoomed to inspect pages for blurriness or pixilation. Pages were replaced as necessary. Landscape pages were rotated so they are readable when opened. All documents were analyzed with Optical Character Recognition software to create searchable PDF documents.

Documents that could be disassembled were taken to a copy works vender to remove the bindings. Documents were then scanned to USB on a Xerox WorkCentre 5845 in the Department of Geosciences. Pages were cropped as necessary. After scanning, all documents were analyzed with Optical Character Recognition software. Each page was inspected visually for quality. If there were oversized maps these were scanned separately on a large format scanner and added manually to the PDF.

The metadata were prepared in accordance with the procedures outlined in the metadata profile for the National Catalog (March 17, 2008; version 1.0) and the NGGDPP Metadata Preparation Guide (01/2013).

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

http://datapreservation.usgs.gov/docs/NGGDPP_MetadataPreparation_CatalogGuide_012013.pdf

All mandatory fields were populated as well as one optional field, AlternateTitle.

Nick Venti performed all the scanning and developed a detailed protocol for the operation. Dr. Mabee and Nick Venti conducted quality control and Dr. Mabee completed all the metadata preparation and submissions.

Comparison of Accomplishments to Project Goals

The project goal was to scan 45 documents in the Contribution Series collection. A total of 50 documents were scanned, exceeding project expectations. All work was completed on time and on budget.