NGGDPP FY 2010 Final Technical Report

State Geological Survey: Alaska Division of Geological & Geophysical Surveys

Award Number: G10AP00122

Project Title: Inventory, Metadata Creation, & Data Rescue for Alaska's Geological Collections

Principal Investigator(s): Jennifer Athey, Geologist III
Alaska Division of Geological & Geophysical Surveys
3354 College Road
Fairbanks, AK 99709-3707
Phone: 907-451-5028
Fax: 907-451-5050
Email: jennifer.athey@alaska.gov

Authorized Institutional Representative: Rodney A. Combellick, Division Operations Manager
Alaska Division of Geological & Geophysical Surveys
3354 College Road
Fairbanks, AK 99709-3707
Phone: 907-451-5007
Fax: 907-451-5223
Email: rod.combellick@alaska.gov

Project Term: July 15, 2010—July 14, 2011

Technical Report Date: October 12, 2011

Abstract

During the FY 2010 phase of the National Geologic and Geophysical Preservation Program (NGGDPP), the Alaska Division of Geological & Geophysical Surveys (DGGS) directly addressed two of the program priorities: creating site-specific metadata for individual items in geologic data collections and preserving critical energy-related and hard-rock digital data and physical samples in danger of becoming lost or destroyed. Our database development team created site-specific metadata for two DGGS geologic data collections: (1) organic geochemistry analyses published by DGGS, and (2) valuable hard-rock mineral core and coal-bed methane core samples stored in deteriorating boxes at the Alaska Geologic Materials Center (GMC). In addition, decaying core samples were re-boxed into new core boxes and relocated to an environmentally controlled area within the GMC. In order to serve the metadata records to the National Digital Catalog, DGGS employs both the Web Feature Service and SiteMap methods. The methods allow the site-specific metadata records to be harvested and synchronized automatically by the National Catalog system, thus freeing DGGS staff members of the manual task of uploading data to an additional database on a regular basis. To keep the online metadata current and error-free, the DGGS database is continually updated via existing business processes with quality control mechanisms programmed to verify the raw data of the sample collections during bulk-uploading processes or within the data-entry interfaces to the database.

Table of Contents

Abstract 2
Table of Contents 3
Introduction 4
Prior Work in Recovering, Archiving, and Cataloging Data 4
Concurrent Data-Rescue and Archiving Projects ..... **Error! Bookmark not defined.**
FY 2010 Metadata Creation—Organic Geochemistry 6
FY 2010 Special Needs (Data Rescue)—Core Samples 8
Conclusion 10
Appendix A: Summary of DGGS Data Preservation Progress and Needs 11

Introduction

The Alaska Department of Natural Resources, Division of Geological & Geophysical Surveys (DGGS), has as its statutory mission the responsibility for collecting, archiving, managing, and disseminating geological and geophysical data on the subsurface energy resources, mineral resources, and geologic hazards of the state. DGGS maintains its own databases of this information on servers at its office in Fairbanks, and a collection of nonproprietary rock samples and archive materials at the Alaska Geologic Materials Center (GMC) in Eagle River representing more than 12 million linear feet of exploration and production drilling. DGGS networks with other state agencies that use these data and that archive related datasets, some of which are proprietary, including the Division of Oil & Gas (DOG), Division of Mining, Land & Water (DMLW), and the Alaska Oil & Gas Conservation Commission (AOGCC). Samples from these collections are examined frequently by private exploration companies and state agencies as a critical step in resource development and management.

Prior Work in Recovering, Archiving, and Cataloging Data

The State of Alaska received federal funds through congressional appropriations to the U.S. Geological Survey (USGS) and Bureau of Land Management (BLM) between FY1998 and FY2004 to recover, archive, catalog, and make publicly accessible minerals-related geologic data and publications that were in danger of being lost or destroyed following the cancellation of federal minerals programs such as the U.S. Bureau of Mines, and the closing of many private mineral-exploration companies over the past few decades. Through the Minerals Data and Information Rescue in Alaska (MDIRA) program, DGGS has made available to the public, via the internet, many geologic reports, maps, and data that were previously unavailable or difficult to find. The following are just a few of DGGS's notable accomplishments through MDIRA:

- Completed a full data assessment focusing (but not exclusively) on non-fuels solid-mineral resources data and samples.
- Led a cooperative program to build and maintain a federal/state interagency, web-accessible, Alaska geoscience bibliography that is geographically indexed and links to online repositories.
- Scanned all previously published DGGS reports and maps.
- Scanned all pre-digital USGS publications and maps with subject material relating to Alaska geology.
- Built a public web-accessible catalog for searching and accessing the scanned maps and reports.
- Documented available DGGS digital GIS data with FGDC-compliant metadata.
- Designed and implemented an enterprise data system to catalog and house digital geologic data and information for Alaska.
- Designed and built a database-driven website to distribute DGGS GIS data via the internet.

- Recovered geochemical datasets published in multiple DGGS reports and integrated them into a web-accessible database system.
- Recovered geochronological data from published and unpublished sources and began integrating them into an electronic database that will be web accessible.
- Began cataloging and archiving field project files and unpublished geologic compilations by DGGS and predecessor agencies.
- Began indexing diverse surface and subsurface industry geologic materials housed at the GMC and integrating the index into a web-accessible database.

FY 2007 NGGDPP Program

DGGS requested and received funding under the FY 2007 National Geological and Geophysical Data Preservation Program (NGGDPP) Program Announcement No. 07HQPA0018 to provide collection-level information on its data holdings. DGGS completed evaluations for 12 of its geologic data and sample collections, and entered detailed information about each one into the NGGDPP online inventory system. See the updated summary table of DGGS collections in Appendix A.

DGGS also received funding under NGGDPP (Order No. 07HQSA0115) to participate in the design and testing of the National Digital Catalog. DGGS helped establish standards for the content, quality, and consistency of metadata describing digital data and physical samples; protocols for searching, transferring, and presenting metadata; clear and consistent ways to explain how users can request access to physical materials; and the extent to which the user interfaces convey branding or meet other organizational needs.

FY 2008 NGGDPP Program

In FY 2008, DGGS received funding under NGGDPP (Order No. 08HQPA0015) to implement a Web Feature Service (WFS) interface to deliver site-specific metadata files to the National Digital Catalog for six of its geologic data collections: (1) core samples and drill cuttings stored at the Alaska Geologic Materials Center (GMC), (2) glass slide collection of processed samples at the GMC, (3) data reports on core samples that have been borrowed from the GMC for analysis, (4) geochemical analyses of rock, soil, and stream sediment samples collected during projects involving DGGS geologists, (5) geochronology analyses of samples collected during projects involving DGGS geologists, and (6) hard-rock surface samples collected by DGGS staff.

The database development team installed GeoServer, a Java-based, Open Source server software that allows publishing of geospatial data to the web, on the DGGS's web server to provide the WFS interface. Each database view contains metadata records of an individual collection and is registered as a feature type within GeoServer. DGGS collection information is generated from the database; the unique identifier of each sample or location is translated into a unique and stable URL for each locality in which DGGS stores data. The OnlineResource metadata element provides this dynamic URL to the National Digital Catalog.

Concurrent Data-Rescue and Archiving Projects

Several datasets compiled under the MDIRA program have become unavailable, are at risk of becoming unavailable, or were never completed and made publicly available. In 2011, DGGS received additional MDIRA funding to migrate three mineral-resource-related database projects to DGGS's Oracle database and create search and data management applications to allow easy access to the data. In DGGS's Oracle environment, the Alaska Geologic Map Index, Alaska Mineral Industry Data Index, and Alaska Paleontological Database projects are ensured regular maintenance, back-up, continued data expansion, and consistent public internet access. Fully integrated map- and text-based search tools will enhance the applications' interfaces.

In 2011, the Geologic Materials Center (GMC) received additional funds from the MDIRA program to allow GMC staff to finish indexing and detailing the remaining 30 percent of the facility's hard-rock core inventory. The funds are being used to (1) research and identify missing information related to the hard-rock mineral core, such as unknown prospect or property locations, missing borehole numbers, and unknown property owners, consultants, or drill companies by working with staff, scientists and geologists at DGGS, BLM, ARLIS, and Alaska Earth Sciences, (2) perform a major re-boxing effort (in conjunction with NGGDPP funds) on core samples in order to prevent major data loss due to substantial box deterioration, and (3) research and understand the potential radiation impacts of several prospects.

FY 2010 Metadata Creation—Organic Geochemistry

Methodology

DGGS completed a Collection Inventory (P1694) and is preparing NGGDPP-compliant metadata for this high-priority dataset by finishing a number of specific tasks. The organic geochemistry dataset is critical for an assessment of technically recoverable petroleum resources in major source-rock systems of the Alaska North Slope led by the USGS scheduled for late 2011 and 2012. The data are commonly requested by researchers and private industry. The tasks included:

- *Data analysis.* Energy geologists reviewed several publications containing organic geochemical data and determined which values needed to be extracted to create a comprehensive dataset, and how the data should be formatted for public ease of use. Geologists also suggested which data fields would most appropriately be presented in an abbreviated online view with sorting options to enhance data searchability. The full dataset will be available for download through a web application.
- *Database programming*. DGGS's analyst/programmer and data manager worked with energy geologists to map the data elements from lab reports and publication text to the Oracle database fields. DGGS had data structures in place for most of the proposed collection site-specific metadata elements and analysis values, and the analyst/programmer created and modified the database structure as needed.

- *Locality Data*. A Student Intern harvested location information from data tables in the publication text and, as necessary, digitized locations from georeferenced sample location maps using the ArcGIS desktop application.
- *Metadata Extraction/Formatting*. A Student Intern extracted the metadata elements and analysis data values from specified data sources and formatted them into an Excel spreadsheet that was later converted into a delimited text file for bulk uploading to the centralized Oracle database. The majority of the publications containing organic geochemistry are older and the data were not available in digital format. In addition, the data were produced by several different laboratories that reported values in different table formats and a variety of charts. Some values were obtained through original digital media from the laboratories. The data manager cross-checked about 80 percent of the extracted values with the original publications to ensure data quality.
- *Data Upload Programming.* Our analyst/programmer created a simple and secure loading script to bulk upload .csv files of formatted metadata elements and analysis data values. The script contains quality control mechanisms to check for specific accuracies and consistencies in the data, and integrates the incoming data with any existing information in the central database such as locations, sample descriptions, etc. By writing this reusable script, future publications with organic geochemistry data can be added to the central database and then directly served to the National Digital Catalog. As part of our current business process, DGGS continually updates the database as new data are generated.
- *Web Output of Analysis Data.* Energy geologists, the analyst/programmer, and the data manager designed a web-based module for displaying the sample-level analysis data that will be available in 2012 to the online community. The module will contain an abbreviated dataset with user options for sorting the data and downloading the full dataset. The analysis data will be generated dynamically from the DGGS database based on the unique and stable URL of the sample and (or) location.
- *WFS, SiteMap and National Digital Catalog (NDC).* For the NDC to be able to auto-harvest the metadata, DGGS will make the organic geochemical data available on the WFS. As in previous years, DGGS utilizes GeoServer, a Java-based, Open Source server software that allows publishing of geospatial data to the web, on DGGS's web server to provide the WFS interface. After the spatial objects are automatically created for each sample location, a database view queries the data tables within the Oracle database and presents the metadata elements in accordance with the Metadata Profile established by the NGGDPP.

DGGS will also follow the methodology described in https://www.sciencebase.gov/xwiki/bin/view/NDC/SiteMapMethod. DGGS will create a sitemap specifically for the NGGDPP that is generated dynamically from DGGS's central database, but stored statically. The WFS and sitemap XML contain the unique and stable URLs for each item within the collection. The metadata are thus available to the NDC at any time, and in multiple formats. The metadata include the date and time of upload, which also describes the currency of the data.

Results

Eight publications were identified with a total of 273 organic geochemical sample analyses among them. The analyses were compiled and uploaded via a script into Oracle and will be available to the National Digital Catalog as both a WFS and sitemap. The WFS interface and sitemap allow the site-specific metadata records to be harvested and synchronized automatically by the National Catalog system, thus freeing DGGS staff members of the manual task of uploading data to an additional database on a regular basis. The following metadata elements, in accordance with the Metadata Profile established by the NGGDPP, will be available for harvesting: CollectionID, Title, AlternateTitle, Abstract, DataType, SupplementalInformation, Coordinates, AlternateGeometry, OnlineResource, BrowseGraphic, Date, DatasetReferenceDate, and VerticalExtent.

Metadata creation differed in three unanticipated ways from the plan outlined in the FY 2010 proposal. DGGS decided that a reusable loading script was adequate to update organic geochemical data in DGGS's Oracle database. Although the proposal stated that a simple data-loading application would be created for the staff to upload their data, DGGS obtains new organic geochemical datasets so infrequently that it was deemed unnecessary effort to create an application when the data loading script would easily fulfill the same function.

The web-based module displaying the organic geochemical dataset will not be available until 2012 due to multiple demands on our analyst/programmer's time. The online dataset view has been designed and FGDC metadata has been written for the dataset. The application itself, including an abbreviated online data view with user sorting options and a full data download option, still needs to be developed. Until the web-based module is available, the organic geochemical data will be available to the public through the WFS and the National Digital Catalog.

Due to the changing needs of the National Digital Catalog, DGGS also decided to provide the metadata to the NDC as XML in the NGGDPP sitemap, in addition to the WFS as in previous years. DGGS tested a methodology similar to the sitemap method for the NGGDPP beta program in 2008 and was already familiar with the technology. Since the NGGDPP suggested using this method in 2011, DGGS thought making this change from the original proposal was appropriate. We anticipate that the metadata will be available for harvesting from the WFS and sitemap by October 31, 2011.

FY 2010 Special Needs (Data Rescue)—Core Samples

Methodology

Geologic Materials Center staff attempted to salvage approximately 10,300 ft of drill core that was on the verge of total loss due to deteriorating written metadata on severely damaged boxes and damage to the samples from multiple freeze/thaw cycles and moisture. The Amchitka boreholes include UA-1, UAE-1, UA-2, UAE-2, UA-3, UA-6 and UAE-6h; unlike other cores

from Amchitka Island, Alaska collected after underground nuclear tests between 1965 and 1971, these cores were not acquired by the U.S. government. Amchitka core is categorized in Collection Inventory P1076. Coal-bed methane cores (Collection Inventory P1285) include Sheep Creek #1, Kashwitna Lake #1, Houston #1, Slats #1 and Little Su #1. Preservation of the core and metadata creation was performed through the following tasks:

- *Detailed inventory of core samples.* Alaska Geologic Materials Center (GMC) staff captured metadata from the deteriorating core boxes.
- *Re-boxing core samples*. Samples in damaged boxes were transferred into new core boxes and barcoded. Samples particularly sensitive to moisture and freeze/thaw cycles were relocated to the indoor facility.
- *Population of GMC database.* The GMC inventory Access database was updated with metadata recorded from the core boxes, allowing staff to help users of the facility find information more quickly and pave the way for a future web interface to query the available materials at the GMC.
- *Population of DGGS database.* Inventory data were quality controlled and uploaded into the DGGS database via a reuseable script, as part of an ongoing development project to upgrade the GMC inventory database system.
- *WFS, SiteMap and National Digital Catalog (NDC).* Our analyst/programmer has transformed the inventory data into a database view of metadata records for each proposed collection, which will be registered as a feature type with the DGGS WFS. The data will also be made available on the NGGDPP sitemap. These data will be available for harvest by the National Digital Catalog by October 31, 2011.

Results

Geologic Materials Center (GMC) staff were able to rescue the hard-rock Amchitka boreholes and a major portion of the coal-bed methane cores. Preservation of the Amchitka went smoothly. Staff re-boxed 75 percent of the 717 boxes, and barcoded and indexed all of the boxes. Preservation of the coal-bed methane core, however, required additional work. In July 2010 GMC staff started a plan to curate, re-box, and inventory 818 boxes of coal-bed methane core located in one the facility's unheated shipping containers. The coal-bed methane core collection contained HX-sized whole core from five different wells that contained young sections of tan-colored lignite, sections of bituminous-B coal, coarse-grained conglomerates, and sandstones.

The project took nine months to complete due to (1) limited staff resources, (2) the logistical difficulties involved in moving 20- to 40-pound disintegrating boxes of core, and (3) the time needed to clean, curate, and re-box the core. The boxes of core containing known sections of coal (approximately 75) were given a higher priority and curated first. All of the “high-priority” core was curated, re-boxed, barcoded, and indexed into the GMC’s Access database.

Fibrous white mold was discovered to be growing on the surface and in the pore spaces of approximately 50 percent of the core because of exceedingly high humidity in the shipping

container and the unique lithology of the core. Black mold spores were also found to be growing on the cardboard boxes used to hold the core. The worst boxes were almost totally disintegrated and this made moving and cleaning the core extremely difficult (Figure 1). Fibrous white mold was cleaned from the external surfaces of the core with a medium-stiff brush in a vacuum hood. Research was conducted to determine how to remove the mold that was discovered growing in the pore spaces of the core, but no safe method was found that wouldn't potentially damage or alter the core.

Figure 1. Core box from coal-bed methane well Kashwitna Lake #1, 848–858 feet. Box on left shows white mold and cardboard deterioration. Box on right shows same core, cleaned and re-boxed.

GMC staff decided that moving all 818 boxes into an environmentally controlled storage area would at least remove the excess moisture content that would contribute to additional mold growth and core disintegration. Approximately 367 boxes of core were moved inside the facility, while 402 boxes of core were cleaned and re-boxed. When the project was completed in May 2011, GMC staff had successfully moved, cleaned, re-boxed, barcoded, and transferred 769 boxes of the 818 boxes of core. The project only lost 49 boxes of core (6 percent) in cases of completely disintegrated core and core box, which prevented the physical movement of those boxes out of the shipping container. Final results include saving approximately 94 percent of the 818 boxes of coal-bed methane core from five wells that would have otherwise been completely

destroyed and rendered useless. This effort has made the core accessible once again, allowing geoscientists an opportunity to study the young, unique coal sections found throughout the core.

Amchitka and coal-bed methane core metadata collected in the GMC Inventory database were uploaded into DGGS's Oracle database via a reusable script. The metadata will be formatted into a database view and fed to the WFS and NGGDPP sitemap. The data will be available to the National Digital Catalog for harvesting by October 31, 2011.

Conclusion

As DGGS builds the premier disciplinary repository focused on Alaska Geology, we are concentrating on the three essential functions of a repository: data archiving, ensuring data integrity, and providing public access. Ensuring that physical and digital data are properly cataloged and archived should be a critical priority for any organization that provides data to the public. Documentation and ensuring data quality for legacy datasets is crucial to make the datasets meaningful and usable. Once physical data collections are converted to digital form, become easily searchable, well documented, and organized by dataset or collection type, users can then focus on merging the data into their own projects and spend more time on analysis and understanding the implications of their scientific data and observations.

The NGGDPP program is an important funding vehicle for data preservation projects. The Organic Geochemical Metadata Creation and GMC Core Sample Data Rescue projects are highly important on DGGS's list of datasets that need preservation. DGGS has many other datasets that need physical or digital preservation (listed in Appendix A), and we will continue to make inroads and expand our repository as time, staffing, and funding allow.

Appendix A

Summary of DGGS data preservation progress and needs. Shaded collections have received funding from the NGGDPP in the past

	Paper (p), Digital (d), Physical Sample (s)	Assess Collection	Catalog Items / Records	Locality Metadata	In NGGDPP	Secure Storage and Archive	Digitize Data	Available Online	Direct Internet Link to Sample Level	Establish Process for additions
Publication collections										
DGGS publications	pd	C	C	IP	NF	C	C	C	C	C
GMC data reports	d	C	C	C	C	C	C	C	C	C
USGS Alaska publications	pd	C	C	IP	NF	C	C	C	C	C
U.S. BLM/BOM Alaska publications	pd	C	50% IP	NF	NF	IP	IP	IP	IP	C
UAF MIRL publications	pd	C	C	NF	NF	C	C	C	C	C
Alaska theses	pd	IP	IP	IP	NF	IP	NA	P	P	NF
Map Index outlines	pd	NF	NF	NF	NF	IP	NF	IP	IP	IP
Published GIS files	d	C	C	C	NF	C	NA	C	C	C
Legacy GIS files	d	NF	NF	NF	NF	NF	NA	NF	NF	NF
Reference Library	pd	IP	IP	IP	NF	IP	IP	IP	IP	IP
Published and unpublished data collections										
DGGS geochemical data	pd	C	C	C	C	C	C	C	C	C
AVO geochemical data	pd	C	IP	IP	NF	IP	IP	IP	IP	C
Geochronology	pd	C	C	C	C	C	C	P	C	NF
Porosity and permeability data	p	NF	NF	NF	NF	NF	NF	NF	NF	NF
Organic geochemical data	pd	C	C	C	C	C	C	P	C	P
Petrographic analyses	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Paleontology reports	pd	NF	NF	NF	NF	IP	NF	IP	NF	NF
Microfossil reports	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Sandstone composition reports	p	NF	NF	NF	NF	NF	NF	NF	NF	NF
Seal capacity analyses (MICP)	d	P	NF	NF	NF	NF	NA	NF	NF	NF
Apatite fission track analyses	pd	P	NF	NF	NF	NF	NF	NF	NF	NF
Energy-related inorganic geochemistry	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Geothermal data	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Sediment grain-size analyses	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
		Status: C: Complete; IP: In Progress; P(+): Planned (funded); NF: Not funded; RF: Requesting Funds; NA: Not Applicable								

	Paper (p), Digital (d), or Physical Sample (s)	Assess Collection	Catalog Items / Records	Locality Metadata	In NGGDPP	Secure Storage and Archive	Digitize Data	Available Online	Direct Internet Link to Sample Level	Establish Process for additions
Measured stratigraphic sections	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Gravity data	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Radiocarbon ages	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Permafrost depth and ice content	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Slope stability data	p	NF	NF	NF	NF	NF	NF	NF	NF	NF
U-Pb detrital zircon analyses	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Paleoseismic data	p	NF	NF	NF	NF	NF	NF	NF	NF	NF
Tephra analyses	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Pebble counts	pd	NF	NF	NF	NF	NF	NF	NF	NF	NF
Field photographs	d	NF	NF	NF	NF	NF	NA	NF	NF	NF
Unpublished field data	pd	NF	NF	C	NF	NF	IP	NA	NF	NF
Geophysical data	d	IP	IP	C	NF	IP	NA	IP	IP	IP
Physical sample and legacy paper collections										
GMC oil & gas well samples	s	C	IP	C	IP	IP	P	80% IP	P+	IP
GMC processed slides	s	C	IP	C	IP	P	P	P+	P+	IP
DGGS surface samples	s	C	IP	90% IP	IP	P+	90% IP	P+	P+	IP
GMC mineral exploration core	s	C	IP	45% IP	P	P	P	P+	P+	IP
U.S. BLM/BOM physical samples	s	C	99% P	35% IP	P	P	NA	P+	P+	NA
DGGS petrographic slides	s	C	NF	NF	NF	NF	NF	NF	NF	NF
AVO rock samples	s	C	C	IP	NF	NF	IP	NA	IP	C
AVO processed samples	s	C	C	IP	NF	NF	IP	NA	IP	C
Geo-engineering core samples	s	NF	NF	NF	NF	NF	NF	NF	NF	NF
Legacy paper air photos	p	NF	NF	NF	NF	NF	NF	NF	NF	NF
Legacy mineral industry data	pd	C	C	C	NF	C	C	IP	C	IP
Legacy DGGS project files	p	P	IP	P	NF	NF	P	IP	IP	IP
		Status: C: Complete P(+): Planned (funded) RF: Requesting Funds IP: In Progress NF: Not funded NA: Not Applicable								