U.S. Geological Survey Grant Award Number G12AP20129
Final Technical Report

UGS FY2012 GEOLOGIC DATA PRESERVATION PROJECT

Steve D. Bowman, Grant C. Willis, and Peter J. Nielsen

Utah Geological Survey
1594 W. North Temple
P.O. Box 146100
Salt Lake City, Utah 84114-6100
(801) 537-3304, fax (801) 537-3400
stevebowman@utah.gov
grantwillis@utah.gov
peternielsen@utah.gov

geology.utah.gov

Project Period: August 15, 2012 to August 14, 2013

October 17, 2013

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number G12AP20129, and the Utah Geological Survey. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official opinions or policies of the U.S. Government. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Government.

CONTENTS

ABSTRACT

In response to the U.S. Geological Survey FY2012 National Geological and Geophysical Data Preservation Program (NGGDPP) announcement, the Utah Geological Survey (UGS) proposed a collaborative project in three internal programs that focused on four goals of the UGS data preservation plan. The major focus of this project is conversion of existing paper records to digital formats. The Geologic Hazards Program proposed to continue organizing, scanning, and creating metadata for our geologic hazard and engineering geology document collection and our historical aerial photography housed at the UGS. The Geologic Mapping Program proposed to continue scanning, cleaning, georeferencing, and creating metadata for historical geologic maps housed at the UGS. The Utah Core Research Center proposed a pilot project to organize, scan, and create metadata for our coal drill-hole data. All three internal programs collaborated as a single NGGDPP project and all quantities of paper to digital conversion and metadata creation stated in the original proposal were exceeded. Specifically, the UGS created metadata for 2958 engineering geology and geologic hazard reports (over 49,443 pages), 17,055 aerial photographs, 228 geologic maps of Utah, and 151 coal drill holes. To be as cost-effective as possible, Utah Correctional Industries (UCI) at the Utah State Prison, Draper, was used for the majority of digital scanning services. Because UCI uses these types of projects to provide work experience for prison inmates, they are able to perform the work at a small fraction of in-house or commercial rates.

INTRODUCTION

Scope of Project

The Utah Geological Survey (UGS) began geologic data preservation activities in the U.S. Geological Survey (USGS) National Geological and Geophysical Data Preservation Program (NGGDPP) with the FY2007 grant award, and has continued in the program with subsequent awards in FY2008, FY2009, FY2010, FY2011, and FY2012.

In recognition of the value and importance of physical and digital geologic resources, the UGS has created a long-range data preservation plan for inventorying, archiving, and preserving geologic, geophysical, and engineering data, maps, well logs, and physical samples for future use by government agencies, industry, academia, and the general public. This project is a part of our long-range data preservation plan activities.

In response to the USGS FY2012 NGGDPP announcement, the UGS proposed a collaborative project in three internal programs that focused on four goals of the UGS data preservation plan. The major focus of this project is conversion of existing paper records to digital formats. The Geologic Hazards Program proposed to continue organizing, scanning, and creating metadata for our geologic hazard and engineering geology document collection and the historical aerial photography housed at the UGS. The Geologic Mapping Program proposed to continue scanning, cleaning, georeferencing, and creating metadata for historical geologic maps housed at the UGS. The Utah Core Research Center (UCRC) proposed a pilot project to organize, scan, and create metadata for our coal drill-hole data. All three internal programs collaborated as a single NGGDPP project and all quantities of paper to digital conversion and metadata creation stated in the original proposal were exceeded.

RESULTS

The FY2012 Utah NGGDPP project consisted of a collaborative project in three internal programs. Table 1 lists the UGS collections that were part of the FY2012 Utah NGGDPP project, along

with their respective original National Digital Catalog (NDC, https://www.sciencebase.gov/catalog/?community=NGGDPP+-+National+Geological+and+Geophysical+Data+Preservation+Program) collection identification numbers, item quantities originally proposed to have metadata created for, and the actual number of items that were processed and for which metadata was created. The goals set forth in our original proposal (April 12, 2012) that included the proposed quantities of paper to digital conversion and creation of metadata were all exceeded during the course of this project.

Table 1 – FY2012 NGGDPP Award Progress Summary

Collection		Item Quantities	
Name	**NDC ID[1]**	**Proposed**	**Completed**
UGS GeoData Archive System (engineering geology / geologic hazard documents)	4777844	40,000 pages	49,443 pages
UGS Aerial Imagery Collection	4777842	10,000 frames	17,055 frames
UGS Geologic Map Collection	2203971	225 maps	228 maps
UGS UCRC Rock Core Collection	2369695	Modify Integral2 database.	Integral2 database modified; 151 records added.

[1] – Original (pre-latest USGS ScienceBase system) NDC identification number.

Metadata created as part of this project followed the NGGDPP specifications (Data Preservation Working Group, 2006; USGS, 2008, 2013) and an XML template (http://datapreservation.usgs.gov/catalog_and_metadata_reference_info/NGGDPP2SampleMetadata.xml) developed by the NGGDPP. The UGS provided metadata for the UGS Aerial Imagery and GeoData Archive System Collections to the USGS in XML format for upload into the NDC, and in CSV format for the UGS Geologic Map and UCRC Rock Core Collections, with one file for each collection.

UGS Aerial Imagery Collection

The UGS Aerial Imagery Collection contains aerial photography of Utah originally taken from 1935 to the present, and includes over 120,000 frames and associated indexes, orthophotomaps (semi-controlled orthophotos), and other materials. Aerial photographs are highly sought after for use in geologic, geotechnical, and environmental investigations to document geologic hazards, land-use, geomorphologic, and other changes that may have occurred in a particular area, along with mapping geology, cultural, and vegetation features. The Geologic Hazards Program inventoried, scanned, and created metadata on 17,055 photographs (frames) in the collection.

In 2009, as part of the metadata creation process, the UGS developed the ImageryManager database system to manage the collection and store associated metadata. The system uses a Microsoft SQL Server 2008 back-end database and a C# Windows front-end application developed specifically for storing aerial photography and other imagery metadata. The system was developed using modern programming tools and methods, and allows for continued expansion of metadata records and future functionality improvements. In 2013, the UGS developed the UGS Aerial Imagery Collection (http://geology.utah.gov/databases/imagery/) web-based search application for public access to frames in the public domain. Currently, over 78,000 frames are available online for searching, viewing, and download.

During the metadata creation process, individual aerial photography frames were sorted by the original acquiring agency, project code, and frame number, creating project sets. These project sets were then assessed for importance in current geologic projects at the UGS and public needs, and a priority list

was developed. Project sets in priority order were then sent to Utah Correctional Industries (UCI) at the Utah State Prison, Draper for digital scanning using UGS-owned and maintained scanners (Epson Expression 10000XL and Creo EverSmart Select II). To be as cost-effective as possible, UCI was used for the majority of digital scanning services. Because UCI uses these types of projects to provide work experience for prison inmates, they are able to perform the work at a small fraction of in-house or commercial rates.

Due to degradation from use, potential for loss and theft, and replacement issues (high cost or irreplaceable), a goal of the UGS Geologic Data Preservation Plan is to digitally scan frames for general use by end-users. This allows end-users access to a high-quality reproduction of the original paper photograph or film positive/negative that can be printed or used in software applications without damaging the original. UCI scanned the paper photograph frames using color-profiled Epson 10000XL scanners and film photograph frames on a Creo EverSmart Select II scanner with an Adobe RGB (1998) color profile. Paper photograph frames were scanned at full-size, with a resolution of 800 dots per inch (dpi), 8-bit color or grayscale depth, and saved in TIFF format. Film photograph frames were scanned at full-size, with a resolution of 1200 dpi, 8-bit color or grayscale depth, and saved in TIFF format. In addition, UCI entered basic metadata into the TIFF files using Adobe's XMP metadata system for subsequent UGS review and upload into our ImageryManager database.

After scanning each frame, a quality control check at UCI was performed on each digital file to verify the correct filename based on a pre-defined file naming system, and that the scan was of high quality. Digital files that did not meet the quality control requirements were sent back for rescanning. During file transfer from UCI to the UGS, a second quality control check was performed to verify correct filenames, metadata, and file quality. The UGS found very few digital files produced by UCI that did not meet our quality requirements and that required subsequent rescanning.

Index maps associated with each project set, where available, were scanned on a Contex Chroma HS 42 scanner, with a resolution of 600 dpi, 8-bit color or grayscale depth, and saved in TIFF and Adobe PDF formats. These index maps were then enhanced in Adobe Photoshop to remove scanning artifacts, dust, and other defects in the scanned image. After enhancement, each index map with coordinate data was georeferenced in ESRI ArcGIS 10.1 software, after which frame center points were digitized if possible given the index map format. Each index map was georeferenced using the North American Datum of 1927 (NAD27), the native map datum, and was then converted to the current NAD83 datum. Metadata on the index maps was not collected as part of this project, and data is not included within the metadata XML file uploaded to the USGS.

Metadata within the XML file for the historical aerial photography follows the NGGDPP specification and includes the following fields with a description of each field:

- Collection ID – A unique identifier originally assigned by the NGGDPP for each collection within the NDC. The USGS assigned an identification of 4777842 to the UGS Aerial Imagery Collection. However, this identification is no longer used in NDC.

- Title – The title field was populated with the individual aerial frame project year, project code, and roll number, flight line number, and frame number, where available. Example: Aerial Photograph: 1937 AAJ-AAK 10-1-1

- Alternate Title – The alternate title field was populated with the name of the original acquiring agency. Example: Original Acquiring Agency: USDA, Agricultural Adjustment Administration

- Abstract – The abstract field was populated with a standard description of the individual record. Example: This item represents a single aerial photography frame that is part of the Utah Geological Survey (UGS) UGS Aerial Imagery Collection, housed at the UGS offices in Salt Lake City. Frames may be available from the DNR Library (801) 537-3333.

- Data Type – The data type field was populated with a common descriptor from the controlled list. Example: Photographs

- Supplemental Information – The supplemental information field was populated with the location of the digital file at the UGS. Example: UGS Digital Location: I:\Imagery\1937_AAJ-AAK\10-AAJ_1-1.tif

- Coordinates – The coordinates field was populated with geographic longitude and latitude frame center point coordinates (NAD83). Example: -110.60219, 40.20762

- Online Resource – The online resource field was populated with a URL address of a web page describing the UGS Aerial Imagery Collection. Example: http://geology.utah.gov/ghp/consultants/aerial_compilations.htm

- Date – The date field was populated with the date of the aerial photograph, as printed on the frame. Example: 1935-04-12

- Dataset Reference Date – The dataset reference date field was populated with the date of the digital scan. Example: 2009-01-02

UGS GeoData Archive System

The UGS has collected numerous unpublished reports, maps, memorandums, field notes, and other geologic-hazard and engineering geology documents since the formation of the Site Investigation Section (now Geologic Hazards Program) in 1980. Few copies were ever produced of most of the documents in the collection. These documents are used in geologic-hazard investigations, geologic and engineering geologic mapping projects, during emergency response activities, and in response to public inquiries. The UGS incorporated brief metadata for some of these documents into the now defunct UGS-developed HAZBIB (Microsoft Access) database, starting in 1985. Due to limited metadata entries (title, author, general location, and geologic-hazard category), lack of text-searchable digital files, and use of outdated database technology, the defunct HAZBIB database was replaced in 2010, by the UGS GeoData Archive System (https://geodata.geology.utah.gov/), a system built with the open-source ResourceSpace digital asset management system (http://www.resourcespace.org/) to enter, store, query, and retrieve metadata and scanned documents. Metadata describing each resource is searchable, along with spatial searching for resources that are local or site-specific in nature. In response to the needs of other UGS programs and the UGS Geologic Data Preservation Project, the UGS GeoData Archive System was extended to collections beyond the original HAZBIB database in 2011. The Geologic Hazards Program inventoried and created metadata on 49,443 pages (2958 reports) in the collection.

The UGS collected and sorted geologic hazard and engineering geology documents from our data holdings and sent them to UCI for scanning (using the Adobe PDF format), optical character recognition (OCR), and partial metadata creation. UCI entered basic metadata fields, such as title, author, date, publisher, report number, county/state location, document type, and media type into XMP compliant metadata within the PDF file. A UGS geologist or geological technician evaluated each document, reviewed UCI's metadata, and entered additional metadata fields of geologic hazard type (keywords),

geographic coordinates, copyright status, and other technical information or notes as needed in the UGS GeoData Archive System during the document upload process.

Metadata within the XML file for the engineering geology and geologic hazard reports follows the NGGDPP specification and includes the following fields with a description of each field:

- Collection ID – A unique identifier originally assigned by the NGGDPP for each collection within the NDC. The USGS assigned an identification of 4777844 to the UGS GeoData Archive System Collection. However, this identification is no longer used in NDC.

- Title – The title field was populated with the individual report title. Example: Report; Geotechnical/Engineering Geology Reconnaissance Study; Proposed Approximate Eight-Acre, One-Lot Subdivision; North and East of Tomahawk Drive and North of Limekiln Gulch; Salt Lake City, Utah.

- Abstract – The abstract field was populated with a standard description of the individual record and report author. Example: Report contained in the UGS GeoData Archive System. Author: Gordon, W.

- Data Type – The data type field was populated with a common descriptor from the controlled list. Example: Reports

- Supplemental Information – The supplemental information field was populated with the ResourceSpace (UGS GeoData Archive System) resource identification number. Example: Resource Space/UGS GeoData Archive System Resource ID 436

- Coordinates – The coordinates field was populated with geographic longitude and latitude of the approximate center of the report-referenced project area (NAD83). Example: -110.60219, 40.20762

- Online Resource – The online resource field was populated with a URL address of the UGS GeoData Archive System. Example: http://geodata.geology.utah.gov

- Date – The date field was populated with the date of the report. Example: 1994-01-03

- Dataset Reference Date – The dataset reference date field was populated with the date of the digital scan. Example: 2011-06-23

UGS Geologic Map Collection

The UGS Geologic Map Database (called MAPBIB) contains an inventory of approximately 2900 geologic maps covering various parts of Utah, that date from about 1890 to present, and that vary from formally published USGS and UGS Map and Bulletin Series maps, to informal consulting reports and internal unpublished “sketch” maps. About half of these maps were published in production runs in which hundreds of copies were printed – most of these maps are still widely distributed and readily available. However, over 1000 of the maps in our MAPBIB database were produced in only very limited numbers, are completely out of print, are in obscure or remote libraries, or are limited-distribution “gray” literature that is very hard to locate. For a few hundred maps, we do not know the exact number, only a single copy may exist. Our long-term goal is to locate and preserve the best possible copy of all maps in the database, and to produce and archive a high-resolution digital copy of each map.

For the FY2012 Utah NGGDPP project, we archived 228 geologic maps. To achieve this, we completed the following steps: (1) assigned priorities to the geologic maps in the MAPBIB database (priorities were based primarily on scale, detail, completeness, scarcity, uniqueness, and/or perceived geologic value of the map, and "best available" coverage of an area); (2) searched out the best paper or Mylar copies of the selected maps; (3) cleaned and flattened the copies; (4) scanned the copies at 600 dpi resolution on a Contex Chroma HS 42 large-format scanner that had been carefully cleaned and calibrated; (5) scanned all accompanying plates on the large-format scanner (booklets and other text were scanned by UCI); (6) imported scans into Adobe Photoshop software where images were checked for accuracy and clarity, cleaned, and straightened; (7) georeferenced the maps in ESRI ArcGIS 10.1 software using at least 16 control points to make sure that high precision was maintained, and saved the georeferenced maps as TIFF and low-compression JPEG files; (8) imported or digitized the map footprint into a spatial database for future spatial indexing and locating; (9) created Adobe PDF files of all of the maps and supporting materials to allow for easier distribution to the public; (10) created metadata that recorded dates, sources, procedures, and georeferencing information; and (11) prepared files for public distribution. These files are currently in the UGS review process.

Once the files are reviewed and approved, they will be served to the public through the UGS website (http://geology.utah.gov/maps/geomap/index.htm) using searchable lists and interactive maps (http://geology.utah.gov/maps/geomap/interactive/index.htm). Most important of all, the archive map files are stored in our UGS digital archive, and copies are sent to State Archives where valuable data is stored and managed. This latter step assures that the data will be migrated forward periodically to make sure file integrity is maintained and map files are preserved for future generations.

Metadata within the CSV file for the geologic maps follows the NGGDPP specification and includes the following fields with a description of each field:

- Collection ID – A unique identifier originally assigned by the NGGDPP for each collection within the NDC. The USGS assigned an identification of 2203971 to the UGS Geologic Map Collection. However, this identification is no longer used in NDC.

- Title – The title field was populated with the full title of the geologic map. Example: Tertiary stratigraphy of the Goose Creek district, Cassia County, Idaho, and adjacent parts of Utah and Nevada.

- Alternate Title – The alternate title field was populated with a shortened title containing key identifying terms. Example: Goose Creek district, Idaho, Utah, and Nevada. Scale: 63360

- Abstract – The abstract field was populated with a brief explanation of the archive map file contents and purpose. Example: This digital file is a high-resolution scan or copy of a published geologic map. It was created to archive a geologic map that could be lost to future generations.

- Data Type – The data type field was populated with a common descriptor from the controlled list. Example: Maps

- Supplemental Information – The supplemental information field was populated with the remaining components of a formal bibliographic reference for the map, including authors, publisher, and plate or figure number. Example: Mapel, W.J.; Hail, W.J., Jr.; Utah Geological Society, Guidebook to the Geology of Utah no. 11, p. 1-16, fig. 2.

- Coordinates – The coordinates field was populated with a calculated center point of each map in longitude and latitude (NAD83). Example: -114, 42.0354

- Online Resource – The online resource field was populated with a URL address of a web page that contains either the map, or a searchable index or database that will help any future user to find the digital map files. Example: http://geology.utah.gov/maps/geomap/index.htm

- Date – The date field was populated with the publication date of the geologic map; generally, this is only known to the nearest year. Example: 1956-01-01

- Dataset Reference Date – The dataset reference date field was populated with the date that the map was scanned for the data preservation archives. Example: 2010-03-31

UCRC Coal Drill-Hole Data

For the FY2012 Utah NGGDPP project, the Utah Core Research Center (UCRC) organized, scanned, and created metadata for 151 coal drill holes. For this project, the UGS manually collected and entered the metadata into the new Integral2 database. New metadata that was generated by this project was exported as a CSV file in conformance with NGGDPP specifications for upload into the NDC.

Metadata within the CSV file for the UCRC Core Collection follows the NGGDPP specification and includes the following fields with a description of each field:

- Collection ID – A unique identifier originally assigned by the NGGDPP for each collection within the NDC. The USGS assigned an identification of 2369695 to the UCRC Core Collection. However, this identification is no longer used in NDC.

- Title – The title field was populated with the combined information of the well operator, well name, well number, and the oil field/district name, if applicable. Example: Royal Land North Horn NH-23-1

- Alternate Title – The alternate title field was populated with the American Petroleum Institute (API) number assigned to each well. If no API number was available for a well, a false API number was used in prior data entry for identification purposes. Example: 43-15-00001-C000 [real API number used here]

- Abstract – The abstract field was populated with a description of the individual record. Example: Coal Company, geophysical log, reports

- Data Type – The data type field was populated with a descriptor from the controlled list that was most applicable (fluid sample, hand sample, rock core, rock cuttings). Example: Report

- Supplemental Information – The supplemental information field was populated with the UCRC phone number as a resource to obtain more information about sample holdings. Example: Phone, (UCRC): (801) 537-3359.

- Coordinates – The coordinates field was populated with geographic longitude and latitude coordinates in NAD83. Example: -111.22147, 39.2401

- Alternate Geometry – The alternate geometry field was populated with the Public Land Survey System (PLSS) section, township, and range, as an alternate method for locating well bores. Example: NWSW 23 18S 6E SLM Emery UT

- Online Resource – The online resource field was populated with a URL address of a web page describing the UCRC collections. Example: http://geology.utah.gov/emp/ucrc/index.html

- Date – The date field was populated with the date of metadata creation, as the well drilled date was not available at the time of this project. Example: 1981-08-26

- Dataset Reference Date – The dataset reference date field was populated with the date of the donation of the geological sample. If no donation date was available, the date of the creation of the metadata document was used. Example: 2013-07-23

- Vertical Extent – The vertical extent field was populated with sample depth information and defined as the unit of measurement, maximum value, and minimum value. Example: feet, 2340, 0

DATA AVAILABILITY

General metadata and other information about the UGS Aerial Imagery Collection, GeoData Archive System, Geologic Map Collection, and UCRC Collections is available at the UGS website (http://geology.utah.gov/) or by contacting the UGS at (801) 537-3300 and asking for the Geologic Information and Outreach Program.

The UGS Aerial Imagery Collection (http://geology.utah.gov/databases/imagery/) contains over 78,000 frames that may be searched, viewed, and downloaded. Information on using the collection is available by clicking on the Help tab. Additional information on aerial photography, along with four published sets and any accompanying reports, is available at http://geology.utah.gov/ghp/consultants/aerial_compilations.htm.

The UGS GeoData Archive System containing geologic hazard and engineering geology-related documents and other materials is available at https://geodata.geology.utah.gov/. Information on using the system and metadata descriptions are available by clicking on the Help & Advice link.

Information about the UGS Geologic Map Collection is available at http://geology.utah.gov/maps/geomap/index.htm and GIS data for some of the maps is available at http://geology.utah.gov/maps/gis/index.htm and on the AGRC State Geographic Information Database (SGID) server (http://gis.utah.gov/data/how-to-connect-to-the-sgid-via-sde/). The UGS recently launched an interactive geologic map on the UGS website (http://geology.utah.gov/maps/geomap/interactive/index.htm) that serves many of the maps scanned and archived through this program. About half of the maps are now available, and we will continue to add new maps during the coming year.

Information about the UCRC Rock Core Collection is available at http://geology.utah.gov/emp/ucrc/index.html. Interested users may obtain additional information about the samples and/or arrange to view the samples by contacting the UCRC at http://geology.utah.gov/emp/corecenter/index.htm or (801) 537-3359.

ACKNOWLEDGMENTS

This work was funded by U.S. Geological Survey Grant G12AP20129 as part of the National Geological and Geophysical Data Preservation Program and the Utah Geological Survey. The UGS thanks Betty Adrian, Sky Bristol, and Richard Brown, USGS, for their support. Gregg Beukelman, Taylor Boden, Ben Erickson, Adam Hiscock, Adam McKean, Pam Perri, and Starr Soliz of the UGS provided technical support for this project. Audrey Caudill (Supervisor), Inmate Noble, and several other inmates of UCI provided digital scanning support.

REFERENCES

Data Preservation Working Group of the National Cooperative Geologic Mapping Program Federal Advisory Committee, 2006, Implementation plan for the National Geological and Geophysical Data Preservation Program: U.S. Geological Survey, 26 p., available at http://datapreservation.usgs.gov/docs/2006DataPreservation.pdf.

U.S. Geological Survey, 2008, Metadata profile for the National Digital Catalog, March 17, 2008, Version 1.0: U.S. Geological Survey, 7 p., available at http://datapreservation.usgs.gov/docs/NGGDPPMetadataProfile.pdf.

U.S. Geological Survey, 2013, NGGDPP Metadata Preparation Guide (01/2013): U.S. Geological Survey, 5 p., available at http://datapreservation.usgs.gov/docs/NGGDPP_MetadataPreparation_CatalogGuide_012013.pdf