National Geological and Geophysical Data Preservation Program

FINAL TECHNICAL REPORT

October 28, 2009

By

Michael D. Laine, Principal Investigator and Ammon McDonald,
Utah Geological Survey
PO Box 146100
Salt Lake City, Utah 84116
Ph. 801-537-3359; Fax 801-538-4828
michaellaine@utah.gov

Award No. 08HQGR0121
Project Period: 09/15/08 through 09/14/09

Issued by:
Maggie Eastman
U.S. Geological Survey
Office of Acquisition & Grants
12201 Sunrise Valley Drive, MS 205
Reston, VA 20192

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Although this product represents the work of professional scientists, the Utah Department of Natural Resources, Utah Geological Survey, makes no warranty, expressed or implied, regarding its suitability for a particular use. The Utah Department of Natural Resources, Utah Geological Survey, shall not be liable under any circumstances for any direct, indirect, special, incidental, or consequential damages with respect to claims by users of this product.

TABLE OF CONTENTS

ABSTRACT 4

INTRODUCTION 4

PROJECT DESCRIPTION 4

THE UGS INTEGRAL DATABASE 5

METADATA FILE CREATION 5

METADATA FIELD DEFINITIONS 6

METADATA FIELD ELEMENTS 6

RESULTS 8

ABSTRACT

To fulfill the National Geological and Geophysical Data Preservation Program (NGGDPP) FY 2008 Program Priority, the Utah Geological Survey (UGS) proposed using the existing digital inventory of the Utah Core Research Center (UCRC) core collection to create files of metadata conforming to National Digital Catalog Metadata Product Requirements. The UCRC core collection was chosen because it is a manageable size, about 800 wells, and the digital data in the collection has an established schedule for updating and quality review.

At the start of the metadata file creation process, minor conflicts were discovered between several existing sources provided by the NGGDPP that define the minimum mandatory metadata fields required for inclusion in the National Digital Catalog. These initial conflicts were resolved with the release of a document titled "Preparing Metadata for the National Digital Catalog" (May 15, 2009) which took precedence over earlier documents.

The UCRC core collection, which includes both whole core and core chips, generated 1335 metadata files. This number was significantly more then expected despite the fact that some wells in the UCRC core collection did not generate metadata files due to missing data or other deficiencies in required metadata fields. The deficient metadata fields are being corrected manually as part of the normal quality review process.

During the contract period, we decided to generate as many metadata files as possible for other existing digital UCRC collections. This effort resulted in the generation of 3996 additional metadata files as follows: 3845 rock cuttings, 79 fluid samples (oil samples), and 72 hand samples (surface samples). 5331 metadata files were generated and uploaded by the UGS to the National Digital Catalog as part of the NGGDPP 2008 FY Award.

INTRODUCTION

The Utah Geological Survey's Utah Core Research Center (UCRC), established in 1951 and now occupying a 12,000-square-foot warehouse, contains the regions only publicly available and most complete collection of geologic cuttings and core from Utah.

The facility currently holds cuttings from more than 3600 drill holes; core samples from approximately 800 drill holes; a collection of type oils from all producing formations in the state; representative coal samples from Utah's producing coal mines; and miscellaneous samples of metallic minerals, industrial rocks and minerals, tar sands, oil shale, geothermal wells, and surface stratigraphic sections. Great Salt Lake sediment and brine samples are also stored at the UCRC. The UCRC provides service to all interested individuals, universities, and companies requiring direct observation of actual samples for their research or investigations, and acts as a repository for irreplaceable geologic samples that might otherwise be discarded.

PROJECT DESCRIPTION

The National Geological and Geophysical Data Preservation Program (NGGDP) FY 2008 Program Priority was to create and populate the National Digital Catalog (NDC) with archived materials.

To fulfill the program priority, the Utah Geological Survey (UGS) used previously archived UCRC digital inventories to create files of metadata that conform to NDC Metadata Product Requirements. The goals of the UGS Metadata creation project were as follows:

1. Establish the structure of metadata files to be used and develop the methodology and technology needed to begin the metadata conversion process to NDC standards.
2. Determine what data in the existing UCRC digital catalog can be used in the creation of new metadata files.
3. Compile and modify well information as needed to assure metadata files conform to NDC Metadata Product Requirement standards.
4. Create NDC metadata files
5. Data review and quality control.
6. Transmission of the compiled metadata files to the NDC.

The UCRC core collection was chosen as the first collection to be converted to NDC metadata files. We chose the core collection because of its manageable size and it has an established schedule for updating and quality review. The UCRC core collection consists of the whole core, core chips, and existing digital files used to catalog and describe the collection. At the start of the project, we estimated that the UCRC core collection consisted of about 800 wells. Actual UCRC core collection file totals were updated following analysis of the final uploaded NDC metadata file report.

THE UGS INTEGRAL DATABASE

All the data elements used to create metadata files for the NDC were collected from the UGS Integral database. Integral is a Microsoft Access based database which contains all of the existing digital information about the holdings of the UCRC.

The Integral database was developed in the early 1990s by Douglas Sprinkel of the UGS for the integration of data for a single point, hence the name Integral. The Integral database was originally developed to replace an existing Paradox database that used an older database format. Integral is an “in-house” database that links sample information to available subsurface logging information. The UGS converted the original Paradox database into an Access database format and plans to eventually convert the existing Integral database into an ArcInfo Geodatabase format for compliance with Geographic Information System (GIS) platforms.

METADATA FILE CREATION

Digital data elements from Integral used for the creation of new metadata files were parsed from queried data in the Integral database and sorted into the metadata fields required for inclusion in the NDC.

We manipulated each metadata field using Microsoft Excel to conform to the required NDC metadata format and exported the files to a XML document using an eXtensible Markup Language (XML) schema template provided by the USGS. The USGS template is at the following location: http://datapreservation.usgs.gov/docs/collectionMetadataExample.xml

METADATA FIELD DEFINITIONS

All new metadata files designed for inclusion in the NDC contain mandatory and optional metadata fields. At the start of the metadata file creation process, we discovered minor conflicts between several existing sources that define the minimum mandatory metadata fields required for inclusion in the NDC. A document found on the USGS website defining the NGGDPP Metadata Profile for the NDC (March 17, 2008) initially suggested which metadata elements to use and which fields were required. This document is found at: http://datapreservation.usgs.gov/docs/NGGDPPMetadataProfile.pdf

The NGGDPP FY2008 UGS Assistance Award included Attachment F titled "National Catalog Metadata Products Requirements" which included specific requirements for the NDC metadata elements. We quickly recognized discrepancies between Attachment F and the NGGDPP Metadata Profile concerning the minimum mandatory metadata fields. The release of a document titled "Preparing Metadata for the National Digital Catalog" (May 15, 2009) found at: http://datapreservation.usgs.gov/docs/NGGDPP_MetadataPreparation.pdf resolved the initial conflicts. The "Preparing Metadata" document, because of its later release date, took precedence over earlier documents with regard to defining mandatory NDC metadata field requirements as used in the UGS metadata conversion process.

METADATA FIELD ELEMENTS

The process of translating metadata files from the Integral database to the NDC format included the following steps: 1) identification of mandatory metadata fields, 2) selection of optional metadata fields which would be most useful in describing the UCRC collection, and 3) defining which data elements and format would be parsed from the existing Integral database and translated to fill each of the mandatory and optional metadata fields.

The following 11 mandatory and optional metadata fields and format requirements where chosen for the UGS metadata files created and uploaded to the NDC:

1. Collection ID	Required
2. Title	Required
3. Abstract	Required
4. Data Type	Required
5. Supplemental Information	Required
6. Coordinates	Required
7. Dataset Reference Date	Required
8. Alternate Title	Optional
9. Alternate Geometry	Optional
10. Online Resource	Optional
11. Vertical Extent	Optional

For clarification, the data elements used in each of the metadata fields for the UGS created NDC metadata files are defined below:

1.) **Collection ID**: a unique identifier assigned by the NDC to distinguish the separate collections within the NDC. The collection ID contains a single unique value. Example: 1089220.

2.) **Title:** a "human-readable" title that is listed in search results. 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: Mountain Fuel Supply Co Cedar Rim 3; Cedar Rim.

3.) **Abstract**: a short "human-readable" description of the physical sample. The description of the individual record was captured into this one general element. Example: Utah Core Research Center, geological sample. 31 boxes core, not continuous.

4.) **Data Type:** a controlled vocabulary of data types. Each reported physical data record must fall into one of the category types from the provided list. The data type field was assigned a descriptor that was most applicable, e.g. fluid sample, hand sample, rock core, rock cuttings, etc., from a list provided by the NGGDPP Implementation Plan. Example: rock core.

5.) **Supplemental Information:** a field was used to provide specific information on how to access the physical data represented by the NDC metadata record. We chose to provide the Utah Core Research Center phone as a resource to obtain more information about sample holdings**.** Example: Phone, (UCRC): (801)537-3359.

6.) **Coordinates:** simple decimal longitude and latitude geographic coordinates are required for the NDC. The NGGDPP field requirements are that the longitude value be listed first followed by the latitude value, with the two values separated by a comma. Some of the UCRC holdings only had UTM coordinates or Public Land Survey System locations. We converted all coordinate systems into longitude and latitude**.** Example: -110.60219, 40.20762.

7.) **Dataset Reference Date**: includes a reference date indicating the currency of the underlying data record and a proper date format. We used the date of donation of the geological sample as the reference date, if no donation date was available, the date of the creation of the metadata document was used. The date format is a four digit year followed by a two digit month followed by a two digit day. Example: 1994-01-02.

8.) **Alternate Title:** for further identification, collection owners can provide additional identifiers for individual records. The alternate title may include specific sample identifications used by the collection. A record's American Petroleum Institute (API) number can be used for this field due to the fact that oil wells typically have an associated API number designation. If no API number was available for a sample, a bogus API number was created in the Integral database.

We maintained the bogus API numbers because sample information is easily and quickly accessed within the Integral database using the API number. Example: 43-013-30040-0000.

9.) **Alternate Geometry**: an underlying collection resource that uses an alternate method of storing the geospatial information for the individual records. The Alternate Geometry field was an optional field that was used to include the Public Land Survey System (PLSS) township, range, and section as an alternate method for locating well bores; most Utah oil wells have PLSS data included in well descriptions or reports. Example: SWNE 19 3S 6W UM Duchesne UT.

10.) **Online Resource:** a field used with a URL pointer to indicate information about specific records. The UGS has an online interactive map that provides information about curated samples owned by the UCRC. Example: http://geology.utah.gov/emp/ucrc/index.html.

11.) **Vertical Extent**: a field to capture sample depth information used by the UCRC personnel. Specification of the vertical extent contains three elements. For the purpose of the NDC these elements were collected as three values representing the unit of measurement, maximum value, and minimum value. Example: feet, 8508, 7890.

RESULTS

The UCRC core collection, which includes both whole core and core chips, generated 1335 metadata files. This number was significantly more than the 800 expected despite the fact that some wells in the UCRC Core Collection did not generate metadata files due to missing data or other deficiencies in required metadata fields. As part of the normal quality review process, the missing elements of individual metadata fields will be researched and corrected manually.

Due to the success of the core collection metadata file generation process, it was decided to generate as many additional metadata files as possible for other existing digital UCRC collections during the contract period. This effort resulted in the generation of 3996 additional metadata files as follows: 3845 rock cuttings, 79 fluid samples (oil samples), and 72 hand samples (surface samples) metadata files.

The UGS generated and uploaded a total of 5331 metadata files to the NDC as part of the NGGDPP 2008 FY Award.

Personnel of the UCRC continue to update, catalogue, and curate the files of rock core and cuttings, fluid samples, and hand samples, as well as thin sections, paper reports, and special analyses for future inclusion in the NDC.