Final Technical Report

AZ Geoscience Information Network and Interface to the National Data Catalogue

By M. Lee Allison and Stephen M. Richard

Principal Investigators

Prepared for the USGS National Geological and Geophysical Data Preservation Program

under

USGS Award 08HQGR0134

Project Period 09/01/2008 to 08/31/2009

June, 2010

TABLE OF CONTENTS

Introduction..... 3
Scope and implementation..... 3
Metadata 3
Metadata produced..... 3
Metadata Production..... 5
Results..... 5

Introduction

As part of the Energy Policy Act of 2005, the U.S. Geological Survey was tasked with creating a National Geological and Geophysical Data Preservation Program (NGGDPP). This Program is intended to work principally with the nation's geological surveys to: (1) archive geologic, geophysical, and engineering data, maps, well logs, and samples; (2) provide a national catalog of such archival material; and (3) provide technical and financial assistance related to the archival material.

The Program is envisioned as a national network of cooperating geoscience materials and data repositories that are operated independently yet guided by common standards, procedures, and protocols for metadata. The holdings of all collections will be widely accessible through a single, common, and mirrored Internet-based catalog, the National Digital Catalog, thus maximizing the availability of and interconnectedness of all the collections.

Scope and implementation

In 2008 -2009, the USGS provided $24,959.00 in Federal funds which were matched by $25,021.00 from the Arizona Geological Survey to create feature level metadata for boreholes with cuttings or cores, geologic maps, mapped earth fissures, geochronology samples, and gravity stations, and to submit metadata for inclusion in the National Digital catalog.

Metadata

Metadata is "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource."[1] Any resource, for instance, a map, can be counted as data. Metadata content describes the particular artifacts (digital or physical) so that they can be found, retrieved, and cited. Metadata for a map could include its date of publication, author, publisher, scale, size, media (whether it is a digital or print object, for example), geographic area covered, edition, longitude and latitude coordinates, title, type (geologic, topographic, political, etc.), and other terms. Metadata content is structured into a metadata schema, which provides a uniform template for facilitating the retrieval, management, and preservation of resources, regardless of format.

Metadata produced

Metadata was prepared for the NDC by several different approaches. Metadata, including tops for all formations intersected, has been compiled into a database for 1149 wells for which oil and gas drilling permits have been issued by the State of Arizona. Basic information including operator, total depth, date drilled, formation at total depth, and status (shut in, producer, plugged and abandoned…) were available in a flat file table. Additional information including formation tops, and any drill stem testing or pro-

[1] NISO, *Understanding Metadata.* (Bethesda, MD: NISO Press, 2004). http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

duction testing information were entered into a digital database. This information was available for wells older than 1982 from a published AZGS bulletin. For newer wells, the information had to be compiled from paper files.

Boreholes with cuttings in the AZGS repository were located using PLSS (township and range) coordinates, and the existing information table for these holding also listed the depth interval sampled and the well owner. Most of these wells were originally drilled as water wells. Using the available information, we correlated as many of the wells as we could with existing wells in our borehole database. The existing database includes water wells registered with the State Department of Water Resources. This correlation is important to establish a link between the wells with cuttings and water production information. Wells that could not be correlated were added to the borehole database as new records. Metadata for 1232 wells with cuttings samples was produced. We treated the paper files in our library, the digital scans of well logs, and core or cuttings (physical samples) as different, related resources, so a particular borehole may have several related metadata records.

Extents for eighty three geologic maps published by the Arizona Geological Survey or in dissertations were added to the AZGS map index. The map index was synchronized with previously separate publication list and bibliographic databases to avoid duplication of data and insure that all AZGS maps are indexed. Metadata was generated for all maps in the AZGS map index (about 1,700 in number) and inserted into the AZGS public catalog server. Metadata for 188 AZGS published geologic maps was inserted into the USGS NGGDPP National Digital Catalog (NDC). Only metadata for AZGS published maps was inserted in order to avoid potential duplication.

Metadata for 2,187 locations of geochronology samples with associated isotopic age date information was generated from the AZGS geochronology database.

Metadata for 91 gravity reference stations in Arizona was produced with information from the University of Texas at El Paso (UTEP) Regional Geospatial Service Center (http://gis.utep.edu/index.php?option=com_content&view=article&id=69%3Anga-gravity-reference-base-stations-arizona&catid=50%3Agravity-base-stations&Itemid=59) and provided by William J. Bakula, of the National Geospatial Agency. After considerable investigation, it was determined that there is no way (without a very significant and costly effort beyond our capabilities) to acquire station metadata for the 57,000 or so gravity measurements in the PACES gravity data catalog (http://gis.utep.edu/index.php?option=com_content&view=article&id=198%3Apaces-gravity-magnetics&catid=51%3Amain-site&Itemid=59).

The AZGS has an ongoing program to map earth fissures in the state of Arizona. Currently our coverage of mapped fissures includes about 500 linear km of earth fissures, mapped in numerous (about 3,500) discrete segments. The total number of fissures is difficult to determine because of the discontinuous, anatomizing nature of the fissures. The fissure locations and mapping history have been made available through a web service (http://www.azgs.az.gov/efv.shtml) after an extensive review process with the Arizona State Land Department, the AZGS's external Earth Fissure Advisory Group, and our Arizona Land Subsidence Group technical advisory body due to the sensitive nature of the content, which may impact property values.

Geoscience data and samples for Arizona housed by organizations outside of the Arizona Geological Survey but used by AZGS personnel have been partially inventoried. The project suffered a setback when the principal AZGS geologist tasked with the collection inventory and geologic map indexing was stricken with a non-work related injury early in the project, and after several months of temporary disability, left the Survey on permanent disability. Project resources had to be diverted to training new personnel to do the map index work. As a result, collection inventory for external resources remains in an informal state and we plan to complete a final product with state resources.

Metadata Production

The AZGS is migrating to using ISO19115 metadata for in-house registration of resources. This is motivated by requirements to provide online catalog services for the USGIN project, and the available free open-source catalog software works best with the ISO standard for detailed metadata. Since most of the resources we are creating metadata for as part of this project already have associated digital files that provide at least part of the necessary content for metadata records, the workflow we used to generate metadata consisted in updating and inserting records in the existing tables, exporting the tables into a simple xml file, and then using XML transformations (XSLT) to convert the simple xml into ISO19139 xml that can be inserted into the catalog server database.

Export of metadata for NDC import required another XSLT transformation because NDC can not yet harvest metadata from other catalogs. Since we already had transformations targeted at our exported content, we edited these transformations to produce the xml format requested for NDC input. This was relatively simple because the NDC content model is quite simple.

After investing several approaches to the extract-transform-load process (ETL) for mapping existing data into the ISO19139 or NDC xml metadata formats, we settled on using Altova XMLSpy to produce the XSLT programs. The visual interface provided the most functional and usable approach. Conclusions of this investigation, as well as the workflow process were described in a presentation made by Wolfgang Grunberg at the Geoscience Data Preservation Techniques Workshop, Indiana Geological Survey, Bloomington, Indiana in July 2009. The presentation slides are available at http://lab.usgin.org/sites/default/files/group/file/u1/AZGS_to_NGGDPP_ETL_Bloomington_2009_July_v2.pdf.

Results

Table 1 summarizes metadata produced and registered in the national catalog, as well as information that is published through AZGS web interfaces.

Table 1. AzGS feature level metadata registered with National Catalog

Name	Number of records	Description
Borehole metadata	1,232	Metadata describing logs, cuttings, paper files associated with wells inserted in NDC
Geologic map metadata	188	83 new maps added, metadata for 188 AZGS-published maps inserted in NDC
Earth fissures	3,500	Published via WFS from AZGS server; actual number of fissures has not been quantified; records are for individual mapped fissure segments with distinct characteristics. We did not insert metadata for these in NDC after consideration of use cases for the information, deciding that it made more sense to make the fissure information a separate data service.
Gravity base stations	91	Information to characterize sites for all gravity measurements is not readily available
Geochronology sample sites	2,187	Sample locations, references and brief summary of date determinations inserted in NDC

Acknowledgments

Work on this project was completed by Stephen M. Richard, Lee Allison, Helen Ireland, Wolfgang Grunberg, and Richard Trapp. Thanks to Sky Bristol and his team for patience with iterations to get the xml format correct for the NDC import.

Bibliography and links

NISO, *Understanding Metadata.* Bethesda, MD: NISO Press, 2004.
http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

National Geological and Geophysical Data Preservation Program,
http://datapreservation.usgs.gov

Metadata records in NDC from Arizona Geological Survey may be browsed through the NGGDPP portal. The URL to see the Arizona Records is
http://datapreservation.usgs.gov/catalog/item/search?searchExtent=&search=Arizona&contextName=NDC&dataOwner=Arizona+Geological+Survey&itemType=Physical%20Item