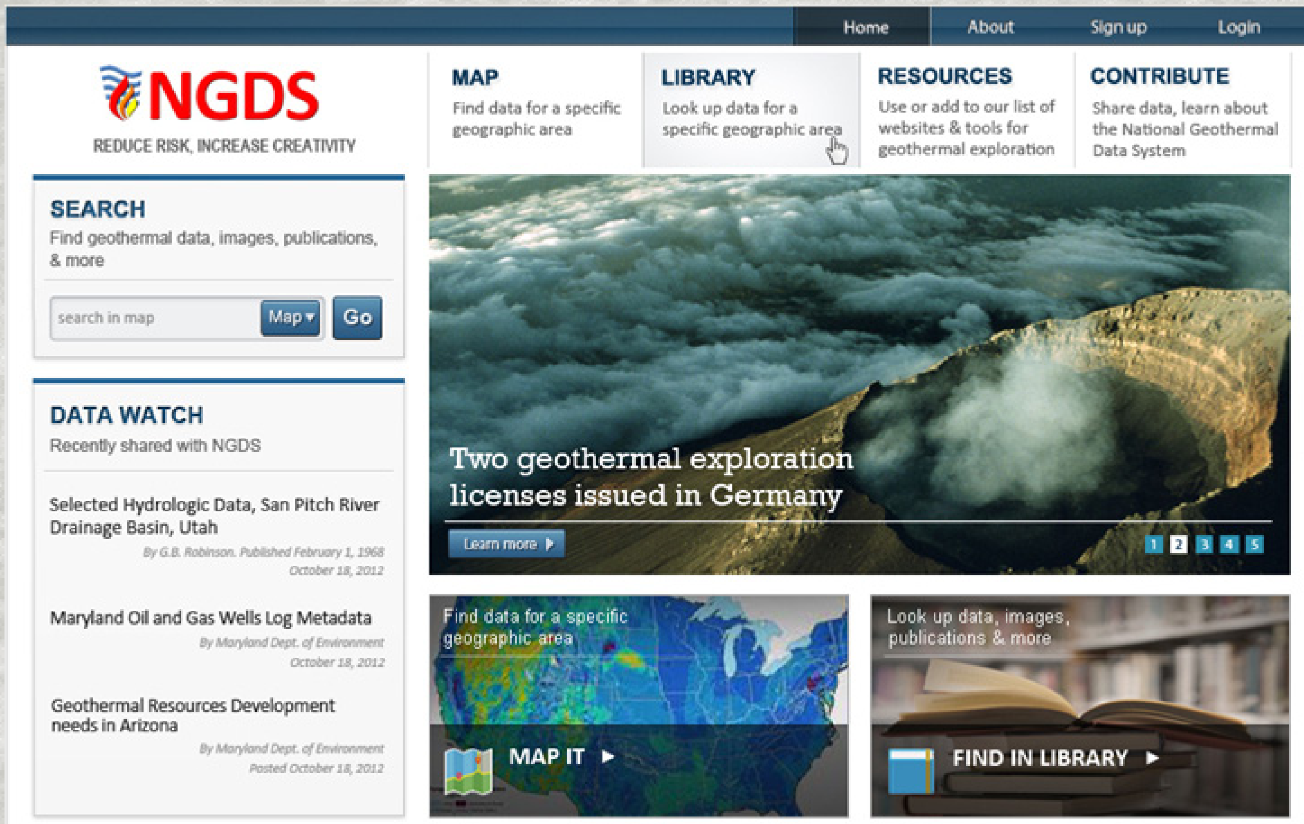
National Geothermal Data System: Transforming the Discovery, Access, and Analytics of Data for Geothermal Exploration

Kim Patten (ed.)
Arizona Geological Survey



OPEN-FILE REPORT OFR-13-06

May 2013

Arizona Geological Survey

www.azgs.az.gov | repository.azgs.az.gov

Arizona Geological Survey

M. Lee Allison, State Geologist and Director

Manuscript approved for publication in May 2013
Printed by the Arizona Geological Survey

For an electronic copy of this publication: www.repository.azgs.az.gov
Printed copies are on sale at the Arizona Experience Store
416 W. Congress, Tucson, AZ 85701 (520.770.3500)

For information on the mission, objectives or geologic products of the Arizona Geological Survey visit www.azgs.az.gov.

Recommended Citation: Patten K. (ed.), 2013, National Geothermal Data System: Transforming the Discovery, Access, and Analytics of Data for Geothermal Exploration: Compendium of Papers from the 38th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, February 11-13, 2013. Arizona Geological Survey Open File Report, OFR-13-06, 37 p.

Cover Photo Credit: Suzanne Boyd, Anthro-Tech, Inc., Xianjun (Sam) Zheng, Siemens Corporate Research, 2013

National Geothermal Data System: Transforming the Discovery, Access, and Analytics of Data for Geothermal Exploration

Compendium of Papers from the 38th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California
February 11-13, 2013
by

Kim Patten (ed.)
Arizona Geological Survey Open-File Report 13-06

May 2013

Arizona Geological Survey
416 W. Congress St., #100
Tucson, Arizona 85701

National Geothermal Data System:
Transforming the Discovery, Access, and Analytics of Data for Geothermal Exploration

Compendium of Papers from the 38th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California
February 11-13, 2013

The National Geothermal Data System (NGDS) is a distributed, interoperable network of data collected from state geological surveys across all fifty states and the nation's leading academic geothermal centers. The system serves as a platform for sharing consistent, reliable, geothermal-relevant technical data with users of all types, while supplying tools relevant for their work. As aggregated data supports new scientific findings, this content-rich linked data ultimately broadens the pool of knowledge available to promote discovery and development of commercial-scale geothermal energy production.

Most of the up-front risks associated with geothermal development stem from exploration and characterization of subsurface resources. Wider access to distributed data will, therefore, result in lower costs for geothermal development.

NGDS is on track to become fully operational by 2014 and will provide a platform for custom applications for accessing geothermal relevant data in the U.S. and abroad. It is being built on the U.S. Geoscience Information Network (USGIN) data integration framework to promote interoperability across the Earth sciences community. The basic structure of the NGDS employs state-of-the art informatics to advance geothermal knowledge.

The following four papers comprising this Open-File Report are a compendium of presentations, from the 38th Annual Workshop on Geothermal Reservoir Engineering, taking place February 11-13, 2013 at Stanford University, Stanford, California.

"NGDS Geothermal Data Domain: Assessment of Geothermal Community Data Needs," outlines the efforts of a set of nationwide data providers to supply data for the NGDS. In particular, data acquisition, delivery, and methodology are discussed. The paper addresses the various types of data and metadata required and why simple links to existing data are insufficient for promoting geothermal exploration. Authors of this paper are Arlene Anderson, US DOE Geothermal Technologies Office, David Blackwell, Southern Methodist University (SMU), Cathy Chickering (SMU), Toni Boyd, Oregon Institute of Technology's GeoHeat Center, Roland Horne, Stanford University, Matthew MacKenzie, Uberity, Joe Moore, University of Utah, Duane Nickull, Uberity, Stephen Richard, Arizona Geological Survey, and Lisa Shevenell, University of Nevada, Reno.

"NGDS User Centered Design: Meeting the Needs of the Geothermal Community," discusses the user-centered design approach taken in the development of a user interface solution for the NGDS. The development process is research based, highly collaborative, and incorporates state-of-the-art practices to ensure a quality user interface for the widest and greatest utility. Authors of this paper are Harold Blackman, Boise State University, Suzanne Boyd, Anthro-Tech, Kim Patten, Arizona Geological Survey, and Sam Zheng, Siemens Corporate Research.

"Fueling Innovation and Adoption by Sharing Data on the DOE Geothermal Data Repository Node on the National Geothermal Data System," describes the motivation behind the development of the

Geothermal Data Repository (GDR) and its role in the NGDS. This includes the benefits of using the GDR to share geothermal data of all types and DOE's data submission process. Authors of this paper are Jon Weers, National Renewable Energy Laboratory and Arlene Anderson, US DOE Geothermal Technologies Office.

Finally, "Developing the NGDS Adoption of CKAN for Domestic & International Data Deployment," provides an overview of the "Node-In-A-Box" software package designed to provide data consumers with a highly functional interface to access the system, and to ease the burden on data providers who wish to publish data in the system. It is important to note that this software package constitutes a reference implementation and that the NGDS architecture is based on open standards, which means other server software can make resources available, and other client applications can utilize NGDS data. Authors of this paper are Ryan Clark, Arizona Geological Survey (AZGS), Christoph Kuhmuench, Siemens Corporate Research, and Stephen Richard, AZGS.

ACKNOWLEDGMENTS

AZGS thanks Roland N. Horne, Thomas Davies Barrow Professor of Earth Sciences at Stanford University, and organizer of the Stanford Geothermal Workshop on Reservoir Engineering, for granting us permission to release these publications as Arizona Geological Survey Open-File Report, OFR-13-06.

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy, Geothermal Technologies Office under Award Numbers DE-EE0001120, DE-EE0002850, and DE-EE0002852.

PROCEEDINGS, Thirty-Eighth Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 11-13, 2013
SGP-TR-198

NATIONAL GEOTHERMAL DATA SYSTEM (NGDS) GEOTHERMAL DATA DOMAIN: ASSESSMENT OF GEOTHERMAL COMMUNITY DATA NEEDS

Arlene Anderson[1], David Blackwell[2], Cathy Chickering[2,] Toni Boyd[3], Roland Horne[4], Matthew MacKenzie[5], Joseph Moore[6], Duane Nickull[5], Stephen Richard[7], Lisa A.Shevenell[8]

[1] United States Department of Energy
e-mail: arlene.anderson@ee.doe.gov

[2] Southern Methodist University
e-mail: blackwell@mail.smu.edu
e-mail: catherine@mail.smu.edu

[3] Oregon Institute of Technology Geo-Heat Center
e-mail: toni.boyd@oit.edu

[4] Stanford University
e-mail: horne@stanford.edu

[5] Uberity Technology Corporation
e-mail: matt@uberity.com, duane@uberity.com

[6] University of Utah, Energy & Geoscience Institute
e-mail: jmoore@egi.utah.edu

[7] Arizona Geological Survey,
e-mail: steve.richard@azgs.az.gov

[8] University of Nevada, Reno
e-mail: lisaas@unr.edu

ABSTRACT

To satisfy the critical need for geothermal data to advance geothermal energy as a viable renewable energy contender, the U.S. Department of Energy is investing in the development of the National Geothermal Data System (NGDS). This paper outlines efforts among geothermal data providers nationwide to supply cutting edge geo-informatics. NGDS geothermal data acquisition, delivery, and methodology are discussed. In particular, this paper addresses the various types of data required to effectively assess geothermal energy potential and why simple links to existing data are insufficient. To create a platform for ready access by all geothermal stakeholders, the NGDS includes a work plan that addresses data assets and resources of interest to users, a survey of data providers, data content models, and how data will be exchanged and promoted, as well as lessons learned within the geothermal community.

INTRODUCTION

Geothermal energy, or literally, the thermal energy of the earth, is often used as a term to refer to conversion of the earth's thermal energy into electricity. Present geothermal power generation comes from high-temperature hydrothermal systems, the 'low-hanging fruit' of geothermal electrical potential. Newer techniques, including Enhanced or 'Engineered' Geothermal Systems (EGS) offer the opportunity to extend use of geothermal resources to larger areas of the western U.S., as well as new geographic areas of the U.S.

Through the American Recovery and Reinvestment Act of 2009, the U.S. Department of Energy (DOE) funded the development of a National Geothermal Data System. In early 2008 DOE issued a funding opportunity announcement to develop a “National Geothermal Database” to overcome barriers to the development of geothermal energy facilities and enable additional investment in conventional and Enhanced Geothermal Systems (EGS). Based on the proposals received and subsequent technical review, DOE funded an effort to create a web-based National Geothermal Data System for all publically accessible geothermal data. Data needs span all geothermal resources and applications including geothermal electricity production as well as direct use applications. Geothermal data is being contributed by industry, academic and national laboratory researchers, and by state and federal agencies. While the focus is on domestic data critical to identifying geothermal potential and characterizing geothermal reservoirs, international data sources may be included especially where such data and information can be utilized or benchmarked to help develop domestic geothermal resources. The system is being implemented using a federated, Service Oriented Architecture (SOA) based on the U.S. Geosciences Information Network (USGIN) (http://usgin.org/). The DOE adopted the US Government Accountability Office (GAO) (http://www.gao.gov/new. items/d06629.pdf) best practices for software development featuring an agile development process that incorporates the latest informatics technology and standards into the system design. NGDS has also adopted the International Organization for Standardization (ISO) (http://www.iso.org) metadata standards for the system catalog.

Data analysis for geothermal resource development presents a highly complex challenge where: "The rate-limiting step for all geothermal development is proving the resource – i.e., having sufficient geoscientific and exploration drilling data to be certain of a certain level of output" (Bloomberg New Energy Finance, 2012).

A variety of data is required to ascertain whether a potential geothermal energy site should be developed for production: composition and hydrologic properties of materials hosting the thermal energy, proximity to existing power grids, and quantity of thermal energy flowing from the interior of the earth are all primary considerations.

The NGDS will provide critical geothermal-related data that can be easily accessed to:

- Help companies be more (cost and time) effective in exploration, development, and usage of geothermal energy.
- Support a knowledge repository and archive for geothermal data, lessons learned, reports.
- Advance earth sciences by identifying gaps in our knowledge and informing new geographic areas of the U.S.
- Provide a reliable base load energy source of knowledge.
- Increase public awareness of geothermal energy.

These goals can only be accomplished if NGDS provides a quality user experience, and is widely adopted by users in the geothermal community.

There are three targeted user communities for NGDS, and each user group has different goals, needs, and tasks when interacting with NGDS.

- **Data providers** expose information to NGDS through standardized, internet-accessible interfaces and standardized formats.
- **End users or data consumers** utilize NGDS to access data to support their work in geothermal energy exploration and development.
- **Application developers** build applications that utilize the data in NGDS, and make it easier for end-users to interact with the system.

An additional NGDS goal is helping users to understand where geothermal investment will have the best opportunity for success. Through NGDS, users will gain access to tools that can improve the usefulness of geothermal data and information.

Providing simple links to geothermal datasources across the country would only improve knowledge enablement to a limited extent. A non-exhaustive list of reasons why making simple links to existing data sources is inadequate includes:

- Data is in multiple formats, layouts, units, paper versions and not searchable via one central index;
- Some database persisted data is difficult to access, visualize and/or interpret, especially for the business/industry user;
- There is no current ability to link some data to additional geological information or datasets;
- There are inconsistent standards for quality assurance or reliability of data.

In order to structure the records so the data can be linked and interoperable, data and metadata content models and interchange formats were created. To date, twenty-eight geothermal data models have been developed, reviewed and adopted. The Geothermal Domain Committee provided expert input on a prioritized list of data models including geothermal drilling and well log data, aqueous chemistry, geophysical data and active fault maps. Expert input on whether data models include the correct information is critical.

NGDS Data Architecture

The NGDS is not a single database. Rather, it is a unified data access system based on the registration of resources in a shared catalog system using standardized metadata. NGDS has a tiered data access scheme accommodating file-based, non structured, and standards-based structured data delivered using standardized web services and interchange formats. A data resource becomes part of the NGDS system when standard NGDS metadata is created, validated, and made discoverable through the NGDS catalog system, and the data resource is accessible via procedures specified in the metadata. Much of the information that is or will be registered in the NGDS is unstructured data. Other resources, such as drill cores, may not be available in electronic format. Especially in such cases, metadata is essential to allow NGDS users to be aware that the resources exist.

Figure 1: An example of a resource (drill cores) requiring metadata records indexable via NGDS (Photo courtesy of Energy & Geoscience Institute).

Additional information about the system design is discussed by Clark et al. (2013).

GEOTHERMAL DATA PROVIDERS

The NGDS provides access to information resources on geothermal energy from a national network of data sources (http://geothermaldata.org/). As of the date of this paper, four project teams are collaborating and leveraging efforts that will culminate in the NGDS launch. This includes the NGDS Design & Testing Project; Heat Flow Data Aggregation; State (geological survey) Contributions to NGDS; and the DOE Geothermal Data Repository. Additional information about the GDR is provided by Weers and Anderson (2013).

Under the leadership of Boise State University (BSU), the NGDS Design & Testing Project includes four key geothermal data providers including: the University of Utah Energy & Geosciences Institute; the University of Nevada, Reno; the Stanford Reservoir Engineering Department, and the Oregon Institute of Technology Geo-Heat Center. The Arizona Geological Survey and Siemen's Corporate Research lead the NGDS informatics development.

Two additional NGDS projects focused on data content are: Arizona Geological Survey (AZGS) "State Contributions to NGDS" including data from all 50 state geological surveys and the "Heat Flow Data Aggregation for NGDS Data Development, Collection and Maintenance" project led by the Southern Methodist University's (SMU) Geothermal Laboratory. The SMU consortium includes: the Bureau of Economic Geology (BEG), University of Texas at Austin; Cornell Energy Institute, Cornell University; the Geothermal Resources Council (GRC); MLKay Technologies; Texas Tech University (TTU); the University of North Dakota (UND), and Siemens Corporate Research (SCR).

The SMU node is focused on improving access to information to allow for new interpretation of data, thereby increasing its usefulness for commercial geothermal energy development. The DOE-GDR is another vital node on the NGDS. The GDR is hosted on the Open Energy Information (OpenEI) Platform.

The U.S. Geological Survey (USGS) and other federal agencies that produce geothermal data are potential NGDS data providers. An interagency agreement with the USGS is designed to provide geothermal resource assessment and classification data. As part of the national geothermal resource assessment, USGS has conducted a comprehensive survey of the available information on geothermal systems and an extensive set of geothermal databases. These databases include chemical analyses of water and gas samples, heat flow measurements, gravity and magnetic surveys, geologic maps, seismicity catalogues, seismic surveys, drilling records and other relevant exploration and development data, including consultants' reports and interpretive studies.

USGS personnel have combined many of these supporting geological, geophysical, geochemical, and hydrologic datasets into Geographic Information Systems (GIS) databases and maps for analysis and publication. The current USGS data provision strategy is to provide online access to data and reports either through direct delivery of data to the DOE GDR or through state geological survey NGDS nodes. The USGS will also provide the information necessary to ensure that geothermal resource data will be current and available. Examples of the data include geothermal resource assessment and derivative products, such as GIS maps, low-temperature data series and related publications. As new data is acquired and geospatial products and data are developed, they will be sent through USGS for review and subsequently provided to the NGDS.

Table 1 summarizes the work plan for deliverable data assets from NGDS Design & Testing Project partners. Abundant data has also been incorporated into the NGDS by the AASG "State Contributions to NGDS" project. The work plans and progress can be monitored at http://www.stategeothermal-data.org/progress/aasg_tracking_map. A list of data delivered by that project is available at http://repository.stategeothermal-data.org/ repository/browse/. An interim search interface for the NGDS is accessible at http://search.geothermaldata.org/.

THE NGDS DATA MODEL

The NGDS is based on a model (Figure 2) that uses the top class ***NGDS_Resource*** to denote any resource within the NGDS system NGDS_Resource's can be further classified as ***Data Resources***, ***Metadata*** or ***Annotations***. NGDS Data Resources represent the actual resources of interest to end users. An example of an NGDS Data Resource might be a spreadsheet, using Comma Separated Values (CSV), showing the temperature of a well at different depths or a physical drill core sample. Every NGDS data resource must be described by at least one metadata record, as described below. The onus of data maintenance is shifted towards organizations having responsibility for data management and preservation. By documenting data schema, encoding formats and practices, data can be put into the 'data integration' format when it is made available on the web. Because of its enhanced utility in a standardized format, management and preservation of the data are more strongly motivated. As previously referenced, the NGDS will initially use twenty-eight data content models and interchange formats, as well as a standard metadata scheme. An integrated data access portal application is in development.

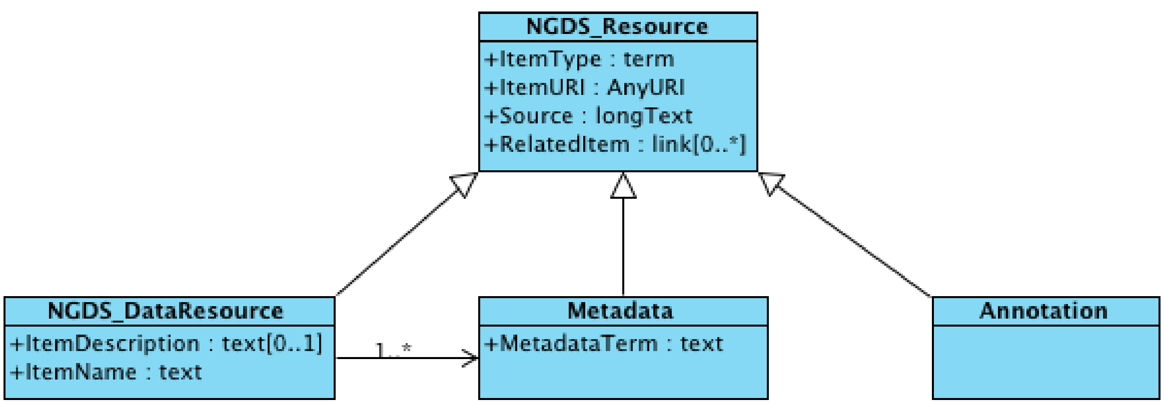


Figure 2: The high level NGDS data model.

Table 1: *Summary of deliverable data items from NGDS development and testing project participants.*

Group	Data Item
Stanford	Bibliographic Database for Proceedings from the annual Stanford Geothermal Workshop count: 2118 metadata records with location keywords
	Metadata Records for 3 Adsorption Data publications
GeoHeat Center, Oregon Institute of Technology (OIT)	717 Technical Papers and bulletin articles online, with NGDS metadata records
	Metadata for 4185 documents in the Geo-Heat Center Library
	Documentation and registration of data set describing 554 Geothermal Wells in Klamath Falls area
	Documentation and registration of data set describing 404 Co-located Sites
	In cooperation with Siemens Corporate Research (SCR) and University of Nevada, Reno (UNR), thermal springs and borehole temperatures will be de-duplicated for the 16 western states, processing non-standard location information, and served in the NGDS content model as the OGC's Web Map Services (WMS, OGC 07-063r1) and Web Feature Services (WFS, OGC 09-025r1 and ISO/DIS 19142).
	Documents and data related to the Klamath Falls #57310 project will be scanned and publicly accessible online with metadata.
	Metadata for GeoHeat software Tools and Spreadsheets.
University of Utah, Energy & Geoscience Institute (EGI)	2635 Scanned well logs indexed in NGDS Well Log Observation Content Model.
	9010 scanned reports, articles, maps, charts and graphs with metadata.
	Geothermal Sample Library samples registered with System for Earth Sample Registration (SESAR - http://www.geosamples.org/), and correlated with well log and well header data sets
	Create metadata for more than 1000 Scanned Documents
	Catalog and scan 20 boxes of well logs.
University of Nevada, Nevada Bureau of Mines and Geology (NBMG)	Metadata *for* more than 400 known publications and grey literature relevant to geothermal exploration and development in Nevada
	More than 2000 documents (notices, permits, gray literature) to be scanned and placed online with metadata records
	Approx. 150 1:24k scale geologic maps to be scanned and geo-referenced, with metadata
	Map and report describing all exploration activity reported in 2012 will be scanned, put online, with metadata
	Metadata for more than 179 existing geologic, geophysical and geochemical data sets relevant to geothermal assessment. Update NBMG Geothermal web map applications to operate with Tier 3 NGDS services.
	NBMG Geothermal map applications will be updated to operate with NGDS services and integrated with NGDS applications being developed by Siemens.

Data Tiers

NGDS was designed to use a tiered data delivery scheme that allows the necessary flexibility to accommodate unmanaged legacy data in whatever form it is available, as well as high value data in standardized content models and/or interchange formats. The system uses a community governance scheme to adopt new interchange formats, and provides a repository where the specifications for each data exchange are available to all.

In order to make the incorporation of a large quantity and variety of data in the NGDS, a tiered data acquisition scheme has been used.

- **Tier 1: Unstructured** — represents file based resources such as unstructured data in text and images, requires a user to extract data for analysis.
- **Tier 2: Structured, but not standardized** — represent data structured in proprietary formats that are not conformant with a standard NGDS content model. Data in this tier would need to be transformed in some fashion by a data consumer in order to integrate with NGDS-standard datasets.
- **Tier 3: Structured, standardized** — data published in the NGDS standardized protocols and interchange formats supported by NGDS content model.

A large part of the available resources are scanned images of legacy reports, maps, and other figures that are registered with metadata and made available as Tier 1 resources. Tier 2 allows registration of existing structured datasets that are not in standard NGDS content models and interchange formats. This is not a preferred data acquisition approach, but is expedient and useful for unique datasets that have only a single instance.

Tier 3 data acquisition is the preferred scheme, but because of the additional effort required to edit and review datasets to get them into the standard interchange format, it has been necessary to prioritize effort. This was done by first surveying the data provider community to determine the types of structured data that they actually have available for inclusion in the NGDS. The team then informally polled geothermal exploration and development practitioners (mostly in the State Geological Survey community) to determine which of these types should be prioritized.

Linking Geothermal Data Providers Through Service Protocols

A protocol is a set of rules used by computers to communicate with each other across a network. Since virtually all of the data types identified for NGDS Tier 3 interchange are geographically located features, the OGC Web Feature Service (WFS) (Vretanos, 2005) is being used as the data service protocol and the OGC's Web Map Service (WMS) (De La Beaujardiere, 2006) is used as a standard protocol for serving geo-referenced map images over the Internet that are generated by a map server using data from a GIS database.

The WFS protocol uses the OGC Geography Markup Language (GML) (Portele, 2007) geometry for location description, and allows feature types to be defined that are expressed by feature-specific eXtensible Markup Language (XML) schemas. Geographic data is also made available for viewing with geographically enabled software as OGC Web Map Services.

Document-based resources use the standard HyperText Transfer Protocol (HTTP) that is the foundation of the World Wide Web. WFS and WMS are implemented on top of HTTP. Some data providers also make files available using the File Transfer Protocol (FTP). These protocols are both widely used within the geothermal community.

CONTENT MODELS

In the NGDS, content models specify the structure and properties associated with an interchange feature, typically including feature-specific metadata allowing documentation of each data item. Content models are specified independent of interchange formats, the latter being a typed expression of the content model. If data cannot be structured using an existing content model, geothermal community members are invited to propose new models.

Development of content models during the first year of the project has been an organic process. The models have evolved rapidly as production scale data compilation has started.

Content Model Inventory

Various approaches have been used to prioritize the kinds of data that will be implemented using Tier 3 services. NGDS consortium members were polled in January and February, 2010 to get an inventory of the resources that they will be contributing to the system, but the results were limited in terms of specifics, mostly recognizing scanned well logs and other kinds of documents. The data resource inventory continues through verbal interviews with information managers at data provider organizations and with geothermal industry practitioners. With the initiation of the AASG "State Contributions to NGDS" project, state geological surveys were polled yielding a larger body of data resources to be made available through the system. The evolution of the Tier 3 information exchange inventory will continue as NGDS participants develop plans for data contributions, and new projects and participants are factored in.

Content models available to date include (see http://geothermaldata.org/page/ngds-content-models for details on the content models):

- Aqueous Chemistry
- Borehole Temperature Observation Feature
- Data Interchange Content Models
- Direct Use Feature
- Drill Stem Test Observations
- Fault Feature

- Fluid Flux Injection and Disposal
- Geologic Contact Feature
- Geologic Unit Feature
- Geothermal Area
- Geothermal Fluid Production
- Geothermal Power Plant
- Heat Flow
- Heat Pump Facility
- Lithology Interval Log Feature
- Metadata
- Physical Sample
- Powell Cummings Geo-thermometry
- Power Plant Production
- Radiogenic Heat Production
- Seismic Event Hypocenter
- Thermal Conductivity
- Thermal/Hot Spring Feature
- Volcanic Vents
- Well Fluid Production
- Well Header
- Well Log Observation

Additional content models under consideration:

- Daily Drilling Report
- Well completion Information
- Well production hardware
- Surface Alteration
- Subsurface Alteration
- Geophysical Survey Results
-

Figure 3 introduces a more in-depth model of the data item content models used for Tier 3 data. The content models are designed based on this pattern, with a distinction between features (Kottman and Reed, 2009) that represent geographically located real-world phenomena, and observations (Cox, 2010) to represent individual measurements of one or more properties of some real-world phenomena. A Feature typically summarizes the results of multiple observations to characterize something like a fault, a geologic unit, a well, a power plant, or a geothermal area. Observations represent the more granular data, 'raw' data like individual temperature measurements, chemical analyses, or heat flow determinations. Observations may have composite results; for instance an individual well log is considered an observation result from a log run event.

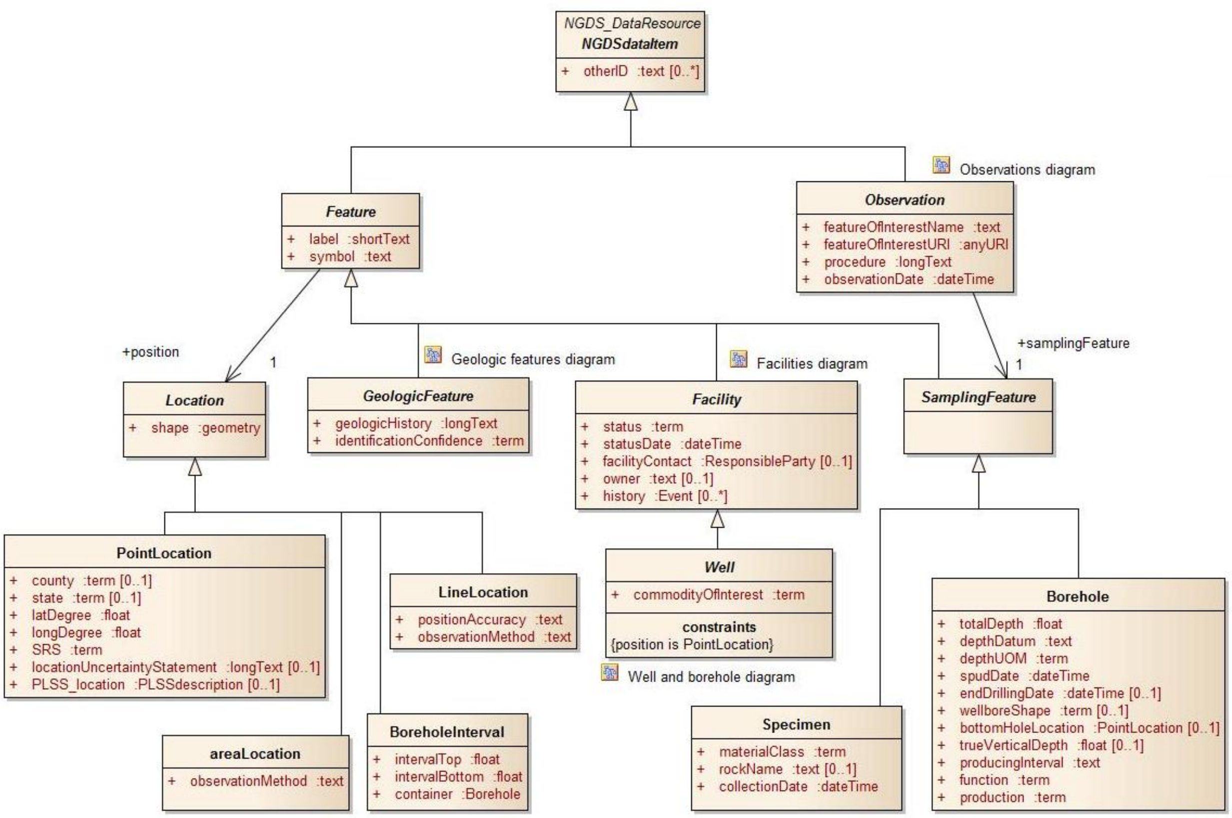


Figure 3: Top level model of NGDS Data Items for Tier 3 data sets.

The key property of a Feature is that it has a spatial-temporal location; only the spatial aspects of location are modeled in Figure 3. Features are subtyped into two broad categories of interest to the geothermal domain. Geologic features are used to represent natural features within the Earth; subtypes include geologic unit outcrop, fault, Quaternary fault, geologic contact, volcanic vent, thermal spring, and geothermal area. Facility is used to represent feature of human origin; subtypes that are currently implemented include well, geothermal direct use site, heat pump facility, and power plant facility.

Another sub-category of Feature, SamplingFeature is used to represent the artifacts that are the immediate target of observations, and serve to geo-locate and contextualize an observation result; specimen (core, rock sample) and boreholes are the dominant sampling features of interest in the NGDS.

Observations represent observed or measured property values that characterize a feature of interest (e.g. a rock unit or geothermal reservoir), have a measurement procedure, are associated with a sampling feature, and have one or more result values. Observation types currently implemented include fluid flux, seismic event, drill stem test, rock chemistry, aqueous chemistry, heat flow, borehole temperature, and borehole lithology interval.

Interchange Formats

In order for information to be exchanged, a content model must be serialized in a form that can be transmitted over a computer network and interpreted by software applications. Interchange formats specify data encoding and internal file structures that can be used to exchange data between different hardware and software applications. A useful analogy can be found in modern printers. Files sent to the printer are exported by computer programs in a format such as the Microsoft Windows Metafile (WMF) or Adobe Systems PostScript formats. WMF and PostScript are interchange formats that can be read by most printers alleviating each software developer having to write instructions for many different printers.

Every service implemented has an associated behavior model and data model. The data model is usually expressed as an interchange format. The use of interchange formats mean data producers and consumers can continue to use their internal data formats that are optimized for their business requirements. Examples might include proprietary data created by a scientific measuring device. Such data formats may be suboptimal for using as interchange formats.

A number of international efforts are under way to develop specifications for data interchange of geoscience information that are applicable to NGDS Tier 3 data types. These include (GeoSciML) (Richard and CGI Interoperability Working Group, 2007; see also http://geosciml.org), and the OGC observation and measurement model (Cox, 2010). These models are very flexible and allow representation of a wide range of content, but are thus correspondingly complex and difficult to use. Thus, in the initial phase of the project, content models have been defined using relatively simple schema in which property values are specified only by string or numeric-valued elements (no nested or complex data types). The content models are designed to be compatible with the more complex and comprehensive models mentioned above to the degree that is practical.

USGIN is currently implementing interchange formats as GML Simple Features (van den Brink et al., 2011), compatible with the service protocol in use (OGC WFS). WFS can be consumed by existing clients like ESRI ArcGIS Desktop and Quantum GIS. As clients are developed for richer-content, the NGDS can adopt more complex, information-rich interchange formats.

Versioning

Another challenge to the geothermal community is the evolution of standardized interchange formats. WFS services have been deployed using interchange formats implemented as the models evolve (an agile process), and iteration of model versions and XML schema for corresponding WFS features can easily result in discrepancies between interchange format versions.

An important part of system operation and maintenance is ongoing review of deployed services and careful validation of new services to maintain conformance with the specifications and a well thought out change mitigation plan. As part of the change management process, XML schemas are versioned and the namespace for the schema elements is unique to that schema version. Thus namespace-aware client applications can determine if an instance document is using a supported version.

METADATA

To meet its main objectives, specifically catalog-based search, discovery and retrieval of resources, the NGDS requires quality metadata describing the information resources. Metadata, in the context of NGDS, is data that describes a physical or electronic resource, provides information about the content of the resource, its origin and processing history, how the content is represented, and how the resource can be accessed. Note that the term resource is used here in a very broad sense to mean any identifiable item of interest to users of the information system.

NGDS metadata content can be generally classified into one of five categories:

Basic metadata provides information that applies to a wide spectrum of resources, and includes the title, a description of the resource, author(s) (originator), the creation (or publication) date, and specification of the natural language of resource content. Information used for metadata maintenance, such as the metadata record ID, update date, point of contact, and metadata specification name are also included in this group.

Guide metadata is used to help users find, evaluate, and access specific items. This group includes access instructions, distributor contact information, bibliographic citation, a unique identifier for the resource, links to access the resource online, keywords categorizing the resource, information about the quality of the resource content, the geographic area described by the content, and any constraints on access or usage of the resource.

Process metadata captures more specific items such as the process used to create the data, the purpose of the data and the context in which the data was created. This could include a description of machines used to measure or sample a physical entity or items like testing processes used to derive the data.

Structural metadata is used to describe the structure of data such as syntax, serialization, tables, columns and indexes. This term can also be used to describe the organization of the data and if electronic, the serialization or Multipurpose Internet Mail Extensions (MIME) type used. This may also include details of physical objects such as drill core samples. The basic NGDS metadata only includes information on the format of an online representation of the resource. The full ISO 19115 (2003, 2006) (schema includes many additional fields for describing resource structure that may be included in NGDS metadata.

Domain specific metadata is often applicable only to a particular data type, and is thus not suitable for inclusion in metadata meant to be applicable across the entire range of NGDS resources. Such information can be included in the description or lineage statement fields in the basic metadata content, if it applies to all records in a dataset. The individual content models developed for NGDS data types include fields for metadata content specific to individual data instances of that type. Metadata at the domain (feature) specific level is accessed through data services for the particular feature type, or may be understood by studying written documentation.

The NGDS adopted the minimum metadata content recommendations for geoscience resources and the metadata content and encoding profile developed by the USGIN project (USGIN, 2011a, 2011b). For NGDS catalog purposes, the USGIN recommendations have been relaxed, allowing some of the recommended mandatory fields to be nilable—i.e. the field must be populated, but a value of 'missing' or 'not applicable' is allowed to indicate that the information is not provided. In the end, the practical minimum metadata requirement is that there is an informative title, some kind of geographic location information, and sufficient information for a user to know how to get the resource. If a document or dataset is not specific to any geographic location, the location keyword 'nongeographic' is used. For geographic location, a latitude-longitude bounding box is the preferred specification, but lacking that, place-name keywords are allowed. Document-based resources registered by project partners are expected to be accessible on the web, in which case the access information will be a web location (Universal Resource Locator or "URL" for short) that will get an electronic version of the document. For physical NGDS referenced artifacts, a contact point to request access to the resource should be part of the metadata.

At a more granular level, individual records (features, objects) in a dataset may include source information, documenting details of observation or measurement procedure and other information specific to a particular data type. This might include information such as location, data and time of observations, and the source of that data. These feature-level metadata are delivered with the data, and only summarized in the work-level dataset metadata that are published to the NGDS catalog. This granularity issue can be difficult because of differing perspectives on what is data or metadata, differing granularity of documentation available, and different use-case priorities.

All geo-scientific data (e.g. geology, geochemistry, geophysics, remote sensing, temperature surveys) require geographic coordinates to place the data in the proper spatial configuration for analysis. The NGDS requires data input to include surface location information, and depths where appropriate, for all data input such that users can query multiple data sets (and publications) to obtain relevant information for their analysis of either site specific or regional geothermal areas. As such, wherever possible, all data is spatially located for ease in locating and using data specific to user's needs. The data included allow work to be conducted in all phases from preliminary, geothermal exploration of regions and target/area identification to site assessments and resource development of specific areas.

Annotations

The Annotation class in Figure 2 represents tags, ratings, event log items, comments or links to other resources that are asserted by NGDS users, as opposed to the data owner, steward, or provider. The ability for users to associate annotations with NGDS resources provides a feedback mechanism resulting in an emergent knowledge base.

Tags are a kind of annotation that consist of plain language text terms assigned by users to categorize resources according to schemes that they find useful. Because individuals think differently, it is useful to enable users to augment metadata by adding such annotation. This *bottom up*, collaborative process produces what is commonly known as a folksonomy (http://en.wikipedia.org/wiki/Folksonomy). The NGDS team believes that incorporation of such crowd-sourced tagging will help improve search effectiveness by combining this approach with the top down controlled keyword approach and using a thesaurus-like functionality.

Metadata Acquisition

The project participants have used various metadata content schemes that must be harmonized to enable an integrated catalog search. Existing metadata includes lists of files compiled in a text document or spreadsheet, various databases constructed by organizations to manage their library holdings, and formal metadata conformant in varying degrees to Federal Geothermal Data Committee (FGDC) or rarely ISO standards, constructed according to locally varying interpretation and practice. In some cases, the metadata collected is not sufficient to conform to the USGIN recommendations. Manual addition of information to complete the metadata could potentially require funding resources beyond what was budgeted.

The challenge facing the team is to minimize the manual data entry required to ensure sufficient metadata content to enable a set of use cases. Making metadata acquisition as simple as possible is a design goal of NGDS. Approaches include user-friendly forms, spreadsheet editing that is familiar to most computer users, transformation processes from existing database metadata, and automated metadata extraction. The metadata requirements were also relaxed somewhat (as noted in the metadata section) to allow ‘missing’ as a value for some required content.

Metadata entry workflows developed and in use for the AASG “State Contributions to NGDS” project that are contributing to NGDS include a web form interface and a spreadsheet template for compiling metadata. The form interface uses background user log-in information to auto-populate some of the metadata, as well as providing pick lists and auto-complete functions in the data entry fields. The date and timestamp of submission can be recorded, saving the data provider from having to create this data for each submission manually.

Use of the spreadsheet allows users to do extract/transform/load processing from their existing metadata table or spreadsheet using familiar cut, paste, search/replace, and fill-down operations supported by the spreadsheet software. The spreadsheet metadata compilation table columns are mapped to the USGIN ISO metadata profile, and metadata entered in each row can automatically be converted to an XML record to import into the NGDS catalog. The software that does the conversion operates on a comma-delimited text (CSV) formatted table, which can be exported from the spreadsheet software or created by a variety of other workflows.

Data providers with metadata expressed in a database schema have a variety of options for publishing the metadata to the NGDS catalog. Standardized Query Language (SQL) views that duplicate the table structure of the metadata compilation spreadsheet can be used to export CSV files that can be converted to XML. A more streamlined approach is to implement a USGIN-ISO XML export function directly against the table in the database. By saving these files in a web-accessible directory that can be harvested by the catalog, the metadata content in the database can be kept synchronized with the NGDS catalog with virtually no user intervention.

Location information

One of the major challenges for metadata acquisition is obtaining the geo-location information for the numerous resources. In order to enable the basic geographic search use cases using a map interface, each resource metadata record must have a latitude-longitude bounding box that delineates the geographic area that is the subject of the resource. The metadata creation form interface allows the user to draw a rectangular box in a map view. With care, this can produce accurate location metadata, yet this is time consuming, typically requiring 3-5 minutes per metadata record. If this is deemed too much effort, locations can be specified using place-name keywords. In some cases, if there is a good correspondence between a named location (mountain range, valley, known geothermal resource area) and the subject area for a resource, this gazetteer approach can yield good results. In many cases it may be possible to correlate the named locations with geographic bounding boxes to enable the map-based geographic search.

A large amount of geothermal data is obtained from wells that are traditionally (in the United States) located with legal descriptions based on survey bases like the Public Land Survey System (PLSS) (http://www.geocommunicator.gov/geocomm/lsis_home/home/lsis-plss-description.html). GIS datasets with the PLSS grids are available from the Bureau of Land Management for many of the western United States (http://www. geocommunicator.gov/GeoComm/lsis_ home /home/ index.htm#plss), and these enable automated mapping of consistently formatted Township-Range-Section-Quarter Section type PLSS locations to a bounding box or center point that can be used in geographic search for wells in a well header feature service.

Automatically Generated Metadata

Some metadata, such as the electronic transfer protocol used to retrieve the NGDS Data Resource (examples: FTP, HTTPS), the methods required (HTTP Post, Get) can be populated by default if the metadata is being uploaded to a repository. Structural Metadata, such as the MIME type, can be inferred during a file upload process as well.

In some cases, a file that is being registered may already contain some useful metadata. Portable Document Format (PDF) documents using version 1.5 or later include a metadata section with content defined by Extensible Metadata Platform metadata standard (XMP) (Adobe Systems, Inc., 2005). The XMP scheme extends Dublin core with a variety of properties. Recent versions of Microsoft® Office® documents also have internal metadata sections. If any of this metadata content was created with the file, a data provider may possess metadata without even realizing it. This sort of metadata be programmatically detected by the NGDS resource registration software using a software toolkit like Apache Tika (http://tika.apache.org/).

Some metadata content can be automatically generated when a resource is registered to the NGDS system. For example, an identifying Uniform Resource Identifier (URI) can be assigned automatically if none is provided, as well as the URL for accessing the resource if the file is uploaded to an NGDS node.

Manual Metadata Entry

When resources are registered in the NGDS, a metadata validation process will be run to determine that metadata requirements are met. This is necessary to ensure a minimal set of metadata to accommodate all the user interface functionality revealed by a User Centric Design (UCD) research project performed as part of the NGDS work. NGDS data resource providers will be requested to complete any missing information. In some cases, there may be several dates associated with the data that must be manually specified, such as the curating date, the creation date and in some cases references to dates or specific tests or observations. Other information about the resource might only be obtainable from the data steward and require manual entry.

For example, the NGDS metadata content recommended more entries than that required for existing metadata from the Oregon Institute of Technology Geo-Heat Center library. The information was incorporated into the compilation spreadsheet. After further refining the content, the catalog import reviewers requested some additional changes in the way Geo-Heat Center data was entered. For example, Geo-Heat Center added a column with location allowing the program to define the bounding box. Keyword entries were separated by a pipe symbol instead of semicolons.

The job of manually creating and verifying the metadata is shared among several roles as described in an NGDS Software Requirements Specification (SRS) document. This alleviates one person from an unfair burden of work and also ensures that quality checks are performed.

METADATA COMPARISON

The NGDS team has compared metadata elements from the different profiles in use to the USGIN ISO metadata profile to determine how compatible the metadata standards are with each other. The metadata models being used within NGDS included:

1. The U.S. Geoscience Information System (USGIN) Profile for ISO 19115/19139 (2003, 2006);
2. The NGDS Metadata Compilation Template v1.3.4, which is a simplified, flat-table view of the USGIN profile;
3. The National Renewable Energy Laboratory (NREL) DOE-GDR Metadata Template (Weers and Anderson, 2013);
4. Dublin Core metadata vocabulary. Dublin Core is a basic set of metadata commonly used to declare citations and associate authors and other attributes with documents. This vocabulary is used by the DOE’s Office of Scientific & Technological Information (OSTI) (http://www.osti.gov/OSTI_OAIrepositorymanual.pdf) and DCAT (http://www. w3.org/ TR/vocab-dcat/);
5. Metadata terminology and taxonomies developed for geothermal data collecting and mapping by the Southern Methodist University (SMU) team as part of a heat-flow data base development project for NGDS;
6. Ordinary plain language folksonomy terminology that arose from the NGDS UCD work.

In the metadata comparison shown in Figure 4 each row has a label for the baseline metadata concept in the left column, and columns for the corresponding metadata content field labels from the schemes to be harmonized listed above. Each row represents a different metadata content element and includes the terms used for that element. The grey shaded boxes indicate places where a model includes no content item corresponding to the concept in that row. Many of the metadata content items were easily mapped to the concepts in column one.

Information in fields that do not map directly to ISO metadata elements can be included in the free text abstract field to be made available to users.

REQUIREMENTS FOR DATA ACCESS

When interviewed, users participating in the UCD study indicated that they prefer to search using a map view as the interface, and would like to know what data exists within the given boundary of a shape on the map. A typical prospector would likely start a search with a map to see the location of data acquisition sites (e.g. wells, outcrops) and access information available from those sites. The ability to filter the data based on whether or not it is within a defined map area would require that the metadata include geospatial location information. The ability to find results meeting the rest of the criteria would require accessing the actual datasets. The metadata would guide discovery and access to the appropriate datasets, which would then need to be analyzed and integrated to respond to identify the target sites.

This will facilitate the types of searches required from the UCD work. To elaborate on this, consider the following use case:

"A geothermal prospector or resource geoscientist working for a land owner or potential developer/investor has a property presented as a geothermal prospect. They need to know what geological, geophysical, land use and other datasets exist that are relevant to the decisions to be made about the viability of a geothermal project. For example, finding data that could indicate there are springs nearby, geochemical geo-thermometer data for the water, data about wells in or near the area including the depth and temperature, the nature of the heat flow gradient and the heat flow of the wells, is a benefit to geothermal prospectors. Additionally being able to locate and retrieve copies of any relevant publications (geological, geothermal, et al) that deal with the area or document it in more detail would be useful."

This requires that the metadata contains latitude and longitude coordinates in order to map them in relationship to roads and power grids. The content model for the dataset must have the temperature data and related depth information. If the data resource is a Tier 1 unstructured resource, the metadata records corresponding to that record should indicate that the data exist in a non-programmatically accessible format so that an individual will be able to review it. Additionally, the metadata for the resource must specify the web location of the NGDS Resource, the protocol used to retrieve it, the resource identifier, and a description of the resource if it is in an electronic format.

A	D	F	H	I
	Metadata Specifications Field List Comparison Worksheet			
BASELINE for comparison Technical metadata interchange content model: Metadata Compilation Template v1.3.4 (and 1.3.5) (fields reordered in groupings)	Plain English metadata capture template (derived from technical metadata interchange content model)	NREL DOE-GDR Metadata Template (Jon Weers)	DCAT Vocabulary (CKAN-supported specification)	DATA.GOV Metadata
Element	Field	Field	Property	Element
title	Data Resource Title	Data Resource Name	title	Title
description	Data Resource Description	Description	description	Description
publication_date	Publication Date	** inferred	release date ; dataset update/modification date	Date released ; Date updated
resource_languages	Data Resource Language	** inferred	language	
resource_id**	Data Resource Name**	** generated	identifier	User Generated ID
resource_url**	Data Resource Path**	Data Resource Filename & Path		Access point
resource_type	Data Resource Type	** inferred	format	File format ;

Figure 4. Metadata comparison spreadsheet

Without such metadata declarations, the NGDS user would not find the data

In some cases a required dataset does not come from the geothermal community. Data such as road information, power grid location and land ownership will be provided by other authorities. Using a Service Oriented Architecture (SOA) accessing map data through standard OGC web services enables mash-ups that include NGDS data with layers added from other sources.

UCD user-research also identified a vocabulary of keyword terms suggested by potential NGDS users to provide guidance on user-expected search methodology. The studies showed that the intended user constituency commonly searched for information using different terminology than those providing the data. In some cases, the words were simply synonyms such as "drill hole" vs. "bore hole". Mapping between the various tag and keyword vocabularies from different organizations and communities is an area of ongoing research.

Access to Large Nationwide Datasets

Facilitating detailed data discovery for highly geographically dispersed datasets is a significant user interface challenge. For example, the SMU heat flow database contains data for thousands of sites throughout the country. The catalog metadata record for this database can only indicate the kinds of information that might be available for each site, and the boundary of the region that contains all the sites. A user must interact directly with the database to learn the precise location of the sites and what information is available from a particular site. In this instance, the catalog would lead the user to a Web Feature Service (WFS), that places the data from the system into an explicitly defined structure (the Content Models referenced previously) that can then be manipulated by one or more front end applications that merge the data with other available resources, such as road information, power grid locations and water resources. The client application must enable the user to navigate from the metadata for a whole regional dataset to the detailed data from individual records in the dataset using a seamless process that requires minimal input from the user.

LESSONS LEARNED

Demonstrating the value of metadata and Tier 2 and Tier 3 data to data providers and users is an ongoing challenge. Many users might think they are referring to metadata when they are actually referring to instance data, or information about a specific field within a data record. The user interface must be carefully designed and constructed to guide users in an intuitive way. By carefully architecting the search, graphical user interfaces for the data retrieval parts of NGDS, users must be able to get the desired results without having to understand the inner working of the system. In order to facilitate this, both the data and metadata models have to be well thought out.

Good metadata is essential to the success of the system and obtaining this information must be made as simple as possible, ideally seamlessly integrated into workflow such that the user is hardly aware that they are 'creating metadata'. Automating metadata creation wherever possible is part of this philosophy. On the other hand, it is also clear that users must be involved in the process to detect errors and omissions. While some metadata can be generated and validated automatically, users should remain involved to ensure the results are both complete and accurate.

CONCLUSION

With the assistance of geothermal domain experts, metadata and Tier 3 data specifications and information exchanges are currently in production mode (see http://geothermaldata.org). Current NGDS development in progress as of the date of this paper is focused on implementation of a portal application for searching all NGDS resources, and 'Node-in-a-Box' software that will simplify deployment of new NGDS nodes and their incorporation into the system. It is a highly complex problem involving both technology and human components. NGDS teams working from both the user centric approach and the data provider side are making progress. The ultimate indicator of success will be known when the NGDS system goes live in early 2014 and real world usage patterns emerge.

A greater geothermal community of practice will emerge as data needs are addressed and the value of an interoperable network is demonstrated. Only then will geothermal community fully engage.

REFERENCES

Adobe Systems Incorporated, (2005), "XMP Specification: Adobe Systems Inc.," San Jose, CA, 112p. (Accessed at http://partners.adobe.com/public/ developer/en/ xmp/sdk/XMPspecification.pdf)

Adobe Systems Incorporated, (1999), "ISBN 0-201-37922-8," (Accessed at http://partners.adobe.com/public/developer/en/ps/PLRM.pdf)

Bloomberg, New Energy Finance, (2012-10-08), "Geothermal Market Outlook, Achilles heels: resource risk and land rights".

Clark, R., Kuhmuench, C. and Richard, S. (2013), "Developing the National Geothermal Data System; Adoption of CKAN for Domestic & International data deployment," Proceedings, Thirty-

Eighth Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, February 11-13, 2013, SGP-TR-198

Cox, S. editor, (2010), "Geographic Information: Observations and Measurements," OGC Abstract Specification Topic 20, v2.0.0,":Open Geospatial Consortium Inc., Document 10-004r3, same as ISO 19156. (accessed at http://portal.opengeospatial.org/files/?artifact_id=41579 2013-01-24).

De La Beaujardiere, J. editor, (2006), "OpenGIS® Web Map Server Implementation Specification," v. 1.3.0: Open Geospatial Consortium Inc., Document 06-042.

Dublin Core Initiative, (2008) "Dublin core Metadata Element Set, Version 1.1: Dublin Core Metadata Initiative," (accessed at http://dublincore.org/documents/dces/).

ISO 19115/Cor.1. (2006), "Geographic information – Metadata," Technical Corrigendum.

ISO 19115. (2003), "Geographic information – Metadata".

Kottman, C. and Reed, C., editors, (2009), "The OpenGIS® Abstract Specification Topic 5: Features, v. 5.0," Open Geospatial Consortium Inc., Document 08-126.

Portele, C. editor, (2007), "OpenGIS® Geography Markup Language (GML)," *Encoding Standard*, **3.2.1**: Open Geospatial Consortium Inc., Document 07-036.

Richard, S. M. and CGI Interoperability Working Group, (2007), "GeoSciML – A GML Application for Geoscience Information Interchange," U. S. Geological Survey Open-file Report 2007-1285, p. 47-59. (accessed at http://pubs.usgs.gov/of/2007/1285/pdf/Richard.pdf, 2013-01-24)

United States of America, Senate and House of Representatives, (2009), "American Recovery and Reinvestment Act of 2009," (accessed at http://www.gpo.gov/ fdsys/pkg/BILLS-111hr1enr/pdf/BILLS-111hr1enr.pdf).

USGIN (2011a), "Metadata Recommendations for Geoscience Resources," U.S. Geoscience Information Network, document USGIN2011-002 accessible at http://repository.usgin.org/uri_gin/usgin/dlio/335

USGIN (2011b), "Use of ISO metadata specifications to describe geoscience information resources," U.S Geoscience Information Network, document USGIN2010-09, accessible at http://repository.usgin.org/uri_gin/usgin/ dlio/337.

van den Brink, L., Portele, C. and Vretanos, P. editors, (2011), "Geography Markup Language (GML) simple features profile (with technical note)," v2.0: Open Geospatial Consortium Inc., Document 10-100r3.

Vretanos, P. A. editor, (2005), "Web Feature Service Implementation Specification v. 1.1.0," Open Geospatial Consortium Inc., Document 04-094.

Weers, J. and Anderson A. (2013), "DOE Geothermal Data Repository (GDR): Fueling Innovation and Adoption by Sharing Data on the DOE Geothermal Data Repository," Proceedings, Thirty-Eighth Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, February 11-13, 2013, SGP-TR-198

PROCEEDINGS, Thirty-Eighth Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 11-13, 2013
SGP-TR-198

NGDS USER CENTERED DESIGN
MEETING THE NEEDS OF THE GEOTHERMAL COMMUNITY

Suzanne Boyd[1], Xianjun Sam Zheng[2], Kim Patten[3], and Harold Blackman[4]

Anthro-Tech, Inc[1], Siemens Corporation[2], Arizona Geological Survey[3], and Boise State University[4]
Boise, Idaho, 83706, USA
e-mail: haroldblackman@boisestate.edu

ABSTRACT

In order to ensure the widest and greatest utility of IT and software projects designed for geothermal reservoir engineering the full consideration of end users' task and workflow needs must be evaluated. This paper describes the user-centered design (UCD) approach taken in the development of a user interface (UI) solution for the National Geothermal Data System (NGDS). This development process has been researched based, highly collaborative, and incorporates state-of-the-art practices to ensure a quality user experience. Work is continuing on the interface, including future usability tests to further refine the interfaces as the overall system is developed.

BACKROUND AND OBJECTIVES

The National Geothermal Data System (NGDS) is a distributed, interoperable network of data repositories and state geological service providers from across all fifty states and the nation's leading academic geothermal centers. This project is sponsored by the United States Department of Energy's Geothermal Technologies Program under Award # DE-EE0001120-#004.

The ultimate goal of the NGDS is to support the discovery and generation of geothermal sources of energy. NGDS will provide critical geothermal-related data that can be easily accessed to:

- Help companies be more (cost and time) effective in exploration, development and usage of geothermal energy
- Support a knowledge repository and archive for geothermal data, lessons learned, reports.
- Advance earth sciences by identifying gaps in our knowledge and informing new knowledge
- Increase public awareness in geothermal energy

These goals can only be accomplished if NGDS provides a quality user experience, and is widely adopted by users in the geothermal community.

There are three targeted user communities for NGDS, and each user group has different goals, needs, and tasks when interacting with NGDS.

- **Data providers** who will expose information to NGDS through standardized, internet-accessible interfaces and interface formats.
- **End users or data consumers** who will utilize NGDS to access data to support their work in geothermal energy exploration and development.
- **Application developers** who will build applications that utilize the data in NGDS, and make it easier for end-users to interact with the system.

This paper focuses on the end users / data consumers.

DESIGN APPROACH

User-centered design

To ensure NGDS is usable and meets users' needs, we followed a systematic and data-driven design philosophy called user-centered design (UCD) (also called the human-centred design process – ISO 9241-210:2010). UCD is based on three underlying principles (Gould and Lewis 1985):

- **Early focus on users and tasks.** Understanding the users of a system, the tasks they perform, and the environment in which they work. Best practices advocate for direct contact between users and the design team throughout the development life cycle.
- **Empirical measurement of product usage**. Emphasis is placed on behavioral measurements such as ease of use early in the design process, through the development and testing of prototypes with actual users.
- **Iterative design.** A product is designed, modified, and tested repeatedly. True iterative design allows the team to shape the product

through a process of design, test, redesign and retest.

The UCD approach ensures usable systems resulting in positive customer experiences, reduced development time, minimized risk for expensive changes to the system post-launch, increased user adoption and an enhanced reputation for the organization responsible for the system (Bias and Mayhew, 2005).

Multi-disciplinary team

Complex systems such as NGDS are best designed and developed by multi-disciplinary teams (Mayhew, 2008). For this project, the Department of Energy brought together experts in geothermal development and production, software development, user-centered design and project management, and outlined a plan that leverages respective skill sets and areas of expertise.

Designing NGDS

We outlined various methods to inform the design of NGDS. Specifically, the team organized:

- A domain committee to gather data requirements and standardize interchange formats for commonly produced data;
- Interviews with expert users that uncovered end-user goals, needs and design requirements;
- Design concept and wireframes to visualize the user experience early on and test the system with end users;
- Usability testing sessions at the Geothermal Resources Council (GRC) annual conference in Reno, NV to further iterate the design concept.
- Use case development to specify how NGDS will support users' key tasks.

The remainder of the project will incorporate additional usability testing throughout the development cycle to continually improve the user experience of NGDS.

DEFINING THE END USERS

Team members worked together define the end-user groups. Specifically, project partners documented NGDS end users' key characteristics, including their domain knowledge, typical organizations they work for, anticipated frequency of use, motivations, goals and tasks for using NGDS. This initial exercise generated well over 26 different user groups.

A classic user experience mistake is to design for every possible user type. Such designs often suffer from busy and overwhelming interfaces that have too many features that too few users care about. To avoid such an experience, the UCD team further analyzed the 26 user groups and clustered commonalities in behaviors, attitudes and motivations.

Based on this analysis, seven key user groups emerged, each with distinct goals for the NGDS. These include:

- **Industry representatives** who focus on the geothermal exploration and development process
- **Researchers** who want to generate scientific knowledge or study a specific area of interest
- **Federal and state agencies** who make land and resource management assessments
- **Legislators** who make policy decisions that impact geothermal exploration and development
- **Interested public** who want to understand implications of local geothermal energy sources
- **Educators and students** who teach, learn or explore the education pipeline for careers in the geothermal energy industry.
- **Financial investors** who determine whether or not invest in a potential geothermal site.

Figure 1 outlines the user profiles for NGDS.

User Personas

Based on user research results, we developed user personas for NGDS. User personas are archetypal representations of users based on data (Cooper, 1999). They are compelling communication tools because they put a personal face on otherwise abstract data about users, and help the development team focus on the real end user, rather than making design decisions based on assumptions. Figure 2 illustrates one of the personas with goals, needs, and desires explicitly spelled out.

EXPERT INTERVIEWS

Goals

We conducted interviews with nine geothermal experts representing the key user groups. The goals of the interviews were to:

- Gather information about the people who use the system; validate/refine the user personas.
- Understand users' current processes and workflow for gathering data when prospecting potential geothermal sites.
- Identify the types of data end users look for and which data is most important for their specific goals.
- Find out where users currently go to locate data.
- Understand users' processes for verifying and analyzing the data.

The interviews were conducted by phone and utilized screen sharing technology. This allowed participants

Ad Hoc User Profiles | NATIONAL GEOTHERMAL DATA SYSTEM

	INDUSTRY	RESEARCHERS	AGENCIES	LEGISLATORS	INTERESTED PUBLIC	EDUCATORS & STUDENTS	FINANCIAL & INVESTOR COMMUNITIES
	Industry representatives or consultants who focus on the exploration and development process • Geologist / geophysist • Engineer • EIS writer or scientist • GIS specialist • Archeologists • Environmental consultants (working on behalf of industry)	Applied researchers and academic researchers • Professor • Research assistant • Geoscientists • Engineers • May work with or for industry	Federal, state, local government organizations • Permit issuer • Land use planner • Policy maker	• Public policy analyst • Lobbyist	People who care about geothermal energy and the environment in a specific area of interest • Homeowner • Technical writer • Environmental interest groups • People who care about the planet	• Elementary school teacher • High school teacher • Undergraduate educator • Graduate students • Students • Job seekers	• Investor • Consultant working on behalf of investor • Companies that do the financial analysis for investors • Investment counselors
MOTIVATION	Determine geothermal potential for a specific site.	Generate scientific knowledge. Scientifically characterize a specific geologic area to contribute to the field's body of knowledge.	Guide exploration and development by establishing, evaluating and modifying policies, rules and regulation that impact land and natural resource use.	Make informed and data-driven policy decisions that impact geothermal exploration, development, land use etc. (examples: endangered species, land withdrawals - wilderness, historical value etc.)	Understand how geothermal activites affect me, my community and the environment.	Find data to inform course work or research (thesis). Understand who the players in the industry are and find job opportunities.	Determine whether or not to invest in geothermal exploration or development. Get an independent / neutral assessment.
DATA USAGE & CONTRIBUTION	Read and potentially contribute data (depends on funding source)	Highest	Read and contribute data (authoritative source for land related information)	Read	Read	Read	Read
EXAMPLES	Ormat, US Geothermal Inc	Universities, National Labs	Federal Bureau of Land Management, National Science Foundation, local agencies that issues permits, Department(s) of Natural Resources, Division of Minerals, State Geological Surveys		Resident of a town that has geothermal power plants.	Graduate student in geological sciences. High school science teacher	Land/mineral development organizations interested in geothermal energy investments. Geothermex
DOMAIN KNOWLEDGE	High	Highest	Medium to low	Medium to low	Low	Medium to low	Medium to low
VISIT FREQUENCY	Frequently (Bi-weekly / Monthly)	Frequently	Periodically	Rarely	Rarely	Periodically	Periodically to rarely
PRIORITY TO NGDS	High	High	Moderate	Moderate	Moderate to low	Moderate to low	Moderate to low

NGDS AD HOC USER PROFILES 5 | 2012

Figure 1: User profiles for NGDS

Meet Matthew!

THE END USER/DATA CONSUMER

Matthew Parker

"My intent when conducting research is to understand the geothermal potential in a specific area for future investments."

Age 48
Profession VP of Resource Management
Location Salt Lake City, Utah

Goals:

Efficiently locate and synthesize quality data to enable my team to produce reports which will provide geotheormal information for specific geographical locations.

Easily share data securely with clients, agencies, and repositories.

Methodology:

- Utilize resources to understand potential of sited locations.

- Gather data such as a well location/ size, geyser temperatures, permeable/ impermeable rocks, property logistics, mineral makeup, etc to determine economic value of a site.

- Verify quality of data.

Resources:

State Bureau of Mines
Agency Websites
Universities
Purchased Propriety Data

Tools:

Coordinate Systems
Numerical Predictive Simulator
Maps

Frustrations:

- Finding data is time consuming and lacks a sense of direction.

- Propierty data should not be so difficult or expensive to obtain from sources.

- Georeferencing is poorly supported by a lack of proper land zoning.

Desires:

- A map system that holistically provides organized quality data of potential geothermal sites.

- NGDS provides an incentive for data providers to contribute data for a lower premium price.

Figure 2: End User Persona for NGDS

to show the facilitator their computer systems, applications and software used for their work, and file structures used to organize data sets.

Interview Findings and Implications for NGDS Design

The expert interviews provided key insights into current work practices and pain points in the data gathering process. We learned that:

- Geothermal exploration and development is highly data-driven process. Gathering quality data is the most cost-effective and efficient way to reduce risk and predict favorability of a specific geothermal site.
- Users value all types of data. While there are key data types that are higher priority, users emphasized that all data, ranging the spectrum of geophysical, geochemistry, well data and more are valuable.
- The search for data is typically geographically oriented. For industry representatives who focus on geothermal exploration it is pertinent that they can visually explore what data exists in a specific geographic area.
- The data gathering process happens over a long period of time, is inefficient, collaborative, and dependent on using many distributed systems.
- Users utilize various approaches to validate the data such as getting expert or peer opinions, reading published research, getting direct measurements, and triangulating data with other sources.
- Users rely on proprietary or third party software to do the analysis. Industry representatives, researchers and financial investors said that they use GEOSOFT, ARC GIS, and other proprietary software to manipulate and analyze data.

Implications for design

The user research findings informed several design requirements and helped the team prioritize features for NGDS. A strong unifying message was the desire to have a "one-stop system." At this point in the program it is not certain how many of the features will be incorporated due to resource constraints. Other key implications for design included:

1. Focusing the core functionality of NGDS on data discovery and validation, and deprioritizing features that focus on data analysis and manipulation. Users indicated that they currently use a variety of proprietary software to analyze, manipulate and visualize data and that much time and frustration goes into discovering the data. NGDS can provide the most value by supporting data discovery and validation. In addition, the system must allow users to export data in a variety of formats that allow for streamlined usage in their own analysis tools.

2. Making geographic search and discovery the most prominent feature of the system. This underscores the importance of continuing to geocode data that will be exposed in NGDS.

3. Supporting multiple ways for users to find and narrow down data set results. This includes providing the ability to do:
    - Simple text searches
    - Geographic searches
    - Narrowing results by different criteria such as topic, user rating, data format and more
    - Sorting data set results by specific criteria

4. Providing the ability to let users rate and comment on data sets to and exposing ratings to all users to support finding and validating data based on peer ratings.

5. Helping users save searches and "favorite" specific data sets so they can retrieve them easily across multiple sessions.

6. Allowing users to share data sets with colleagues to support the collaborative nature of the data collection process.

7. Ensuring multiple formats for each data set so users can export data and utilize it in their analysis tools.

8. Providing key metadata for each data set that helps users determine the validity of the data. These include information about source, author, dates, instruments, related data, and peer rating of the data set.

9. Relying on common web-based standards for the user interface. Users are already familiar with other types of web based systems such as Google, Bing, common e-commerce or library web sites. NGDS can leverage standardized interactions to minimize users' learning curve with the system.

10. Giving users an option to "follow" certain datasets and signing up for email alerts when new data that meets their criteria have been uploaded or changed in the system.

NGDS Design Concept

To visualize these design requirements and get feedback from prospective users, we prototyped the system. The prototype laid out key sections of the system, including:

- **The NGDS homepage** (Figure 3) which provides easy access to top users' tasks.
- **The map** (Figure 4) where users can easily find geocoded datasets by searching geographically or zooming into an area on the map. Map layers let users visualize thematic data such as geological features, environmental data, land use, ownership and more.
- **The library** where users can access all datasets (geocoded or not) by browsing, searching or both.
- **A resources section** which provides links to tools, applications and web sites that help users analyze their data. In addition, users can suggest additional information .
- **The contribute section** where users can contribute data to NGDS.

Moreover, the prototype reflected key features in the system such as interactions that detail how a user can rate a data source, save a search or add a data set to their favorites.

Figure 3: NGDS homepage

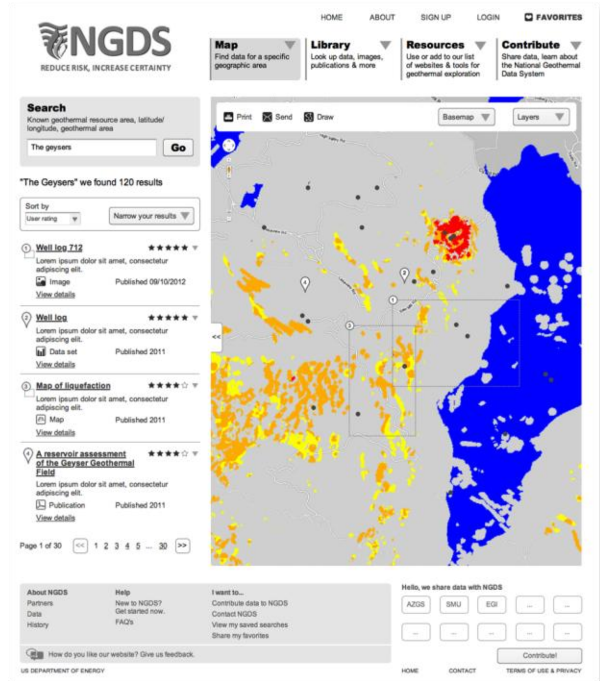


Figure 4: NGDS Map

Usability Testing the NGDS design at GRC

To evaluate the prototype with target audiences, we prepared for usability studies at the Geothermal Resources Council Annual Meeting and Geothermal Energy Association Annual Expo in Reno, NV. Our goal was to collect behavior and perceptual data in a short amount of time, and refine our design concept before finalizing the design requirements and starting system development.

A total of 18 users representing key user groups took part in the study. At the outset of each study, we conducted a basic user interview to further understand participants' role, characteristics, and goals for geothermal exploration. This information was used to further validate the user profiles.

After the pre-study interview, we asked each participant about a recent project they worked on, and what types of data they sought to support the project. We then asked them to use the NDGS system to explore this recent scenario. Following this open-ended scenario we provided users with key predeveloped scenarios and observed as they used the NGDS prototyped system. We captured a variety of notes including their think-a-loud protocol, pathways in the system, errors and task completion.

In this study we learned that:

- Participants worked with the prototype easily. In particular they easily navigated to the correct sections and used a combination of searching, browsing, and sorting to work with results lists. In addition, key features such as sharing datasets with others, selecting favorite data sets and downloading a data set were easy to use.
- Participants were primarily drawn to use the map to find respective data sets and examine potential for a specific geographic area. In most cases, participants zoomed to a geographic area of their interest and expected related datasets to display in the results list.
- Participants have high expectations for the ease of use of the system. They stated that NGDS must save time and money in comparison to the data collection process they are currently practicing.

In addition, the study findings suggested that there were several areas for improvement in the user experience including:

- More controls to manipulate the map such as the ability to draw a precise shape (polygon) on the map and refine results within that area.
- Ability to track how many people had downloaded a specific dataset.
- More file formats for for exporting datasets so they can work with them easily in their proprietary systems.
- Making search more discoverable.
- Auto-suggest or complete search terms.

The design team iterated the design concept to address found usability problems and meet users expectations.

NGDS Use Cases

To ensure the NGDS design concept and user requirements are fully understood by the software development team, we developed use cases. Use case modeling is a common methodology used in software development to identify, clarify, and organize system requirements. A typical use case is a list of steps that defines interactions between a user and a system, to achieve a specific goal (Constantine & Lockwood,1999; Jacobson et al, 1992).

Figure 5 illustrates the different types of users and key goals for the NGDS system. For instance, end-user/data consumers are using NGDS to gather data, validate data, and analyze the data after exporting it. Data submitters are as important as data consumers. The two major use cases for data submitter are batch import of dataset files, and create metadata record through a form.

We developed more than 50 use cases for NGDS. Each use case captures the important user requirements and steps for different users interacting with the NGDS and provides a foundation for the NGDS system requirement specification document.

Initial Website

We constructed a beta website and search engine to provide users with a basic mechanism for finding data while the UCD process is in process. The site is available at www.geothermaldata.org and includes resources for data users (i.e. the area that will transition to the map and library described above), data submitters (i.e. the contribute section), and application developers (i.e. the resources section). The site currently includes a beta search-engine style, map-based search application for data discovery. The search application includes much of the functionality end users identified, including the ability to search by area (bounding box) or by key word, and to access the data in a variety of formats, including Excel, ArcGIS shape files, and Open Geospatial Consortium (OGC) Web Map Services (WMS) or Web Feature Services (WFS). The site and search functionality will be refined as the UCD process concludes and the software is developed.

Next steps

The team is refining the NGDS design concept and testing extended functionality at the 2013 Stanford Geothermal Workshop. In parallel we are prioritizing features and preparing for the agile development process. The UCD and completed NGDS system is scheduled to launch in-full during the Fall of 2013.

REFERENCES

Bias, Randolph G. and Mayhew, Deborah J. Cost-Justifying Usability: An Update for the Internet Age.

Constantine, L. L., and Lockwood, L. A. D. (1999) Software for Use: A Practical Guide to the Essential Models and Methods of Usage-Centered Design. Reading, MA: Addison-Wesley.

Cooper, A. (1999). The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams Press: Indianapolis, IN.

Gould, John D. and Jones, Lewis. (1985). “Designing for usability: key principles and what designers think. Communications of the ACM, 28, 300 -311.

Jacobson, I, Christerson M., Jonsson P., Övergaard G. (1992) Object-Oriented Software Engineering - A Use Case Driven Approach, Addison-Wesley.

Mayhew, Deborah J. User Experience Design: The Evolution of a Multi-Disciplinary Approach. Journal of Usability Sudies, Volume 3, Issue 3. May

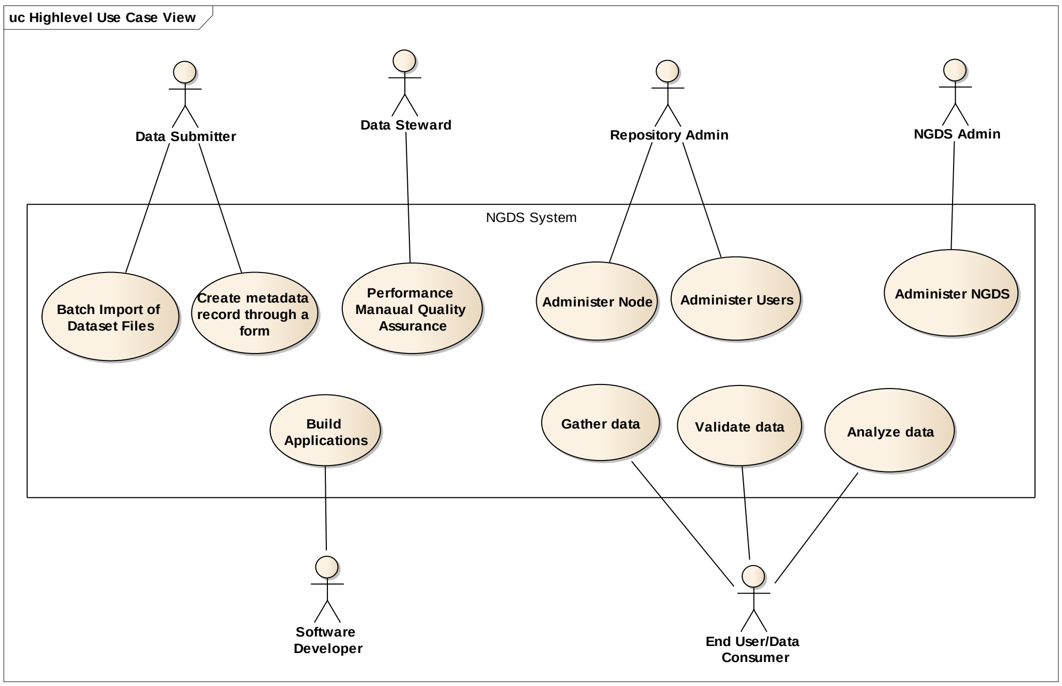


Figure 5: High Level Use Cases for NGDS

PROCEEDINGS, Thirty-Eighth Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 11-13, 2013
SGP-TR-198

DEVELOPING THE NATIONAL GEOTHERMAL DATA SYSTEM ADOPTION OF CKAN FOR DOMESTIC & INTERNATIONAL DATA DEPLOYMENT

Ryan J. Clark[1], Christoph Kuhmuench[2], Stephen M. Richard[1]

[1]Arizona Geological Survey,
416 W. Congress St. Tucson, Arizona, 85701
e-mail: ryan.clark@azgs.com, steve.richard@azgs.az.gov
[2]Siemens Corporate Research
755 College Road East
Princeton, NJ, 08540, USA
e-mail: christoph.kuhmuench@siemens.com

ABSTRACT

The National Geothermal Data System (NGDS) Design and Testing Team is developing NGDS software currently referred to as the “NGDS Node-In-A-Box”. The software targets organizations or individuals who wish to host at least one of the following:

- an online repository containing resources for the NGDS;
- an online site for creating metadata to register resources with the NGDS
- NDGS-conformant Web APIs that enable access to NGDS data (e.g., WMS, WFS, WCS);
- NDGS-conformant Web APIs that support discovery of NGDS resources via catalog service (e.g. CSW)
- a web site that supports discovery and understanding of NGDS resources

A number of different frameworks for development of this online application were reviewed. The NGDS Design and Testing Team determined to use CKAN (http://ckan.org/), because it provides the closest match between out of the box functionality and NGDS node-in-a-box requirements.

To achieve the NGDS vision and goals, this software development project has been initiated to provide NGDS data consumers with a highly functional interface to access the system, and to ease the burden on data providers who wish to publish data in the system. It is important to note that this software package constitutes a reference implementation. The NGDS software is based on open standards, which means other server software can make resources available, and other client applications can utilize NGDS data.

A number of international organizations have expressed interest in the NGDS approach to data access. The CKAN node implementation can provide a simple path for deploying this technology in other settings.

Introduction

Significant growth in the geothermal energy contribution to the national energy portfolio requires reducing the risk and cost of defining resources, characterizing new classes of larger energy resources, and optimizing management and expansion of exploited geothermal fields. Achievement of these objectives largely depends on access to accurate geoscience information. Today, much of the existing information is inaccessible or difficult to locate and access. A DOE report by Deloitte (2008, pg. 27) concluded that "A study conducted in 2000 for NREL (Entingh, D., 2000) revealed that over a 25-year period, numerous geothermal research efforts were conducted with state and federal funding. Despite these efforts, the analysis and information contained in those research documents is difficult to access without significant research efforts. That same study cited that much geothermal resource attribute data also exists but is distributed among numerous locations and often stored in boxes, without any data index or organization."

Most of this data is unstructured and consists of documents such as publications, notes, images, or figures. There is also a rapidly growing body of structured data available that allows for new understanding or industrial exploitation of geothermal systems.

The NGDS is a distributed, interoperable network of data repositories and state geological service providers from across all fifty states and the nation’s leading academic geothermal centers. This project is sponsored by the United States Department of Energy's Geothermal Technologies Program under Award # DE-EE0001120-#004. The ultimate goal of the NGDS is to support the discovery and generation of geothermal sources of energy. NGDS will provide critical geothermal-related data that can be easily accessed to:

- Help companies be more (cost and time) effective in exploration, development and usage of geothermal energy
- Support a knowledge repository and archive for geothermal data, lessons learned, reports.
- Advance earth sciences by identifying gaps in our knowledge and informing new knowledge.
- Increase public awareness in geothermal energy

These goals can only be accomplished if NGDS provides a quality user experience, and is widely adopted by users in the geothermal community. There are three targeted user communities for NGDS, and each user group has different goals, needs, and tasks when interacting with NGDS.

- Data providers who will expose information to NGDS through standardized, internet-accessible interfaces and interface formats.
- End users or data consumers who will utilize NGDS to access data to support their work in geothermal energy exploration and development.
- Application developers who will build applications that utilize the data in NGDS, and make it easier for end-users to interact with the system."

To facilitate utilization of existing data by the geothermal research and development community, the National Geothermal Data System project is developing three types of functionality: (1) data presentation and analysis applications, (2) applications for efficient searching across data repositories, (3) metadata repositories (catalogs) describing geothermal data resources and supporting efficient search, and (4) repositories for geothermal data resources. This system is a full scale deployment of the distributed geoscience information network concepts developed by the USGIN project (http://usgin.org).

Figure 1 depicts the deployment and interaction of the three application types. Data repositories (3) enable storage of research data with associated metadata to support data discovery, evaluation, and utilization. Presentation and analysis applications (1) are specifically designed to make use of structured data. They allow for statistical analysis as well as for human-user friendly presentation of structured data. The Lookup service (2) acts as glue between the analysis applications and the repositories. It allows for searching across multiple repositories. The lookup service includes a Portal application that provides a user interface for searching for and reviewing data. It allows for simple data exploration, addition of annotations on data, and data downloading.

The NGDS is designed based on a service-oriented approach using open standards to support data access by a wide variety of software applications, promote novel approaches to data analysis, and foster the development of tools by third parties. The key protocols for current deployment are the Open Geospatial Consortium (OGC) geographic data protocols (WFS, WMS, and CSW) and the underlying standard Web protocols (HTTP, FTP). HTTP provides functionality for basic file access, while the OGC protocols are used to search across the repositories, access data and to display it in appropriate form.

In this paper we outline the design of a system that enables storing and annotating data with geographic reference, searching data across multiple data repositories and use of this data in simple data browsing tools. As a basis for this system we are using CKAN, an open source, python-based framework application that brings many features required for the file-repository and search function, as well as extensions that implement some basic geographic data capabilities. An important aspect of our system is the implementation of an intuitive, functional user interface that supports search across the grid of data repositories.

The next section discusses example user scenarios. We then analyze the gap between NGDS requirements and the off-the-shelf features provided by CKAN, and finish with a high level view of the node-in-a-box implementation using CKAN.

User Scenario: Geothermal Data Repository

A user has a scanned report on reservoir modeling at a geothermal development prospect, based on a dataset of temperature, lithologic and geophysical logging of 200 boreholes. The report is to be registered as a document resource for the NGDS, the maximum temperatures measured for each well are to be published as a data service and the scanned logs are to be registered a resources indexed to the borehole from which they were obtained

Figure 2 shows the services, features, data instances and metadata that would be involved in describing the data resources from this collection. Services include the metadata catalog service, and data services for well headers (one record for each well), well log

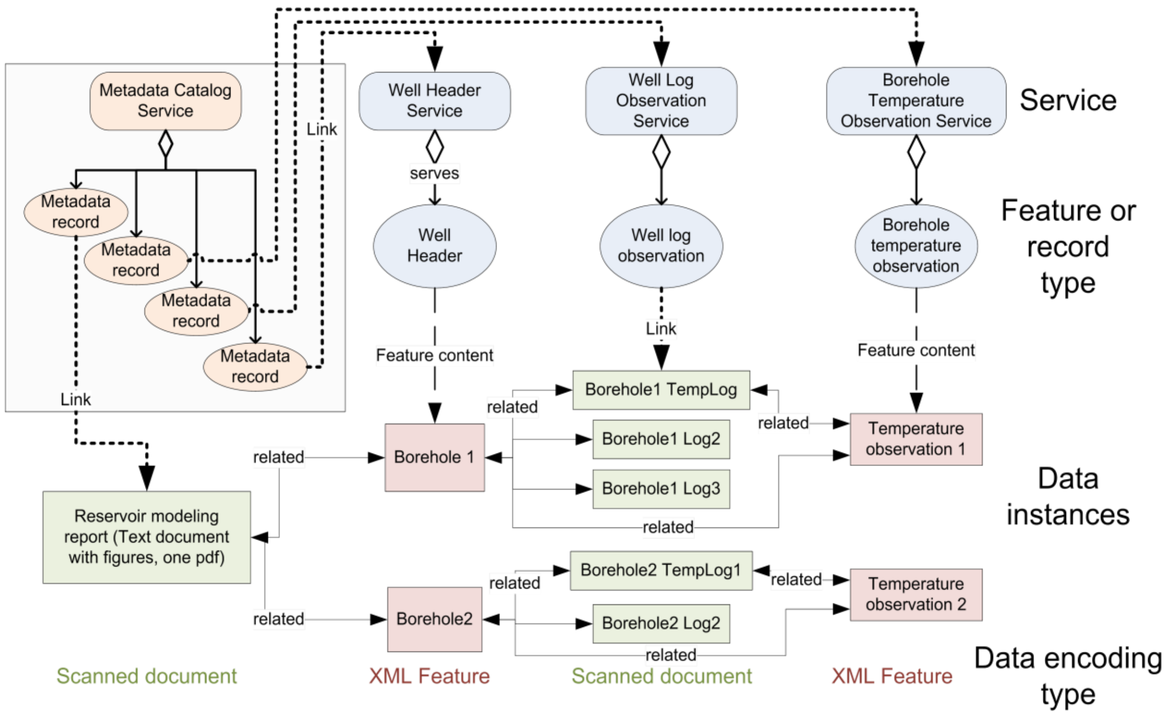


Figure 1. Services, features, and content associated with a reservoir modeling report based on a collection of well log data.

observations (one record for each well log), and borehole temperature observations (one record for each temperature measurement, 0 to many measurements per well at different times or depths). Metadata would include a metadata record describing the scanned report document, with a link to the web location from which the report can be obtained. Ideally, the metadata for the report would also include links to related well header records for that report, and the

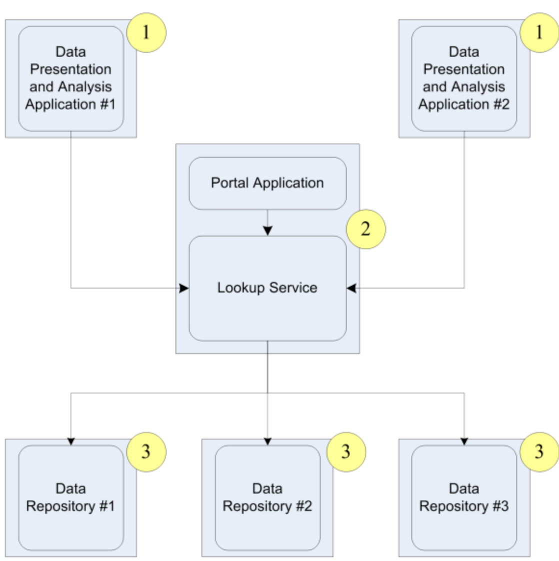


Figure 2. Geothermal Repositories, Lookup Services and Presentation Applications.

well header records would include links to the related report that discusses results from the well. Each well header record also includes links to related well logs and temperature observations from the well.

The fields in the feature or record types are defined by NGDS content models. Users access data and metadata through service interfaces that provide responses in standard XML interchange formats that implement the NGDS content models. NGDS services are defined by OGC specifications (CSW 2.0.2, WMS 1.3.0 and WFS 1.1.1).

The described dataset could be incorporated into the NGDS at several levels. The simplest would be to create a metadata record for the scanned reservoir report and get the scanned report accessible online. The summarized temperature measurements would be included as tables in the paper. This is referred to as Tier 1 data delivery.

Tier 2 data delivery would consist of registering any data tables (well locations, temperature measurement points) that are included in the original data as data sets using the schema from the original data. The dataset could be published via a web service; this does not result in an interoperable data service, but still provides better access to the structured data. Metadata for the services would have to include a data dictionary.

Table 1. Open source frameworks reviewed for project use.

Project	URL	Description
CKAN	http://ckan.org/	Open source project, maintained by Open Knowledge Foundation
GeoNode	http://geonode.org/	platform for the management and publication of geospatial data; mature open source project
DSpace	http://www.dspace.org/	turnkey institutional repository application; mature open source project, maintained under the stewardship of DuraSpace, a not-for-profit organization

The preferred data publication approach is to load data for well locations into the NGDS well header content model, information about the well logs into NGDS well log observation content model, and the temperature measurement data into the NGDS borehole temperature content model, and publish the data using OGC WMS (for map images showing the locations of wells), and WFS (to provide structured data describing the wells, logs, and temperature measurements). This is Level 3 data delivery, and requires data integration (into the standard content model and interchange format) by the data provider, but makes data acquisition and utilization simpler for the data consumer.

User requirements for a data provider in this scenario include a metadata editor to create NGDS-compliant metadata, a publicly accessible web location to place the file for Level 1 access. For level 2 or level 3, a WMS and WFS server software installation must also be available; the user must also have the necessary permissions to copy files or data into the publicly accessible file system or database supporting the OGC service server.

Example User Scenario: Data Discovery

Metadata and data resources can be published from any node in the system, but an important system requirement is that a data consumer can search and access these various data resources from a single point of entry. This requirement will be met by providing a system node that aggregates metadata from the various publishing nodes, and provides a map-based search interface. Searching via this portal application will locate registered resources anywhere in the system; the metadata for the resources will include necessary information to access the resource without the user having to go to a new web location. This scenario implies two sets of requirements. One set encompasses the functions necessary to configure an aggregating node to harvest metadata from other nodes, and the functions necessary to execute the harvest—getting the records, avoiding duplication of metadata, and validating acquired records. The second set of requirements encompass the search application functions, including: maintaining search indexes; displaying search results as extents on a map, in summary lists, and the full metadata record formatted for reading on screen; simple data browsing capabilities for data evaluation; and data download of full or filtered datasets.

The two scenarios discussed here are fundamental to system operation, but do not include a variety of other use cases for system administration, monitoring, and maintenance. These are not considered here.

Framework Candidates

The first major choice facing the development team was to select a development framework. AZGS developers had done significant work on components to support the AASG Geothermal Data project (AASG package), and Siemens developers had done significant work developing a software package to support the Southern Methodist University Heat Flow Database project (GTDA package). Several other software projects were identified that provided some of the capabilities required for the NGDS node-in-a-box concept; these are summarized in Table 1.

Architectural Drivers

A major consideration in the evaluation is the long-term viability of the node-in-a-box application developed for NGDS. Building on an existing, active and widely used open-source project is considered a major asset to assure long term viability of the application.

Other significant factors that were considered of high value in the development framework include: 1) adaptability of the user interface to be compatible with an independently developed user experience concept for NGDS users; 2) ease of extensibility, with a plug-in architecture that allows addition of functionality without having to modify the core codebase; 3) support for geographic data and map-based search and data browsing; 4) support for administrative activities like user management, access control, and activity logging.

Gap Analysis

The five candidate packages (AASG, GTDA, DSpace, CKAN, and GeoNode) were evaluated in a gap analysis comparing out of the box capabilities against a compiled set of requirements for the NGDS node software. The analysis considered feature completeness, based on how many of the required features are supported by the framework, and the difficulty of implementing missing functionality.

The result of this analysis placed CKAN as the best candidate framework. Decision to utilize CKAN was reinforced by its adoption by various government open data initiatives, and the responsiveness of the CKAN community to our requests for information and assistance.

CKAN

CKAN is a sophisticated application framework that provides many functions helping us to implement the application functions as laid out in the introduction. CKAN is written in Python and makes use of a variety of open source frameworks. The base framework is Pylons [http://www.pylonsproject.org/about/pylons], which itself is a combination of various open source frameworks integrated to form the basis for Web-based Enterprise-level applications.

The primary CKAN user scenario is data storage and management. Its core functionality consists of three features:

1. File storage,
2. Metadata management, and
3. Management of structured data.

In addition it offers a plug-in mechanism enabling developers to rapidly extend CKAN's core functionality. There are now a number of powerful extensions available supporting the development of productive data management sites. Most importantly, it provides an extension (ckanext-spatial) that supports geographic features as well as exposing metadata according to the OGC standard catalog service (CSW). Also, it is possible to modify or override CKAN's rather simple default UI in various ways.

Finally, CKAN implements typical housekeeping features (user management, logging, etc.) that are tedious to implement but crucial for the site's usability.

In the following we discuss these features and how we intend to extend CKAN with the remaining missing features we need for our data storage application.

Figure 3 shows the most important components of CKAN. The most elemental functions are grouped in the CKAN core component. A large number of these components provide infrastructure for the two core function blocks *CKAN file storage* and *CKAN Metadata management*. The remaining components inside the core provide basic essential services such as information persistence on the data management layer or session management on the service layer. Often these services are implemented using existing open source frameworks. All of these components expose a well-defined API, accessible in two ways: (1) a RESTful Web-service accessible to Web UI designers. (2) via an extension interface. Developers can use the second approach to add features to a component, or to "hook into" CKAN's core features. Extensions can inject callback functions that are called when certain activities are ongoing (e.g. upload of a new file). The following section describes some of the more important components depicted in Figure 3.

CKAN File Storage

The CKAN file storage feature provides the functionality to store, update, and delete files of arbitrary type. Stored files can be grouped into datasets and exposed under a single URL allowing external reference to them. Similar to a resource management system like Subversion (SVN), CKAN keeps track of the history of files. CKAN provides two file storage implementations: The first one maps the uploaded files to the file system of the underlying operating system. The second one allows for storing files on cloud drives. In both cases a Postgres database is used for persisting the state of the file store.

CKAN Metadata Management

The CKAN metadata feature provides the functionality to associate metadata entries to datasets. Out-of-the-box, CKAN implements a rather simple metadata concept that maintains very limited metadata content about the stored documents. However, CKAN is designed so that the default metadata content and UI can be extended with some Python programming.

The metadata can also be used to search for datasets. CKAN uses SOLR for indexing the metadata and provides an API for searching through that index. Its default UI provides a very simple search bar. Any dataset with metadata containing the entered key words will be returned as a list.

Structured Data

CKAN's file management feature allows for managing arbitrary files, which are handled as atomic blocks that cannot be searched internally or accessed in sub units. However for structured data, e.g. a table stored in a CSV file, it is useful to be able to return individual fields to the UI so that they can be viewed in an appropriate form (e.g. tables, or charts). CKAN's datastore extension takes care of this functionality. It allows for uploading structured documents into a database so that they can searched and search results be returned to the UI. In fact the datastore extension allows for returning such structured data as JSON objects that can be handled in an appropriate way by the UI.

Custom extensions are necessary to take full advantage of the datastore's features; these also require implementation of an appropriate UI. One required extension that must be developed is support for OGC services required by NGDS architecture.

Spatial data

Many data storage applications have to keep track of data associated with geospatial information, i.e. they are associated with areas on a map. CKAN comes with an extension that allows for associating geospatial information with datasets and provides a geospatial search feature, i.e. datasets can be searched on a map. The feature also works closely with map widgets that can be incorporated into the UI. However, there is no default UI for map-based search, and this will need to be developed by the project. The spatial extension also provides the capability of exposing metadata via the standard OGC catalog protocol (CSW) required by NGDS architecture.

Harvesting

Harvesting is an operation in which one node replicates metadata records from another node. The basic NGDS deployment plan is to harvest metadata from all NGDS nodes into a single aggregating node that will host the search application for the NGDS data access portal. CKAN provides an extension that enables harvest functionality. For our geothermal application this is a crucial feature. CKAN offers a UI for managing harvesting but the functionality can also be accessed via a command line tool.

Extensions via the Plugin Interface

CKAN provides an extension interface allowing developers to add functionality to the storage system. (see Figure 3). The extension interface provides for three features: (1) It allows new extensions to "hook" into CKAN's core components, (2) it allows new ex-

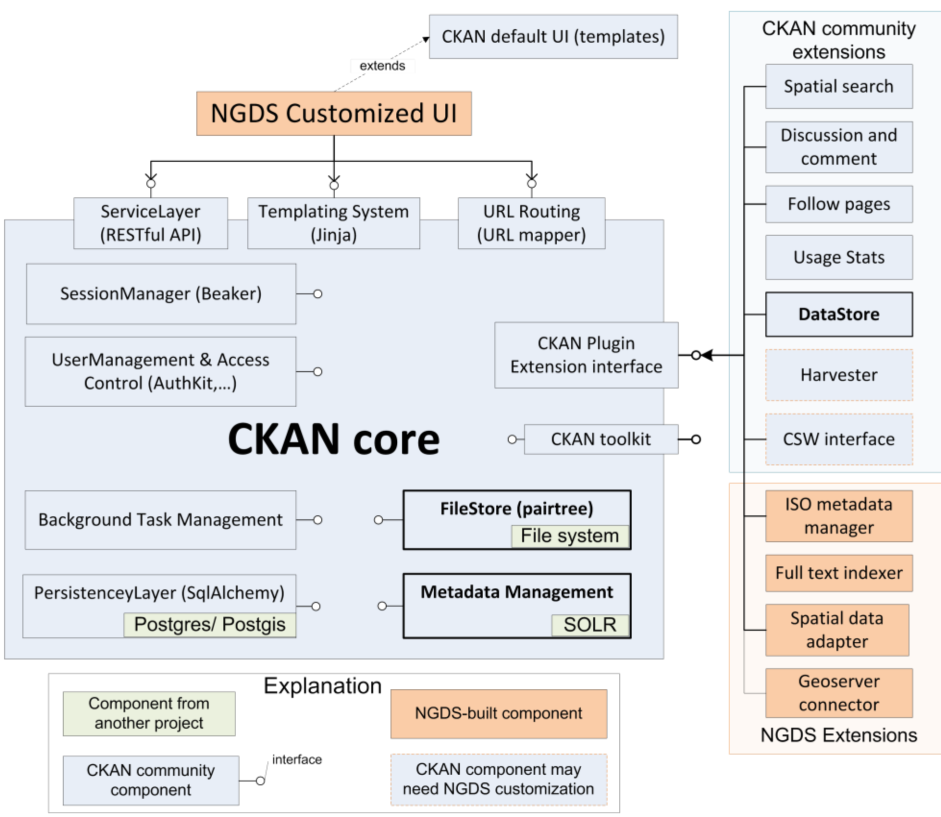


Figure 3. CKAN Architectural components and interfaces. Blue components are developed by the CKAN Community, orange components will need to be be built for NGDS requirements.

tensions to make use of the API of the system which is provided as a “toolkit”, and (3) it allows new extensions to make new “hooks” available.

Hooks are places within the component’s code at which extensions can register callback functions to be called when the relevant code segment is executed. This way an extension may add additional functionality, for example when a file is uploaded to the system. An extension can also provide hooks on its own so that other extensions can register callback functions. The functionality of the CKAN core features is aggregated in a toolkit that provides a well-defined, stable interface with a standardized obsolescence strategy. Using these three features developers can modify or extend all important steps in the data storage workflow.

Extension of the UI

CKAN provides a very basic out-of-the-box UI, mostly for educational purposes as an example to helps developers design UI’s dedicated to their specific application. CKAN incorporates the Jinja 2 HTML templating system (http://jinja.pocoo.org/) for user interface construction. Web interface applications can access CKAN’s core functionality via the CKAN RESTful API. CKAN’s URL routing system also allows mapping URLs to functions within a user extension for additional customization options. UI developers can use any of these three mechanisms to override CKAN’s default UI by providing a path to a UI overriding directory in CKANs configuration file. CKAN will search first in this directory for resources and only afterwards in the default directories.

Adaptation of CKAN for NGDS

CKAN offers many of the features required for NGDS but there are several adaptations and extensions that must be implemented. The simple default CKAN UI needs to be replaced with a UI following the NGDS look and feel and supports the NGDS-specific features. This NGDS user interface design has been developed using wireframes and user feedback, independently of the decision on the implementation software framework. The interface needs to incorporate a very flexible map widget that will be used to search for resources as well as visualize search results. The UI shall also realize faceted search capabilities to facilitate seamless combination of multiple search parameters to filter search results. The look and feel and functionality of the NGDS UI requires a ground-up UI implementation based on CKAN’s template system.

Other features not provided by CKAN or any of its extensions, but required for NGDS include:

1. Creation of NGDS metadata, including spatial extent
2. Uploading structured data with geospatial information
3. Consistency check for well-known structured data files
4. Providing OGC services for uploaded structured files with geo-spatial information
5. Full-text indexing of documents
6. Role-based right for uploading and publishing data
7. User feedback and rating of uploaded data

The most important ones are discussed in the following:

Handle structured data that includes geo-spatial columns

One basic extension we require is the capability to handle structured data that contains geospatial information. Note that the CKAN’s spatial extension can search spatially but it does not support uploading CSV files with geospatial information. It merely assumes that the data are already uploaded. In order to support this functionality we will have to add a new hook into the CKAN datastore and to implement a new extension. The following Figure 4 depicts how the components interact with each other.

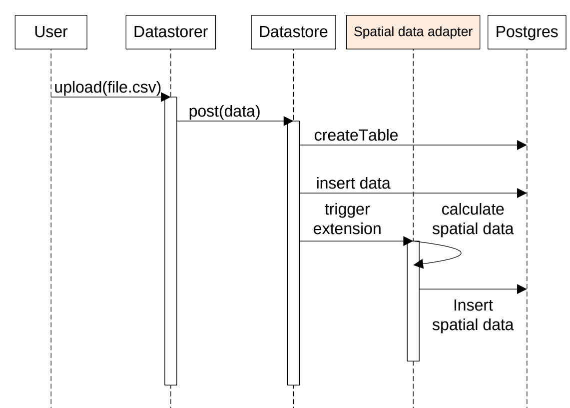


Figure 4: Message Sequence Chart handling structured data with spatial information. ‘Spatial data adapter’ is a custom component to be implemented by the project.

As before the CKAN datastorer and datastore will be used to upload structured data files and to create a table with the data in the Postgres database. The datastore will then trigger the new extension to create the column in the table containing the spatial information required by PostGIS (the FOSS geospatial extension for Postgres) for spatial operations.

Checking consistency of uploaded structured data files

When structured data that use a registered template format are uploaded, CKAN will validate the uploaded data for consistency with the template requirements. This operation can provide automatically calculated metadata such as minimum values, maximum values or averages in a dataset. A new extension is required to realize the validation function. Exactly where this extension will hook into the CKAN workflow remains to be determined.

Providing OGC services

A major gap in CKAN functionality is missing support for the OGC services WFS and WMS. Implementing these services as extension to CKAN would require significant effort, and is unnecessary because the services are implemented by GeoServer [http://geoserver.org/], another FOSS component we will use. An extension is required to connect the CKAN datastore with GeoServer. This extension will use the GeoServer API to dynamically configure and deploy new OGC services on the node GeoServer whenever a structured file is uploaded that contains geospatial data. In its simplest form this extension will provide GeoServer with an SQL view to the newly inserted data. More complex functionality in the future could create more complex OGC services that combine data from various tables or execute complex calculations before deploying a service.

Fulltext Indexing

Although we are putting significant effort into handling structured data the majority of uploaded data will consist of unstructured information such as publications. For such data it is important that we are able to have a full-text index available for searching. We will make use of SOLR in order to realize the full text indexing feature. SOLR will be triggered to index uploaded documents via an extension hooked into the upload process.

Role-based Uploading

In NGDS upload and publishing rights shall be restricted to authorized users. Moreover, it must be possible to separate the right to upload data from the right to publish data. We envision a process with selected users with upload permission (so-called data submitters) and selected users with publishing rights (so-called data stewards). Data stewards may review uploaded resources and decide if they are of sufficient quality for publishing.

It will not be possible to provide this functionality without modifying CKAN core features. First of all we need to introduce the new role "data steward" which does not yet exist. In a second step we have to provide a way for the system to control access to resources. We will do this on a per resource granularity. Hence, we introduce an attribute called "published" which can only be modified by users in the data steward role. We then have to integrate this new attribute into the dataset objects behavior so that they do not list resources that are not yet made public.

User feedback

We intend to enable the user to provide feedback to published datasets. This will be realized via an extension that keeps track of user ratings and annotations. Exactly where this extension will hook into the CKAN workflow remains to be determined.

NGDS Implementation

Figure 5 depicts how we will realize these features on top of CKAN and GeoServer.

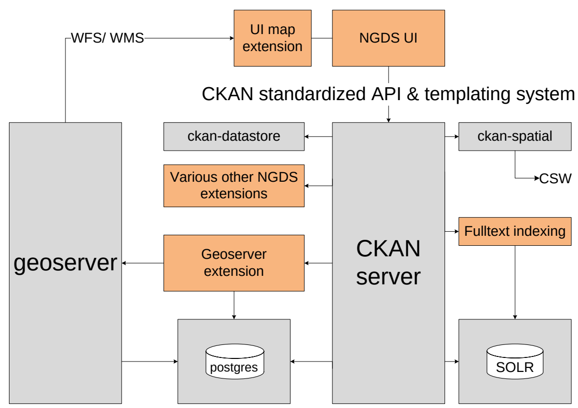


Figure 5: NGDS Component Architecture.

As can be seen GeoServer and CKAN access a single Postgres database, and an NGDS extension manages the GeoServer instance according to the data that is uploaded to CKAN. Other NGDS extensions described above are represented in a single box labeled "Various other NGDS extensions". The full-text indexing extension is represented separately because it will use SOLR as an additional external component.

Conclusions and Next Steps

Implementation and testing of the NGDS node-in-a-box will facilitate participation of new data providers in the system. Deployment of an aggregating node and NGDS portal will promote use of the information accessible through the system. An active community of providers and data consumers is essential to the long term viability of the NGDS. We have outlined an ambitious development program. Use of existing components and participation in active FOSS developer communities will facilitate progress and provides a path for long term utility of the software.

International Deployment

The node software package is being implemented using open source software, so there will be no licensing barriers to deployment of the system at any location in the world. Because of the interest in geothermal energy development in Africa and South America, a number of organizations have expressed interest in using this technology on other continents. To facilitate usage by non-English speakers, one of the design objectives for the user interface is to enable language localization using standard software engineering approaches. Actual translation of interface text and documentation to other languages is not part of the current work plan.

References

Deloitte Consulting, LLP, 2008, Geothermal Risk Management Strategies: Report for Department of Energy - Office of Energy Efficiency and Renewable Energy, Geothermal Program, 44p. http://www1.eere.energy.gov/geothermal/pdfs/geothermal_risk_mitigation.pdf

Entingh, D., 2000, Status of DOE Geothermal Technical Report Collections: Princeton Energy Resources International, LLC, Geothermal Studies and Analyses: Report 6A, for Department of Energy - National Renewable Energy Laboratory, 36p. http://www.perihq.com/documents/-Report%206A.pdf

PROCEEDINGS, Thirty-Eighth Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 11-13, 2013
SGP-TR-198

FUELING INNOVATION AND ADOPTION BY SHARING DATA ON THE DOE GEOTHERMAL DATA REPOSITORY NODE ON THE NATIONAL GEOTHERMAL DATA SYSTEM

Jon Weers [(a)], Arlene Anderson [(b)]

[(a)] National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401-3305
jon.weers@nrel.gov

[(b)] U.S. Department of Energy
1000 Independence Ave. SW
Washington D.C. 20004, USA
Arlene.anderson@ee.doe.gov

ABSTRACT

Data and information collaboration within the geothermal community is greatly needed. The lack of data sharing between the geosciences disciplines presents barriers to geothermal development.

Geothermal market analysts describe the need for data as a main geothermal development obstacle. “The rate-limiting step for all geothermal development is proving the resource – ie, having sufficient geoscientific and exploration drilling data to be certain of a certain level of output.” (Bloomberg New Energy Finance, Q3 2012 Geothermal Market Outlook[1])

In their Geothermal Risk Mitigation Strategies Report from 2008, Deloitte LLP identified the need for a national geothermal database to “provide developers and investors with a much-needed framework for investment evaluation,” and the need for visualization tools that would rely on that data to “reduce the inherent risk in early stages of development and encourage an independent investment market.”[2]

In order to help solve this problem, the U.S. Department of Energy’s Geothermal Technologies Office (DOE GTO) has developed a plan, secured funds, and is supporting the development of the National Geothermal Data System (NGDS). The NGDS is being designed as a system of distributed nodes, all in communication with one another. Each node will collect data and provide access to the collected data to the other nodes. The DOE Geothermal Data Repository (DOE-GDR) is DOE’s flagship node on the NGDS.

This paper describes the motivation behind the development of the DOE-GDR, its role in the greater NGDS, the potential benefits of using the DOE-GDR to share geothermal data of all types, and DOE’s data submission process.

STRATEGIC DIRECTION FOR PROVIDING DOE GEOTHERMAL LINKED OPEN DATA

In its May 2011 Strategic Plan, the U.S. Department of Energy (DOE) highlighted the importance of the success of their projects to include data reusability:

> *DOE’s success should be measured not when a project is completed or an experiment concluded, but when scientific and technical information is disseminated. Beyond broad availability of technical reports, e-prints and multimedia, and publication in peer-reviewed journals, open access to experimental data and analysis codes is increasingly important in policy-relevant research areas. The Department will establish guidelines for use with both grants and contracts to ensure appropriate access to, and retention of, scientific data and analysis methods. In more applied areas, knowledge of what did not work can be of equal value with positive results, for that can prevent the misapplication of significant private resources (DOE Strategic Plan, May 2011[3])*

In line with DOE’s strategic objectives, the DOE Geothermal Technologies Office (DOE GTO) is providing access to its geothermal project information through the Geothermal Projects Database[4] and the DOE-GDR. The DOE-GDR is intended to be one of many nodes on the National Geothermal Data System currently under development.

DOE's Node in the National Geothermal Data System (NGDS)

Figure 1 shows the NGDS is currently being designed as a system of distributed nodes, all in communication with one another. Each node will collect data and provide the other nodes access to the collected data.

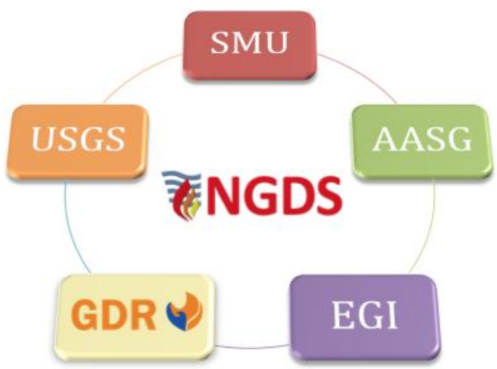


Figure 1: NGDS Conceptual Illustration Showing DOE Geothermal Data Repository Node

The DOE-GDR will be DOE's flagship node on the NGDS, and will be the submission point for all data generated by recipients of DOE GTO funds.

Data Submission Requirements

In general, recipients of DOE GTO funds must, within 90 days of a DOE request, update their data plans to include a list of the specific types of data that will be generated as part of each task and project deliverable. Questions or concerns on individual project submission requirements should be raised with DOE Project Officers.

WHY SHARE DATA?

The sharing of data is essential to advancing the adoption of Geothermal technologies in the marketplace, identifying potential areas for exploration, and fostering innovation. In their Geothermal Risk Mitigation Strategies Report from 2008, Deloitte LLP asserts that "the supply of demonstrated data from the field reduces uncertainty and increases investor confidence"[2]. In addition to simplifying the early stages of exploration and minimizing risk, sharing data can lead to other tangible benefits. Access to shared data allows analysts looking to improve the exploration process to build better models and simulations. Additionally, access to shared data could enable researchers in fields yet unknown to innovate the next groundbreaking technology.

> *Linking information from different sources is key for further innovation. If data can be placed in a new context, more and more valuable applications – and therefore knowledge – will be generated. (Linked Open Data: The Essentials, 2012[8])*

Linking Open Data

Furthermore, linking shared data can significantly increase the usefulness of the data, or in data science terminology, the *re-usefulness* of the data. The science of producing linked open data involves the careful attribution of semantic properties to the data. In the case of the DOE-GDR, these properties are often derived from the metadata collected with each data submission. Metadata is commonly defined as "data about data"[7] and includes those data needed to source each data submission and to describe it in such a way that it can easily be found as the result of a search. Linking the underlying data involves tying key elements of the metadata to existing semantic concepts. These semantic links serve as a structure by which machines can relate like concepts. Similar to the way that HTML, the coding language of the Web, serves information in a human-readable format, linked open data serves information in a machine-readable format. This format becomes the framework by which related information from disparate datasets can be linked together to form comprehensive digital assets.

The DOE-GDR is hosted on the Open Energy Information (OpenEI) platform, which utilizes a semantic wiki to serve more than 56,000 content pages, all of which are available as linked open data.

Integration with the OpenEI platform allows the DOE-GDR to take advantage of existing data models and semantic content. For example, OpenEI contains detailed data on geothermal resource areas in the United States. The information on each area is stored in a well-defined semantic concept, or data model, and served as linked data, which allows the DOE-GDR to import that data directly into the DOE-GDR data submission form. This makes the identification of areas related to submitted data considerably easier and more accurate than conventional data submission forms, which typically require the submitter to define a bounding extent for the geospatial area relevant to the submitted data. The data linkage between OpenEI and the DOE-GDR allows the DOE-GDR form to tap into the previously defined geospatial area on OpenEI and automatically associate it with the submitted data. Once the submission has been made publicly available, this same linkage could then tie the submitted data back to the appropriate resource area on OpenEI, allowing users of OpenEI to quickly view the data that has been submitted for a given area.

Sharing the Right Data

Deciding what data to share can be fraught with erroneous assumptions, unanswered questions and even philosophical debate. It is not uncommon to ask, "*What exactly is data? What isn't data?*" The DOE GTO provides guidance on what data the office wants to collect, as explained on the DOE-GDR: "data generated from geothermal exploration, research, and all data generated from projects funded by the GTO."[5] This includes raw data generated by instrument and measurement, and summary information not already included on the GTO Projects Database. The Projects Database is the companion to the DOE_GDR, providing basic information for each DOE project. Ultimately users will be able to navigate from basic information for each funded geothermal project to the project data submissions. When in doubt, DOE-GDR submitters should consult with their DOE GTO Project Officer to help identify desired data and approve data plans.

The Value of Raw Data

While access to raw data can aid in the understanding of the aggregate findings and summaries derived from it, the raw data also offers the potential for *reusability*. Summary information can be filtered toward a specific topic or focus. Raw data remains free to be used in any context, including topics unrelated to projects for which the data was collected. This is the *re-usefulness* of raw data; its ability to be used again and again in unforeseen ways to illustrate, justify, or innovate new concepts or technologies.

Precautionary Measures for Protected Data

All data submitted to the GDR are destined for eventual public release, pending a review process. Protected data are held in a secure data store with restricted access pursuant to the Intellectual Property Terms and Conditions defined on the DOE-GDR site. This protected data will remain in the secure data store until the previously agreed upon moratorium on its release has expired, at which point it, too, will be made publicly available.

Personally Identifiable Information

Not all data are suitable for submission to the DOE-GDR. Data should be cleansed of Personally Identifiable Information, or PII, prior to submission. The DOE-GDR defines PII as "any piece of information or combination of pieces that could be used to compromise the identity of an individual."[6] More information on this can be found in the frequently asked questions (FAQ) section on the DOE-GDR site.[6]

DOE-GDR DATA SUBMISSION

Submitting data to the DOE-GDR is a 3-step process:

Step 1. Register: The DOE funds recipient must register for an account at the data submission site (https://gdr.openei.org) and log in prior to submitting any data. Existing OpenEI users need only log in.

Step 2. Submit data: The DOE-GDR provides various options for submitting data. All of the options, shown in Figure 2 below, require a small amount of metadata to be collected for each digital asset being submitted. Selecting the appropriate submission option is a matter of personal preference and may depend upon the size of the submission or the number of individual files.

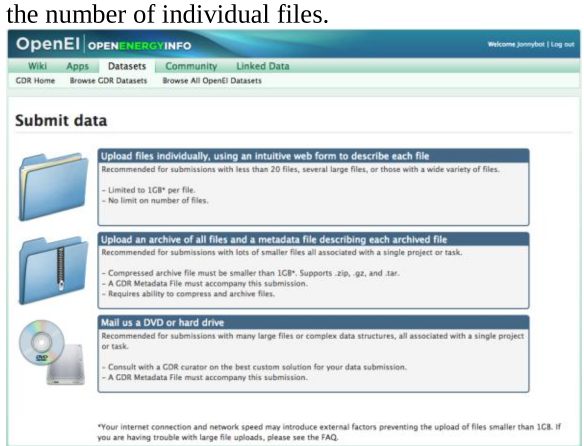


Figure 2: DOE-GDR Data Submission Options

Once a selection has been made, a web form will guide the submitter through the remainder of the submission process.

Step 3. Modify or Cancel and Resubmit: Submitters are not permitted to edit the data provided. However, they may cancel a submission at any time prior to public release. Cancellation purges submitted data from the DOE-GDR system. They may then resubmit the edited data. The metadata provided with a submission may be edited at any time prior to curation, a part of the submission review process. Once in curation, a submitter may no longer edit the metadata associated with a given submission. These restrictions help to ensure proper handling of protected data by making sure no piece of data is able to circumvent any step in the submission process.

CONCLUSION

The distributed node nature of the NGDS should allow users of the DOE-GDR, and all other NGDS nodes, to have access to the greater pool of geothermal information. Routine submission of publicly accessible data to the DOE-GDR can help to minimize the risk of future investment in geothermal technologies. The submission of raw data and the provision of detailed and accurate metadata can fuel innovation in the geothermal sector. Furthermore, the linking of semantic concepts can help insure the reusability of the data and advance understanding of geothermal sciences.

REFERENCES

1. Bloomberg, New Energy Finance, Q3 2012 Geothermal Market Outlook, Achilles heels: resource risk and land rights, 8 October 2012
2. Deloitte LLP, Geothermal Risk Mitigation Strategies Report. 2008.
3. United States, Department of Energy, Strategic Plan, Assure Excellence in R&D Management, (Washington: GPO, 2011) 43-44.
4. "GTP Projects Database." *Geothermal Technologies Program*. U.S. Department of Energy, 9 Jan. 2013. Web 15 Jan. 2013. http://www4.eere.energy.gov/geothermal/projects.
5. "DOE Geothermal Data Repository." *OpenEI: Open Energy Information.* National Renewable Energy Laboratory, 15 Jan. 2013. Web. https://gdr.openei.org.
6. "Frequently Asked Questions." *DOE Geothermal Data Repository on OpenEI.* National Renewable Energy Laboratory, 15 Jan. 2013. Web. https://gdr.openei.org/faq.
7. "Metadata" *Wikipedia* Wikimedia Foundation, Inc., 7 Jan. 2013. Web. 15 Jan. 2013 http://en.wikipedia.org/wiki/Metadata.
8. Bauer, Florian, and Martin Kaltenböck. *Linked Open Data: The Essentials: A Quick Start Guide for Decision Makers*. Vienna: Edition Mono, 2012. Print.