Award No. G09AP00109

National Geological and Geophysical Data Preservation Program

Washington State Metadata Project:

Producing Metadata for Physical Rock Samples and Geotechnical Reports

Final Technical Report

Robert Berwick
Karen Meyers
Washington Department of Natural Resources
Division of Geology and Earth Resources
PO Box 47007
Olympia, WA 98504-7007
(360) 902-1431 / (360) 902-1461
(360) 902-1785 fax
robert.berwick@dnr.wa.gov / karen.meyers@dnr.wa.gov

Award Term: July 1, 2009 to June 30, 2010

Submitted: September 28, 2010

Abstract

The Washington State Department of Natural Resources, Division of Geology and Earth Resources, houses many geologic and geophysical data collections, 15 of which have been inventoried in the National Digital Catalog. For our 2009 project for the National Geological and Geophysical Data Preservation Program (NGGDPP), we selected three collections as high priorities for data preservation, based on their usefulness and perishability: (1) a collection of rock cores, (2) a collection of rock cuttings, and (2) a continually growing collection of geotechnical reports. The objective of our project was to create sample-level metadata for the physical sample collection and a subset of the geotechnical report collection over the course of 12 months, beginning July 1, 2009.

To accomplish this objective, in-house staff collected the required information on the core and cuttings, which are stored in our off-site storage facility, and on a subset of the geotechnical reports located at our headquarters. As a result, we created metadata records for 47 boreholes that produced rock cores, 309 boreholes that produced rock cuttings, and 56,658 boreholes represented in 10,000 geotechnical reports. This information was compiled into XML format in accordance with the National Digital Catalog specifications, and submitted for inclusion in the catalog.

Table of Contents

Introduction

During the first year of the National Geologic and Geophysical Data Preservation Program (NGGDPP), the Washington State Department of Natural Resources, Division of Geology and Earth Resources (DGER) inventory identified a total of fifteen collections in its archives. For this year's project, DGER produced metadata for three of those collections and loaded them into the National Digital Catalog:

- Rock cores (collection P901)
- Rock cuttings (collection P902)
- Geotechnical reports (collection P961)

Our reasons for selecting these collections for our first metadata project are several-fold. Not only do these collections contain a wealth of high-value geologic and geophysical information, but they have all been at imminent risk of degradation or loss, either through physical deterioration, lack of documentation, or disposal. Also, despite their usefulness for a variety of purposes, the information in these collections has largely not been readily available or easily accessible.

Both of these collections contain valuable subsurface geologic data that are not available elsewhere, and prior to this project, were difficult for potential users to discover and access. Washington State has an abundance of geologic hazards and natural resource issues, so demand for quality subsurface data is high. For example, Washington has the second-largest population in the nation at risk from earthquakes; the data targeted in this project have been successfully used by geologists to physically locate faults in the subsurface by geologists. Quality subsurface data are used by governmental agencies, consultants, scientists, engineers, planners, private companies, and entrepreneurs to manage natural resources—including potential underground CO_2 sequestration and natural gas storage reservoirs, groundwater, aggregate, coal, base and precious metals, and potential hydrocarbon and geothermal resources—in the best interest of the citizens of the Pacific Northwest.

Unfortunately, all three collections selected for this project are highly perishable, although somewhat less so now that they have been better documented. DGER's rock sample collections are stored off-site several miles from DGER's headquarters, in a storage facility that is undersized and due for replacement. These collections are not readily available to users, due to the physical difficulty in accessing samples (as some boxes are stacked in front of other boxes, and some boxes are difficult to access due to the combination of shelf height and box weight) as well as a lack of a consistent organization system. In addition to this lack of accessibility, for some samples, documentation was sketchy at best; in a few of these cases, the source of the

best available information on the origins of a sample was the personal knowledge of a specific individual.

Geotechnical reports are an indispensable source of subsurface geologic and geophysical information, and are produced at considerable expense. Unfortunately, geotechnical reports are generally not made available for long-term use; most local governments commonly dispose of these reports at the conclusion of the project for which the borings were drilled, and geotechnical firms typically keep their records of geotechnical investigations private. A countless number of geotechnical reports submitted to local governments have been already been lost. Over the past several years, DGER has made an effort to collect geotechnical reports for borings drilled in Washington to preserve their long-term value. DGER has collected these reports over the years as supporting information for mapping and hazards studies, but has stepped up efforts to collect as many reports as possible for preservation and access by contacting local jurisdictions and engineering firms and requesting copies of any reports they have. Notably, DGER recently acquired the electronic geotechnical report collection of the Pacific Northwest Center for Geologic Mapping Studies (GeoMapNW); this collection was at risk of loss due to the uncertainty of continued funding for the organization, and preserving this collection became a priority for DGER.

Project Goals

The purpose of this project is to preserve and enhance access to geologic information in DGER's archives and storage facilities by providing metadata to the National Digital Catalog. The specific goals for FY 2009 were to produce metadata for our entire rock core and rock cuttings collections, and a subset of our geotechnical report collection. At the beginning of this project, DGER estimated that the number of items for which we would attempt to create metadata would be as follows:

- Rock cores (collection P901): entire collection (about 200 boxes, unknown number of holes)
- Rock cuttings (collection P902): entire collection (about 700 boxes/containers, unknown number of holes)
- Geotechnical reports (collection P961): approximately 10,000 reports (approximately 40,000 to 50,000 records, since a single report can have information on several boreholes)

The project continued over a one year period beginning July 1, 2009 and ending June 30, 2010. It was anticipated that metadata for the rock cores and cuttings would be completed in about three months, and generation of metadata for the geotechnical reports would be completed during the remaining project period.

Methodology

Rock Core and Cuttings—Inventory and Metadata

For the two physical sample collections included in this project, DGER first undertook an initial inventory of samples in our off-site storage facility. For the purposes of this project, a "sample" generally refers to a set of physical samples associated with a single borehole. The samples in the warehouse are stored in various types of containers, and with various degrees of organization and labeling. Sample boxes and containers are stored on tall shelving units, with each shelf having a unique identifier.

For each sample, we recorded the shelf number, type of sample (core or cuttings), and as much of the following information as was available via labels either on the container or on the physical samples within the container:

- oil and gas permit number
- well name
- company responsible for having the hole drilled
- well location
- range of sample depth
- any comments on the samples (such as miscellaneous information included on the label)

The extent to which samples were labeled varied widely, and hence, the amount of information we were able to record on each sample during the inventory was variable.

At the conclusion of this inventory, we had identified rock core from 56 boreholes, and rock cuttings from 525 boreholes. These numbers differ from the numbers put forth in our original proposal, for two reasons: (1) the numbers in the proposal reflected the estimated number of boxes/containers in the warehouse containing rock core or cuttings, rather than the estimated number of boreholes; and (2) the number given in our proposal for boxes of core was in error due to transposition of a decimal point, as we had estimated about 200 boxes of core (instead of the 2,000 mentioned in the proposal).

The next step was to try to match each sample with borehole logs and locations. Many of the samples are from oil and gas wells documented in DGER Information Circular 75, "Oil and Gas Exploration in Washington, 1900-1982" (McFarland, 1983), and an addendum that covers the years 1983 to present; for each of these wells, DGER maintains a file containing all known information regarding the drilling of that well. The information circular is organized by oil and gas well permit number, and contains information on the location, name, and company associated with each well. Where the samples in the warehouse were labeled with a permit number, it was usually a straightforward process to cross reference those with the data in the

information circular to obtain the information needed for the National Digital Catalog. Occasionally, however, information collected from the warehouse for a particular permit number appeared to be in conflict with the data in the information circular; in these cases, research in the oil and gas files themselves was needed to resolve the discrepancies. Documentation in the oil and gas files also allowed us to assign permit numbers to a large number of samples that were not labeled with permit numbers in the warehouse; these samples were instead labeled with "well numbers", and an index found in the oil and gas files allowed us to cross reference these particular well numbers with oil and gas permit numbers.

For samples that were not represented in Information Circular 75 or the oil and gas files, we attempted to discover any information we could about them through discussions with long-time DGER employees and through research in our Washington Geology Library. Our library research was guided by leads from DGER employees, or by information included on sample labels in the warehouse (for example, project names, location names, dates). Unfortunately, a significant number of these samples remain unidentified, because of either poor sample labeling, or difficulty in connecting labeled samples with a particular location or project. Samples of rock cuttings were much more likely to be poorly labeled than core samples, as is reflected in our project results (see below). Should any of these unidentified samples be identified at a later time, it is our intention to update the metadata in the National Digital Catalog to reflect the new information.

Geotechnical Reports—Metadata

Preserving and maintaining access to the large GeoMapNW collection of geotechnical reports was our first priority, so we focused our efforts this year on producing metadata for a significant part of this collection. The reports in this collection were documented in a digital database, which contained identification and location information for each of the boreholes described in each report.

In order to produce metadata suitable for the National Digital Catalog, we needed to extract information from various parts of this database, which was divided into several related tables, as well as construct metadata for elements that were not represented in the database (for example, the metadata elements "Abstract" and "SupplementalInformation"). Another part of the processing involved creating active links to the PDF file of each report from each borehole record. Throughout the process, we discovered various errors in the data that needed to be corrected—for example, we discovered that the document titles had been cut off at 50 characters, and had to open each document to find the full title and correct the data. Most of these tasks, and the exporting of the metadata to an XML file formatted for importing into the catalog, were accomplished using Microsoft Excel and a complex series of computer programming scripts that we custom-built for this purpose.

Results

Below is a comparison of our project goals to our actual accomplishments during the grant period (July 1, 2009 to June 30, 2010). The delivery date of the metadata to the contracting official, while not within the grant period, was within 90 days of the end of the grant period, thus still satisfying contract requirements. This delay was because we were awaiting sufficient development on our online interactive mapping application (*see* Geotechnical Reports, below) to allow us to house our electronic geotechnical reports in a permanent location for access by both the interactive mapping application and the National Digital Catalog.

Physical Samples: Rock Cores

The inventory identified about 350 boxes of core from 56 boreholes. Of these, we were able to obtain sufficient identification and location information to produce metadata for 47 of those boreholes.

Comparison of goals to actual accomplishments for the rock cores collection (P901):

Goal:	Total number of holes for which core is available	**56**
Results:	Number of core holes with complete metadata submitted to the National Digital Catalog	**47**
Difference:	Number of core holes with insufficient identification or location information	**9**

Physical Samples: Rock Cuttings

Our inventory identified about 960 boxes (or other containers) of cuttings from 525 boreholes. Of these, we were able to obtain sufficient identification and location information to produce metadata for 309 of those boreholes.

Comparison of goals to actual accomplishments for the rock cuttings collection (P902):

Goal:	Total number of holes for which cuttings are available	**525**
Results:	Number of holes with complete metadata submitted to the National Digital Catalog	**309**
Difference:	Number of holes with insufficient identification or location information	**216**

Geotechnical Reports

We collected and processed information, and produced metadata compliant with National Digital Catalog standards, for 10,000 geotechnical reports, representing 56,658 boreholes.

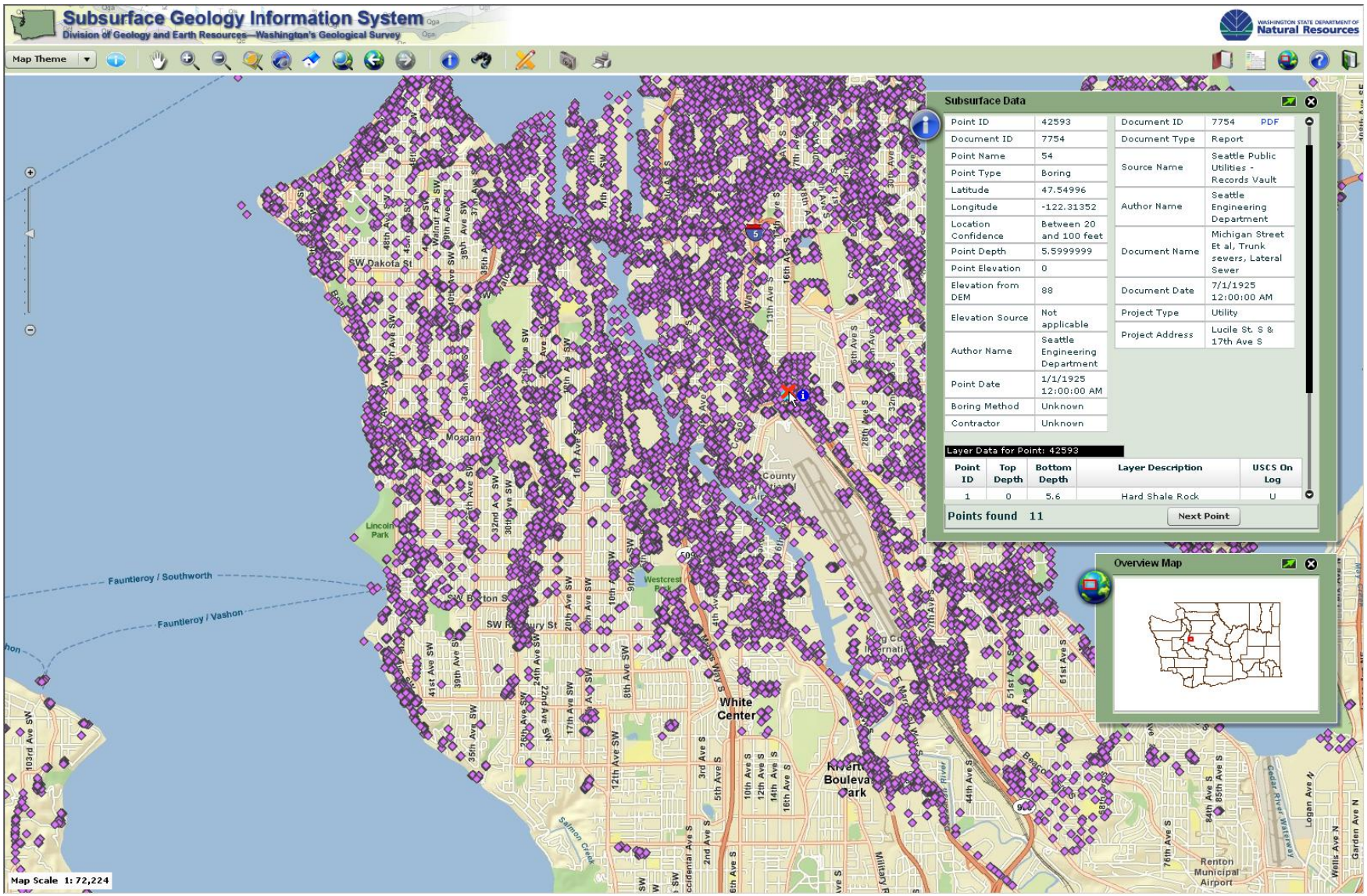
Comparison of goals to actual accomplishments for the geotechnical reports collection (P961):

Goal:	Total number of boreholes (and associated reports) for which we intended to produce metadata	**40,000–50,000 (10,000 reports)**
Results:	Number of boreholes (and associated reports) with complete metadata submitted to the National Digital Catalog	**56,658 (10,000 reports)**

In addition to producing metadata for the National Digital Catalog, we also produced a geospatial dataset (feature class) for use in a geographic information system (GIS). This feature class consists of point features, each representing a borehole; tied to each point is a set of attributes describing the borehole in detail, as well as a link to an electronic copy of the geotechnical report from which the attributes were derived. This enabled us to include the data collected for the National Geological and Geophysical Data Preservation Program in a map layer as part of our online interactive mapping application, which is currently in development. This map layer will enable users to access information on any of the boreholes in this database, including the associated geotechnical report in PDF format. Below is a screen shot of the application, showing the map layer in question; this online application will be available to the public shortly (fall 2010).

Reference

McFarland, Carl R., 1983, Oil and gas exploration in Washington, 1900-1982: Washington Division of Geology and Earth Resources Information Circular 75, 119 p.
[http://www.dnr.wa.gov/Publications/ger_ic75_oil_gas_exploration.pdf; addendum for the years 1983 forward, http://www.dnr.wa.gov/Publications/ger_ic75_oil_gas_exploration_addendum.pdf]



Screen shot of online interactive mapping application (in development), showing the map layer containing geotechnical borehole information, including digital geotechnical reports.