NGDS Final Report

DOE Award Number: DE-EE0001120-#004

Name of Recipient: Boise State University

Project Title: National Geothermal Database (NGDS)

Name of Project Director/PI: Harold S. Blackman

CO-PI: Joseph Moore

Consortium Teaming Members: University of Utah, University of Nevada-Reno, Stanford University, Oregon Technology Institute, Arizona Geological Survey

Executive Summary

Research Contributions to Understanding

The ultimate goal of the National Geothermal Data System (NGDS) is to support the discovery and generation of geothermal sources of energy. The NGDS was designed and has been implemented to provide online access to important geothermal-related data from a network of data providers in order to:

- Increase the efficiency of exploration, development and usage of geothermal energy by providing a basis for financial risk analysis of potential sites
- Assist state and federal agencies in making land and resource management assessments
- Foster the discovery of new geothermal resources by supporting ongoing and future geothermal-related research
- Increase public awareness of geothermal energy

It is through the implementation of this distributed data system and its subsequent use that substantial increases to the general access and understanding of geothermal related data will result. NGDS provides a mechanism for the sharing of data thereby fostering the discovery of new resources and supporting ongoing geothermal research.

Technical effectiveness and Economic Feasibility

The NGDS has been demonstrated and is fully functional. It consists of a redistributable, open-source system. The system provides an effective user interface supporting the search of data resources, the ability to establish an NGDS node, the ability to upload and share data, as well as the ability to search for and then download specific data sets. Access is supported through an active website that also supports the access to other tools. The NGDS now provides access to well over 40,000 data records.

Benefits to Public

The NGDS is a free online resource that is available to the public. It has the potential to support research as well as governmental, and educational activities. To that end the National Geothermal Data System consists of three linked communities:

1. Data providers who will expose information to the system through standardized, internet-accessible interfaces and interchange formats

2. Software developers who will build applications that utilize the data in the system, and make it easier for end-users to interact with the system.

3. End-users who will utilize the software and information provided by the system in order to understand and develop geothermal resources.

The NGDS includes data covering a wide range of topics, from well logs and drilling data to temperature, geochemical, and geophysical measurements. Standardized data access to frequently available datasets will facilitate utilization of these information resources.

A key component of the system is the catalog service through which data providers register the availability of resources, and through which users discover, evaluate and access resources. A resource will be considered part of the system when it can be located by searching the catalog service, which will return a metadata record describing how the resource can be accessed. Data providers maintain nodes in the network, connected through the use of standardized metadata for describing resources, content models for geothermal data, and common web-service protocols for exchanging information. These standards have been developed in conjunction with the US Geoscience Information Network (USGIN), thereby providing interoperability with a wider range of geoscientific information. The node contains the data that is shared across the system with other participants and is maintained and operated by the owner.

The software itself is a redistributable, open-source software package created to give data providers a simple way to register data sources, load data and expose those data as a node in the NGDS. The software supports batch import and upload of shared datasets in supported formats adhering to standard content models. The use of this software is not required in order to participate as a node in the network; data providers may use whatever tools they wish to expose their data, as long as they utilize interchange formats and web-service protocols conforming NGDS specifications.

End-users may interact with the system through the NGDS website that is designed to provide information about the NGDS. It serves as an entry point to the system, allowing users to discover data and applications that utilize NGDS resources. The site includes information on the project’s progress, NGDS specifications, access to the Map-Centric Search Application described below as well as other software applications utilizing NGDS data, presentations, documentation and tutorials, a catalog of NGDS nodes, and any other results as they become available.

The search application is a user-friendly, web-based application to support finding, visualizing, mapping, and acquisition of data by end-users. This application allows users to discover and access resources made available across all NGDS nodes, and to search for data across the system based on topic, location, and or other criteria. Standardized metadata describing each dataset provides the user with the information necessary to determine the utility of that dataset for their purposes. Geographic datasets can be visualized through a map interface that allows users to inspect the details of individual data points (e.g. wells, temperature measurements, etc.) from properly formatted datasets. In addition to visualization within the application, the interface provides the information necessary for users to access the data from other applications.

Summary Of Project Activities

In this section each task, as defined by the Project Implementation Plan (PIP) and referenced to the Statement of Project Objectives (SOPO), is described including deliverables.

PIP Task 1.0 Data Assessment (SOPO Task 8)

Purpose: Determine the amounts and types of data that will be made available to the NGDS under this project; determine progress on current digitization efforts.

Approach: Work with each project consortium member on their inventory of existing and proposed data products for NGDS. Assist institutions to prioritize data types based on recommendations from the Project Domain Steering Committee (see Management Section). Based on these priorities, work with the consortium members to plan workflows for:

- Serving arbitrary data files via web-servers annotated with meta-data in a catalog (**Tier 1**)
- Serving structured tabular data via (geospatial) web-services annotated with meta-data in a catalog (**Tier 2**), and

- Serving structured tabular data via geospatial web-services according to agreed upon NGDS wide content models annotated with meta-data in a catalog (**Tier 3**).

Site visits to data provider institutions are anticipated.

Milestone: Submission of NGDS consortium member data inventory and work plans for inclusion of targeted datasets in the NGDS. Review this inventory with the Project's Domain Steering Committee. Discuss this inventory with the Software provider to assess difficulty in automating steps of workflows to bring data online.

Outcome: Accurate understanding of data integration work remaining to be done during project lifetime that will inform the domain steering committee and PI in developing concrete work plans to incorporate information resources into the NGDS in the time frame available.

Actual: All work has been executed as planned, documented in two reports: "Data Acquisition for National Geothermal Data System", and "NGDS Design and Testing Subrecipient Data Inventory" found in Appendix 1a and 1b respectively. This work resulted in the acquisition of nearly 30,000 data records representing millions of data points. Each data subrecipient was able to complete their plans resulting in a meaningful data set for inquiry by the NGDS user community.

PIP Task 2.0 User Research and Experience Concept (SOPO TASK 3)

Purpose: (a) Determine priority user groups and understand key-user characteristics, goals and needs. (b) Create a low-fidelity concept for the user interface (UI) of the geothermal community user interface and communicate it to the software developers to implement.

Approach: (a) We'll use multiple methods to understand key user characteristics and needs, including an online survey and contextual interviews with representative users from key audience groups. (b) Create low-fidelity (non-interactive) mockups for the UI and evaluate it in a series of usability studies at the GRC annual meeting. Iterate on the concept working with NGDS partners. (c) Prioritize the features and functionality of the user experience concept based on project status and risks. The user experience concept informs the functional and non-functional requirements (task 3.0).

Milestone: Data-driven user profiles / personas, key usage scenarios, priority data needs, user experience concept and UI feature requirements for the geothermal community user interface.

Outcome: Shared understanding of priority users and their needs and prioritization of user scenarios that will be implemented.

Actual: As indicated the development of the user interface was a multi-staged process. The effort began with a series of one on one interviews conducted with different user types from the commu This resulted in an initial concept documented in Appendix 2a. This initial set of wireframes was then further tested with potential users and then improved upon for a second version also shown in Appendix 2b. A third user test was conducted to further refine the concepts resulting in the next iteration. This iteration really formed the "final" interface although improvements have continued to be made, and will continue throughout the life of NDGS. This version is also shown in Appendix 2c. These improvements occurred as a part of the software development task described below.

PIP Task 3.0 Software Requirement and Design Specification (SOPO TASKS 1 and 2)

Purpose: Describe the functional and non-functional requirements of the node software including data import, data dissemination, and user interaction (for both NGDS data providers and NGDS data consumers in the public, including industrial and research communities). Describe the design of the software including which open source components to utilize and how to deploy the complete software package.

Approach: The requirements will be defined based on the design of the NGDS (e.g. interoperable geospatial web services) and incorporate results from Task 1.0 as to which data types and formats need to be imported and which web services will be required and Task 2.0 as to which user functionality needs to be supported. The design specification of the software will address the requirements and incorporate lessons learned from other similar geo-informatics projects and will heavily rely on existing open source software packages. A Software Test Plan will also be developed and will adhere to standard software testing practices.

Milestone: Requirements specification document and design specification document available.

Outcome: Software specification.

Actual: An initial software specification was prepared and underwent extensive review and comment with our customer and technical reviewers. The final version 2.7 was approved and delivered and is attached in Appendix 3a. In addition a software architecture document was prepared and delivered that describes the NGDS design and architecture. It may be found in Appendix 3b. All documentation as well as the code itself is housed in the project Github located at github.com/ngds.

PIP Task 4.0 Project Website (SOPO TASKS 11 and 12)

Purpose: Provide a website for collaboration among project participants and to inform the public about project activities. This site is not intended to serve as the user interface for discovery and access of data.

Approach: A NGDS project website will be designed and published to provide a portal for all relevant information on the NGDS including a roadmap and any results as they become available including but not limited to NGDS standards, available software packages for data providers and consumers, presentations, documentation and tutorials, and a listing of all NGDS nodes. This website will provide access through links to at least one NGDS catalog and NGDS user interface for data discovery and access.

Milestone: Up-to-date and operational website showing the status of each project participant, links to interfaces, and guidelines and resources for participating in NGDS.

Outcome: A functional, informative website that engages the community and public.

Actual: Website was developed and is available and functioning at (https://geothermaldata.org). The website has been updated as needed with respect to user feedback.

PIP Task 5.0 Software Development, Iteration and Usability (SOPO Tasks 1, 2, and 3)

Subtask 5.1 Node Software Development

Purpose: Develop an open source software package (reference implementation) that supports serving data to the NGDS as one node of the distributed network, along with a geothermal-community user application providing key discovery, browse and access functionality to utilize NGDS.

Approach: The software will be developed in an iterative process based on the requirements and design defined in Task 3.0. After each iteration, the results will be presented to the project manager who will support the prioritization of features to be implemented in the next iteration. The software will include but is not limited to the following major components:

- **User interfaces**: we will implement the user interface as web applications for end-users, including both NGDS data providers and NGDS data consumers in the public, including industrial and research communities, according to the requirements developed in Tasks 2 and 3. These applications will access services provided by the data-provider node reference implementation.
- **Metadata:** components for creating, storing, managing, searching and delivering metadata describing NGDS resources using NGDS service protocols and interchange formats. Allow editing of metadata and tagging of content by keywords.
- **Data management**: component for storing data to support data access through NGDS services.
- **Data processing pipeline**: parse, transform, and import data into the data management component and deploy WFS and WMS web services in the data dissemination component.
- **Data dissemination**: components for publication of specific data (based on Task 1 and 2 requirements) as geospatial web-services using NGDS interchange formats and web service profiles.

- **Data analytics**: Include analytics and business intelligence tools for data review and quality control
- **Federation**: component to enable automated federation by discovery of other NGDS nodes and replication of metadata catalog between nodes.
- **System services**: components to provide central services for URI dereferencing, annotation vocabularies, and quality of service monitoring.
- **Security**: components to provide network-wide authentication and access control as defined in the requirements and to perform security review and penetration testing.
- **Administration**: components to provide user management, usage monitoring, and services to synchronize, backup, and restore the node data.

Milestone: Software releases as defined in the project plan and milestones defined in sub-tasks.

Outcome: NGDS software package that supports data import and dissemination to the NGDS and allows searching, browsing, mapping, analyzing and downloading the hosted data.

Actual: An agile software development process was employed by the team to produce the software products. Eighteen major sprints were completed resulting in a fully functional NGDS in accordance to the Software specification. Several smaller sprints were employed by the team in final two months to further improve usability and functionality of the system. The software code is tagged at version 1.0.1 and is available at https://github.com/ngds/ckanext-ngds/. In the sustainability phase of the project future versions of the code will be pushed to the Github master branch with newer versions being tagged. Current users will be notified by RSS fed, email, or other notification system, along with a list of upgrades and changes indicated for that version.

Subtask 5.2 Iterative Design and Usability Testing

Purpose: Optimize the user-experience of NGDS through iterative design and testing.

Approach: We will work closely with project partners to iteratively test and design the user experience for NGDS. This includes:

- Designing and implementing rapid usability tests with small user groups to evaluate the usability of the evolving NGDS system (2-3 cycles).
- Conducting research to inform design of features for next sprint cycle.
- Validating design changes made based on prior user testing
- Providing recommendations on how to refine the user interface design based on usability study results

Milestone: User experience evaluations and improvements

Outcome: A highly useful and usable interface that supports how key users expect to work with NGDS and encourages adoption, continued use and data contributions.

Actual: Three user experience tests were conducted through the development phase of the project. Each time results were incorporated into the developing user interface. In addition team members continued testing and revision on an ongoing basis, and a user test and demonstration was completed in December, and again in January in Washington D.C. In summary, the system uses project developed graphical user interfaces for a uniform look-and-feel for web applications. The project-developed applications provide online help explaining how to perform user-related functions. Finally, the system underwent review for compliance with section 508 amendment to the Rehabilitation Act of 1973 and ISO/TS 16071 “Ergonomics of human-system interaction- Guidance on accessibility for human-computer interfaces”.

PIP Task 6.0 Pilot Data Import, Integration & Testing (SOPO TASKS 9 and 10)

Subtask 6.1 Get file-based resources registered and online (Tier 1)

Purpose: To enable the discovery and download of annotated files with geothermal data.

Approach: Support each project consortium member to make file-based resources accessible online using the NGDS node software from Task 5.0 either hosted by the consortium member or using a host of another project partner. Based on the results of Task 1.0 determine best approach to getting metadata into the NGDS catalog for example by using the NGDS node software import modules if the format is supported or creating scripts to convert the metadata to the required format. Metadata will be converted incrementally based on the plans laid out in Task 1.

Milestone: 90% of Tier 1 resources registered in system by end of Q2 2013, 100% by project completion.

Outcome: Files with geothermal data discoverable and accessible from network of connected NGDS nodes.

Actual: The revised content model repository has the most up to date versions, and total data amounts to over 40,000 metadata records. 100% of Tier 1 resources are registered, including data, scanned maps, publications, or documents, plus a little over 350 records for tier 3 services.

Subtask 6.2 Structured data in web services (Tier 2 & 3)

Purpose: To enable the discovery and access of annotated web-services with geothermal data.

Approach: Support each project consortium member to make structured data available via web services using the NGDS node software from Task 5.0. The selection of data to be served via standardized web service (Tier 2) and if available standardized interchange formats (Tier 3) will be based on the assessment of Task 1.0 and feedback from Project’s Domain Steering Committee (PDSC). Based on the results of Task 1.0 determine best approach to getting data into the NGDS web services for example by using the NGDS node software import modules if the format is supported or creating scripts to convert the data to the required format.

Milestone: Registered Tier 3 resources based on PDSC recommended priorities. The minimum target is to have at least one Tier 3 service from each data provider.

Outcome: Interoperable web-services with geothermal data discoverable and accessible from network of connected NGDS nodes.

Actual: Metadata conversion and addition from each of the sub-recipients is now complete, interoperable web-services are functional and geothermal data discoverable from NGDS nodes.

Subtask 6.3 Develop automated data import routines

Purpose: Supports tasks 6.1 and 6.2 by adding data import routines to the NGDS node software to enable automated and repeated import of metadata and data.

Approach: Analyze the data formats of sample files by different consortium members and determine mapping of information elements to the NGDS content models

Milestone: Software releases of Task 5.0 with additional import capabilities.

Outcome: Automated import of selected data sources and formats will be supported by the node software to enable easy, reproducible, and updatable publication of geothermal data via the node software.

Actual: Routines were developed and applied to the data as needed. Automated data import features were integrated into the online node submission process.

Subtask 6.4 Pilot testing

Purpose: To demonstrate that pilot deployment meets requirements

Approach: Utilize developed test suite in accordance with the test plan to test and verify client/server functionality and operational integrity. A single instance of the software node will be brought online and tested with prototype user applications for search and map-based browsing.

Milestone: Single NGDS node online with pilot data, operates with web-based user application for search and discovery using NGDS catalog and data access services.

Outcome: Demonstration of pilot functionality for product owners.

Actual: Testing continued throughout the project with fixes planned and executed through a process of prioritization. All issues are maintained on the Github (https://github.com/ngds) and will be worked in the sustainability phase of the project. Testing was successfully completed with the sing NGDS node online with pilot data and operating in a web-based application in late 2013, with the user interface released as v1.0 at the Geothermal Energy Association Annual Exhibit..

PIP Task 7.0 Distribution (SOPO Tasks 1, 2, and 3)

Purpose: Make the NGDS node software redistributable as open source software.

Approach: The software will be compiled into a redistributable package, for example, as an executable installer package or a downloadable virtual machine image. The source code will be uploaded to a publically accessible repository with instructions on how to build and extend the software. The license will be an open source license acceptable by the DOE.

Milestone: Binary software, source code, and documentation downloadable by the public.

Outcome: NGDS software openly available and extensible.

Actual: A final version installer was completed and tested and is available at https://github.com/ngds/install-and-run, including in installation documentation

PIP Task 8.0 Documentation (SOPO TASKS 1, 2, 3, and 4)

Purpose: Document the NGDS software including an administration manual to operate a node, a developer's manual to extend the software, and tutorials for users.

Approach Document how to administer, extend and use the NGDS software.

Milestone: Administration Manual, Developer's Manual, and user tutorial available.

Outcome: Use and extension of NGDS software documented.

Actual: Development of documentation has been completed including the developer manuals and th administrator and online user tutorials. There are three documents supporting the developer. The fir is a Developer's guide found in appendix 4a, the second is a Systems Overview found in Appendix 4b, and the third is a document describing how to be an NGDS node found in Appendix 4c. Administrator and User tutorials are available on https://geothermaldata.org under tool tips and other various programs, and user videos and help are also available at www.youtube.com/geothermal data, and at http://usgin.org/tutorials . These provide support and information how to operate the systems from all users including administrators.

PIP Task 9.0 Production Data Integration (SOPO TASK 10)

Purpose: Remaining information resources deemed relevant will be made accessible according to priorities established by the project's DSC.

Approach: More data will be made available similar to Tasks 6.1 and 6.2 by supporting project consortium members in converting, annotating, and importing data into an NGDS node.

Milestone: Register Tier 1 or deploy Tier 2/3 resources based on DSC-recommended priorities.

Outcome: Interoperable web services with geothermal data discoverable and accessible from network of connected NGDS nodes.

Actual: The NGDS central aggregator is functioning at geothermal.org, hosted by Arizona Geological Survey (AZGS) harvesting metadata from http://catalog.usgin.org/geothermal (AASG data, Geoheat data), CKAN instances at New Mexico Bureau of Geology and Mineral Resources (NBMG), Arkansas Geological Survey (ArGS), and University of Utah's Energy and Geoscience Institute (EGI), USGS Science Base, Southern Methodist University (SMU) node, and NREL GDR.

PIP Task 10.0 Federated Rollout (SOPO TASKS 1, 2, and 3)

Several instances of the production data provider reference implementation software will be brought online to test the federation capability and bring online the remaining data deemed relevant.

Purpose: To demonstrate federated network operation

Approach: Several instances of the production data provider reference implementation software will be brought online, and federation capability tested and verified.

Milestone: Three nodes online, harvesting metadata between them, and separately accessible by the same user-search application.

Outcome: Interoperable federated system.

Actual: Four nodes are online, fully functional and harvesting data. Nodes are functioning at NMBG ArGS, and EGI.

PIP Task 11.0 Service Testing (SOPO TASK 2)

Purpose: To monitor system operations and ensure conformance with specifications

Approach: Develop and deploy service monitoring capabilities and automated service-based conformance tests

Milestone: Operational service availability report page; validation service for metadata records; validation service for 2 other selected data services.

Outcome: Better system reliability.

Actual: System tests have been implemented and are informing continued development. Validation services are in place for continued use in sustainability phase. All project source code has comments, at least on a per-class level. The system's architecture and configuration parameters are well documented and the system's source code is covered by unit tests to at least 50 % coverage.

PIP Task 12.0 Software maintenance (SOPO TASK 6)

Purpose: Fix operational problems in software, and add features as necessary to keep system up to date and meeting user requirements for duration of project.

Approach: Maintain online bug reporting and feature-request system using standard open source tools, prioritize issues and allocate resources to resolve them.

Milestone: Operational bug reporting and feature-request system that is monitored and issues are acted on promptly.

Outcome: Better system operation.

Actual: Maintenance is continuing and a list of operational issues and desired features is being maintained and worked within the available resources. The list of operational issues is being maintained on the Github (and will be provided as a part of the total package for sustainability.

PIP Task 13.0 Sustainability Plan (SOPO TASK 7)

Purpose: Ensure the continuity and viability of NGDS beyond the project-funding period.

Approach: Develop a business model to sustain system maintenance and support addition of new data and services by engaging stakeholders who have financial resources available and an interest in continuing system operation. Educate system data providers and users through workshops and tutorial documentation. Develop policies and procedures for node mirroring and backup. Develop a plan for tools and processes to collect user feedback on an ongoing basis to continually monitor user satisfaction, understand what is working and what is not, foster a collaborative online community, and establish priorities for updating NGDS based on user feedback.

Milestones: Plan for data and operations maintenance and user feedback system for institutional participants, and a business plan for sustaining the system infrastructure.

Outcome: A plan is in place for NGDS operational continuity beyond the duration of the project.

Actual: The sustainability plan is under implementation with the creation of a 501c3, USGIN Foundation Inc., and final documentation is complete and is attached in Appendix 5. Board members are being identified, and the corporate by-laws are under review. A multifaceted business model is being pursued with several pathways for financial support including, but not limited to products and services, grants and contracts, foundation support, appropriations, and contributions and donations.

PIP Task 14.0 Project Management and Reporting (SOPO TASK 12)

Purpose: Successfully manage and report on the project to DOE.

Approach: The project will be led by a Principal Investigator (PI), and informed by the recommendations of 3 committees. The PI will direct the work of the team through a Project Manager that closely supervises the work of each of the team organizations. Reports and other deliverables will be provided in accordance with the Federal Assistance Reporting Checklist following the instructions included therein. Comprehensive reports at the end of each project phase will serve as the basic milestones for each of the tasks and their subtasks. These reports will serve as the basis for go / no-go decisions for the project as a whole and for the various tasks. It is anticipated that the review of the reports will also lead to revisions of the tasks and subtasks as necessary.

Milestone: Participation in all required reporting documents and presentations.

Outcome: Effective project implementation. Informed project participants and funding agency.

Go / No-Go Decision Points from the Statement of Project Objectives

Actual: Project closeout is underway and will be complete with this report. We have executed work according to the revised plan, initiated in June of 2012. The resource loaded schedule was managed through biweekly meetings (Full-Project, Software Development, and User Concept Development) where work progress and schedule were addressed. Our DOE technical monitors and their advisors routinely attended these meetings to ensure open and complete communication. We have produced a fully functional NGDS as envisioned and laid the groundwork for its sustainability.

Problems: Since the project was re-planned June of 2012, no new issues have emerged.

PRODUCTS / DELIVERABLES

Training and Professional Development: None

Publications, Conference Papers, and Presentations

> "NGDS User Centered Design: Meeting the Needs of the Geothermal Community", by Suzanne Boyd, Sam Zheng, Kim Patten, and Harold Blackman, Proceedings, Thirty-Eighth Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford California, February 11-13, 2013
>
> "Developing the National Geothermal Data System", by Ryan Clark, Christoph Kuhmuench, Stephen Richard, Proceedings, Thirty-Eighth Workshop on Geothermal Reservoir Engineering,Stanford University, Stanford California, February 11-13, 2013
>
> "National Geothermal Data System (NDGS) Geothermal Data Domain: Assessment of Geothermal Community Data Needs", by Stephen Richard, Arlene Anderson, David Blackwell, Cathy Chickering, Toni Boyd, Roland Horne, Matthew MacKenzie, Joseph Moore, Duane Nickull, Lisa Shevenell, Proceedings, GRC Transactions, Vol. 37, Thirty-Eighth Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford California, February 11-13, 2013

“National Geothermal Data System: Transforming the Discovery, Access, and Analytics of Data for Geothermal Exploration”, Kim Patten (Ed.), Open-file Report OFR-13-06, Arizona Geological Survey, May 2013

“National Geothermal Data system (NGDS) Geothermal Data: Community Requirements and Information Engineering”, Arlene Anderson, David Blackwell, Cathy Chickering, Toni Boyd, Roland Horne, Matthew MacKenzie, Joseph Moore, Duane Nickull, Stephen Richard, and Lisa Shevenell, Geothermal Resources Council 37th Annual Meeting, Las Vegas, Nevada, September 29 - October 2, 2013

“NGDS User Centered Design: Meeting the Needs of the Geothermal Community II”, by Suzanne Boyd, Sam Zheng, Kim Patten, and Harold Blackman, Geothermal Resources Council 37th Annual Meeting, GRC Transaction Vol. 37, Las Vegas, Nevada, September 29 - October 2, 2013

“Sustaining the National Geothermal Data System: Considerations for a System Wide Approach and Node Maintenance”, Lee Allison, Arlene Anderson, Cathy Chickering, Kim Patten, and Stephen Richard, GRC Transactions, Vol 37, 2013, Geothermal Resources Council 37th Annual Meeting, Las Vegas, Nevada, September 29 - October 2, 2013

National Geothermal Data System: An Exemplar of Open Access to Data”, Kimberly Patten, Lee Allison, Stephen M. Richard, Harold Blackman, and Arlene Anderson, Geological Society of America Abstracts with Programs, Vol 45, No 7, p648, October 2013.

Website

The NGDS website was revamped and put on line on October 1, 2012. The web address is geothermaldata.org. The website provides the portal to the program and serves as the access point for all data providers, and users of the system. It provides lists of project participants, links to interfaces, and how to participate in NGDS. It also provides access to products and some data at this juncture. The key sections of the website include:

The NGDS Homepage

The Map

The Library

Resources Section

Contribute Section

Patents and IP: None

Other Products / Deliverables: None

1 15

Appendix 1a

Data Acquisition for National Geothermal Data System

ARIZONA GEOLOGICAL SURVEY
416 W. CONGRESS, SUITE 100
TUCSON, AZ 85701

DATA ACQUISITION FOR NATIONAL GEOTHERMAL DATA SYSTEM

A Report to:
Harold Blackman, PI
Scott Keneman, PM

helpdesk-admin

ABSTRACT

The National Geothermal Data System is based on a collection of protocols for metadata and data inter-change. There is no single system data model, rather there is a metadata interchange format and a col-lection of data interchange formats (the NGDS content models, see Data Access Phase 1 deliverable document). Each interchange format defines a data model for an information item of interest, and is de-signed to access particular information tailored to some use case or use cases. Data providers and client applications are linked through open source interfaces that decouple clients and servers such that they can evolve independently without breaking the system. The OpenGeospatial Consortium Catalog Service for the Web (CSW), currently at version 2.0.2 is proposed for catalog search and discovery service. The USGIN profile for ISO metadata defines the metadata interchange format for the NGDS, and is already in use by the NGDS catalog deployed for the AASG Geothermal Data Project. Initial data services are being implemented using WFS 1.1.1 simple feature services, with interchange formats defined for key geo-thermally interesting datasets. Interchange content is encoded as GML Simple Features, an XML application scheme. File-based data are incorporated into the NGDS by creating a metadata record conforming to the metadata specification and inserting it into the NGDS catalog. Files containing the data must be in a web-accessible location specified by a URL in the metadata record.

Introduction

The goal of the National Geothermal Data System (NGDS) is to provide online access to important geothermal-related data from a network of data providers in order to:

- Increase the efficiency of exploration, development and usage of geothermal energy by providing a basis for financial risk analysis of potential sites
- Assist state and federal agencies in making land and resource management assessments
- Foster the discovery of new geothermal resources by supporting ongoing and future geothermal-related research
- Increase public awareness of geothermal energy

As part of the NGDS Design & Testing project reboot, held June 2012 at Boise State University, the Arizona Geological Survey was tasked to produce an inventory of the data to be incorporated into the NGDS by the Design and Testing project partners (see accompanying document), and a report assessing the data available and outlining plans for incorporating data into the NGDS. This document presents a framework for National Geothermal Data System (NGDS) architecture focused on the network data publication and data access aspects of the system. The data acquisition plan builds on strategies and techniques developed under the aegis of the Association of American State Geologists (AASG) contributions to the NGDS project and the NSF US Geoscience Information Network project.

NGDS-A distributed data system

Approach

A major time consuming aspect of bringing disparate datasets together is data integration. This process involves matching field or element names in the schema for various data sets, selecting those that contain the information of interest, and then merging content into a single data set with consistent usage of vocabulary and units of measure in a standardized collection of fields or elements. Data integration may be done by data providers who choose to deliver data in standardized interchange formats, by data consumers who acquire data in heterogeneous formats and schema and figure out how to extract what they need, or data integration may be done by middleware layers that implement transformations between known formats and schema.

Data integration in our current system of scientific information interchange is mostly left to the data consumer. Until recently, the most common approach was has been for an investigator to collect various datasets and integrate them into a single database that was used for some analysis; some small part of the data might get published, and the compiled dataset was subsequently committed to oblivion. Centralized data aggregation schemes have also been developed and deployed, but rarely outlive project funding or are not maintained and rapidly grow stale due to out-of-date data or use of retired technology. A tremendous amount of effort has been made towards developing systems to promote the management of data such that it may be reused without having to repeat the same integration and cleanup processes over and over.

The US Geoscience Information Network project has implemented an approach to data access promoting reuse through the use of standard formats and access protocols to discover, evaluate and access common data sets (e.g. borehole temperature data, heat flow measurements). The onus of data maintenance is shifted towards organizations that are tasked with data management and preservation. By documenting data schema, encoding formats and practices for vocabulary usage, 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. Maintenance of the data by its originator or a data steward with a direct interest in the data increases likelihood that data will be updated to correct errors, add new data, and migrate data to new technology platforms.

In such distributed information systems there is no central database or single data model. Data interchange is based on standard interchange content models and encoding schemes that are like database views designed to present particular information. The same information may be presented in more than one interchange format, tailored for different application use cases. Interchange content models are optimized to facilitate filtering data through service interfaces and utilization by end-user applications. Thus many of the interchange formats are denormalized (information is duplicated), and may require the use of standardized units of measure to avoid client complexity handling different possible units in interchange content. One of the goals of the system is to allow development of client and data server applications independently.

The use of schema and encoding specifically designed for data integration and interchange means data producers and consumers can continue to use internal data formats that are optimized for their business requirements. Use of the community interchange formats reduces the amount of work required because only one transformation from internal to interchange format has to be engineered for each interchange format in use.

This requires education of the data providers/publishers on the use of the interchange formats, but shared interchange formats results in a larger community of IT personnel who know how to get data into and out of the integration format. Mapping data into an interchange format is likely to be done more accurately when those who originate the data work closely with data managers who understand the interchange formats. The net effect is a greater likelihood that the distributed information system using the documented interchange formats will outlast any particular researcher, data provider, project, or agency. HTML on HTTP, NetCDF, and XML are examples of data integration formats that have achieved wide usage and long term usefulness.

Data integration by providers introduces additional costs into the data delivery process, and this cost dictates that there must be consideration of the benefits obtained. For data that are not provided using documented interchange formats, detailed metadata describing the schema and encoding of the data will be necessary to enable reuse. The NGDS steering committees will need to develop policies to determine what data should be presented in the data integration format (along with any other formats that the data publisher wants to use), and what data are specialized to a degree that data integration effort by the providers is not warranted. Criteria for such decisions will likely include how many providers have a particular kind of data, how often that kind of data are known or expected to be used, the cost of obtaining or reproducing the data, and the expected useful lifetime of the data.

System Architecture

The NGDS is implemented as a loosely coupled system with public specifications for interfaces between components. The system architecture can be described in terms of the functional components shown in Figure 1. Information resources are stored in repositories that are implemented as file systems or databases. Server components include document repositories, data servers, infrastructure servers, and the NGDS Catalog. Client side components are grouped into catalog search clients, and business applications. The server applications might be tightly coupled to local databases and file systems using proprietary interfaces, but will implement standard NGDS server interfaces for interaction with client side business and catalog search applications. NGDS is a network that is layered on top of the World Wide Web, defined not by the implementation of particular hardware and software components, but by the protocols and interchange formats that link the functional components together. If interaction between components uses standard interfaces, the internal implementation of any particular component is invisible and irrelevant to other components.

The basis for information access is the interchange formats used by the data services to present data to client applications. There is no single system data model, rather there is a collection of data models (the NGDS content models, see Data Access Phase 1 deliverable document), which are designed to access particular information tailored to some use case or use cases. Business applications interact with data provider databases via queries based on the data attributes exposed in the interchange content models. The use standard protocols for data access likewise enables data access using off the shelf software, both commercial and open-source. OpenGeospatial Consortium (OGC) Web Map Service (WMS) and Web Feature service (WFS), and OpenDAP NetCDF services are currently in use for NGDS data services.

Conventions for metadata content and encoding are the foundation for an interoperable discovery system—essentially the content model and interchange format for metadata. Metadata can be harvested freely between catalog nodes, and searches can be federated across multiple nodes. Strict adherence to conventions for encoding service connection information in metadata allows automated access to services with minimal user input. Use of standard protocols for searching metadata catalogs enables client applications to be reused by multiple portals, and allows individual search client instances to search multiple catalogs simultaneously.

The framework for implementing data handling requirements is a community of data providers exposing information through standardized internet-accessible interfaces (services), a community of software developers building applications that will utilize the information resources available to the community, and a community of users taking advantage of the software and information to develop geothermal resources. The system service inventory is focused on entity services that provide information resources. As used here, an entity service is a service that provides a requested resource packaged in some interchange format in response to a request, as opposed to a functional service that takes some input package of information and produces an output response according to some processing logic operating on the input information. The system functional components shown in Figure 1 are discussed in the following sections.

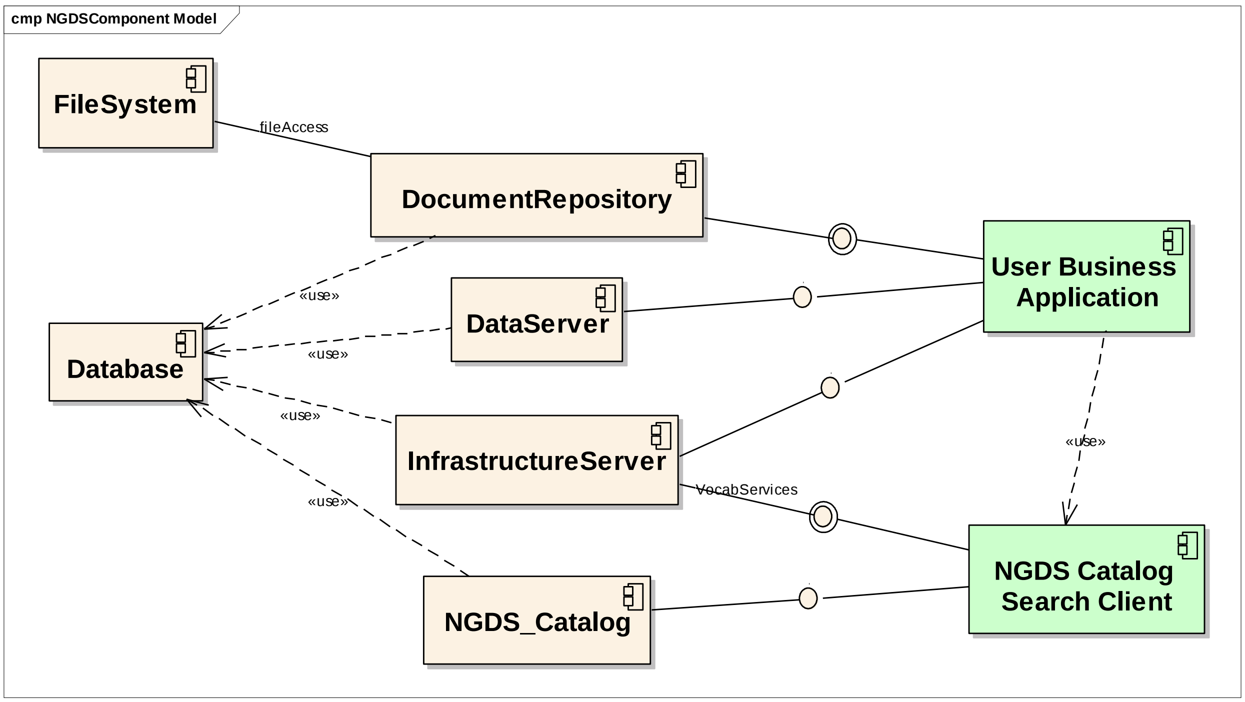
CATALOG

A NGDS catalog component implements the Open Geospatial Consortium Catalog Service for the Web (CSW) v.2.0.2. Other protocols such as the Open Archive Initiative Protocol for Metadata Harvest (OAI-PMH) or the OpenSearch protocol may also be offered. The CSW was selected for initial development work because it operates in the same framework as the other Open Geospatial Consortium services being tested for data delivery (the Web Map Service and Web Feature Service), is designed for geospatial data, and has a variety of free, open-source software projects developing clients and servers for the protocol, as well as a variety of commercial products (including ESRI ArcGIS) that are implementing the protocol.

The NGDS specifies that catalogs must provide search results encoded in conformance with the USGIN ISO metadata profile (http://lab.usgin.org/USGIN-ISO-metadata-v1-1). This profile is based on the ISO 19115/19119/19139 metadata specifications and XML implementation. Use of this metadata schema allows richer metadata content that enables greater automation of access to resources. The CSW service s

simple XML encoding of the Dublin Core Elements and Terms (csw:record), but this is a subset of the ISO profile and will require no additional effort for data providers. The CSW specification defines a collection of metadata content elements as core queryable and returnable elements (see OGC 07-006r1). The base CSW specification adds a bounding box as a core queryable requirement for any CSW catalog. Any CSW server must be able to search using criteria based on core queryable elements, and must include the core returnable elements in csw:record XML response documents (although element values may be nil).

Figure 1. Functional components of National Geothermal Data System. A variety of implementation choices are available for each of the components. Components on the left are mostly hosted by system servers, and interact with the client components on the right through a collection of interfaces defined by the service profiles.



DOCUMENT REPOSITORY

Data in documents will be accessed via URL from document repositories, which are applications that manage a collection of web-accessible files. In this context, 'document' is used in a very general way as a packaged body of intellectual work with an author (or editor, compiler, or similar originating role), a title, and some status with respect to Review/authority/quality. Documents can be packaged in a single file (e.g. a MS Word document) or a group of related, linked digital files (e.g. ESRI Shape file). Documents provide a straightforward path to get data online quickly and easily for the data provider, but if this approach is used for datasets (e.g. Excel spreadsheets, Microsoft Access databases), it requires the data consumer to do all data integration work themselves. In addition, for the datasets to be useful for data consumers, the metadata descriptions must clearly define the entities and attributes (or features and properties) of the datasets such that users can understand their meaning.

Many options are available for implementing document repositories, including DSpace (FOSS, http://www.dspace.org/), OCLC ContentDM (commercial), Fedora (FOSS, http://fedora-commons.org/), CKAN (FOSS, http://CKAN.org) and the Drupal-based document repository developed in collaboration with the USGIN project (http://repository.usgin.org). The NSF

DataOne project is working on an environmental data community repository architecture based on FOSS software.

The bottom line is that documents meant to be accessible on the World Wide Web over an extended period of time (never say forever…) need to be curated, with a scheme for assigning permanent identifiers to the document that will dereference to get a file representing the document. Actual web locations typically change over time, so document identifiers need to be engineered with a redirection scheme. The identifier scheme must have a well-known and reliable dereferencing service (e.g. the Domain Name System on the Web), and a way for a resource (i.e. file…) steward to update the web location associated with the identifier. PURL and the Handle system are examples of technology for implementing such permanent identifiers using the http identifier scheme. The USGIN project has implemented a URI redirection service at http://resources.usgin.org that is currently used by AASG projects for the NGDS.

In order to register (make known) resources in document repositories, a repository must make metadata for contained resources available using the USGIN ISO metadata profile in a way that can be harvested into the NGDS catalog system. This metadata must contain the required minimum content to allow discovery and access to any document in an NGDS repository, including a URL that will retrieve the resource.

Data Servers

A Data Server is any component that implements a service providing data using at least one protocol and interchange format conforming to an NGDS specification. Data service delivery of content differs from the simpler document-based delivery because it requires that the format and content delivered will conform to some know set of rules (a protocol), allowing software to interact directly with the data server to facilitate user acquisition and integration of data into their work environment.

Infrastructure Server

An operational NGDS will require a collection of functions that must be available on a system wide basis. Scoping and description of these infrastructure capabilities is outside the scope of this document.

Database and File System

Various databases and file systems accessed by server applications will house the actual system resources. For security and simplicity, these will probably not be directly accessible for system users, but will be accessed through NGDS service interfaces implemented by other components. Many user applications may also have local data stores in databases or file systems, used to cache resources obtained from the system for offline usage, better performance, and reliability.

Clients

The client applications implement most of the desktop analytical and search functionality required by the system. These are outside the scope of this data-access system architecture except for the provision that they operate with the NGDS catalog for resource discovery and evaluation, and utilize NGDS services and repositories for data access.

Data Delivery to the NGDS

An information resource will be considered part of the NGDS system when it is locatable using an NGDS catalog search, and accessible via the web according to procedures described in the metadata record obtained from the NGDS catalog. The first step to make any information resource part of the system is to create a metadata record describing the resource and how it can be obtained. For guidance on metadata preparation, see Metadata section. Data access in the NGDS is discussed from two points of view. First is the data delivery platform, which may be file-based, via web-application, or service-based. Second is the data formatting and integration, which is ranked in three tiers: Tier 1-unstructured; Tier 2-structured; and Tier 3-structured in a standard content model.

Data Delivery Platform

File-based:

This is the simplest approach. A file-based representation of the resource is made available at some web location. The file should be assigned a URI that is persistent and guaranteed to dereference to get the file on the web for the duration of the file lifetime.

Web applications

Information resources can be made available to users via web-browser based applications that allow users to browse, view, process, analyze, or download data in various ways. Such approaches can provide useful functionality, but do not lend to interoperability or resource reuse, because the application functionality is typically tightly coupled to a particular data source. In such a case, application function cannot easily be applied to other data sources, and the data cannot be accessed directly by other applications. Download function commonly allows some filtering and selection of data subsets, but these operations required direct human interaction. The acquired files may be structured or unstructured (see Data Formatting and Integration Tiers).

It is the intention of the NGDS that applications use service interfaces to acquire NGDS data that they utilize. In this fashion the work of the data compiler, data hosting, and application development agents in the system can be decoupled. It is entirely conceivable that applications may acquire and cache data of interest locally for performance or privacy concerns.

Example web applications:

- Nevada Bureau of Mines and Geology Geothermal Web Application
- Geothermal Prospector
- EarthChem
- System for Earth Sample Registration

Service-based approaches

The ultimate objective is to make all structured NGDS data available through web services using well documented community specifications for service protocols and data interchange formats. This sort of data delivery requires more sophisticated client and server software stacks, and more rigid quality control of the content. Because of the additional up front cost, only widely available and useful data are deemed of sufficient value to warrant web service delivery for the foreseeable future. Web services may be deployed using custom data schema corresponding to existing database holdings, but the goal is to facilitate use by data consumers by delivering data of a particular type in a consistent schema. NGDS interchange format currently deployed use simple feature data interchange/exchange schema for service-based delivery--these are 'flat' file formats that can be represented as simple spreadsheets or text tables with no information loss. These are developed as content models independent of a particular implementation, documented using Excel workbooks.

NGDS Project Partners can implement web services on their own servers (preferred approach), or by arrangement with another project partner or organization. Data delivered by NGDS services should conform to one of the NGDS content models.

Data Formatting and Integration Tiers

Tier 1. The information is made available in an unstructured form, such as narrative text, graphic images (raster or vector), sound recordings, movies, etc. Users must be able to determine the correct software to access file content, and must be able to understand and interpret the file content for the information to be useful. This is the easiest for the data submitting agent, but requires the most effort by the data user. It is still the norm for much data acquisition, but one of the purposes of the NGDS is to take data delivery to higher tiers to realize more value.

Tier 2. The information is loaded in a structured format such that the information can be processed using computer software. Structured data can be made available as a file, or exposed through a web service. To be useful, the metadata must fully describe the data structure, e.g. the tables and fields in each table are defined, or the elements in the content provided by the service.

Tier 3. The information is presented using a documented, community data structure and made available as a file. In this case the metadata can point to the data model specification that is used.

Data Acquisition Plan

This data acquisition plan is a road map for bringing data into the information infrastructure that is the foundation of the NGDS. The intention is to get the NGDS off the ground, with useful data content, as quickly as possible by using existing, tested Open Geospatial Consortium services, particularly Web Map Service (WMS) and Web Feature Service (WFS). In a nutshell, the steps in this plan are:

1. Identify the kinds of information to be made available through the system.
2. Prioritize acquisition according to availability, importance for geothermal resource evaluation and development, and difficulty of acquisition.

3. Make data resources accessible
    a. For document based resources and datasets that do not have specifications for interchange protocols, data schema, and file format: create metadata for resource and make resource available in a web-accessible location linked to from the metadata (Tier 1).
    b. Data types for which NGDS data acquisition services and interchange formats have not been specified will be made available in user-defined data files that will be described by metadata in the system catalog and placed in web-accessible servers (Tier 2).
    c. For high value datasets with sufficient volume, design and implement xml schema based on any applicable standards to use as an interchange format in WFS service response documents, and make the data available through WFS service (Tier 3). Metadata describing service function and content go in catalog.
    d. Gridded data sets (depth to bedrock, Bouguer gravity anomaly…) will be served using NetCDF and OpenDAP on THREDDS servers.
    e. Map-based portrayals of information can be made available as documents, and as WMS service layers. Metadata describing map content and distribution points go in catalog.

Standardization of automated, interoperable data acquisition via services and community interchange formats will be developed incrementally, starting with highest priority data types. Priority will be determined by data availability and requirements from application developers in the user community working on client software useful for geothermal resource development.

For interoperable data to be presented to the system using standardized protocols, interchange formats, and vocabularies, the development team will need to work with the user community (data providers and consumers) to determine a useful starting collection of attributes for entities or features that will be delivered, including units of measure and required controlled vocabularies. Interoperability means in practice that software will use the same access protocol for a given kind of information from any NGDS data provider, without any provider-specific customization. Some important requirements include:

1. Ensure interoperability among data sets with members adopting common standards and protocols.
2. The data schema must be vetted with stakeholders
3. Data schema for interchange formats, and instance documents based on these schema must be versioned, such that expanded or modified versions can be introduced without disrupting working systems.

The process of identifying kinds of information to be made available will be pursued on two fronts. NGDS consortium members were polled in August and September, 2012 to get an inventory of the resources that they will be contributing to the system, mostly comprising scanned well logs and other kinds of documents (Tier 1). State geological surveys were polled under the auspices of the AASG geothermal data project yielding a variety of data items to be made available through the system. A list of identified data items (resource types, entities…) of interest developed through the AASG project is presented in Table 1. Content models have been developed for many of these data items.

Table 1 Summary of data items compiled from AASG data providers (5/27/2010). This inventory of data items will be updated and revised based on continuing input from NGDS consortium members, state data providers, and other geothermal data producers. The categorization scheme is explained in the next section.

Data Item	Category	Notes for data product
Borehole lithology log dataset	Feature	Lithology log consists of collections of intervals defined by top and bottom coordinate in borehole trace, and association with lithology description; each of these intervals is treated as a feature analogous to a geologic unit outcrop feature for geologic map data. Lithology descriptions will include original recorded text, lithology categories from CGI vocabulary, and other properties. We anticipate using the GeoSciML portrayal schema (https://www.seegrid.csiro.au/wiki/CGIModel/GeoSciMLThematicView) for these descriptions, with extensions if necessary. Logs will be associated with a well header identifier, and metadata for the original description. A paper copy of a log or a scan of a paper log is considered a kind of document, and would be integrated into the system by inserting a metadata record for the document in the catalog.
Digital well log	Coverage	A dataset that consists of a collection of measurements of some physical property as a function of depth in a borehole. WITSML (Energistics), NetCDF, LAS are possible interchange formats.
Temperature depth log	Coverage	This is a kind of well log coverage, with the sampling frame corresponding to a borehole, the spatial reference is length measured along the borehole track, measured property is temperature.
Aquifer temperature map	Document	See geologic map. If spatial data are points with temperature measurements, should be considered a
Document	Document	A document is a packaged unit of content with a single authorship (which may include several people). Examples include books, reports, journal articles, geologic maps, other kinds of maps. Internal content within a document is generally not individually identifiable (unlike records in a dataset).
Geologic map	Document	Geologic maps will be made available through one or more of several mechanism: Download of image file (tiff, jpg, or pdf), ideally georeferenced; an OGC Web map service based on a map image (no getFeatureInfo) or better on vector data; or as a vector data set. Map images as files may be published in repositories
Geothermal map	Document	see geologic map
Gravity map data	Document	See geologic map. Spatial data may be grid or contours. Service-based delivery of grid data will probably use NetCDF or OGC WCS.
Resource suitability map	Document	See geologic map. Spatial data could be coverage or polygons.
Active Fault	Feature	A GeoSciML GeologicStructure feature or or GeoSciML Portrayal fault feature; attributes should include at least a statement of evidence for fault being active; ideally includes orientation information, time since last displacement, magnitude of displacement, and displacement rate.
Geologic Unit feature, geothermal characterization	Feature	A geologic unit description specifying properties important for geothermal energy evaluation; includes standard aquifer properties like lithology, permeability, porosity, as well as thermal properties like thermal conductivity and specific heat.
Geologic unit feature, Alteration description	Feature	Base on GeoSciML alteration description. Properties that are useful for interpreting nature of paleo hydrothermal systems that have effected rock body.

Data Item	Category	Notes for data product
Geothermal system feature	Feature	Feature characterizing a geothermal system; properties include reservoir volume, average temperature.
Geothermal system feature, Enhanced	Feature	Similar to geothermal system feature but including information characterizing any enhancement processes and their results.
Hot spring description	Feature	A hot spring is a kind of water source with temperature data for the water produced, along with other properties associated with a water source (location, flow rate, water chemistry data, time series for flow and chemistry?).
Intrusive body with heat	Feature	Treat as GeoSciML Geologic feature with geothermal characterization
Sample	Feature	Associated with site; becomes sampling feature that is an accesses point for variety of other observations, analogous to a borehole as a smapling feature.
Volcanic vent feature	Feature	Volcanic vent has a location that may be represented by a point or polygon. A complex data model could be imagined, including eruption history, magma composition, fluid and gas compositions, eruption rates, associated heat flow measurements, associated magma body… For NGDS purposes, we need to identify the key properties of interest as a starting point.
Water source feature	Feature	possible characteristics of interest include chemical, physical, and flow-related properties. A soft-typed approach with {property, measured value} pairs is recommended. We will confer with partners providing this kind of information to establish practices for common kinds of characterization. We anticipate being able to use CUASHI practices for much of this kind of data.
Bottom hole temperature	Observation	Include supporting information for the borehole, including location, type of hole (petroleum, mining, groundwater), etc, depth of the measurement, time since between stopping circulation and measurement, the diameter of the borehole at the measurement point.
Chemical analysis (whole-rock chemistry)	Observation	Individual records will be a collection of {measured constituent, abundance pairs}, with identifiers for the analyzed sample, and analysis procedure. The procedure is considered to include the who, how, with what equipment information for the analysis.
Crustal Stress data	Observation	Associated with a site sampling feature. Various observation types may be necessary to account for different experimental procedures.
Drill stem test	Observation	An observation feature that includes the results of a drill stem test. Key observation results, include equilibrium pressure, fluid composition, temperature.
Earthquake hypocenter	Observation	Treat as observation because hypocenter (or epicenter) location is always the result of a measurement and analysis process that depends on a collection of seismometer recordings. The hypocenter (epicenter) can also be conceived as a feature, with the observation as metadata for definition of the feature.
Flow rate	Observation	Always an attribute of a water source feature or of a water channel feature.
Fluid inclusion data	Observation	Associated with a sample; needs content model
Gravity Station data	Observation	Associated with a site sampling feature. Include raw observation data, corrections, observer, instrument.

Data Item	Category	Notes for data product
Heat flow measurement	Observation	Heat flow measurements are based on a temperature gradient measured over some interval, and a thermal conductivity value for the material between the two temperature points. The location and temperatures defining the gradient and estimated conductivity must be reported in a complete heat flow report, along with procedure metadata.
Permeability	Observation	May be reported through observation service associating individual samples with permeability measurements. Permeability may also be reported associated with a geologic unit in a
Temperature	Observation	Attributes will include temperature, units, X, Y, Z coordinate, borehole identifier, and measurement procedure. Different measurement procedures will need to be documented. Bottom hole temperature data is one kind of borehole temperature data. Temperature log datasets (as opposed to scanned log documents) are treated as a kind of digital well log. Typically is measured in a borehole to be geothermally interesting.
Thermal conductivity measurement	Observation	Observation feature, attributes include identification of sample used for measurement, procedure, and result with uncertainty.
Trace constituent chemistry dataset. (Water Chemistry, trace-element data)	Observation	Trace element chemical analyses report concentrations of constituents that do not form a significant part of the total material. Most water quality or water chemistry data fall in this category, as well as rock trace-element data. Individual records will be a collection of {measured constituent, concentration pairs}, with identifiers for the analyzed sample, and analysis procedure. The procedure is considered to include the who, how, and with what equipment information for the analysis.
Production statistics record		Content will need to be worked out by experts.

The data acquisition process will be planned to focus on delivering information to enable use cases being implemented by the Geothermal Desktop application in order to make utilization of implemented functionality immediately useful.

Data Item Categories

The term 'Data item' used here to denote a kind of information object; in other contexts these might be called entities or objects. Each item is associated with a 'type' that specifies a collection of properties. The data items are grouped into broad categories to group types that will use the same data schema, delivery mechanisms, and metadata schema. Initial data types were identified and draft content models developed by surveying project partners at the beginning of the project. The approach was bottom up, based on the kinds of information project partners reported in their holdings. The Data Items are grouped into higher-level information categories that correspond to the 4 main categories of services in the system architecture. These are:

Observation – an information resource representing the event of observing or measuring and recording properties of some feature (Open Geospatial Consortium, Observations and Measurements (O&M),http://www.opengeospatial.org/standards/om). Observations represent the basic data that are the foundation for the other information categories. The content model includes: a result, which is the measured or observed value; a feature of interest, which is the feature the observer wishes to characterize; a procedure, which includes information on who made the observation and how it was made. A sampling feature may be specified to record what part of the feature of interest was the actual target of the observation. The observation model allows modeling composite observations, which may represent the aggregation and interpretation of one or more input observations. At least initially we will be using a flattened, simplified version of the full O&M schema that can be served using OGC simple features.

Feature – an information resource representing some identifiable thing of interest in the world. A feature is described by a collection of attributes that are typically each the result of one or more observations. Features present a more aggregated or interpretive view of the world than observations (although a feature can be modeled as the result of an observation). Features will be delivered via OGC Web Feature Services (Open Geospatial Consortium, Web Feature Service, http://www.opengeospatial.org/standards/wfs) or other similar services. Typically, features have a geographic location.

Document -- A packaged body of intellectual work; has an author, title, some status with respect to review/authority/quality. 'USGS peer reviewed' would be a 'status property'. We will have to account for gray literature, unpublished documents, etc. A document may have a variety of physical manifestations (pdf file, hard-bound book, tiff scan, Word processor document…), and versions may exist as the document is traced through some publication process. They may be map, vector graphics, or text. Sound, moving images are included as document types. Documents will typically contain or summarize information from a variety of features and observations. This information may be unstructured text or image, non-standard structured data, or standard structured data.

Coverage – A dataset that reports the values of some continuously varying property over some spatial or temporal extent. Examples include well logs that report the values of resistivity, density, or some other property along the well bore, gravity maps that report measured (perhaps by extrapolation) values of gravity over some geographic region (e.g. a digitize log in an LAS file, not a scanned log image, which would be a document). A coverage may be the result of one or more observations. A coverage may also be associated with an individual feature, such as a map showing the thickness distribution of a geologic unit or the average temperature at some depth in a Geothermal system.

In general, measured quantities should be reported in the units of the original measurement, as well as in units that will be requested for data integration. We propose that the original data provider is most likely to provide accurate unit conversions, as opposed to leaving unit conversions to the data consumer or service provider.

METADATA

Metadata is data that describes an information resource. The metadata content is intended to support resource discovery, evaluation, and acquisition by users of the NGDS. Metadata typically includes a title, text summary of the information content of the resource, when the resource was published (what constitutes publication depends on the kind of resource), who is the authority responsible for the accuracy of the resource content, information about the geographic area that the resource content is related to, and keywords that categorize the resource in various ways to increase search efficiency. These are basic parameters used for information search. To be useful, the metadata must also inform a user on how to access the described resource. Ideally this access information is provided in a machine-interpretable format so that client software can take a user directly to the resource without any user intervention beyond invoking a 'get data' operation. More in depth metadata provides documentation on the provenance and quality of the resource that can be used to assess the resource for fitness before the user goes to the effort of acquiring it.

It should be noted that metadata in the NGDS catalog is scoped to describe resources at the level of documents or datasets. Feature level metadata that describes provenance, acquisition processing, and quality properties at the level of individual observations or features may also be available, but this information is delivered through a data service, not a metadata catalog service.

METADATA PRACTICES FOR NGDS

Metadata should be created and submitted to an NGDS catalog for any resource that is meant to be accessible individually via the web. Individual data items are accesses through data services; the dataset that contains the item is described by a metadata record, which also indicates the services or files that may be accessed for item-level processing.

Individual documents require one metadata record per document. Some document types may consist of a collection of files, e.g., ESRI shape file. In general these should be bundled into a single file like a zip archive or UNIX tar file. The metadata must include the URL at which the document can be accessed. These documents might be scans of well logs, scanned reports or publications, or data in a spreadsheet, such as an Excel file.

Datasets (data products) are typically considered as individual works (see FRBR), unified by the compilation activity that brings information together into a single data structure, editing and verifying content as necessary. This approach is based on our interest in facilitating data access by users, and the observation that such a user is first interested in a particular kind of data, and upon finding a fit-for-purpose dataset, will next want to know how to get the data. From this perspective a dataset work (data product, resource) will have a single metadata record that may include specification of multiple distributions of the resource. For instance a borehole temperature dataset may be available with all the records in an Excel spreadsheet, or visualized through a web map service, or individual observations may be accessed through a web feature service (see for example Montana Thermal Springs).

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 the data. These feature-level metadata are delivered with the data, and only summarized in the metadata records 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.

Table 2. Minimum required metadata content in USGIN metadata content recommendations.

Content item	Cardinality	Scope notes
Title	**1 entry**	Succinct (preferably <250 characters) name of the resource.
Description	**1 entry**	Inform the reader about the resource's content as well as its context.
Originators	**1 to many**	Authors, editors, or corporate authors/curators of the resource.
Publication Date	**1 entry**	Publication, origination, or update date (not temporal extent) for the resource. Use a "year" or ISO 8601 date and time format. Alternative date formatting must be machine readable and consistent across all datasets. If no publication date is known, estimate the publication date range, enter the oldest year as the publication date, and include the estimated date range in the Description field.
Geographic Extent - *Horizontal*	**1 entry**, point or minimum bounding rectangle	Values given in decimal degrees using the WGS 84 datum. Required if resource has location Some resources may not be usefully described by an extent; if no extent is specified the default is Earth.
Access Instructions	**1 entry**	Text description of how to access the resource.
Distribution Contact Party	**1 entry**	The party (name of organization or person, etc.) to contact about accessing the resource.
Distribution Contact Email	**1 entry**	How to contact the party responsible for distribution
Metadata Date	**1 entry**	Last metadata update/creation date-time stamp in ISO 8601 date and time format. This may be automatically updated on metadata import if a metadata format conversion is necessary.
Metadata Contact Party	**1 entry**	The party (name of organization or person, etc.) to contact with questions about the metadata itself
Metadata Contact Email	**1 entry**	How to contact the party responsible for metadata content and accuracy
Metadata Specification	**1 entry**	Identifier for metadata specification used to create a metadata record encoding this content

Table 3. Recommended metadata content. This table includes the minimum content requirements (from Table 1) with recommended content to produce useful metadata to describe resources, credit the originator of the resource, and inform users how to obtain or access a resource. The resource description should provide sufficient information to assist in discovery of the resource through an online search, and to allow users to evaluate the fitness of the resource for an intended purpose. Explanation of fonts used: *Terms in italics are groupings of metadata properties*; **required**, conditional, and optional metadata content; (number of values that can be specified are in gray).

Content item	Cardinality	Scope notes
Resource		
Title	1 entry	Succinct, preferably <250 characters, name of the resource.
Description	1 entry	Inform the reader about the resource's content as well as its context.
Originators	1 to many	Authors, editors, or corporate authors/curators of the resource.
Publication Date	1 entry	Publication, origination, or update date not temporal extent) for the resource. Use a "year" or ISO 8601 date and time format. Alternative date formatting must be machine readable and consistent across all datasets. If no publication date is known, estimate the publication date range, enter the oldest year as the publication date, and include the estimated date range in the Description field.
Geographic Extent - Horizontal	1 entry, point or minimum bounding rectangle	**North Bounding Latitude**, **South Bounding or Point Latitude**, **East Bounding Longitude**, **West Bounding or Point Longitude.** Values given in decimal degrees using the WGS 84 datum. Some resources may not be usefully described by an extent; if no extent is specified the default is Earth. This convention would have to be modified for systems describing extraterrestrial resources. For resources for which the geographic extent is not applicable, the recommendation is to include the keyword ‘nonGeographic’ in the subject [spatial] keywords element. If a particular encoding scheme requires a bounding box, a very small bounding rectangle will be created if only a point coordinate is given; the asserted point will be the southwest corner of the bounding box.
Contact - Author or Intellectual Originator	0 to 1	The primary party responsible for creating the resource. **Organization Name**, **Person Name**, **Street Address**, **City**, **State**, **ZIP Code**, **Email**, **Phone**, **Fax**, **URL**. If contact information is provided, include at least the organization or author name.
Bibliographic Citation	0 to 1	Full bibliographic citation if the resource has been published.
Subject Keywords	0 to many	Thematic, spatial and temporal free-form subject descriptors for the resource. A keyword may be assigned on metadata import if none are present. If possible, submit keywords in separate Thematic, Spatial, and Temporal keyword categories.
Resource Language	0 to 1	Use three letter ISO 639-2 language code [defaults to "eng" for English].

Content item	Cardinality	Scope notes
Resource ID	0 to many	Resource identifier(s) following any public or institutional standard. Identified consists of an identifier string and if applicable a **Resource ID Protocol** identifier string that specifies the protocol for the resource ID standard. For example: undefined, ISBN-10, ISBN-13, ISSN, URN, URI, IRI, DOI, HTTP, SSN, etc. Examples: doi:10.1000/182; isbn:0-671-62964-6; issn:1935-6862; azgs:OFR-10-02 Many protocols build the identifier for the protocol into the identifier string.
Geographic Extent – Vertical	0 to 1*	**Datum Elevation**, **Upper Boundary, Lower Boundary, CRS**. The upper and lower boundaries of the extent are reported relative to the datum elevation. This will typically be the Earth surface at the location of the resource, or mean sea level [MSL]. The coordinate reference system CRS value [normally a URI] specifies the datum for vertical measurement and the conventions used to define the coordinates of the extent boundaries relative to the datum positive up, positive down, units of measurement). The upper boundary is farther from the center of the earth than the lower boundary. *The same vertical extent may be reported using different coordinate reference systems in the same record to accommodate different conventions for searching. A borehole with collar elevation 4787 feet above sea level has core extracted at depths between 100 and 470 feet. The vertical extent of the core could be reported in either of the following ways: {0, CRS datum MSL positive up meters, 1420, 1308} or {1450.6, CRS datum Earth surface positive up meters, -30.3, -142.4}.
Temporal Extent **Start Date** **End Date**	 0 to 1 0 to 1 required if start date exists	Temporal range over which the resource was collected or is valid. If the resource pertains to specific named geologic time periods, those terms should be entered as keywords (preferable as part of Temporal Keywords), use ISO 8601 date and time format: YYYY-MM-DDTHH:MM:SS.
Quality Statement	0 to 1	Text specification of the quality of the resource.
Lineage Statement	0 to 1 entry	Text description of the resource's provenance.
Resource Access information		
Access Statement	1 entry	Text instructions for how to access the resource.
Distribution Contact	1 entry	The party to contact about accessing the resource. **Organization Name**, **Person Name, Street Address**, **City**, **State**, **ZIP Code**, **Email**, **Phone**, **Fax**, **URL**. In general, a contact for distribution should be required for physical resources.

Content item	Cardinality	Scope notes
Link to the resource	0 to many entries	A URL pointing to a resource or resource webpage. **URL**, **Link Function**, **Representation Format**. URL is minimum content required if a link is included. A Link Function term from a controlled vocabulary to specify what a HTTP GET using the URL will invoke is recommended. The link might return an html page, electronic document in some other format, an end point for a service, an online application that requires user interaction, etc. Representation Format is a controlled vocabulary term specifying the format [MIME media types] of a file-based response if applicable.
Constraints Statement	0 to 1 entry	describe the resource's legal and usage constraints.
Distribution Keywords	0 to many entries	keywords describing the physical form of the resource (core, rock sample, digital file, book, journal article), formatting of resource content (file format, e.g. tiff, xls, MIME type), or physical distribution media (film, floppy disk, online service, hard copy). Table 6 in USGIN ISO metadata profile includes a vocabulary for distribution format for use with the ISO19115 distributionFormat name property. Use of these keywords allows users to search for particular kinds of artifacts
Metadata management information		
Metadata Date	1 entry	Last metadata update/creation date-time stamp in ISO 8601 date and time format. This may be automatically updated on metadata import if a metadata format conversion is necessary.
Metadata Contact	1 entry	The party to contact with questions about the metadata itself. **Organization Name**, **Person Name, Street Address**, **City**, **State**, **ZIP Code**, **Email**, **Phone**, **Fax, URL**.
Metadata Specification	1 entry	Identifier string for the specification to which metadata record encoding is declared to conform. Should indicate the base standard and version, as well as any profile that applies to the content or encoding. Ideally the identifier could be dereferenced to obtain information about the applicable specification. Identifiers for metadata encoding specifications to be used in the USGIN and NGDS systems will have to be formally defined and registered for such identifiers to be broadly useful.
Metadata UUID	0 to 1 entry	A Universally Unique Identifier [UUID] will be assigned during the metadata import process if one is not provided. Unique identification of each metadata record is included to reduce duplication of entries across multiple metadata catalogs. The UUID format provides unique identification without centralized coordination.

The required metadata content is listed in Table 1, recommended metadata content is summarized in Table 2. The metadata content requirements are explained in detail in USGIN metadata recommendations. These requirements proscribe the content of the metadata, but not the delivery format. ISO19139 xml is the preferred encoding based on its expanding adoption in the community. USGIN guidelines for implementing the content recommendations in ISO19139 XML are available, as well as a detailed USGIN profile for interoperable metadata using ISO19139.

FGDC XML metadata records can be transformed to USGIN ISO 19139 format and harvested to the NGDS catalog. FGDC XML should be tested to validate against the official XML schema. Please confer with the NGDS developer team about metadata formatting to facilitate import of metadata into the NGDS catalog.

GeoSpatial indexing

User research as part of the NGDS Design and Testing project indicates that location-based search is the primary approach by which NGDS users will access information. The USGIN metadata recommendations (Table 1) require a bounding box in geographic coordinates (latitude-longitude) to associate a resource with location. In practice, there are many resources of interest for the NGDS that do not have metadata including geographic coordinate locations. Assignment of bounding boxes can be a time consuming process. In the near term, it is expedient to update existing metadata using location keywords relevant to geothermal resources. In the Phase one Data inventory report, work activities have been added estimating effort required to add location keywords for metadata that does not currently have any usable location information. Data submitters will be requested use the Geothermal Area names from NREL's OpenEI (http://en.openei.org/wiki/Geothermal_Areas) when applicable.

Content Models and XML schema

Each feature or observation type that is prioritized for delivery using a community data interchange format has an associated content model that defines the information contained in documents that deliver data from the data system servers to client users. The content model in its simplest form is a statement that some particular entity (feature or observation) will have some list of properties. For example a content model for a record describing a book feature might include: title, author, publisher, publication date, publication place, number of pages, and an ISBN number (a URI).

To document these content models, we are using MS Excel workbooks. Each workbook includes a description of the entity type, a list of the elements (fields, attributes) in a data instance for that entity presented both as a table and a list, and lists of vocabulary terms meant as a start or example for the kinds of vocabularies that will be required in the long run for semantic interoperability. Use of the terms in these example vocabularies will facilitate migration to formal vocabularies later.

In order to actually use the content models, they must be implemented using an encoding scheme that can be parsed by computer. The NGDS is currently using eXtensible Markup Language (XML) to implement interchange content models, but in view of the rate of technology evolution, other schemes may be adopted as the system matures. The important thing is that the information is encoded consistently in a structured, well documented format. This enables computer programs to parse the interchange documents and extract the desired information, and as newer encodings are adopted, conversion between formats can be automated using software.

Since many of the data types are associated with geographically located features, the Open Geospatial Consortium Web Feature Service (WFS) is proposed as the starting point for implementation of feature services. This protocol uses GML geometry for location description, and allows feature types to be defined that are characterized by feature specific xml schema.

A number of international efforts are under way to develop specifications for data interchange of geoscience information (GeoSciML), and basic observation and measurement data (ISO19156). These xml schema are very flexible to allow representation of a wide range of content, but are thus correspondingly complex. Currently there are no client applications that can do more that transform complex xml to html for display.

Thus, project services are defined using simple xml schema with string and numeric-valued elements. These services can be consumed by existing clients like ESRI ArcGIS and Quantum GIS. Simple feature schema will be compatible with GeoSciML, ISO specifications, and other complex standard schema to the degree that is practical. As clients are developed for richer-content complex feature services, the NGDS can adopt more complex, information-rich schema. There are also a number of other data formats in use in related communities for geoscience information interchange, including WaterML in use by the CUAHSI project, and NetCDF in use for large numeric data sets in the atmospheric and remote sensing communities. Where ever possible, NGDS data providers should reuse existing schema to take advantage of tools developed to consume data in these formats.

Development of content models has been an organic process. The models have evolved rapidly as production scale data compilation has gotten under way. WFS services have been deployed as the models evolve, with several iterations of model versions and XML schema or some models. XML schema for NGDS data interchange WFS features are being placed in a USGIN repository. 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 known schema version.

If data to be served are not accounted for by an existing content model, network participants are invited to propose new models. A document with guidelines for construction of a template workbook is available online. Content models may be developed under the auspices of a particular project, but they are submitted for review and adoption for use by other project partners in the NGDS and broader USGIN community. To view applications that are compatible with the NGDS, please visit the NGDS Applications page.

Table 4. Summary of current content model inventory. XML schema for content models in use are at http://schemas.usgin.org/schemas

Content Model	Status	Description
Active Fault	Currently in Use	Content for features representing active fault traces. Note that fault features require a geometry field that contains the geographic location of the fault trace (typically named 'shape' in ESRI feature classes). The actual template for deploying services must be a GIS feature class, thus cannot be implemented as an Excel spreadsheet. Design is focused on faults portrayed as line segments. The content model is based on the GeoSciML thematic view for ShearDisplacementStructure (see https://www.seegrid.csiro.au/wiki/bin/view/CGIModel/GeoSciMLThematicViewModel), extended with some convenience properties specific to young fault structures. Active in this context denotes that there is compelling evidence of fault displacement during the Quaternary.
Aqueous Chemistry	Currently in Use	Content requested for basic data characterizing a chemical analysis of an aqueous fluid for geothermal data. Typically water temperatures are recorded. A basic content model for sample characterization, location, and analysis metadata is proposed, along with several 'suites' of analytes representing common analysis results. A feature for reporting single analyte results is also proposed; use of this approach would decompose an analysis into one feature for each analyte for each analysis.
Borehole Temperatures	Currently in Use	Content model for borehole temperature observation features; specifies content elements for an interchange format for temperature measurement data obtained from boreholes. Fields in the spreadsheet view will become XML elements in interchange documents for WFS simple features/geothermal data web services. Typically bottom-hole temperatures (BHT) are recorded from log headers, and this information will be provided through a borehole temperature observation service. Subsurface temperature data may also be obtained by measurements targeted specifically to obtain thermal data. The HeaderURI for a particular borehole (well for simple wells) is the cross-referencing link (foreign key) used to associate the header record, well logs, temperature measurements, and other information from a particular borehole.
Direct Use Sites	Currently in Use	Content requested for basic description of sites utilizing geothermal energy for space heating, greenhouses, resorts and other direct use applications.
DrillStemTests	Currently in Use	Content model for the interchange of drill stem test observation results by the AASG geothermal data project for the National Geothermal Data System. The HeaderURI for a particular borehole (well for simple wells) is the cross-referencing link (foreign key) used to associate the header record, well logs, temperature measurements, and other information from a particular borehole. This model is in the process of being updated to generalize for any kind of well-flow test.
Seismic Event Hypocenters	Currently in Use	Content requested for features representing Earthquake hypocenter observations. Design is focused on EQs hypocenters portrayed as points. Hypocenter data for geothermal data system is intended as a tool to identify seismically active areas that are often associated with hydrothermal activity, thus the content model does not include detailed information that would be important for seismological analysis; such information should be accessed by including related resource links. As such,

Summary

The central idea of the data access architecture proposed here is that data providers and client applications should be linked through open source interfaces that decouple clients and servers such that they can evolve independently without breaking the system. The hypertext transfer protocol (http) and hypertext markup language (html) are the established protocols and interchange formats in use on the internet, and in the near term these will probably continue to be the mainstay of most interaction in the NGDS.

The OpenGeospatial Consortium Catalog Service for the Web (CSW), currently at version 2.0.2 is proposed for catalog search and discovery service. The USGIN profile for ISO metadata is proposed for metadata interchange in the NGDS, and is already in use by the NGDS catalog deployed for the AASG Geothermal Data Project.

Initial data services are being implemented using WFS 1.1.1 simple feature services, with interchange formats defined for key geothermally interesting datasets. Interchange content is encoded as GML Simple Features, an XML application scheme.

File based data are incorporated into the NGDS by creating a metadata record conforming to the metadata specification and inserting it into the NGDS catalog. This may be done by creating the metadata through a forms interface that is directly connected to a catalog database, or by creating an XML metadata record and placing it in a registered web-accessible folder to be harvested by an NGDS catalog. Files containing the data must be in a web-accessible location specified by a URL in the metadata record.

Gap Analysis:

Conversion of Public Land Survey System (PLSS) locations to latitude-longitude. Many wells have locations reported in PLSS coordinates (Township-Range-Section-Section part). If the syntax for the PLSS location is consistent within a given dataset, these can be matched to center points from BLM PLSS shape files by automated processing.

Boundaries for named geothermal areas. We plan to use the Geothermal Areas from NREL OpenEI. Most of these seem to have point locations; need to investigate whether polygons could be drawn to more accurately locate the areas. Areas without at least a point location will need to be located.

Geologic map index. Need to figure out how to link USGS National Geologic Map Database to NGDS catalog so geologic maps appear in search results. We may need to mine their web pages.

Glossary of Terms for Geothermal Data System

Aquifer: A geologic unit that is a hydrologically connected body of material and contains water.

Artifact: A thing created by humans, usually for some practical purpose.
(Source: http://www.merriam-webster.com/dictionary/artifact)
Attribute: A binding between a property, a data type, and a data item; an implementation of a property.
Cardinality: A constraint on the number of instances of assigned property values associated with an individual data item. A cardinality of 1 indicates exactly one value is required; 0..1 indicates an optional single value; 1..n indicates that one or more values is required; 0..n indicates that a value is optional, and multiple values may be specified.
Content model: A model that identifies and defines the data items and the properties (with cardinality) associated with each data item.
Data item: An identifiable unit of information. Generally represents some entity in the world.
Data type: A specification of the representation of a single value in an information system, using integer, floating point, string, Boolean.
Dataset: An identifiable collection of information, characterized by its scope, authorship, theme, purpose and data schema.
Document: a packaged body of intellectual work with an author (or editor, compiler, or similar originating role), a title, and some status with respect to Review/authority/quality. A document has 1 to many manifestations in FRBR terminology (see http://www.loc.gov/cds/downloads/ FRBR.PDF for details), which in web terminology would be called representations.
Feature: An information resource representing some identifiable thing of interest in the world.
Feature type: Type for representing a feature.
Geologic structure: A configuration of matter in the Earth based on describable inhomogeneity, pattern, or fracture in an Earth material. The identity of a geologic structure is independent of the material that is the substrate for the structure. Geologic structures are more likely to be found in, and are more persistent in, consolidated materials than in unconsolidated materials. Properties like "clast-supported," "matrix-supported," and "graded bed" that do not involve orientation are considered kinds of geologic structure because they depend on the configuration of parts of a rock body.
(Source: http://www.geosciml.org/geosciml/2.0/doc/GeoSciML/GeologicStructure/Geol...).
Geologic unit: A body of material in the Earth whose complete and precise extent is inferred to exist (NADM GeologicUnit, Stratigraphic unit in sense of NACSN or International Stratigraphic Code), or a classifier used to characterize parts of the Earth (e.g. lithologic map unit like "granitic rock" or "alluvial deposit," surficial units like "till" or "old alluvium").
(Source: http://www.geosciml.org/geosciml/2.0/doc/GeoSciML/GeologicUnit/GeologicU...).
Geothermal system: A body of material in the Earth from which energy may be extracted as heat in a fluid circulated through the body and transported to an external point of use. A geothermal system does not have to be exploited or exploitable. (provisional definition).
Information resource: A resource that can be transmitted electronically.
Property: A phenomenon that is inherent in the nature of some other phenomenon, and may be used to characterize it by specifying a value.

Protocol: A set of rules which is used by computers to communicate with each other across a network (http://en.wikipedia.org/wiki/Network_protocol).
Representation: A binding between a symbol (in language, text, graphics, computer bits, etc.) and a human concept or resource.
Resource: An identifiable thing that fulfills a requirement. Usage here is close to definition used in RDF (http://www.w3.org/TR/REC-rdf-syntax), generalized from ISO19115, which defines resource as an 'asset or means that fulfills a requirement' without defining asset or means. "An object or artifact that is described by a record in the information model of a catalogue." (OGC 07-006r1)
Schema: A formally structured representation of a conceptualization. A model presented using some specific notation.
Specification: A document that describes the technical characteristics of an artifact, possibly including a description of what it should do, or an explicit set of requirements that it must satisfy. (Based on http://en.wikipedia.org/wiki/Specification).
Type: Specification of a collection of attributes used to represent a data item, along with constraints including cardinality and value domain for those attributes

REFERENCES

Dublin Core Initiative, 2008-01-14 Dublin core Metadata Element Set, Version 1.1: Dublin Core Metadata Initiative, accessed at http://dublincore.org/documents/dces/.

Franklin, Michael, Halevy, Alon, and Maier, David, 2005, From databases to dataspaces: a new ab-straction for information management: ACM SIGMOD Record, V. 34, No. 4, ISSN:0163-5808.

Open Geospatial Consortium Catalog Service for the Web (CSW): http://www.opengeospatial.org/standards/cat

Open Archive Initiative Protocol for Metadata Harvest (OAI-PMH): http://www.openarchives.org/pmh/

OpenSearch protocol: http://www.opensearch.org/Home

ISO 19115:2005, Geographic information - Metadata

ISO 19115/Cor.1:2006, Geographic information – Metadata, Technical Corrigendum

ISO 19119:2005, Geographic information - Services

ISO 19119:2005/Amd 1:2008, Extensions of the service metadata model ISO 19108 designates:

OGC 07-006r1¸ OpenGIS Catalog Services Specification version 2.0.2, Corrigendum 2 release, 2007

OGC 07-045, OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile, Version 1.0.0, 2007

USGIN Team, 2010, Use of ISO 19139 xml schema to describe geoscience information resources v1.1, Current version at http://repository.usgin.org/uri_gin/usgin/dlio/337.

USGIN Team, 2010, Metadata Recommendations for Geoscience Resources, current version available at http://repository.usgin.org/uri_gin/usgin/dlio/335.

Appendix 1b

NGDS Design and Testing Subrecipient Data Inventory

ARIZONA GEOLOGICAL SURVEY
416 W. CONGRESS, SUITE 100
TUCSON, AZ 85701

Data Acquisition for National Geothermal Data System

A Report to:
Harold Blackman, PI
Scott Keneman, PM

helpdesk-admin

09/28

Background

The National Geothermal Data System (NGDS) will provide online resources to make it easy for users to extract, assess, and synthesize data according to criteria they select. As part of the NGDS Design & Testing project reboot, held June 2012 at Boise State University, the Arizona Geological Survey agreed to work with NGDS Design & Testing subrecipients (henceforth referred to as "subrecipients") to inventory the data for submission to the NGDS, and work with the subrecipients to develop plans for incorporating their resources into NGDS. We employed similar strategies and techniques developed during the course of the Association of American State Geologists (AASG) contributions to the NGDS project.

Goal

Determine the amounts and types of data currently compiled under the NGDS Design & Testing project, the format of that data, and identify any remaining data (and format) available for inclusion in the system. Data will be discussed in three tiers of interoperability that are described in greater detail in the accompanying data assessment report:

1. Tier 1 -- Serving arbitrary data files via web-servers annotated with metadata in the NGDS catalog;
2. Tier 2 -- Serving structured data in files or via (geospatial) web services annotated with metadata in the NGDS catalog;
3. Tier 3 -- Serving structured tabular data via geospatial web services according to NGDS-adopted community content models annotated with metadata in the NGDS catalog.

The National Geothermal Data System is based on a collection of protocols for metadata and data inter-change. There is no single system data model, rather there is a metadata interchange format and a col-lection of data interchange formats (the NGDS content models, see Data Access Phase 1 deliverable document). Each interchange format defines a data model for an information item of interest, and is de-signed to access particular information tailored to some use case or use cases. Data providers and client applications are linked through open source interfaces that decouple clients and servers such that they can evolve independently without breaking the system. The OpenGeospatial Consortium Catalog Service for the Web (CSW), currently at version 2.0.2 is proposed for catalog search and discovery service. The USGIN profile for ISO metadata defines the metadata interchange format for the NGDS, and is already in use by the NGDS catalog deployed for the AASG Geothermal Data Project. Initial data services are being implemented using WFS 1.1.1 simple feature services, with interchange formats defined for key geo-thermally interesting datasets. Interchange content is encoded as GML Simple Features, an XML application scheme. File-based data are incorporated into the NGDS by creating a metadata record conforming to the metadata specification and inserting it into the NGDS catalog. Files containing the data must be in a web-accessible location specified by a URL in the metadata record.

Inventory Methods

Data providers were contacted using email and phone calls. Webinars were utilized to share views and screens of existing databases and metadata records.

In order to complete the inventory we requested each of the subrecipients to evaluate their data, both structured (tabular, database, spreadsheet) and nonstructured (scanned documents, pdfs, paper files). We requested:

1. An Inventory of the data provider's holdings that are to be included as an NGDS resource, broken down into documents and datasets, providing a best-estimate inventory of how many of these resources already have structured metadata descriptions in some form.
    a. Documents should be categorized into "already online," "in files but need to get online," "to be scanned," and "to be converted to datasets."
    b. Datasets should be categorized into "file-based data," "file-based, using NGDS content model," and "data for deployment as a data service according to an NGDS content model and interchange format." Datasets in NGDS content models may be in various formats, including the Excel workbook templates that have been in use by the AASG project, as long as the information can be exported as a comma-delimited (csv) text table for processing.
2. Provide a description of the data provider's method for compiling metadata, and indicate the application file type (e.g. MS Excel, MS Word, MS Access, FileMaker…) in which the metadata is stored.
3. A face-to-face, telephone conversation, or webinar to review the inventory and current status of metadata, make a plan (process and timeline) for (1) the process of converting existing metadata to a form suitable for import into the catalog, (2) the creation of metadata for remaining resources and its import into the catalog, and (3) the methodology that will be used to deploy NGDS services for selected datasets (if any).

We provided subrecipients with an Excel workbook to assist in the inventory process. The accompanying spreadsheet SOW_Sheets_BoiseNGDS_I.xls (located in the gdsdpwg.net SharePoint site shared documents 'Go_No-Go Drafts in Preparation for 25-October' folder) includes completed statements of work for the inventory compiled by each of the subrecipients. The Results by Recipient section of this document describes the completed inventory and steps required for each subrecipient to submit or host data in the NGDS.

Data Submission

Data Interoperability Tiers

This section expands upon the three tiers of data interoperability in an effort to clarify the workload associated with the subrecipient data inventory. Table 1 provides introductory information on input (data submission) and output (data use) of data in the NGDS.

Data Tier	Type of Data	Input Requirements	Output Requirements
Tier 1	Scanned Documents	Upload the file to a web accessible location; Create metadata for record	Accessible via HTTP GET; Metadata points to the URL accessible via CSW
Tier 2	Structured (tabular) data not conforming to an NGDS content model. Examples include Excel/csv tables, Microsoft Access databases)	Upload files to a web accessible location; Create metadata for record. Metadata should include descriptive information about the column headers and data types that the documents contain.	Accessible via HTTP GET; Metadata points to them via CSW. Optional – data may be exposed as a WFS/WMS.
Tier 3	Structured data conforming to an NGDS content model	Upload files to a web accessible location; WFS and WMS services are deployed to serve data; Create metadata pointing at all endpoints (because the content model defines entity/attribute information the creation of metadata is less cumbersome).	Accessible via HTTP GET; Data exposed as WFS/WMS services; Metadata points to all endpoints accessible via CSW.

Content Models

In order to accomplish Tier 3 data interoperability the use of standardized interchange formats is required. The NGDS content models document collections of properties used to characterize a particular data type of interest to the NGDS; these content models are implemented via XML schema. Data exchange between data providers and consumers is via XML documents that will validate against the XML scheme for the content model instantiated by the document. There are currently over twenty active content models in use for the NGDS system. Content models are developed for data sets that are expected to be submitted by more than one data provider. Data types (e.g. bottom-hole temperatures, heat-flow measurements, rock thermal conductivity) have been identified and implemented for the AASG Geothermal Project. The design and testing project Domain Steering Committee (DSC) has conducted two user focus group meetings to review the geochemical and well log observation content models, and will continue to collect user community input on data requirements to implement system use cases.. The DSC will prioritize data items for data collection or development of new content models and interchange formats; these recommendations will be used to determine the final work plan for each of the NGDS Design & Testing subrecipients.

If a data submitter wishes to register a structured dataset that does not conform to an existing content model, then a good metadata record describing what is in the data, how it is structured, and how to access the dataset are required for inclusion in the NGDS. Such unique datasets will be available as downloadable files or through existing data access applications and will not be distributed as a web service. Such a dataset might be deployed as a web service is if it was required as input for a user application that NGDS is developing, or for an existing application like ArcMap that we know the community uses.

From experience, we recommend caution in determining whether a content model is required for a particular dataset, since the process of researching existing data structures and receiving feedback from the user community content model production requires considerable time. The importance of a content model can be estimated based on two factors:

1. What data are required for Geothermal User applications that we decide to build for the NGDS (search capabilities will be developed by SCR)
2. What data do our data provider partners have or are able to compile in a structured format in the remaining project time.

Process for Approving Content Models

1. **Build on Existing Models:** To develop content models for the AASG Geothermal project, existing data sets have been surveyed, and field headings in existing data are incorporated wherever possible (e.g., Aqueous Chemistry Content Model, derived from USGS Chemistry Suites found at: http://hotspringchem.wr.usgs.gov/datatype.php). Based on this survey a draft content model is developed in a content model Excel Workbook.
2. **Request Community Comments:** The draft content model is sent to subject matter experts for vetting. Comments and suggestions are incorporated into the content model.
3. **Submit for Working Group Review:** After preliminary comments are received, the content model is placed on the DOE Geothermal Data System Domain and Population Working Group (GDSDPWG) working group site for additional comments and review. The new model is announces to WG members with a request for comment and a deadline for the comment period. At least 4 weeks is recommended for the comment period.
4. **Make Content Model Available for Data Providers:** After comments are received and addressed, the GDSDPWG takes a vote on recommending the content model to project management for NGDS data services. If the recommendation is accepted, an XML schema for the interchange format is posted to http://schemas.usgin.org/schemas, and the content model is considered 'live'.
5. **Revise When Necessary:** Comments received after preliminary GDSDPWG approval may be considered. If changes are adopted after a content model is put into production, the content model is reposted on GDSDPWG for review and acceptance.

Results by Subrecipient

This section documents the results of the data inventory conducted, beginning in August 2, 2012 and concluding September 29, 2012. This is a complete inventory of data available; data priorities will be set by the Domain Steering Committee.

STANFORD

Stanford has a database of bibliographic citations for geothermal publications. The data base contains 12,809 records (as of 09/15/2012). Of these, 2118 are for publications in the proceedings from the annual Stanford Geothermal Workshop, and are unique to the Stanford Geothermal Program library. Metadata for these records will be added to the NGDS catalog. 4832 entries are for publications in the GRC library; these are duplicates of records that will enter the NGDS catalog through GRC participation in the SMU project. 5859 records are for publications that are not likely to be accounted for in other metadata compilation activities as part of the broader NGDS program. As part of Stanford's NGDS activities over the past three years, many of these 5859 records have been scanned from legacy documents found not to exist within OSTI or GRC databases, or added from new conferences. With a few exceptions, all entries in the Stanford database have accompanying downloadable PDF files of the full document.

Metadata for Stanford Workshop proceedings documents and other documents that are not duplicated by other NGDS projects will be created by a database-to-ISO 19139 XML output view from the Stanford database, using a process similar to that used to generate HTML views of the metadata records from the database using a PHP script. Metadata records will conform to the USGIN ISO metadata guidelines (http://gdsdpwg.net/Technical%20Documents/Candidate%20Specifications/USGIN_MetadataRecommendationsGeoscienceResources_v1.04.docx, http://repository.usgin.org/sites/default/files/dlio/files/2012/u11/implementationusgin_recommendationisometadata.pdf). Metadata records will be placed in a web-accessible folder for harvest by NGDS catalog servers. We estimate that coding for the transformation to ISO XML should be achievable with about 80 hours of effort, including time to study the target XML schema. Content issues with the metadata will likely emerge as a result of the computer processing, requiring some metadata updates.

Stanford also has an Excel spreadsheet with adsorption sample data, a paper that interprets this data, and 3 other publications related to adsorption data. The dataset in the spreadsheet will be documented in a metadata record (created using repository.stategeothermaldata.org), and metadata records will be generated for the associated scientific papers. We estimate about 2 hours of effort for this activity.

Documents in the Stanford geothermal library for which metadata will be published to the NGDS catalog will need to have persistent identifiers. Stanford will assign persistent URIs for all such documents. An identifier scheme will need to be determined, ideally based on primary or alternate primary key values in the existing database. We estimate about 10 hours of effort to develop a URI scheme, and assign values to records.

To maximize utility, metadata records for papers that are about specific locations should be geolocated. As has been outlined in the USGIN metadata recommendations, the first choice would be a latitude/longitude bounding box, but this will probably be too time-consuming in the time available. Stanford will add NGDS-adopted location keywords for each geolocated document for which metadata will be published to the NGDS catalog. The USGS geothermal area keyword list is being reviewed for use in this context. For resources that are not about any particular location, the location keyword 'non-geographic' should be applied. We estimate approximated 10 minutes for each document, which would total 1330 hours for the 7977 records in question. This work might be simplified by only assigning location keywords for records within the US.

GeoHeat Center, Oregon Institute of Technology

OIT's GeoHeat center has an extensive library of scanned documents related to geothermal energy resources. About 4,400 of these have been scanned and have metadata records. Individual documents will be assigned URI's either by DOI registration, or using the USGIN URI scheme and dereferencing service. GeoHeat will work with SCR to de-duplicate their metadata collection against other document collections already registered with NGDS; only metadata for documents not already registered will be exposed for harvest to the NGDS catalog.

1. **Documents (Library)**

4,185 files total. 185 are online with metadata, 3,500 have metadata compiled and the files need to go online, and 500 need to be scanned and metadata collected. Files are placed online through the Oregon Institute of Technology Library. Files are not currently geolocated. As part of the metadata compilation process, bounding boxes will be determined when possible, and location keywords will be added to the metadata for all geolocatable documents using local place names and the USGS Geothermal Area vocabulary. URLs for file download are assigned by the OIT library, so metadata records cannot be completed until the library completes its document curation process and provides URLs for the online files. The metadata will be loaded into the NGDS metadata compilation spreadsheet (http://gdsdpwg.net/Lists/System%20Proposals/Attachments/4/AASG_GeothermalDataMetadataCompilationTemplate1.3.4.zip), a python script will be provided by AZGS or SCR to convert to ISO19139 XML and the records placed in an OIT-hosted web accessible directory for harvest to the NGDS catalog. We estimate that scanning and metadata creation will take about ½ hour per document, or about 250 hours. Processing of the OIT metadata spreadsheet to produce ISO metadata records for catalog import, including review and edit time, should take about 10 hours per spreadsheet for transformation and error correction. The number of spreadsheets to be processed is uncertain, but we estimate a total of about 60 hours work on metadata import and error correction. Location keyword assignment is estimated at about 10 minutes per document, for a total of 700 hours of additional work for 4185 records.

2. **Documents (Technical Papers and bulletin Articles)**

717 documents online, with metadata. For each scanned document not already registered with NGDS catalog, an ISO19139 XML metadata record conforming to the USGIN metadata profile (http://lab.usgin.org/USGIN-ISO-metadata-v1-1) will be created and placed in the OIT web accessible metadata directory for harvest into the NGDS catalog. Metadata is already available and needs to be transformed and errors corrected to ISO XML for incorporation in the NGDS catalog; assuming the metadata have already been quality reviewed and updated, this should take about 10 hours of work.

3. **Direct Use Sites Database**

The Geo-Heat Center maintains a database of direct-use sites in the western US that currently includes records for 500 sites. These data will be loaded into the NGDS Direct Use Site Content Model (http://gdsdpwg.net/Lists/System%20Proposals/in%20progress-delivery%20template%20Direct%20Use%20Feature). In some cases locations may have to be geocoded from addresses, or town centers, and so may have low position accuracy. Best data quality would be assured if each site were reviewed to determine current activity, verify location, and other attributes required for the Content model. We are guessing this would take on the order of 1 hour per site, for a total of 500 hours, based on compilation of each site into the content model as it is reviewed.

4. **Klamath Falls Well Information**

Data for 554 geothermal wells in the Klamath Falls area. It is unclear what information is available for these wells, but well header records will be created for wells that have associated logs, scanned files, temperature data, engineering information, or economic data, and related data will be placed on line More information and sample data will be needed to fully determine the best way to incorporate these wells (location data, document types, metadata quality). Alternatively, this dataset may be documented with a single spreadsheet or database file and registered with a metadata record for this file. Documentation and registration as a single data set should take less than one hour.

5. **Co-located Sites**

The GeoHeat center maintains a database documenting the number of wells with measured water temperatures above 50 deg. C that are located with 8 kilometers of a community. 404 sites are documented in this database, but location information may be hard to obtain. If location information cannot be extracted in the time available, this will be treated as a single dataset and registered to the NGDS catalog. The thermal well and spring data compiled for the NGDS project should provide a relatively complete data set for locating thermal sources in proximity to populated areas. It is unclear what location information may be hard to obtain—the thermal sources or the communities. Registration as a single dataset should take less than one hour. Reviewing and updating the dataset is likely to take on the order of an hour per site, or about 400 hours.

6. **35 mm slides/ Pictures**

8,422 slides. These slides are not online at this time, and location accuracy for the pictures varies. Metadata may be limited. Create a metadata record for the entire collection for harvest to the NGDS catalog, which should 2-4 hours. The difficult part will be assigning meaningful location keywords or a bounding box for the collection.

7. **GeoHeat Wells and Springs Database for 16 Western States**

The GeoHeat center maintains a database of thermal wells and springs in the 16 western states. Much of this data is probably already being served by AASG geothermal data partners AZ, CO, UT, CA, OR and WY. The GeoHeat data will be loaded into the borehole temperature template and thermal spring template as appropriate and deployed as NGDS WMS and WFS services. GeoHeat will work with SCR to identify duplicated wells and thermal springs and flag the records that are duplicated in other datasets. Service hosting may be done in cooperation with UNR (Nevada HUB for

AASG project). Data loading should take 20-30 hours, including time for processing non-standard location data; deduplication with respect to existing state data will be a process requiring several weeks of effort.

8. **Klamath Falls #57310 OIT DOE project to develop geothermal on campus**

Documents and data related to the Klamath Falls #57310 OIT DOE project will be scanned or loaded into NGDS content models as appropriate. Scanned documents will be placed online, metadata describing these resources will be created and places in the OIT web-accessible metadata directory for harvest to the NGDS catalog. Estimate 40 hours maximum for this activity, but the total quantity of documents is unclear. No structured data has been identified as a product of this project.

9. **Geo-Heat Center Software**

These tools/spreadsheets can be incorporated into the NGDS system with metadata and links to download the software. Estimate 5-10 hours creating individual metadata records for tools and other data files.

University of Utah, Energy and Geoscience Institute

Geothermal Data

General work plan: Metadata for well logs and scanned documents will be created, converted to ISO19139 XML conforming to the USGIN metadata profile (http://lab.usgin.org/USGIN-ISO-metadata-v1-1). Wells will be geolocated.

1. **Well Logs**

EGI has 2,635 scanned well log documents, with metadata compiled in an excel spreadsheet. Metadata for well logs will be published via the NGDS Well Log Observation content model (http://gdsdpwg.net/Lists/System%20Proposals/Attachments/102/WellLogObservationContentModel0.8.zip). Well URI's will be generated following the USGIN URI scheme. Well header information will be compared with well headers records already registered with NGDS to identify duplicate records and assign existing URI's to wells already registered.

Well locations are currently described using Public Land Survey (PLSS) descriptions, but in order to create Well Log Observation records, a Latitude/Longitude coordinate location is required. Inspection of sample data from EGI indicates that PLSS locations are described using widely varying syntax, and are not currently amenable to machine processing. EGI will edit the metadata spreadsheet to make PLSS location descriptions consistent enough to enable machine processing to determine corresponding geographic coordinates. SCR will provide text processing support and software to convert uniformly formatted PLSS location descriptions to geographic coordinates. SCR and EGI will coordinate on appropriate formatting for PLSS locations to enable machine processing. We estimate conversion of PLSS locations to latitude/ longitude points will take 100 hours, including editing PLSS locations strings and a QA review of the determined geographic locations.

Transformation to the content model for web service publication is estimated to take another 40 hours.

2. **Reports and Articles & Maps, Charts and Graphs**

EGI has 9,010 Scanned reports, articles, maps, charts and graphs in individual Adobe Acrobat, TIFF, or PNG files. Metadata describing these documents has been compiled in an Excel spreadsheet. The metadata will be loaded into the NGDS metadata compilation spreadsheet (http://gdsdpwg.net/Lists/System%20Proposals/Attachments/4/AASG_GeothermalDataMetadataCompilationTemplate1.3.4.zip), and edited as necessary to comply with USGIN metadata recommendations (http://gdsdpwg.net/Technical%20Documents/Candidate%20Specifications/USGIN_MetadataRecommendationsGeoscienceResources_v1.04.docx, http://repository.usgin.org/sites/default/files/dlio/files/2012/u11/implementationusgin_recommendationisometadata.pdf). AZGS python script will be used to convert the tabular metadata into ISO19139 metadata records and these will be placed in a Web accessible directory hosted by EGI for harvest into NGDS catalog. This should take about 40 hours for setup, processing and QA review. Assignment of location keywords is estimated at 10 minutes per document, or about 1500 hours total.

Geothermal Sample Library

EGI has a significant collection of core, cuttings, and chip boards from geothermal wells from the US as well as other countries. Sample location descriptions will be converted to geographic (lat/long) coordinates using the same procedure used for the well logs. Wells with samples will be correlated with wells containing logs and the same identifiers assigned. Sample description data will be loaded into the sample registration content model, and registered with the System for Earth Sample Registration (SESAR, http://www.geosamples.org/services/registration). Conversion of PLSS locations can be combined with the work on other well location geographic coordinate assignment. Samples originate from total of about 3000 wells; this collection of wells probably overlaps to some degree with the collection of wells from which EGI has logs. Correlation of wells with samples with wells that have logs should be done as part of a location keyword pass. We estimate about 20 minutes per sampled well to check correlation with existing registered well (Well header from state geothermal data project, well from EGI well log collection, or new well) and loading of metadata with location keywords in to the NGDS physical sample template. Correlation of wells between EGI log and sample datasets, and with existing NGDS header data, and setting up a physical sample spreadsheet for registration with SESAR is thus estimated at about 1000 hours of work.

Core, Cuttings, Chipboards

Core: 217 boxes representing 1,455 boreholes. Core from each distinct borehole will be treated as a single sample for registration. Well header records will be created for the boreholes with 'Related Resources' links to logs (if applicable, see above) and samples from that hole.

Cuttings: 1,198 boreholes represented. Cuttings from each distinct borehole will be treated as a single sample for registration. Well header records will be created for the boreholes with 'Related Resources' links to logs (if applicable, see above) and samples from that hole.

Chipboards: 363 Chipboards. Chipboards from each distinct borehole will be treated as a single sample for registration. Well header records will be created for the boreholes with 'Related Resources' links to logs (if applicable, see above) and samples from that hole.

ADDITIONAL ITEMS

Scanned Documents, scanned, not catalogued

More than 1,000 files have been scanned but do not have metadata in the EGI metadata compilation spreadsheet. For each scanned document, a metadata record will be created loaded into the NGDS metadata compilation spreadsheet following the procedure for metadata already compiled. These records will be converted to ISO19139 XML records and places in the EGI web-accessible metadata directory for harvest to the NGDS catalog. Additional data entry may be necessary to comply with USGIN metadata standards (keywords, descriptions/abstracts, originator metadata, geographic bounding boxes, and links to the documents on the EGI server). Estimate 15 minutes per file, including location keywords, or about 250 hours.

3. **Documents, not scanned, not catalogued**

Twenty (20) boxes of documents have not been scanned. Inspection by EGI indicates that the boxes contain well logs. These documents will be scanned, metadata created and published to the NGDS catalog following procedures developed for existing scanned documents and metadata compilation. Well locations will be determined from PLSS locations where possible, using the automated PLSS to geographic coordinate workflow developed for Task 1. Wells will be cross referenced with existing wells in NGDS well header services and with other EGI wells already processed; well header records will be created for wells not already registered in the system. Based on a guess of 30 logs per box, for a total of 600 logs, and 30 minutes per log to scan and create metadata, plus 20 hours for transformation and loading of well log observation data into the NGDS content temple to deploy the well log observation service.

EGI well header and well log observation services will be hosted by an NGDS partner in the near term (UNR Hub), but EGI plans to deploy their own WFS/WMS services before the end of the project.

UNIVERSITY OF NEVADA, RENO: NEVADA BUREAU OF MINES & GEOLOGY

NBMG is providing data to both the AASG project and the BSU project. Original distribution of data deliverables was data from the Great Basin Center for Geothermal Energy (GBCGE) would be submitted to BSU and that data from NBMG would be submitted to the AASG project. Upon significant organizational changes at UNR and closer inspection of the data significant overlap has occurred. A significant amount of geothermal data is already on line at the NBMG web site (http://www.nbmg.unr.edu/geothermal/index.html), much of which is inherited from the GBCGE and thus includes data from parts of the Great Basin in adjoining states (Oregon, Idaho, Utah, Arizona, California). We have inventoried resources accessible on the NBMG web site and identified in a statement of work summarizing remaining data at NBMG. We met with co-PI Gary Johnson and discussed the deliverables and determined that a sensible subdivision of labor between the projects would be to work on document scanning, and metadata compilation, QA/QC and updating under the auspices of the BSU project. Work on data manipulation and update to deploy NGDS data services for the existing NBMG data for which there are NGDS interchange formats will be done under the auspices of the AASG Geothermal Data project. Gary Johnson has also done considerable work on various web map applications for viewing the Nevada data; if BSU project management deems appropriate, part of the NBMG project resources could be applied to updating these applications to utilize NGDS services and better integrate them with NGDS applications being developed by the Siemens team. Under this scheme, the following items would be submitted as deliverables under the NGDS Design & Testing award. All UNR data will be served on UNR servers and contributed to the NGDS by submitting USGIN-compliant metadata for the catalog.

1. **Documents (bibliography)**

More than 2,000 bibliographic entries for documents showing the source of all known publications relevant to geothermal exploration and development in Nevada. The complete bibliography will be registered as an NGDS resource and a file containing the bibliography made available online. The bibliography will also be analyzed to determine which citations are for documents not already accessible through the NGDS that NBMG has made or will make accessible; metadata for these documents will be loaded into the NGDS Metadata Content Model and submitted to the NGDS catalog. Estimate about 1 hour to create metadata for the bibliography using the http://repository.stategeothermaldata.org metadata forms. Processing of the bibliography to create metadata for documents that have not been registered with NGDS from other sources is estimated at 80 hours.

2. **Scanned Documents**

More than 2,000 documents including notices, permits, changes, and gray literature will be scanned and made accessible online from the NBMG repository, with metadata records created to describe them submitted to the NGDS catalog. Estimate about 30 minutes per document, for about 1000 hours work.

3. **Geologic Maps**

Approximately 100 1:24,000-scale geologic maps will be scanned and georeferenced, USGIN-conformant metadata will be created and loaded into the NGDS catalog. Estimate about 30 minutes per document, for about 50 hours work.

4. **Exploration Activity 2012**

A map and contributed report of all exploration activity reported in 2012 will be scanned, documented with NGDS metadata, and made available online. Estimate 2 hours work.

5. **Map Application Development**

NBMG geothermal map applications will be updated to operate with NGDS services such as hot spring features and water chemistry observation data, and to integrate with other NGDS applications being developed by the Siemens team. NGDS will not support development of applications that do not utilize NGDS services. Each map application takes on the order of 1 weeks to set up, debug and deploy; this is an open ended item to which hours should be according to NGDS requirements and priorities by project management.

SUMMARY

Group	Data Item	Effort
Stanford	Bibliographic Database for Geothermal Publications Proceedings from the annual Stanford Geothermal Workshop count: 2118 Legacy Documents (not within OSTI/GRC) count: 5859 TOTAL Documents = 7977	80 hrs.
	Assignment of Persistent Identifiers for Metadata records (URIs)	10 hrs.
	Adding NGDS-adopted Location Keywords in the metadata for each geo-located document (7977 records)	1330 hrs.
	Adsorption Data Metadata Records 3 publications Spreadsheet with adsorption data	2 hrs.
GeoHeat Center, OIT	*1. Documents* 4185 metadata records to be placed online with location keywords and/or bounding boxes for each record. 3500 of the 4185 records need to go online 500 to be scanned, and metadata collected 185 already online	1010 hrs.
	2. Documents (Technical Papers and bulletin articles) 717 documents online, with ISO 19139 XML metadata record placed online for harvesting.	10 hrs.
	3. Direct Use Sites Database 500 sites loaded into the NGDS Direct Use Site Content model, with locations (may be geocoded with position accuracy statements)	500 hrs.
	4. Klamath Falls Well Information 554 Geothermal Wells in Klamath Falls area. Documentation and registration as a single data set in a metadata record.	1 hr.
	5. Co-located Sites 404 sites - may be registered as a single record, or location information can be extracted.	1 hr. (single record) 400 hrs. (with review and update/location extraction)
	6. 35 mm slides/Pictures 8422 slides. Create a single metadata record for the entire collection; assign location keywords or bounding box.	2 - 4 hrs.
	7. GeoHeat Wells and Springs Database for 16 Western States In cooperation with SCR and 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 WMS/WFS services.	140 hrs.

	8. Klamath Falls #57310 OIT DOE project to Develop Geothermal on Campus Documents and data related to the Klamath Falls #57310 project will be scanned and loaded into NGDS content models as appropriate. Scanned documents to be publicly accessible online with descriptive metadata. Metadata will be placed online and harvested to the NGDS catalog.	40 hrs.
	9. Geo-Heat Center Software Tools and Spreadsheets to be incorporated to the NGDS system with metadata and links to download the software.	5 - 10 hrs.
University of Utah, EGI	*1. Well Logs* 2635 Scanned documents with metadata to be placed in NGDS Well Log Observation Content Model. Well header information to be compared with registered datasets to identify duplicates and assign URIs. EGI to standardize PLSS location descriptions for each record, and SCR to provide processing support for conversion to lat/long. EGI and SCR to coordinate.	140 hrs.
	2. Reports and Articles & Maps, Charts and Graphs 9010 scanned reports, articles, maps, charts and graphs in Adobe Acrobat, TIFF or PNG format. Metadata for these resources will be loaded into the NGDS metadata content model, hosted by EGI online in tabular format, and harvested to the NGDS catalog.	1540 hrs.
	3. Geothermal Sample Library Sample locations to be converted to geographic coordinates using the same procedure as the well logs (item 1). Samples to be correlated with well log and well header data sets using assigned identifiers. Samples to be registered with SESAR.	1000 hrs.
	4. Scanned Documents (scanned, not catalogued) >1000 files. Metadata records to be created and loaded into NGDS metadata content model and converted to ISO 19139 XML and placed online for harvest to NGDS catalog. Metadata should comply with USGIN metadata standards.	250 hrs.
	5. Documents (not scanned, not catalogued) 20 boxes of well logs to be scanned, metadata created, and published to the NGDS catalog. Well locations to be determined from PLSS locations using the methods developed for Task 1, and loaded to the NGDS Well Log Content Model. The scanned documents will be hosted online, and header information extracted for those wells not duplicated in the NGDS system.	820 hrs.
University of Nevada, NBMG	*1. Documents (bibliography)* >2000 entries for sources of all known publications relevant to geothermal exploration and development in Nevada, not already accessible in the NGDS system. Metadata to be loaded into the NGDS Metadata content model and submitted to the NGDS catalog.	81 hrs.
	2. Scanned Documents >2000 documents (notices, permits, gray literature) to be scanned and placed online from NBMG repository, with metadata records created and submitted to the NGDS catalog.	1000 hrs.

	3. Geologic Maps Approx. 100 1:24k scale geologic maps to be scanned and geo-referenced, with USGIN-conformant metadata created and loaded to the NGDS catalog.	50 hrs.
	4. Exploration Activity 2012 Map and report describing all exploration activity reported in 2012 will be scanned, documented and USGIN-compliant metadata to be loaded to the NGDS catalog.	2 hrs.
	5. Map Application Development NBMG Geothermal map applications will be updated to operate with NGDS services and integrated to other NGDS applications being developed by Siemens. No support for applications that do not use NGDS services.	40+ hrs. (to be determined by project management)

Appendix 1: Items found on the NBMG web site:

Data

Power plant production

Data Tables and Graphs of **Geothermal Power Production in Nevada, 1985-2011** (1.6 Mb) data also available as a downloadable MS Excel spreadsheet (.xls) (352 Kb)

Power plant production data can be delivered using the Power Plant Production Observation content model, which is still in review on the gdsdpwg web site.

Site Descriptions

Descriptions of more than 200 geothermal sites based on updated information from NBMG Bulletin 91, ***Thermal Waters of Nevada***.

Sites appear to include a variety of different features that are represented using well header, thermal spring, direct use site, power plant, geothermal area, and possibly other NGDS content models. From the point of view of the NGDS architecture, it would be best to explore how the site information fits into the evolving content model framework. The site reports should be linked to these various features as an online representation of the feature

Detailed Maps Large-scale maps for many geothermal sites can be accessed by clicking on their links in the appropriate site descriptions.

These maps appear to be customized maps showing roads, PLSS section lines, well and spring locations and other political or cultural features in the vicinity of a geothermal area. They might be registered with metadata records for each map, or simply linked to well, power plant or geothermal area features.

Photos More than 300 photos are available and are sorted by geothermal site.

Currently the AASG project is not individually registering photos with metadata records, although this would probably be useful for particularly useful or informative images. In the near term, links from other features to the photos as 'related resources' seems sufficient to incorporate photos into the information system.

Downloadable Geothermal Databases

Geothermometers: Spreadsheet with about 4000 geochemical geothermometer determinations. Location information, analytical metadata, is highly variable. Would be difficult to put in a more structure database with consistent location information. Spread sheet needs entity-attribute metadata to describe fields. Probably best registered as a single structured but not standardized (Tier 2) dataset.

Production and Injection Data for Geothermal Wells in Nevada Links from http://www.nbmg.unr.edu/geothermal/Geothermal/geo_production/ are broken, so it's not clear what form these data are in, but should be appropriate for NGDS delivery using the fluid flux observation content model.

GEOTHERM Database This is a legacy database maintained by the USGS from 1974-1983, with data from many states. The database has 120 fields for 8082 records, but is sparsely populated. Probably best registered as a single dataset. Needs entity attribute metadata to clarify what the fields represent.

GREAT BASIN GROUNDWATER GEOCHEMICAL DATABASE

Great Basin Groundwater Geochemical Database

This would appear to be the same data as the Aqueous geochemistry data listed in

DOCUMENTS –

Scanned documents are on an ftp site; thus don't get indexed by google/yahoo/bing etc. It would be better if they were moved to an http accessible repository.

Scanned geothermal papers, reports and various gray literature (also cataloged in NGDS).

The bibliography file could be made accessible from NBMG file repository, and registered with metadata in the NGDS catalog. The online search application ideally would be searching a subset of the full NGDS catalog, restricted to Nevada related resources if that were determined to be a useful resource. If the search used NGDS protocols and interchange formats, the search client software developed for the NGDS portal could be adopted for a Nevada-centric portal.

Bibliography - A searchable list of about 1,400 geothermal references, and a downloadable file in MS Word.
Click here for complete listing in Microsoft Word

PUBLISHED MAPS, AVAILABLE AS DOWNLOADS

These should be described by metadata records in the NGDS catalog. It looks like these maps were all produced using GIS; if they are available as ArcGIS projects it would be quite simple to make them accessible via OGC WMS and ESRI map services

NBMG Map 151 Geothermal Potential Map of the Great Basin, Western United States Coolbaugh, et al. (2005)

NBMG Map 161 **Nevada Geothermal Resources Update** Penfield et al (2010)

NBMG Open-file Report 09-10 **Geothermal Favorability and Exploration Activity Map of Nevada** Zehner et al (2009)

NBMG Map 141 (2nd Edtion) **Nevada Geothermal Resources** Shevenell and Garside (2005)

VARIOUS INTERACTIVE MAPS

If the data supporting these maps is in GIS feature classes and the map layouts are in ESRI ArcMap projects, it would be quite simple to make them accessible via OGC WMS and ESRI map services.

Map 141

01 - Location Data

02 - Land Status Data

03 - Geologic Data

04 - Geochemical Data

05 - Geothermal Data

06 - Geophysics Data

07 - Groundwater Data

08 - Geothermal Favorability Data

09 - Geothermal Favorability Layer Data

10 - Geothermal Exploration Projects

DATASETS THAT HAVE NGDS CONTENT MODELS:

Table 5. Feature based datasets amenable to delivery using NGDS content models and interchange formats.

Dataset	description	file format and size	file name	NGDS content model
Quaternary Faults Western US	Modified USGS Quaternary fault data base with known or estimated slip rate data	Shapefile (7.8 Gb)	QuaternaryFaultsWesternUS	Quaternary fault content model
Nevada Cinder Cones	Cinder cones < 6 MA in Nevada	Shapefile (9.1 Mb)	NVCinderCones	These four datasets would be accounted for using the Volcanic Vent model
Quaternary Mafic Vents	Known Quaternary basaltic vents and cinder cones in the Great Basin	Shapefile (67.7 Mb)	QuaternaryMaficVents	
Quaternary Rhyolite Vents	Known Quaternary dacite to rhyolite vents within the Great Basin	Shapefile (12.1 Mb)	QuaternaryRhyoliteVents	
Quaternery Volcanic Ages	Radiometric ages of Quaternary volcanic rocks of the Great Basin	Shapefile (20.6 Mb)	QuaternaryVolcanicAges	
04_Geochemical_Data	This .zip file contains all of the location data below	Shapefile (145 Gb)	04_Geochemical_Data	Aqueous Chemistry
Hot Springs and Wells	Hot springs and wells of the western United States, from the Geo-Heat Center	Shapefile (219 Mb)	Hot_Springs_and_Wells	data would be factored into Thermal Spring, Well Header, Borehole Temperature models
Great Basin Geothermal Power Plants	Geothermal Power Plants in the Great Basin	Shapefile (23 Mb)	Great_Basin_Geothermal_Power_Plant	Power Plant Facility

NetCDF / OpenDAP and WMS services:

The datasets in this table are in gridded datasets that could be served as numeric array data using NetCDF/OpenDAP from the free, open-source, THREDDS server. This server can also provide WMS and WCS services on the gridded data. These sort of services would offer application independent access to these datasets.

Table 6. Grid based datasets amenable to NetCDF, WCS and WMS services

Dataset	Description	Format	Dataset name
Great Basin Crustal Thickness	Crustal thickness of the Great Basin as estimated by seismic refraction.	ESRI GRID (69 kb)	GBCrustal
Great Basin shallow crustal residual gravity anomaly - upward continued 20 km	Shallow crustal residual gravity anomaly produced by removing 20 km upward continuation of isostatic residual anomaly from isostatic residual anomaly	TIFF Image (17.0 Mb)	GBGrIsoRes
Shallow crustal isostatic gravity anomaly for the Great Basin	A shallow crustal residual calculated by removing broad regional anomaly patterns without regard to source.	TIFF Image (4.9 Mb)	GBShall
Great Basin isostatic residual gravity anomaly	Isostatic residual gravity anomaly for the Great Basin reduced at 2.67 g/cc density on a shaded relief map.	TIFF Image (6.7 Mb)	GBGrShRes
Great Basin residual gravity upward continued	Color isostatic residual gravity anomaly upward continued 10 km on shaded relief map with 1 km grid cells	TIFF Image (9.2 Mb)	GBGrIsoResUp
Great Basin bouger gravity anomaly	Complete bouger gravity anomaly reduced at 2.67 g/cc density on shaded relief map with 1 km grid cells	TIFF Image (6.1 Mb)	GBGrBouger
Basement isostatic residual gravity field	Basement isostatic residual gravity field for Nevada Quakes, calculated by estimating the variations in thickness of Cenozoic deposits on the observed isostatic residual gravity field, then removing it to derive a basement isostatic residual	ESRI GRID (301 kb)	NBMGiso

	gravity field		
Earthquake moment release map for the Great Basin	Earthquake moment release map of the Great basin, derived from earthquakes occurring in the ANSS earthquake catalog	ESRI GRID (289 kb)	Moment1
Magnetic total field intensity anomaly	Magnetic total field intensity anomaly (IGRF corrected) of the Great Basin, reduced to 305 meters above ground	TIFF Image (10.5 Mb)	GBTtlMag
Geothermal Potential Map of Nevada	A four-category geothermal potential model of Nevada; the base layer used in NBMG Open File Report 09-10.	ESRI Integer Grid (72 kb)	Expl2010
NVLR5	2002 logistic regression model of Nevada using 5 geologic and geophysical evidence layers unaffected by groundwater or shallow impermeable cap rock layers.	ESRI Integer GRID (91 kb)	NV_LR5
2005 Favorability Map	Logistic regression model using 4 evidence layers (through-going faults, heat flow, fault and geodesic strain, and earthquake seismicity), adjusting for the effects of high lateral-flow aquifers. Base layer to NBMG Map 151.	ESRI Integer GRID (644 kb)	Favorability05
Gravity/DEM Gradient	Evidence layer for 2005 favorability map uses DEM and 'basins only' gravity gradients as a proxy for large faults having major vertical displacements.	ESRI Real GRID (2.46 Mb)	GBGravDEM
Temperature Gradient	Evidence layer for 2005 favorability map uses temperature gradient derived from heat flux data, using thermal conductivities assigned to grouped geologic rock units.	ESRI Real GRID (110 kb)	GBTempGrad

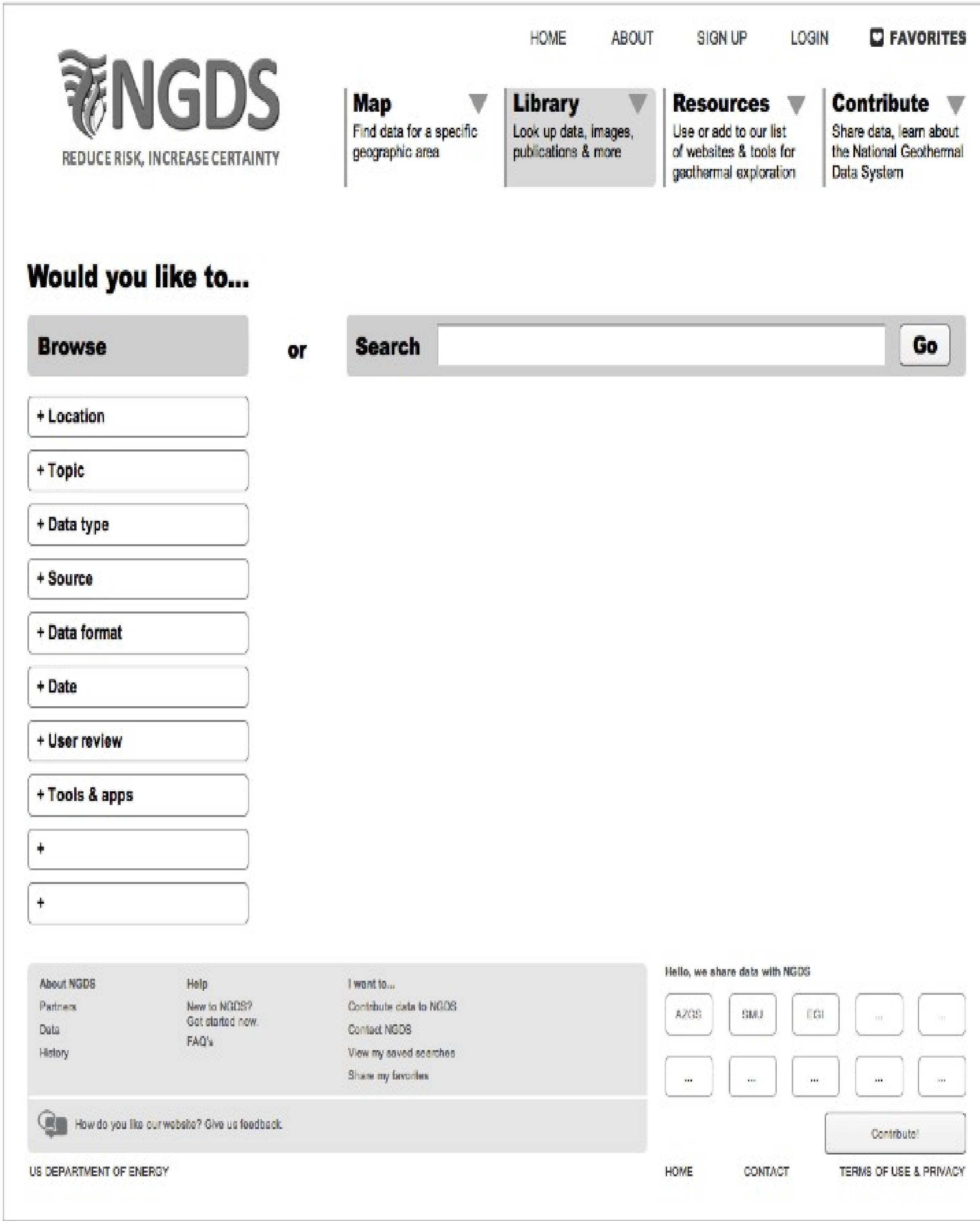
Combined Dilational Strain	Evidence layer for 2005 favorability map combines strain tensors from geodesy (GPS points) together with those from known Quaternary faults.	ESRI Real GRID (2.75 Mb)	GBDilStrain05
Earthquake Epicenters	Evidence layer for 2005 favorability map based on earthquake seismicity weighted by distance from the earthquake epicenter.	ESRI Real GRID (570 kb)	GBDistQuakes

Data submitted under the AASG project includes 1,500 well log observation point features, 1,000 borehole temperature observations, geothermal power production records for 20 power plants, IGSN/SESAR sample registrations and NGDS metadata for 100 geothermal core samples, 300 hot spring features, 1,000 age date observations, 4,000 rock chemistry observations, a OneGeology conformant GeoSciML-portrayal service for the Geologic Map of Nevada, at least 3 digitized historic geologic maps of regions with geothermal significance, approximately 20 direct use site features will be de-duplicated with OIT data and served by UNR, and monthly pumping and Injection records values reported by individuals wells at each power plant site in the NGDS fluid flux borehole observations content model. Datasets will be made available in summary data files (basically how they are now published on the NBMG geothermal data web sites) and NGDS web services. Metadata records will describe each dataset, with the file based and service-based access as alternate distributions.

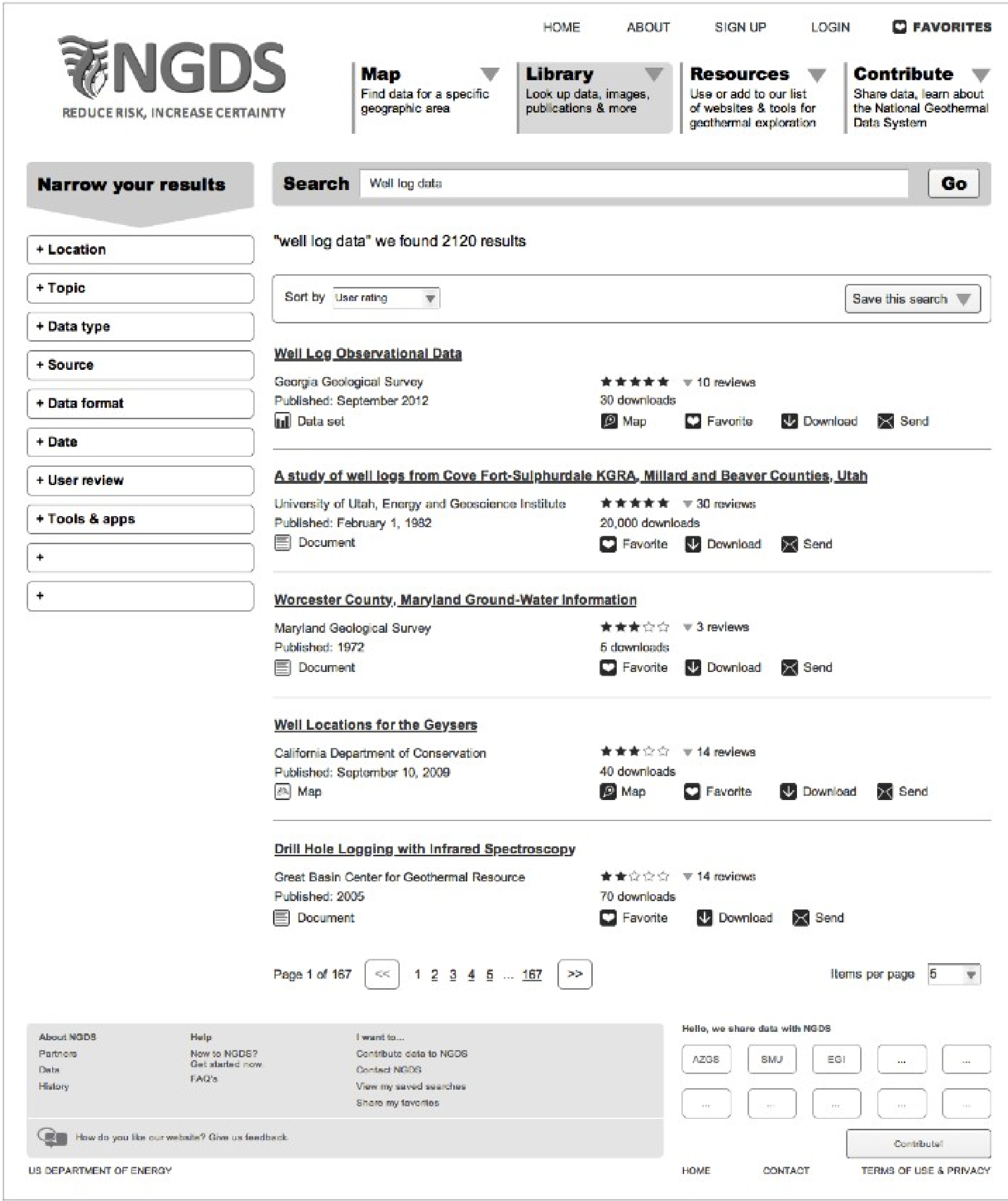
Appendix 2a

Interface Design Initial Concept

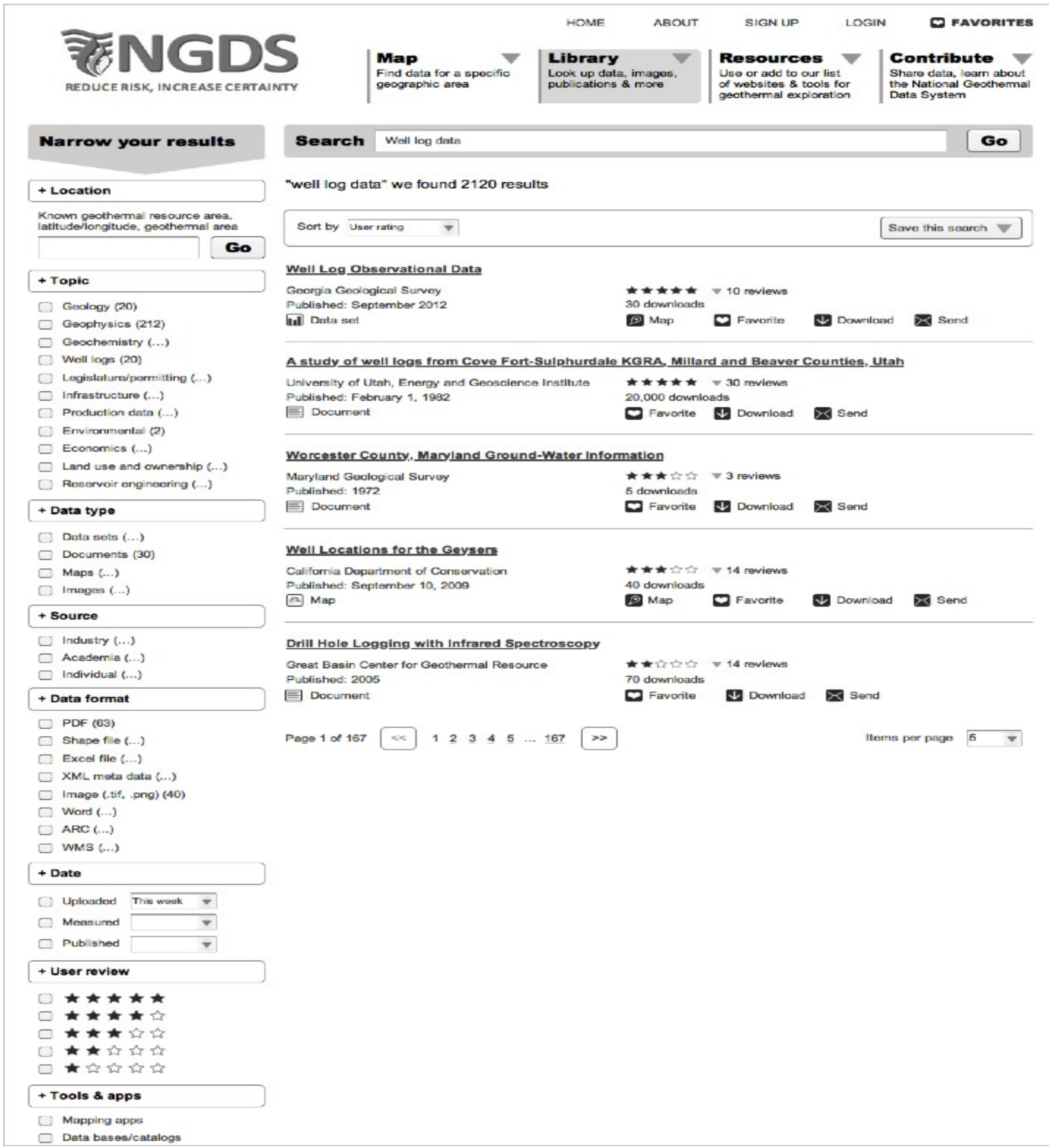
NGDS Design Concept, version 3



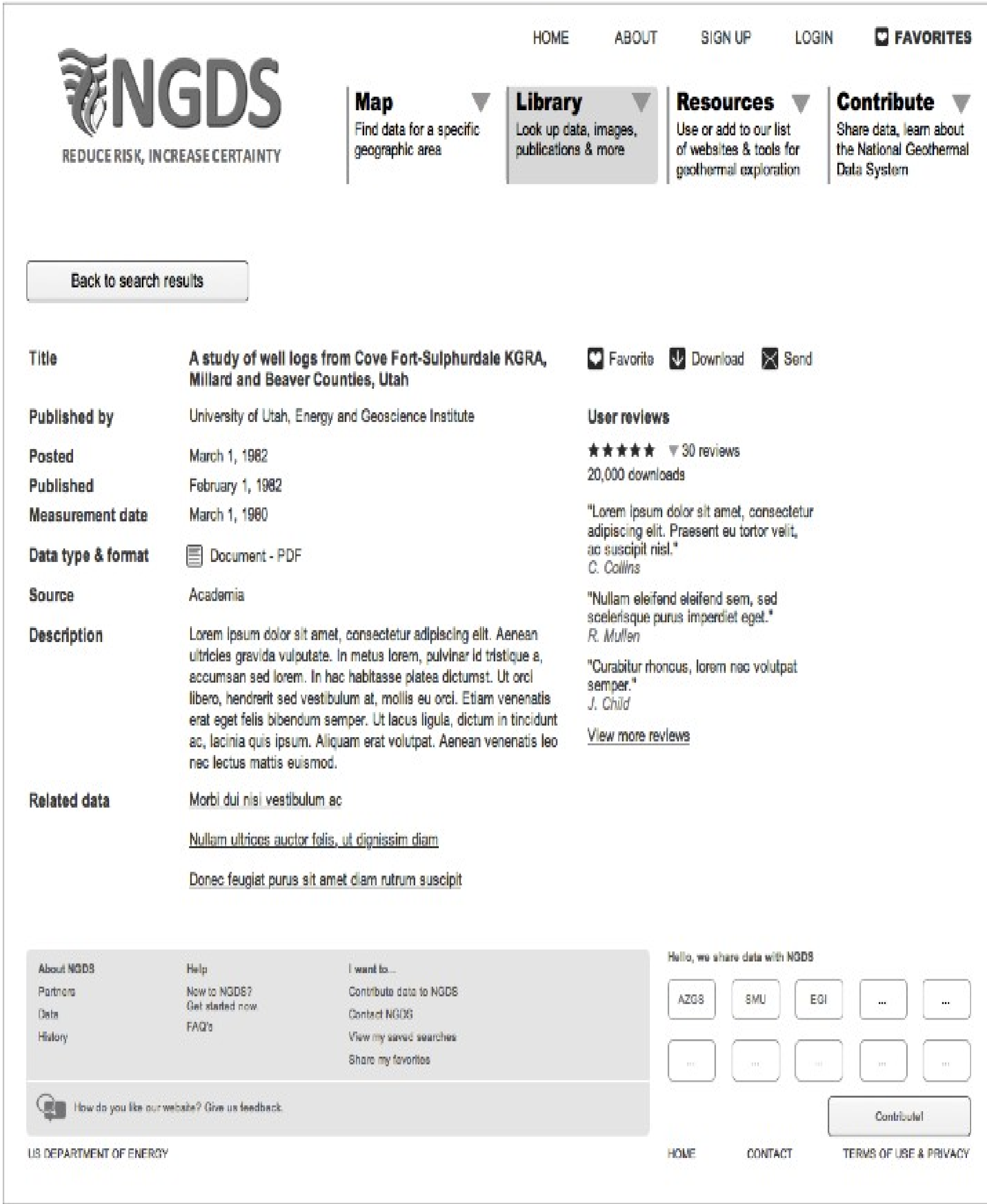
NGDS Design Concept, version 3



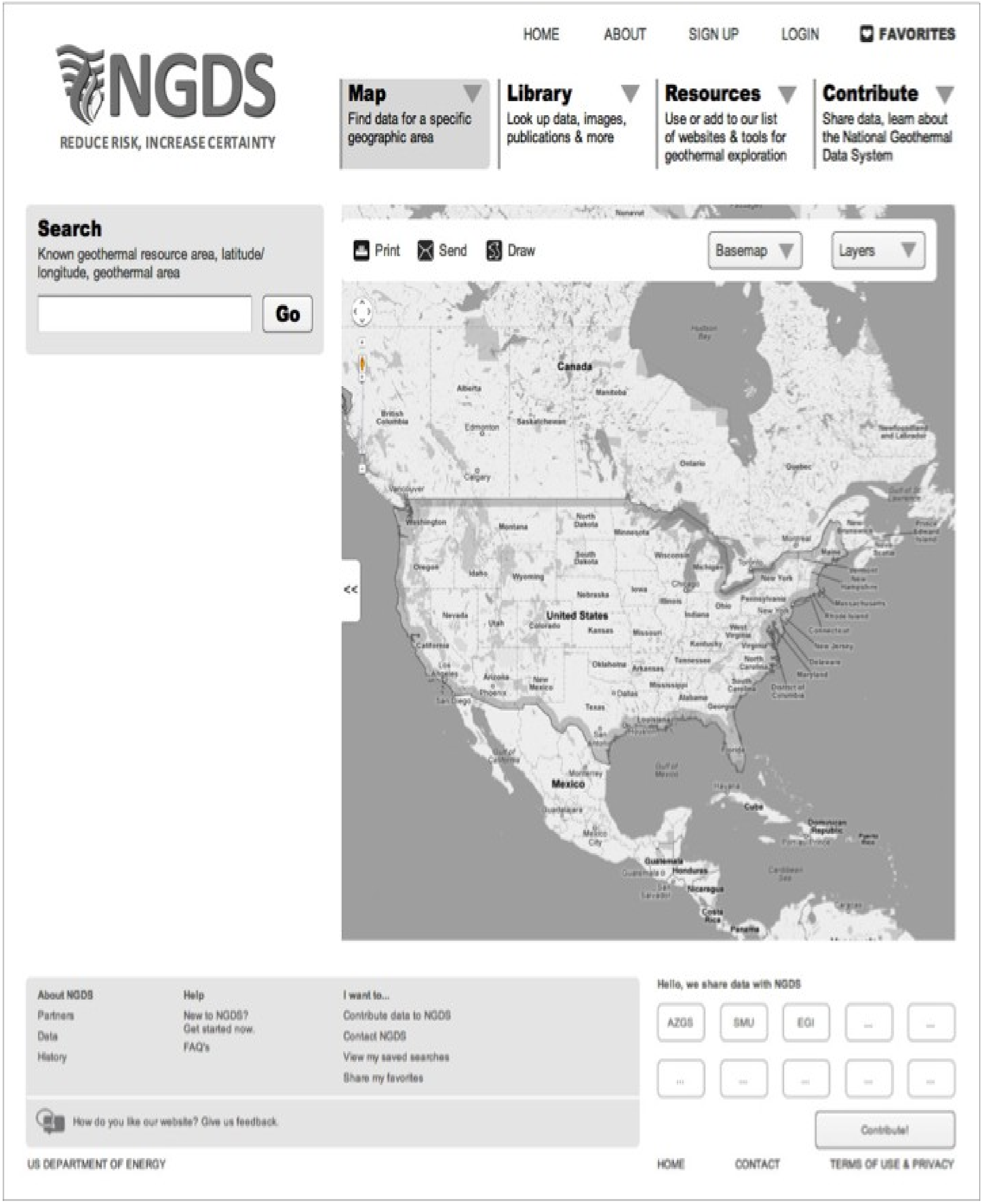
NGDS Design Concept, version 3



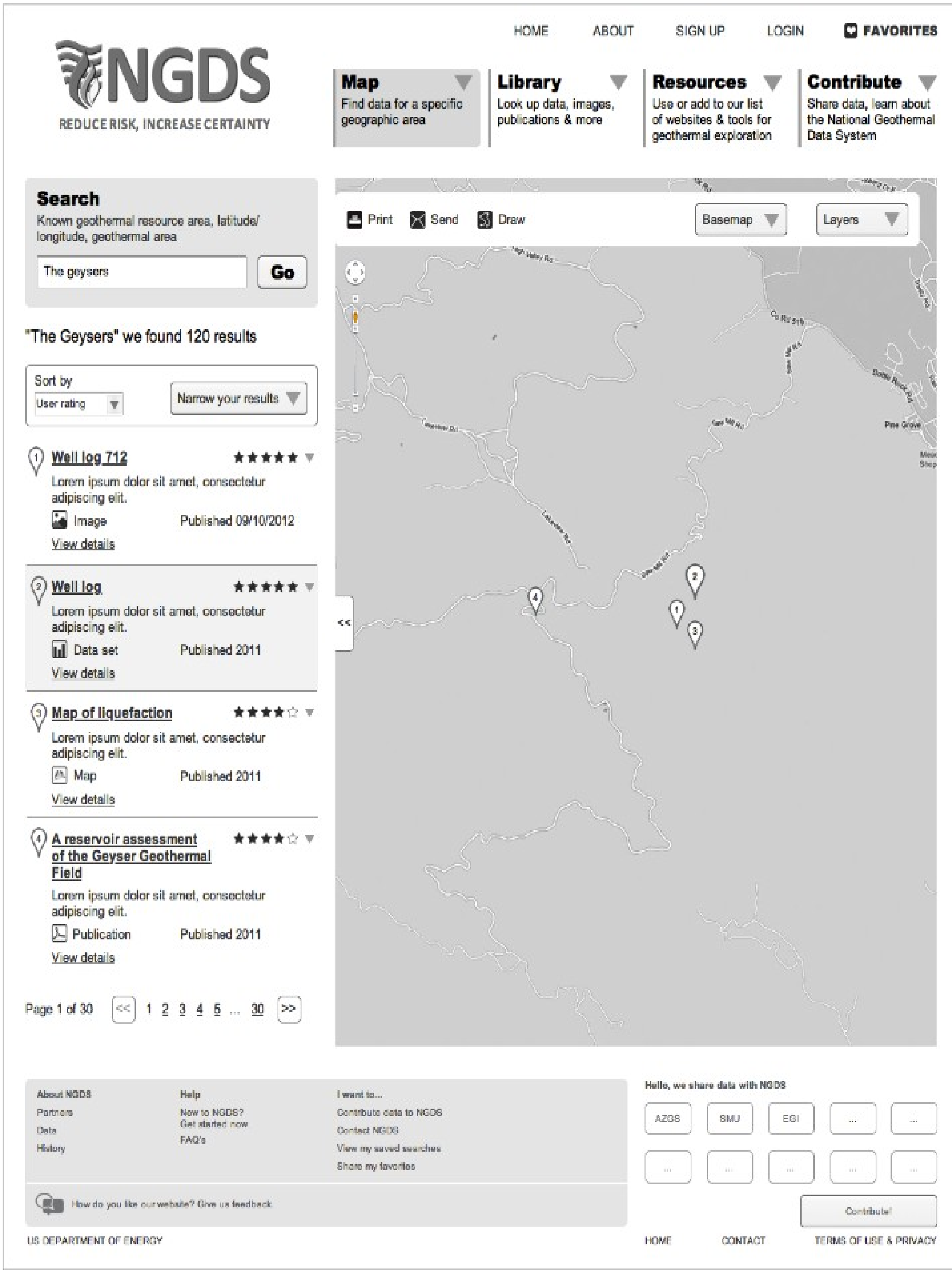
NGDS Design Concept, version 3



NGDS Design Concept, version 3

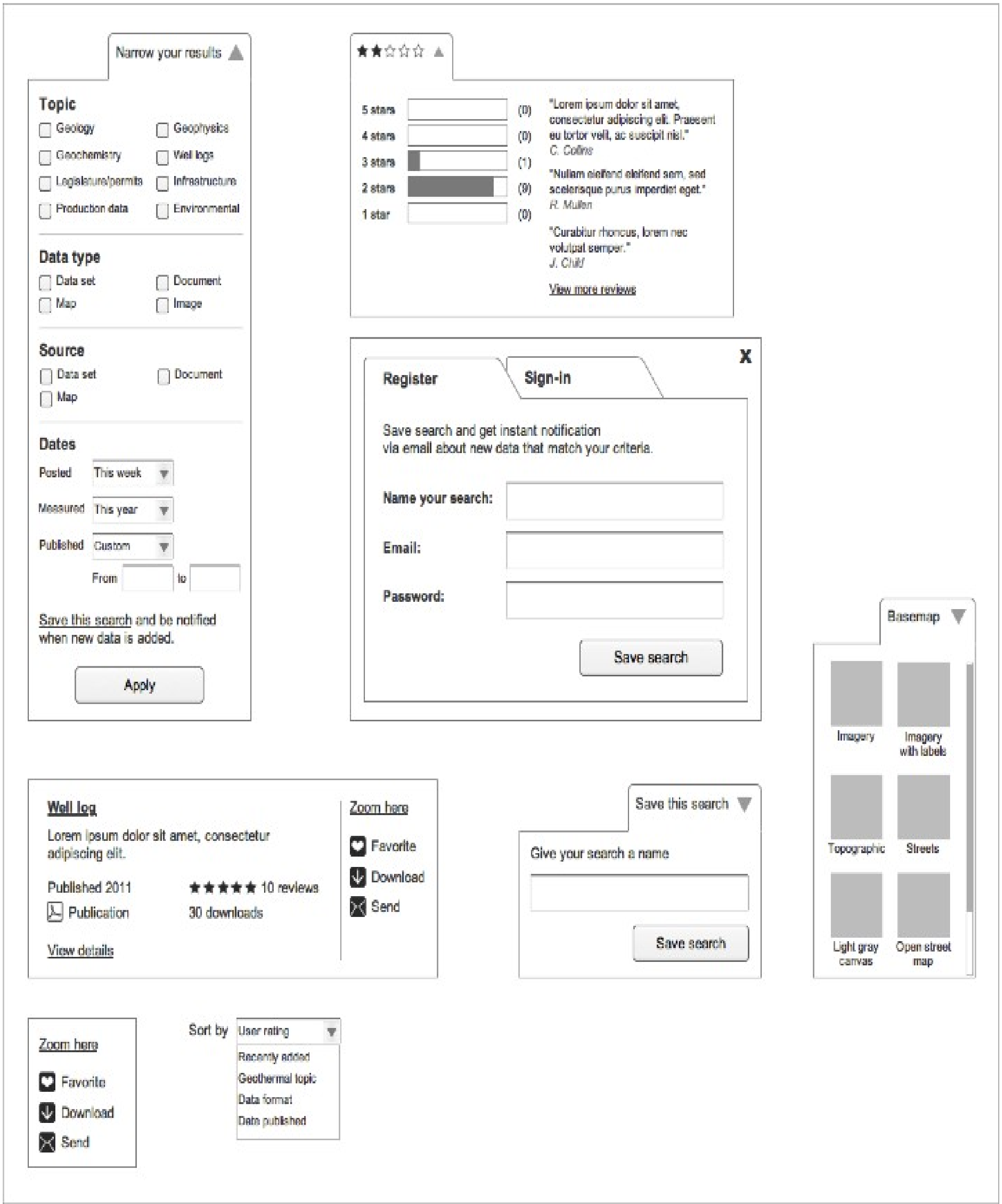


NGDS Design Concept, version 3

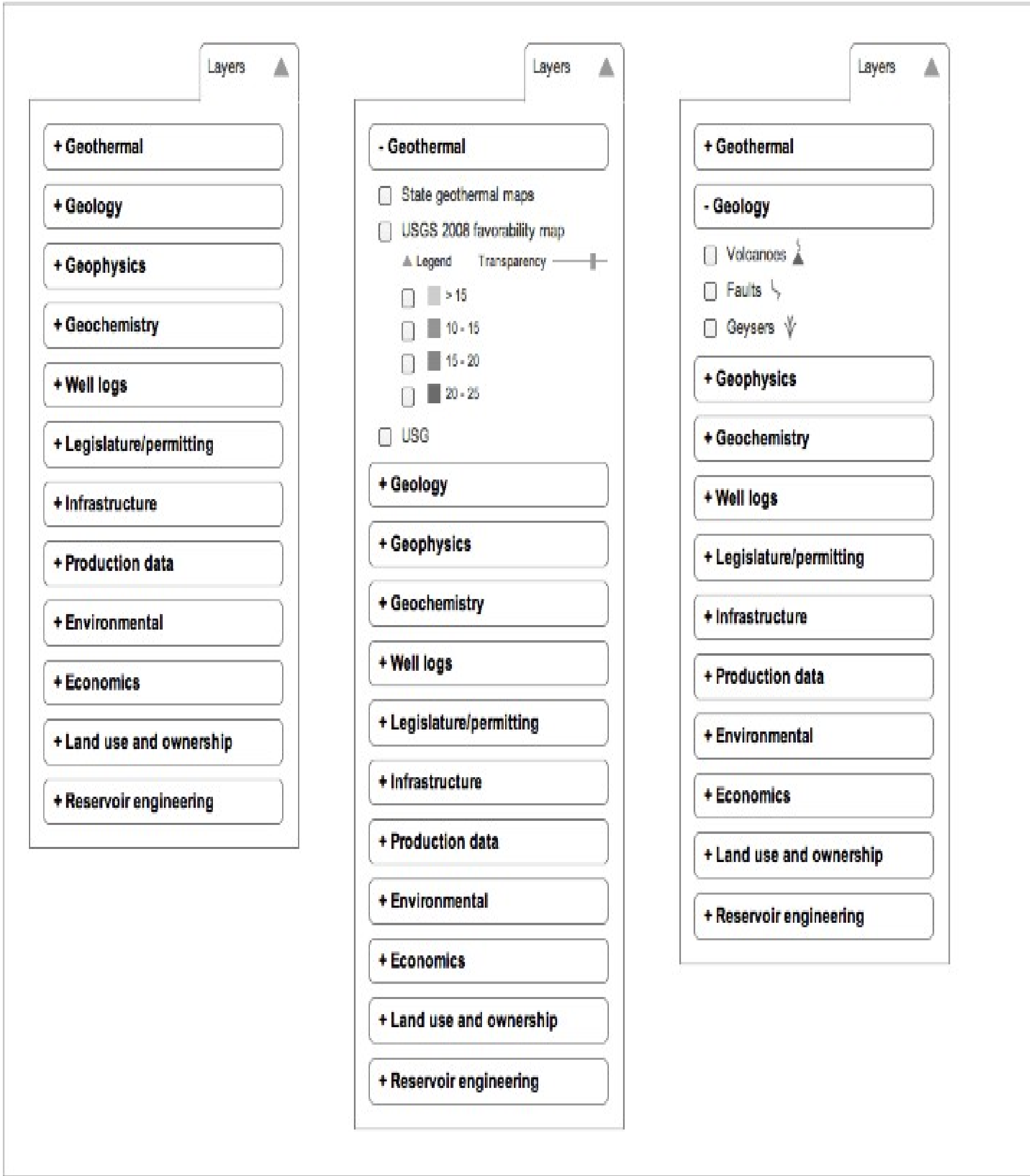


HOME
ABOUT
SIGN UP
LOGIN
FAVORITES
NGDS
REDUCE RISK, INCREASE CERTAINTY
Map
Find data for a specific geographic area
Library
Look up data, images, publications & more
Resources
Use or add to our list of websites & tools for geothermal exploration
Contribute
Share data, learn about the National Geothermal Data System
Search
Known geothermal resource area, latitude/ longitude, geothermal area
The geysers
Go
"The Geysers" we found 120 results
Sort by
User rating
Narrow your results
Well log 712
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Image
Published 09/10/2012
View details
Well log
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Data set
Published 2011
View details
Map of liquefaction
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Map
Published 2011
View details
A reservoir assessment of the Geyser Geothermal Field
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Publication
Published 2011
View details
Page 1 of 30
<< 1 2 3 4 5 ... 30 >>
Print
Send
Draw
Basemap
Layers
<<
About NGDS
Partners
Data
History
Help
New to NGDS? Get started now
FAQ's
I want to...
Contribute data to NGDS
Contact NGDS
View my saved searches
Share my favorites
How do you like our website? Give us feedback
Hello, we share data with NGDS
AZGS
SMU
EGI
Contribute!
US DEPARTMENT OF ENERGY
HOME
CONTACT
TERMS OF USE & PRIVACY

NGDS Design Concept, version 3



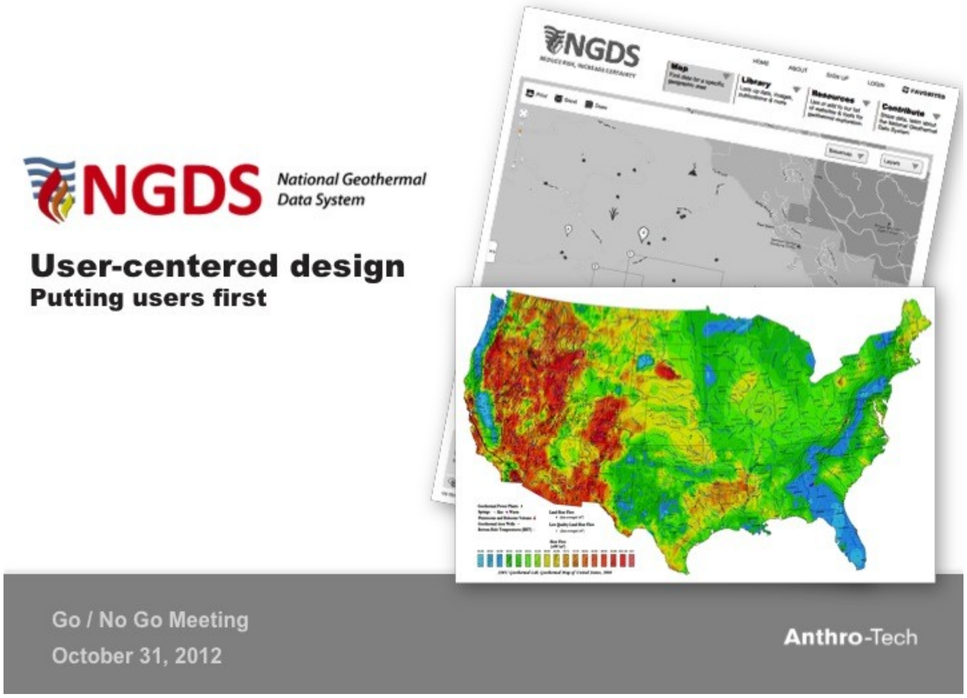
NGDS Design Concept, version 3



NGDS Design Concept, version 3

Appendix 2b

Interface Design Version 2



UCD ROADMAP

Topics for today

- **GOALS** | Success for NGDS
- **UCD ROADMAP** | Steps taken to inform the NGDS user experience
- **USER RESEARCH FINDINGS** | Insights from user interviews and concept usability testing.
- **UX CONCEPT** | A vision for the user experience of NDGS
- **RECOMMENDATIONS** | A path forward for NGDS
- **QUESTIONS**| Discussion and feedback

GOALS

"As early as 1904 there was geothermal production and it has been on a constant rise ever since. The U.S. produces the largest amount of geothermal energy of any country, so I'm really proud of the industry and pleased to see that the federal government is trying to find a way to continue the development."

GOALS

Successful systems align user & business goals

Users want to …

Explore geothermal potential

Mitigate risks for my investments

Contribute research data

NGDS

DOE wants to …

Produce geothermal energy

Support a knowledge repository and archive for geothermal data, lessons learned, reports.

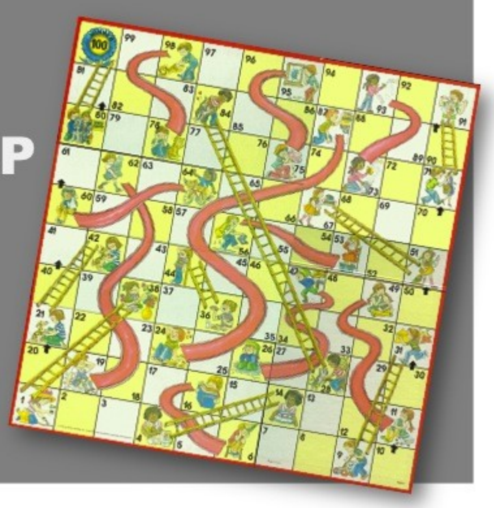
Advance earth sciences by identifying gaps in our knowledge and informing new knowledge.

Help companies be more (cost and time) effective in exploration, development and usage of geothermal energy.

Increase public awareness in geothermal energy

U.S. DEPARTMENT OF ENERGY

UCD ROADMAP



UCD is about designing useful and usable experiences

Characteristics of UCD:

- Early focus on users & business goals
- Data-driven
- Iterative
- Collaborative
- Scalable

ISO 9241-210:2010

UCD ROADMAP

Synthesize user research to date

User profiles

User scenarios

Comparative site analysis

Guidelines for design

One-on-one expert interviews

Interview guide
Interview recordings
Interview notes
Summary of findings
Recommendations for design

Prototype

UX concept – V. 1

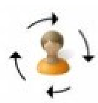
Data needs

UCD ROADMAP

GEA GEOTHERMAL ENERGY ASSOCIATION

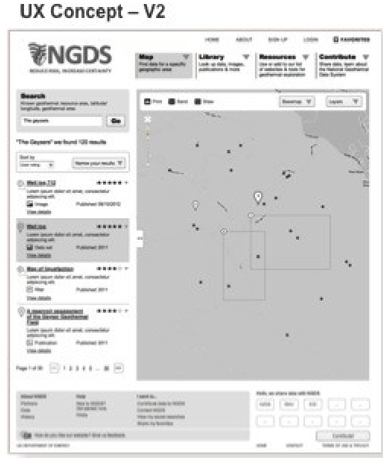
UX concept usability tests

Data card sort

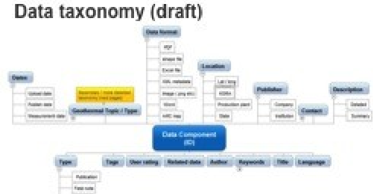


Test with users (18)
Team feedback

UX Concept – V2



Data taxonomy (draft)



USER RESEARCH & TESTING

USER RESEARCH

Top user goals

1. Determine geothermal potential for a specific area.
2. Scientifically characterize a specific geological area to contribute to the field's body of knowledge.

USER RESEARCH

Geothermal exploration is a data-driven process

42%

As much as 42% of all expenses associated with geothermal energy production can be attributed to **exploration.**[1]

1. http://en.wikipedia.org/wiki/Geothermal_exploration

> *"Data is like the holy grail for us."* - P3

> *"We're making high dollar decisions and are trying to predict something complex from anything you get."* - P2

> *"Due diligence is a legal term and it means that if you are investing in something, and you're leading other people to invest their money in it, you should do your due diligence to figure out that the project will do what they say it will do."* - P8

USER RESEARCH

Data discovery

- Time consuming
- Inefficient
- Cumbersome
- Dependent on visiting **many** different sources and systems
- Collaborative
- Happens over time
- Geographically driven
- Supported by users' internal processes / tools

> *"The biggest challenge, of course .. is simply finding out if the data exists."* -P2

> *"You know, there's no good clearinghouse, (which is) probably why NGDS was envisioned. Might get past data from one, then some from another... no good single point."* – P7

> *"Everything we do is geographically driven."* - P8

USER RESEARCH

NGDS: The sweet spot

GEOTHERMAL EXPLORATION PROCESS

NGDS: provide quality data from all stages of the exploration process

Identify area of interest — Productive power plant

DATA COLLECTION

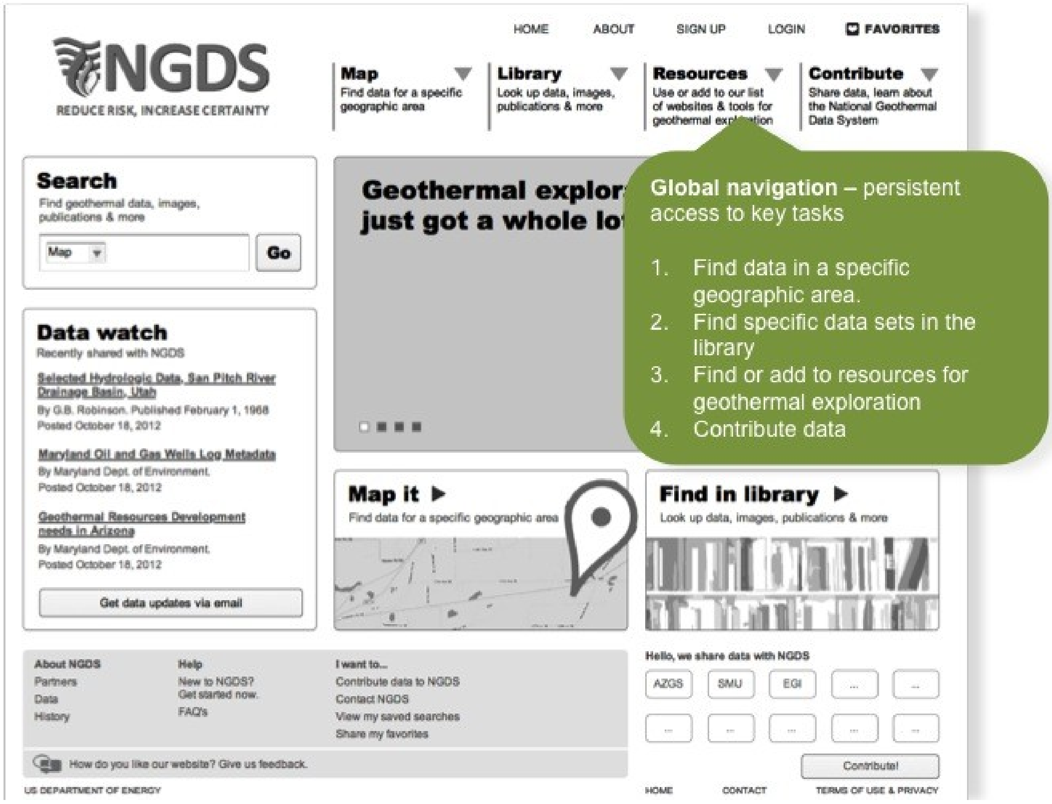
NGDS: focus on data discovery and validation

Gather — Validate — Analyze

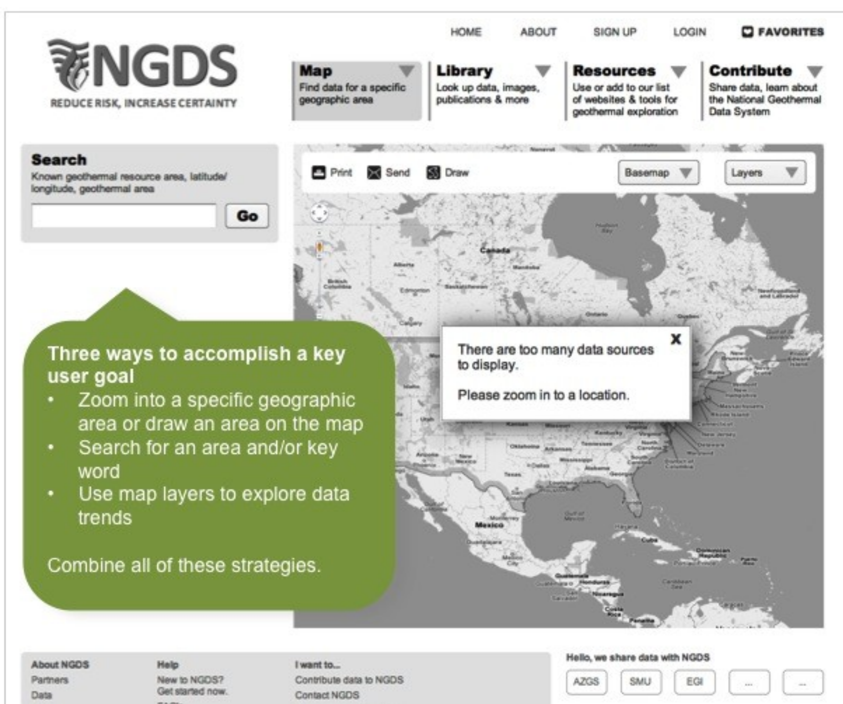
NGDS User Experience Concept

“*Give us a common interface that links the data behind the scenes.” – P3*

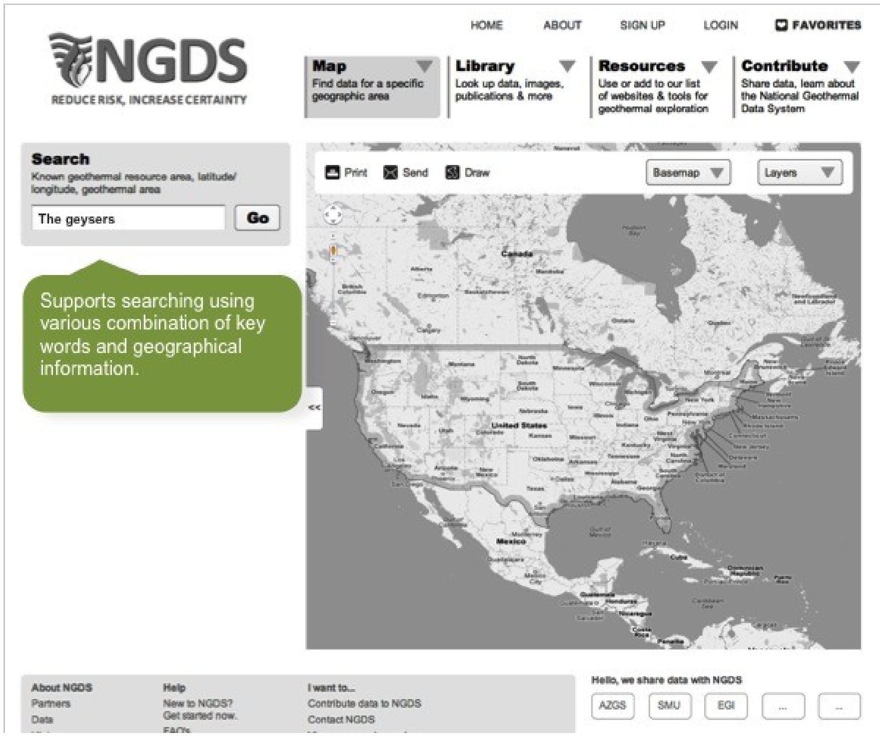
UX CONCEPT



UX CONCEPT

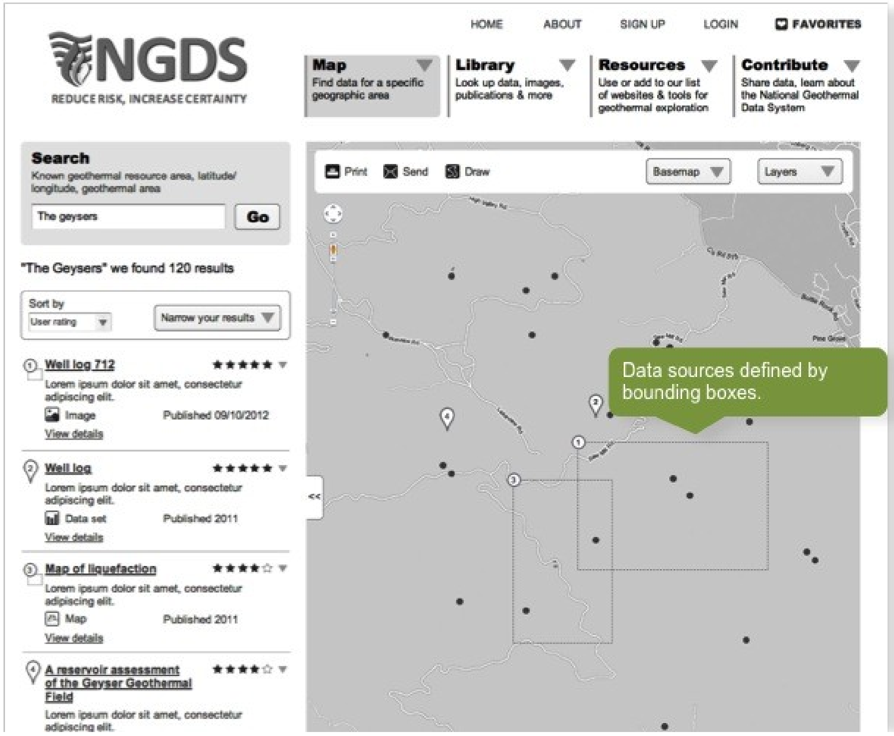


UX CONCEPT

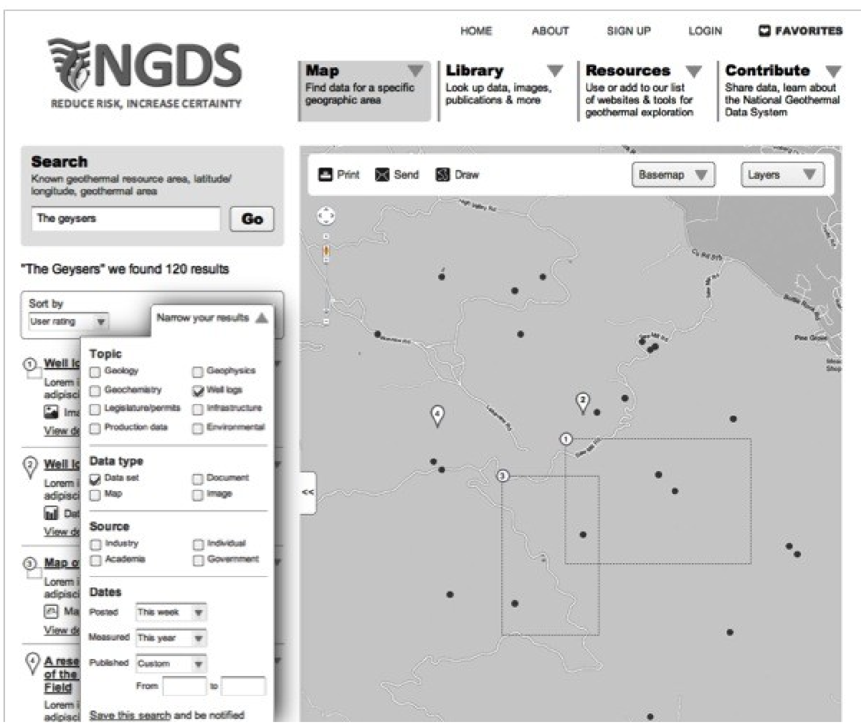


HOME
ABOUT
SIGN UP
LOGIN
FAVORITES
NGDS
REDUCE RISK, INCREASE CERTAINTY
Map
Find data for a specific geographic area
Library
Look up data, images, publications & more
Resources
Use or add to our list of websites & tools for geothermal exploration
Contribute
Share data, learn about the National Geothermal Data System
Search
Known geothermal resource area, latitude/ longitude, geothermal area
The geysers
Go
Supports searching using various combination of key words and geographical information.
Print
Send
Draw
Basemap
Layers
About NGDS
Partners
Data
Help
New to NGDS? Get started now.
I want to...
Contribute data to NGDS
Contact NGDS
Hello, we share data with NGDS
AZGS
SMU
EGI

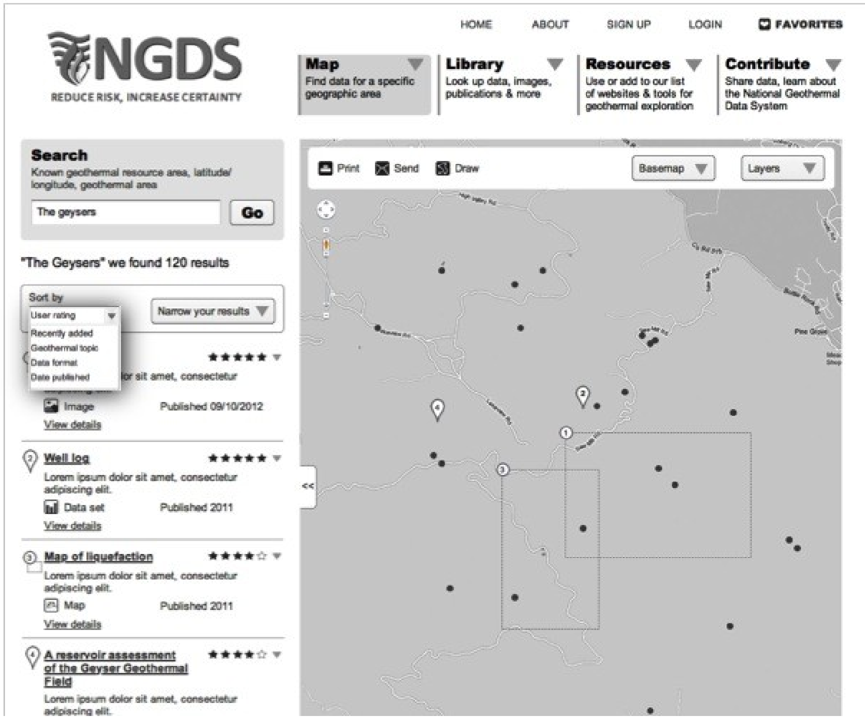
UX CONCEPT



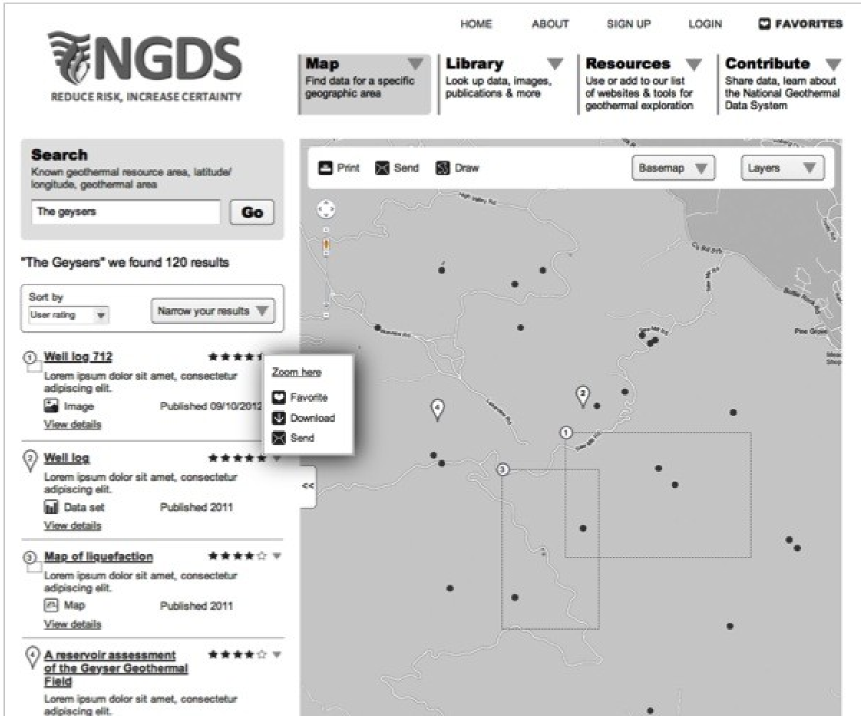
UX CONCEPT



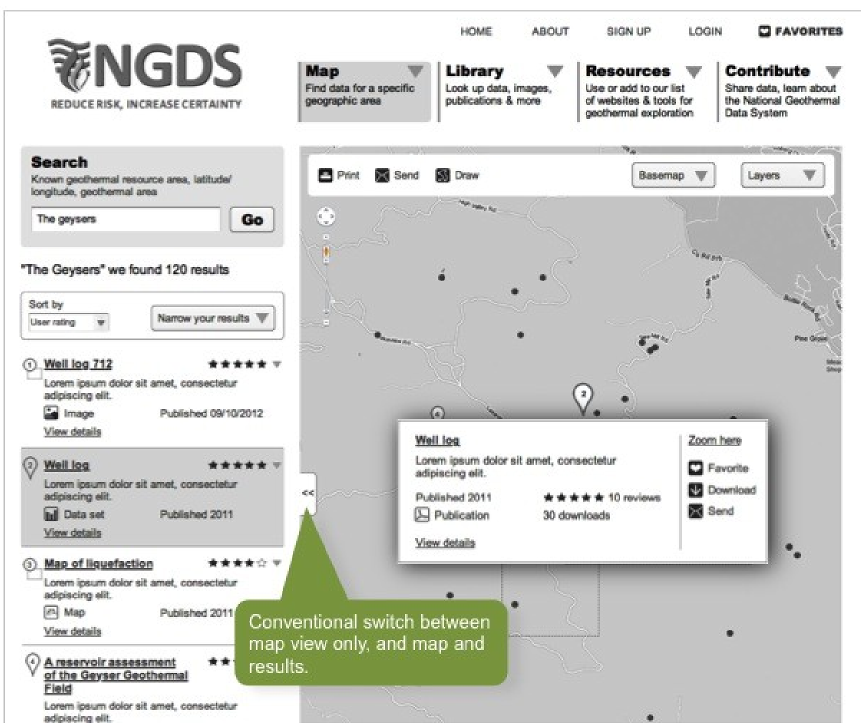
UX CONCEPT


HOME
ABOUT
SIGN UP
LOGIN
FAVORITES
NGDS
REDUCE RISK, INCREASE CERTAINTY
Map
Find data for a specific geographic area
Library
Look up data, images, publications & more
Resources
Use or add to our list of websites & tools for geothermal exploration
Contribute
Share data, learn about the National Geothermal Data System
Search
Known geothermal resource area, latitude/ longitude, geothermal area
The geysers
Go
"The Geysers" we found 120 results
Sort by
User rating
Recently added
Geothermal topic
Data format
Date published
Narrow your results
Image
Published 09/10/2012
View details
Well log
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Data set
Published 2011
View details
Map of liquefaction
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Map
Published 2011
View details
A reservoir assessment of the Geyser Geothermal Field
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Print
Send
Draw
Basemap
Layers

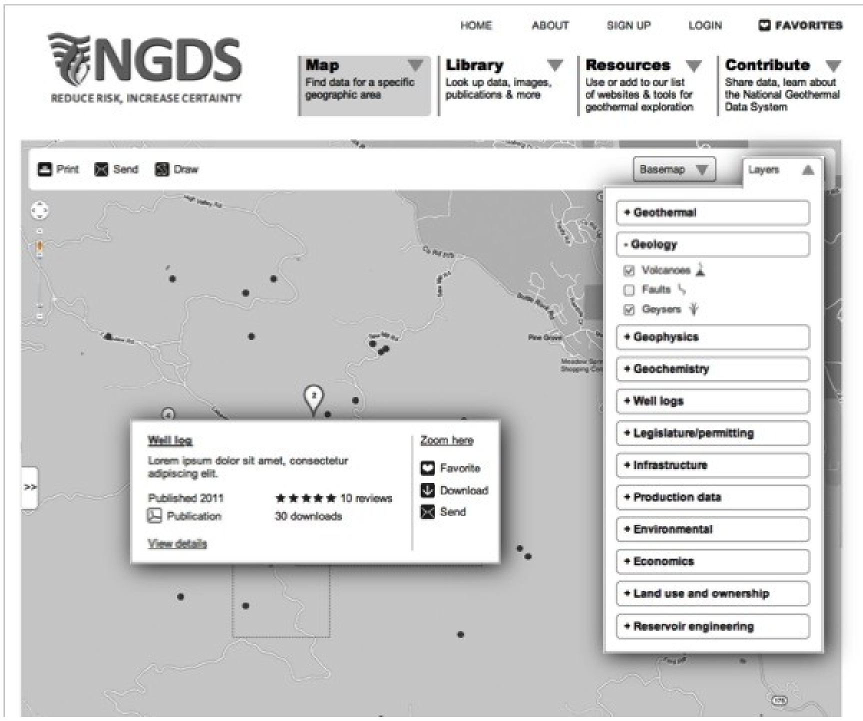
UX CONCEPT



UX CONCEPT

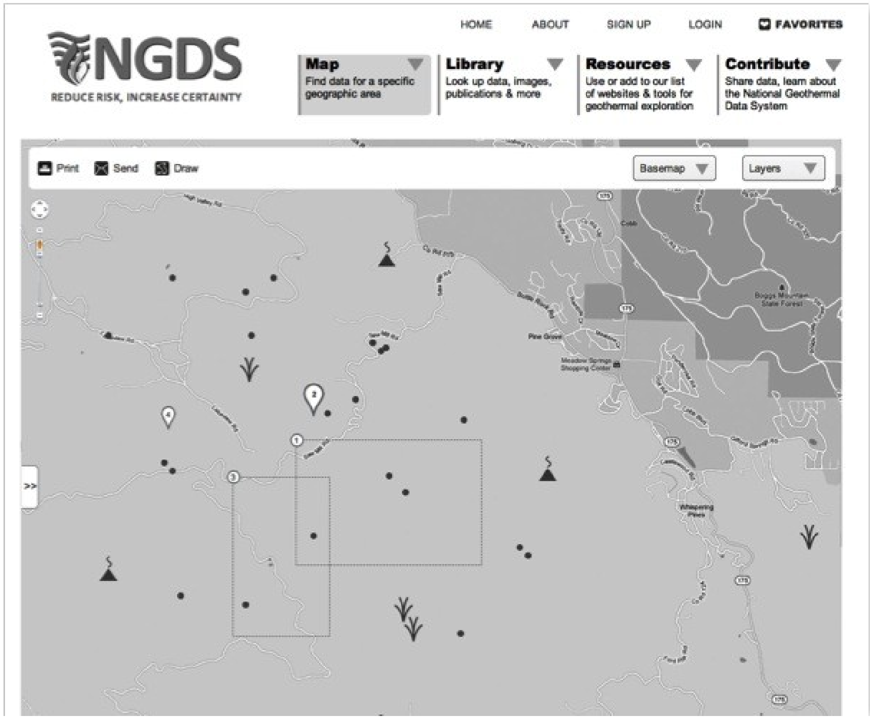


UX CONCEPT

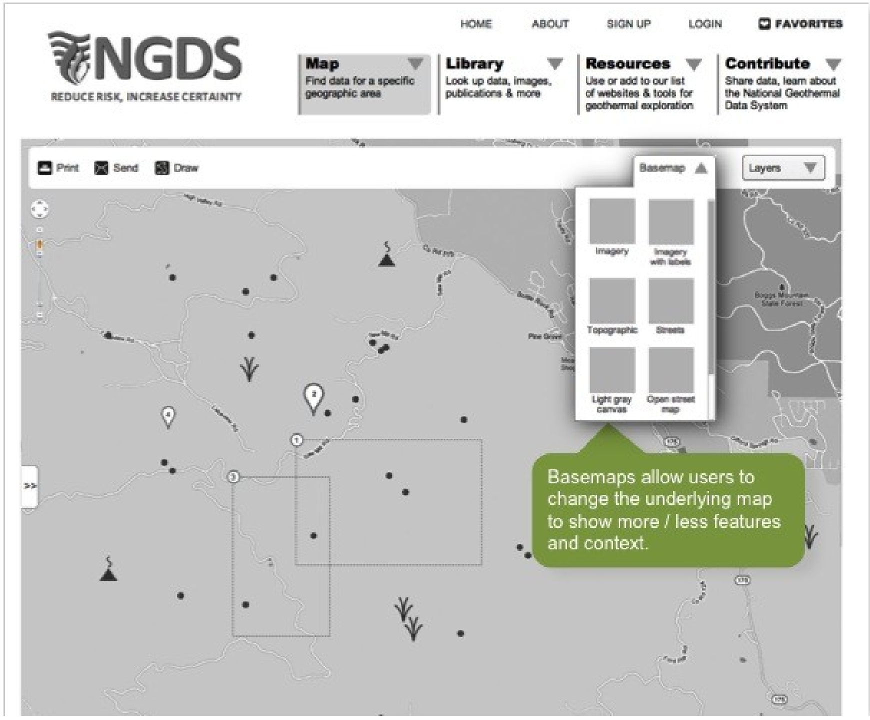


HOME
ABOUT
SIGN UP
LOGIN
FAVORITES
NGDS
REDUCE RISK, INCREASE CERTAINTY
Map
Find data for a specific geographic area
Library
Look up data, images, publications & more
Resources
Use or add to our list of websites & tools for geothermal exploration
Contribute
Share data, learn about the National Geothermal Data System
Print
Send
Draw
Basemap
Layers
+ Geothermal
- Geology
Volcanoes
Faults
Geysers
+ Geophysics
+ Geochemistry
+ Well logs
+ Legislature/permitting
+ Infrastructure
+ Production data
+ Environmental
+ Economics
+ Land use and ownership
+ Reservoir engineering
Well log
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Published 2011
10 reviews
Publication
30 downloads
View details
Zoom here
Favorite
Download
Send

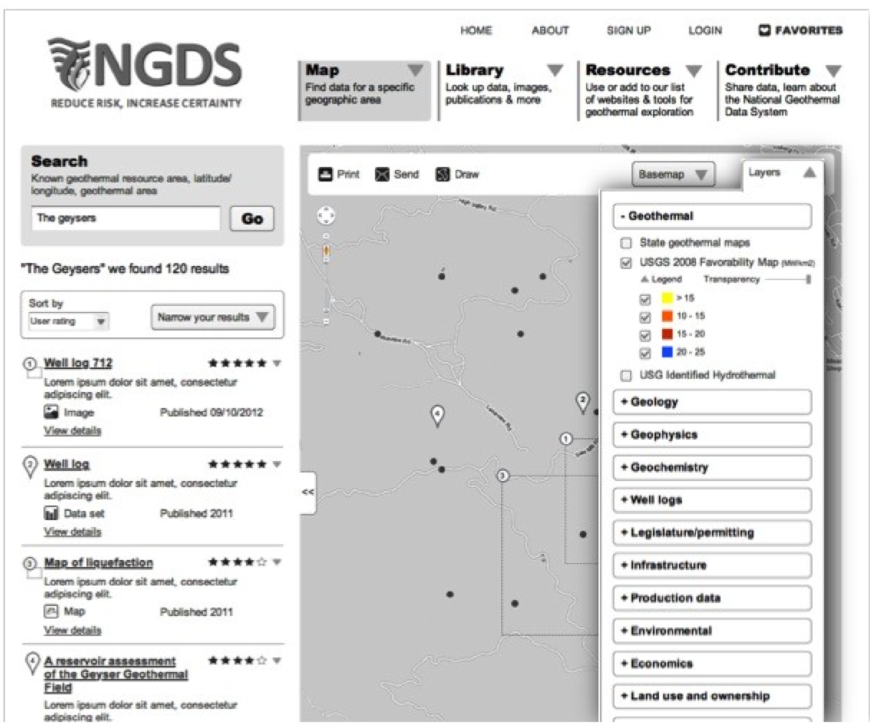
UX CONCEPT


HOME
ABOUT
SIGN UP
LOGIN
FAVORITES
NGDS
REDUCE RISK, INCREASE CERTAINTY
Map
Find data for a specific geographic area
Library
Look up data, images, publications & more
Resources
Use or add to our list of websites & tools for geothermal exploration
Contribute
Share data, learn about the National Geothermal Data System
Print
Send
Draw
Basemap
Layers

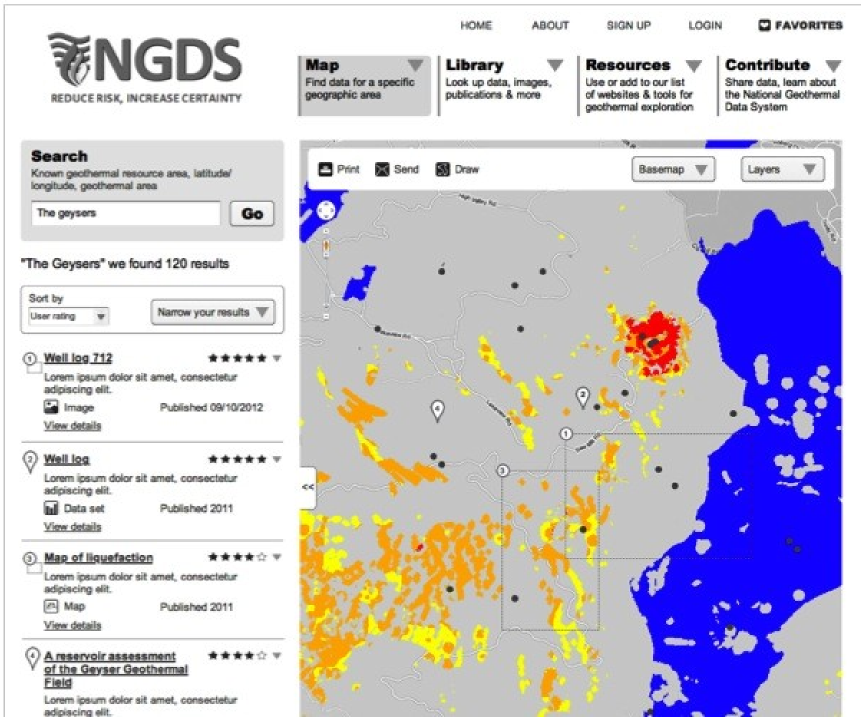
UX CONCEPT



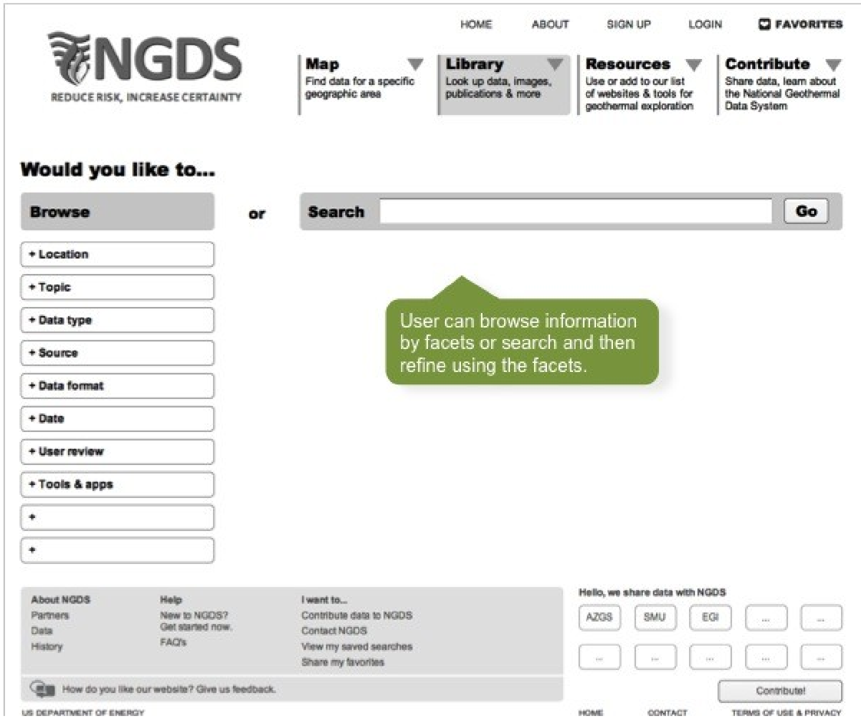
UX CONCEPT



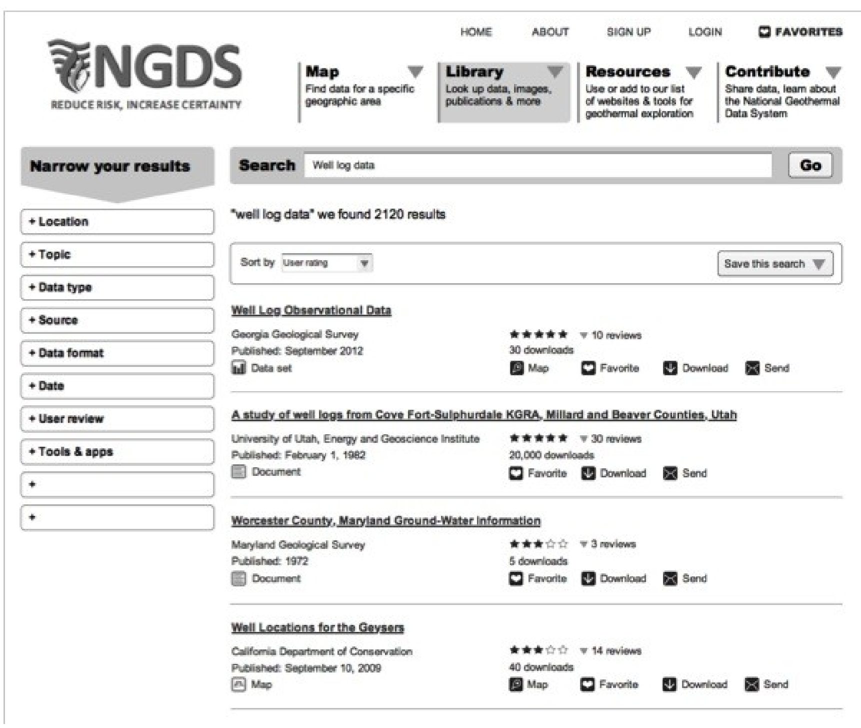
UX CONCEPT


HOME ABOUT SIGN UP LOGIN FAVORITES
NGDS
REDUCE RISK, INCREASE CERTAINTY
Map
Find data for a specific geographic area
Library
Look up data, images, publications & more
Resources
Use or add to our list of websites & tools for geothermal exploration
Contribute
Share data, learn about the National Geothermal Data System
Search
Known geothermal resource area, latitude/ longitude, geothermal area
The geysers
Go
"The Geysers" we found 120 results
Sort by
User rating
Narrow your results
Well log 712
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Image
Published 09/10/2012
View details
Well log
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Data set
Published 2011
View details
Map of liquefaction
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Map
Published 2011
View details
A reservoir assessment of the Geyser Geothermal Field
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Print
Send
Draw
Basemap
Layers

UX CONCEPT



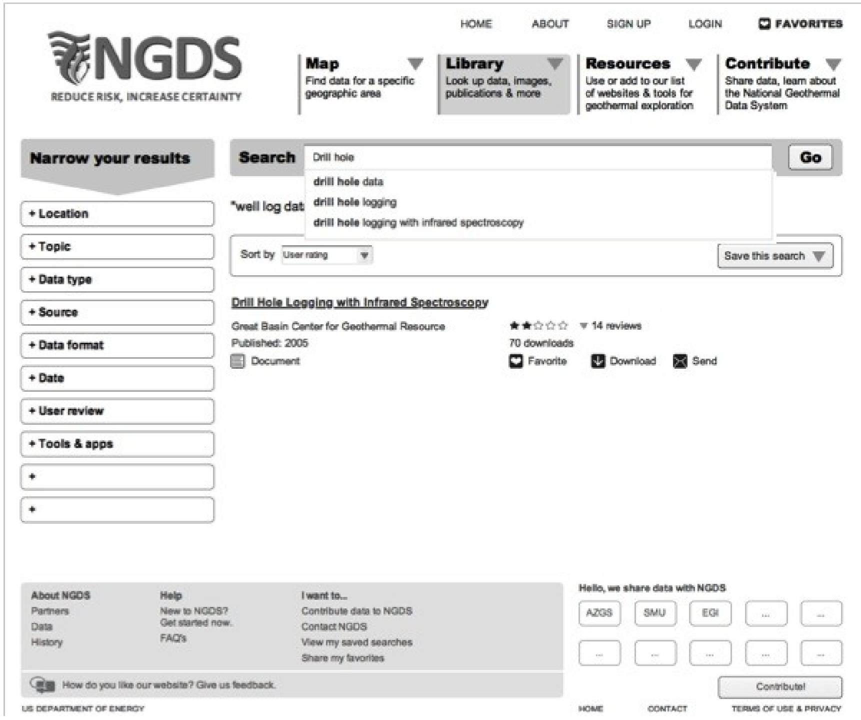
UX CONCEPT



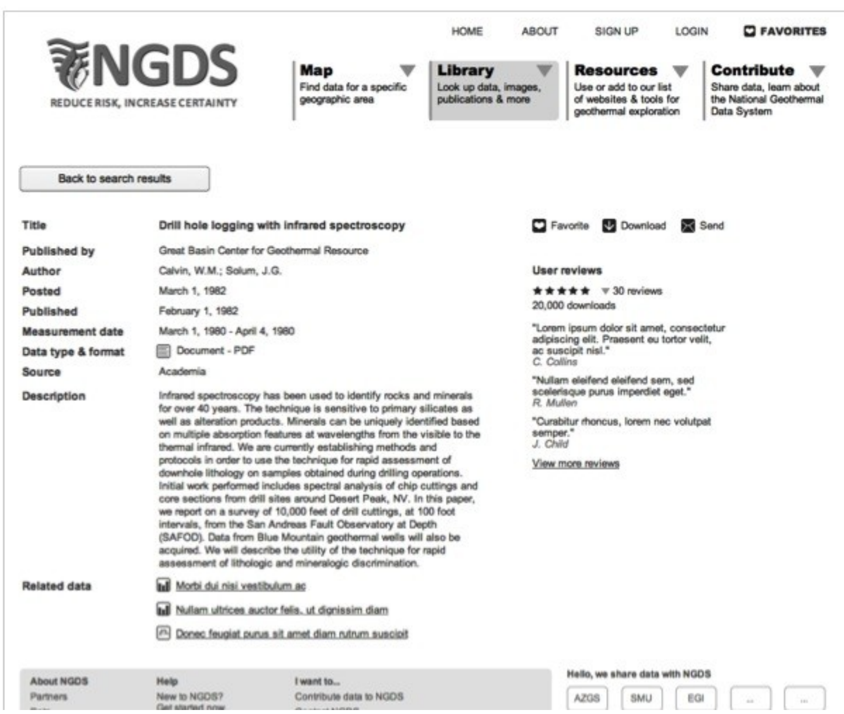
UX CONCEPT

HOME
ABOUT
SIGN UP
LOGIN
FAVORITES
NGDS
REDUCE RISK, INCREASE CERTAINTY
Map
Find data for a specific geographic area
Library
Look up data, images, publications & more
Resources
Use or add to our list of websites & tools for geothermal exploration
Contribute
Share data, learn about the National Geothermal Data System
Narrow your results
Search
Well log data
Go
"well log data" we found 2120 results
+ Location
Known geothermal resource area, latitude/longitude, geothermal area
Go
+ Topic
Geology (20)
Geophysics (212)
Geochemistry (...)
Well logs (20)
Legislature/permitting (...)
Infrastructure (...)
Production data (...)
Environmental (2)
Economics (...)
Land use and ownership (...)
Reservoir engineering (...)
+ Data type
Data sets (...)
Documents (30)
Maps (...)
Images (...)
+ Source
Sort by
User rating
Save this search
Register
Sign-in
Save search and get instant notificat
via email about new data that match
Name your search:
Email:
Password:
Save search
Well Locations for the Geysers
California Department of Conservation
Published: September 10, 2009
Map
14 reviews
40 downloads
Favorite
Download
Send

UX CONCEPT



UX CONCEPT



Recommendations

GOALS

Considerations | Near term

- Estimate development effort based on current understanding of UX concept. Scope to fit budget, time and quality.
- Consolidate meta data taxonomies to support faceted search, filtering and browsing on map and library.
- Test low-fidelity concept with additional user groups, especially financial investors and government representatives.
- Develop and test UI concept for data contributors
- Develop user stories from use cases to inform upcoming sprints .
- Proceed with iterative design and testing, in agile cycles.

GOALS

Considerations | The big picture

Consider

- The mobile context.
- Scalability of NGDS to other renewable energy data repositories and GIS based systems.
- Content / support for developers as part of www.geothermaldata.org.
- Online help community tools such as Get Satisfaction

Thank you

Suzanne Boyd
suzanneboyd@anthro-tech.com

Camy Naasz
camynaasz@anthro-tech.com

Jenny Greeve
jennygreeve@anthro-tech.com

Anthro-Tech

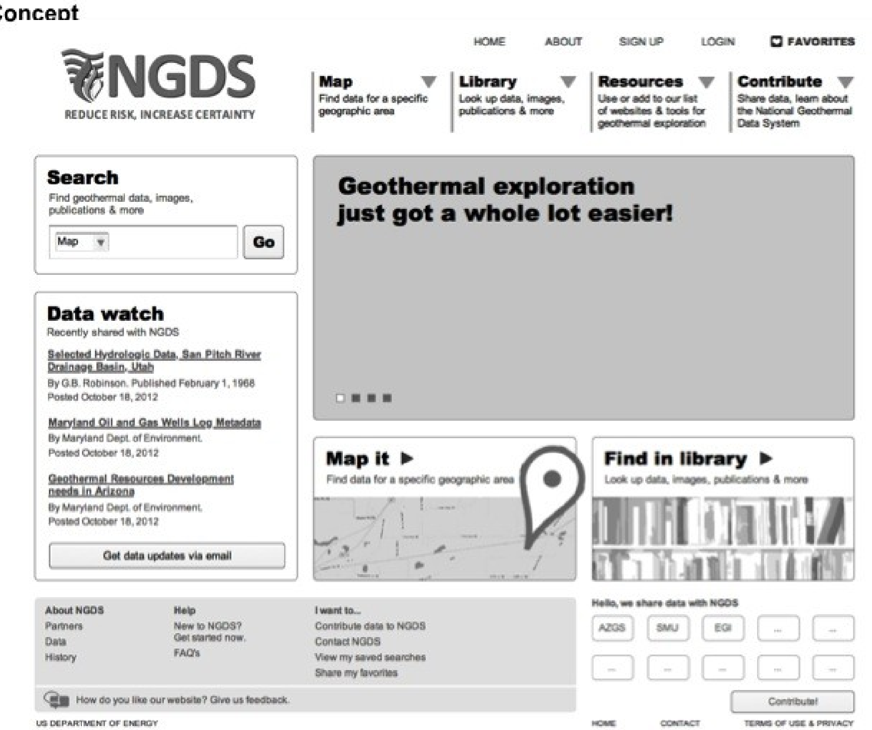
Appendix 2c

User Interface Final Design

User-Centered Design Process

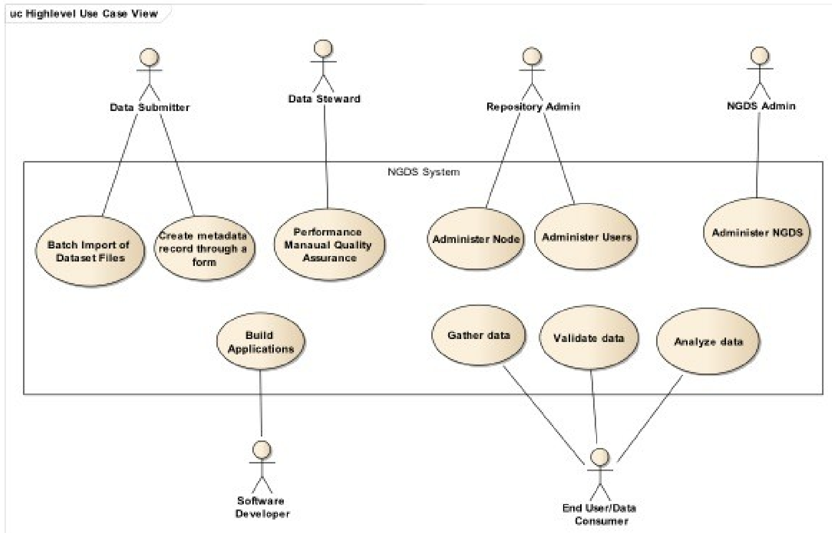
SIEMENS

UX Concept



SIEMENS

High-level Use Cases



SIEMENS

User-Centered Design Process (Cont.)

Interactive UI Mockups → Stanford Usability Testing → Define UI "Look & Feel"

Developed Interactive Mockup and Tested at Stanford Workshop

SIEMENS

SIEMENS

Usability Testing Results: Summary

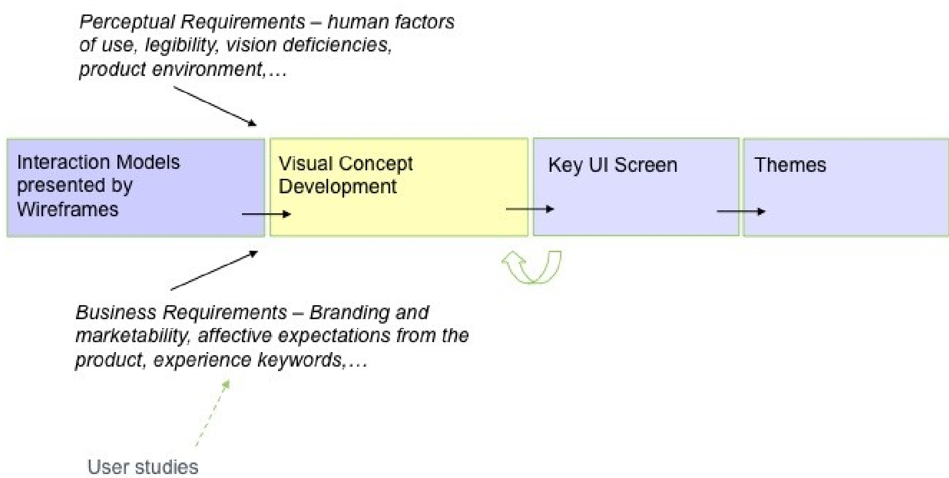
- Thanks to Roland's kind support, and to Arlene and Jay's great recruiting effort, **in total twenty one people, from Industry, Research and Government,** participated the NGDS UI Usability Testing for both the End User/Data Consumer UI and the Data Contributor UI.

- In general, all participants expressed that "**They would like to use the NGDS system frequently**" once the system is up and running. (On average **4.5 out of 5,** where 1 means very unlikely, 5 means very likely).

- And the participants thought "**The system was easy to use**". (**4 out of 5,** where 1 means very difficult to use, and 5 means very easy to use).

SIEMENS

Define NGDS " Look & Feel"



SIEMENS

Modern UI Trends

Flat UI

Skeuomorphism

Context sensitive navigation

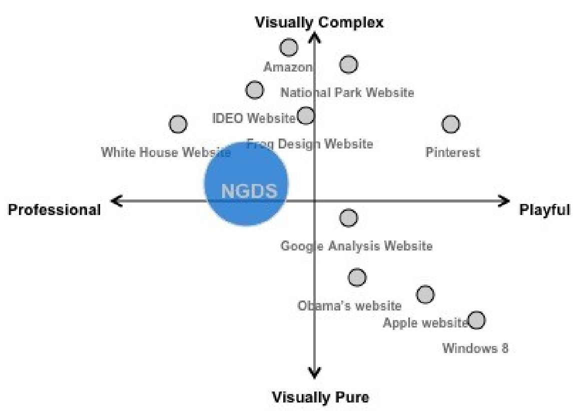
Collapsed content

Content chunking

Long pages

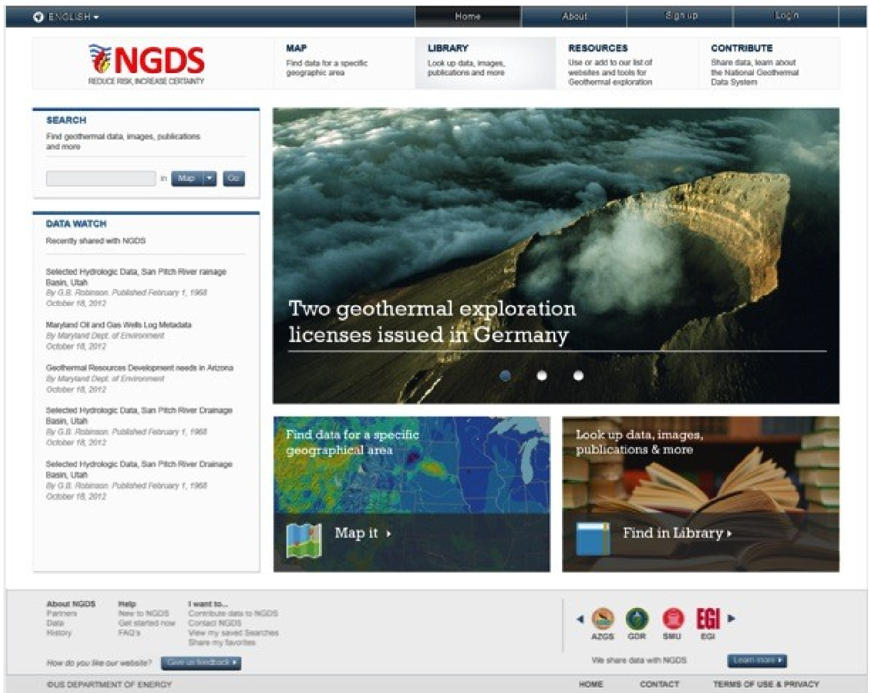
SIEMENS

Defining the NGDS UI Design Language

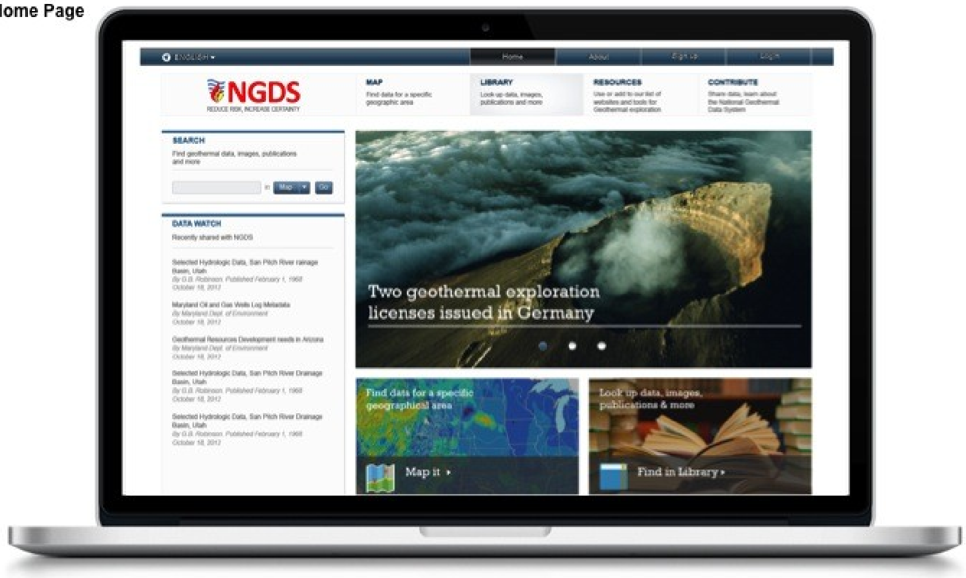


- Action page: all in one screen long; Information page: allows scrolling long page
- Content First. Hide non-essential parts.
- UI Style in between "Skeuomorphic" and "Digital(Flat)", Professional rather than Playful
- Awareness to different screen sizes

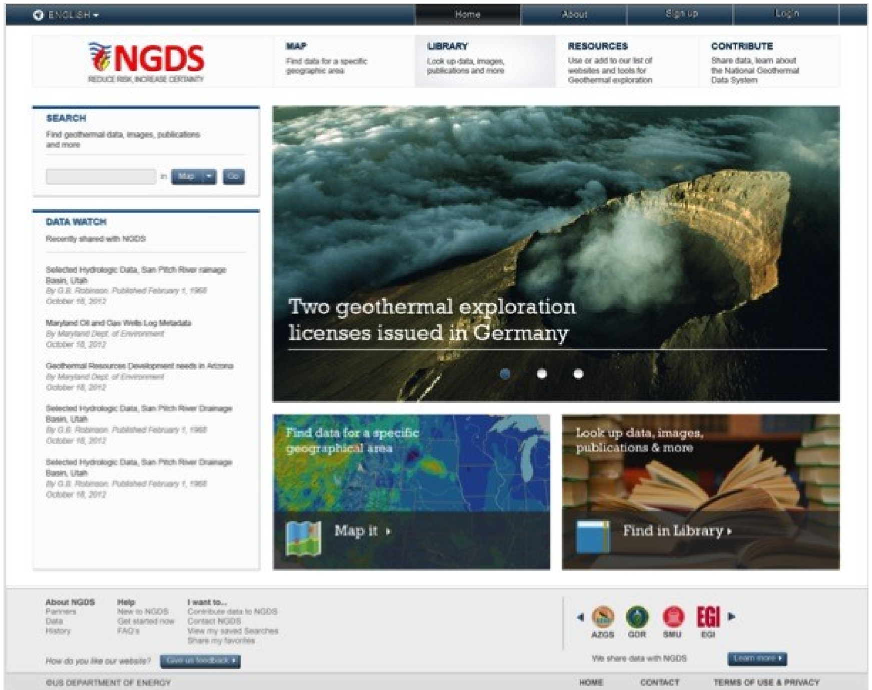
Final Design



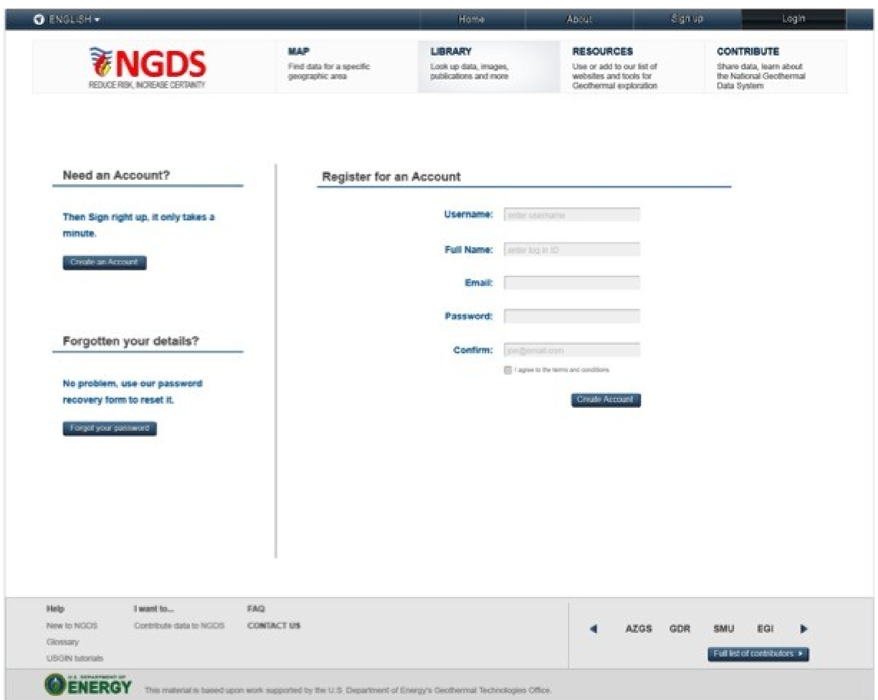
Home Page


ENGLISH
Home
About
Sign up
Log in
NGDS
REDUCE RISK, INCREASE CERTAINTY
MAP
Find data for a specific geographic area
LIBRARY
Look up data, images, publications and more
RESOURCES
Use or add to our list of websites and tools for Geothermal exploration
CONTRIBUTE
Share data, learn about the National Geothermal Data System
SEARCH
Find geothermal data, images, publications and more
in
Map
Go
DATA WATCH
Recently shared with NGDS
Selected Hydrologic Data, San Pitch River rainage Basin, Utah
By G.B. Robinson. Published February 1, 1968
October 18, 2012
Maryland Oil and Gas Wells Log Metadata
By Maryland Dept. of Environment
October 18, 2012
Geothermal Resources Development needs in Arizona
By Maryland Dept. of Environment
October 18, 2012
Selected Hydrologic Data, San Pitch River Drainage Basin, Utah
By G.B. Robinson. Published February 1, 1968
October 18, 2012
Selected Hydrologic Data, San Pitch River Drainage Basin, Utah
By G.B. Robinson. Published February 1, 1968
October 18, 2012
Two geothermal exploration licenses issued in Germany
Find data for a specific geographical area
Map it
Look up data, images, publications & more
Find in Library

Home Page

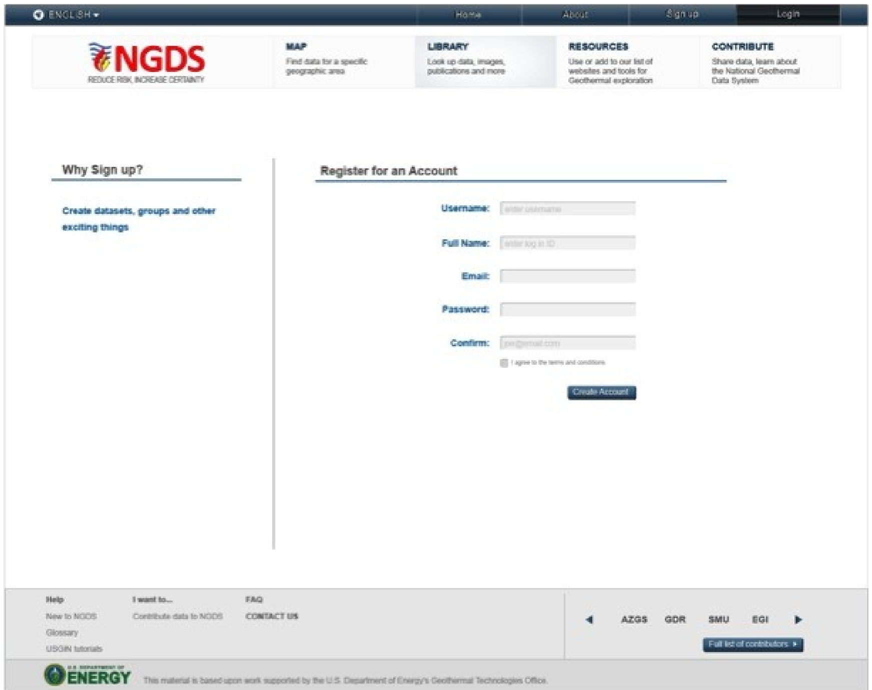


Login Page

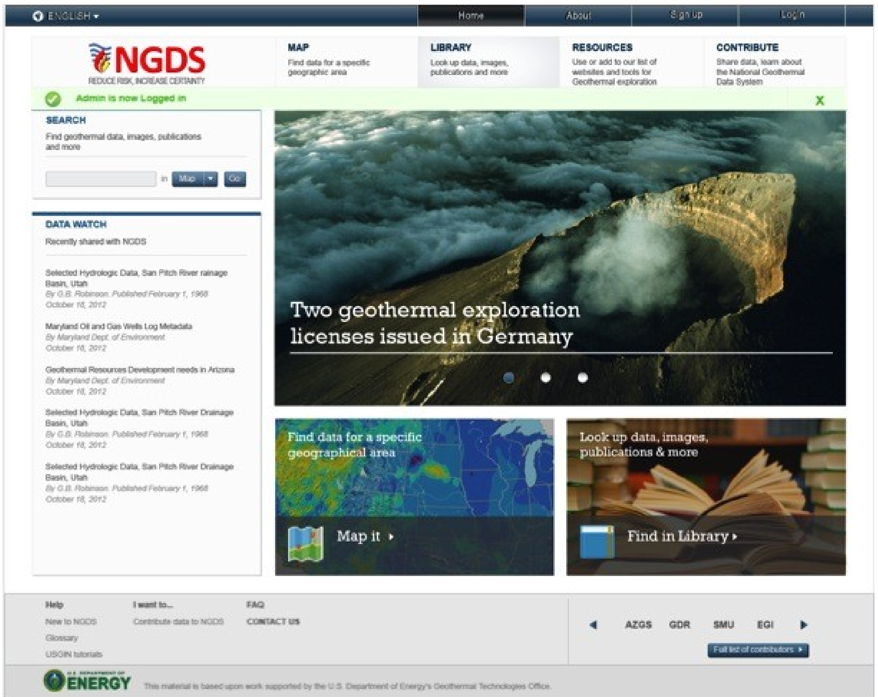


Login Page-

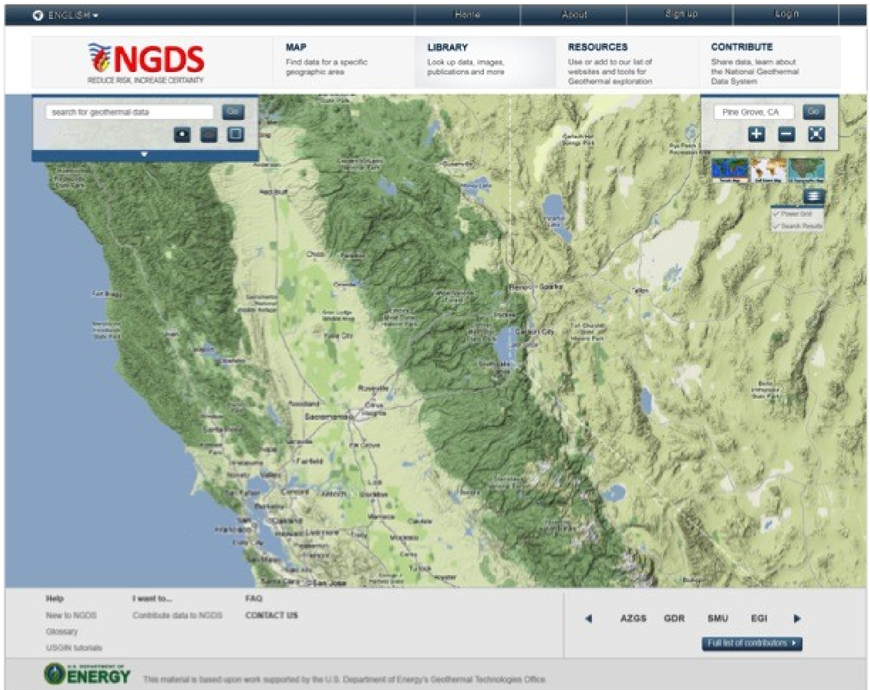
Register Account



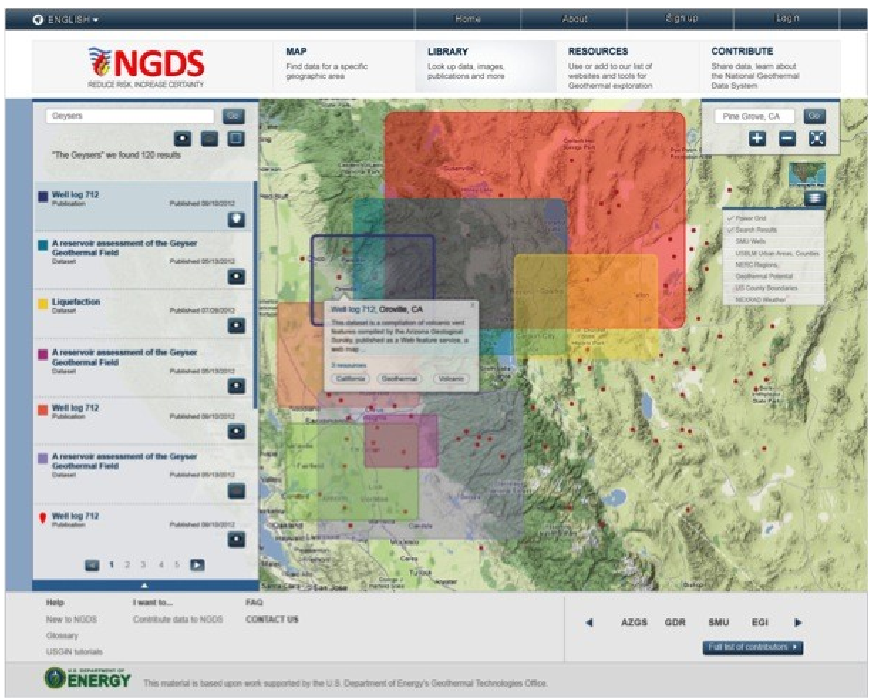
Logged in message



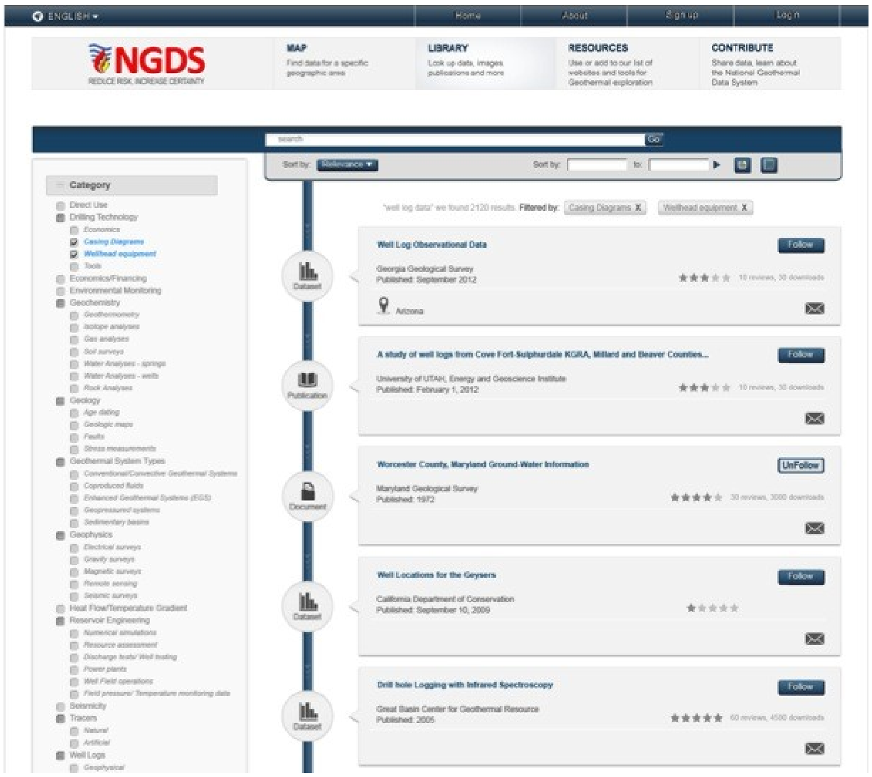
Map Page
1/2



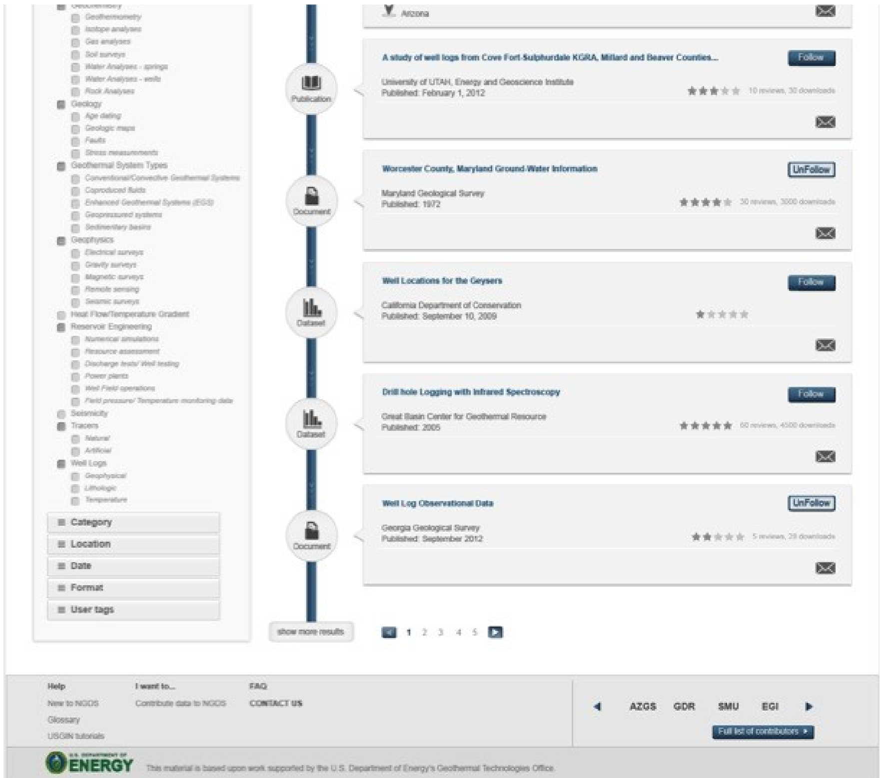
Map Page
2/2



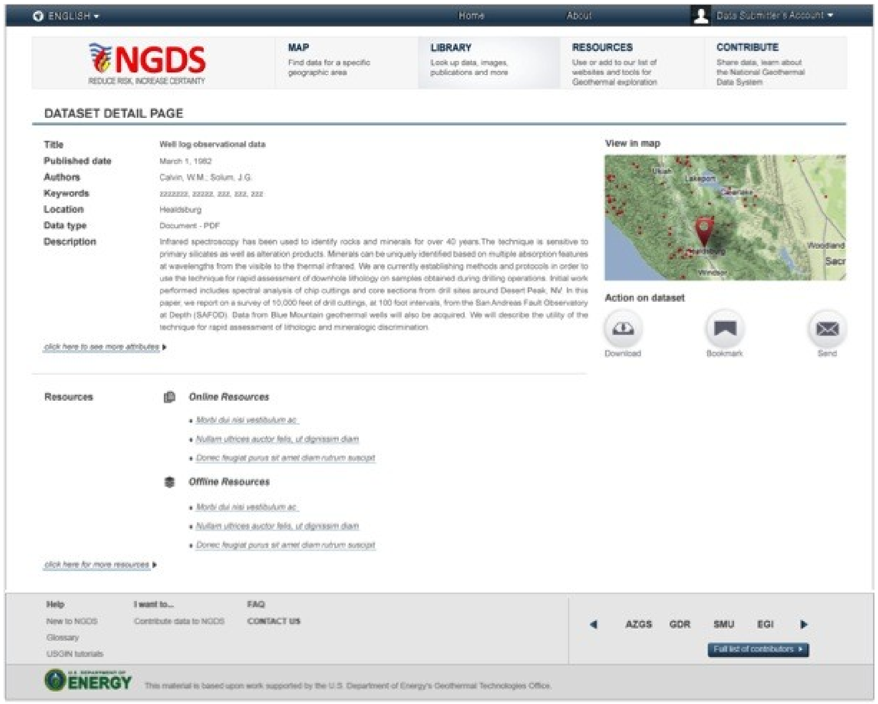
Library page
-after search
1/2



Library page -after search 2/2

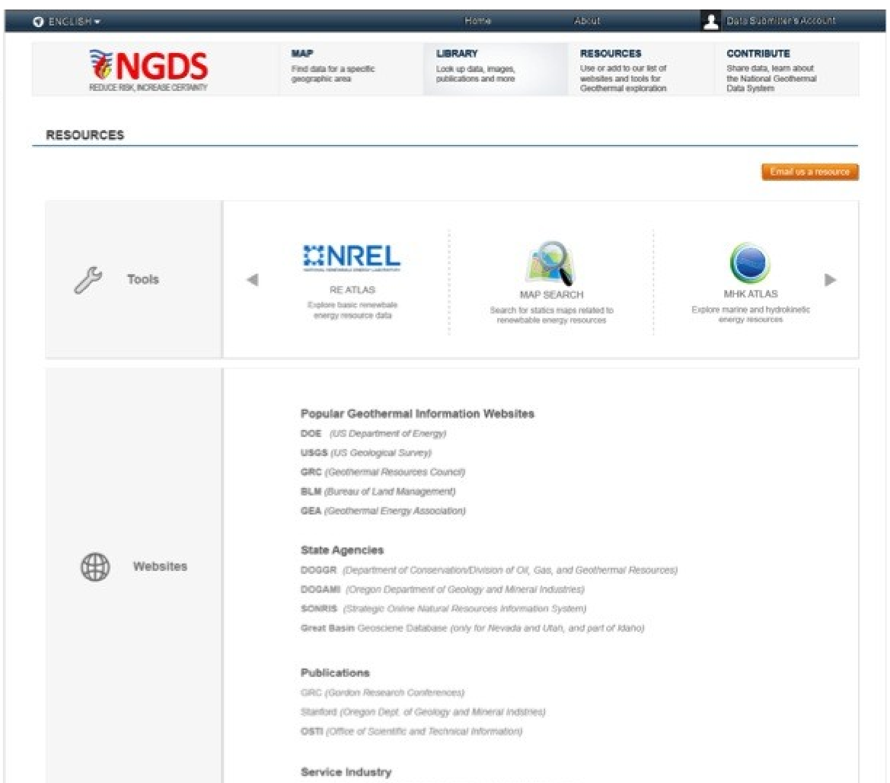


Library page
-Details



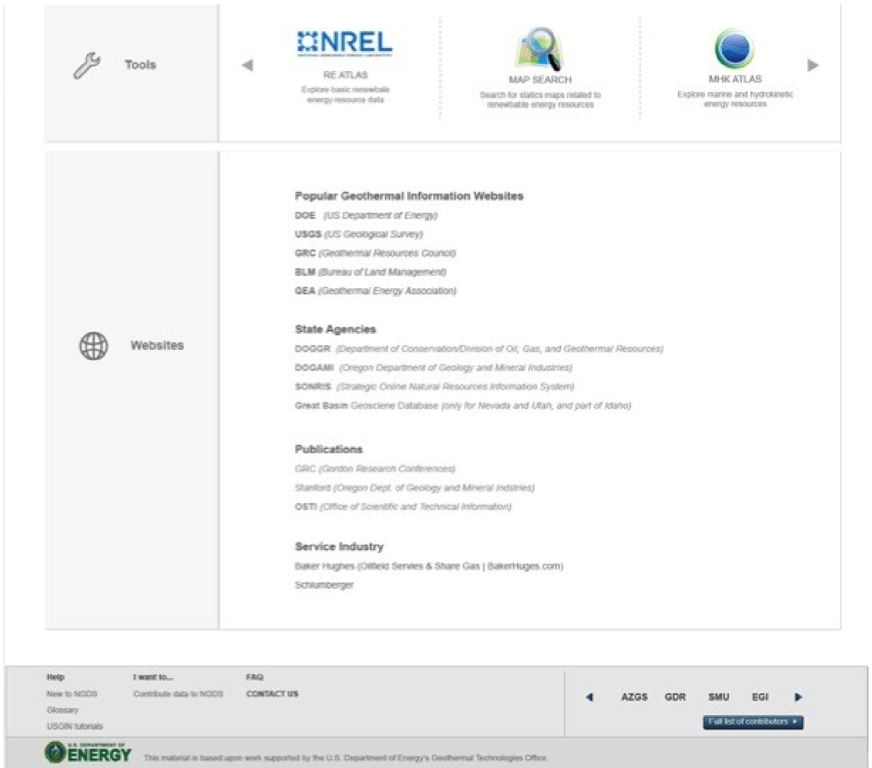
Resource page

1/2

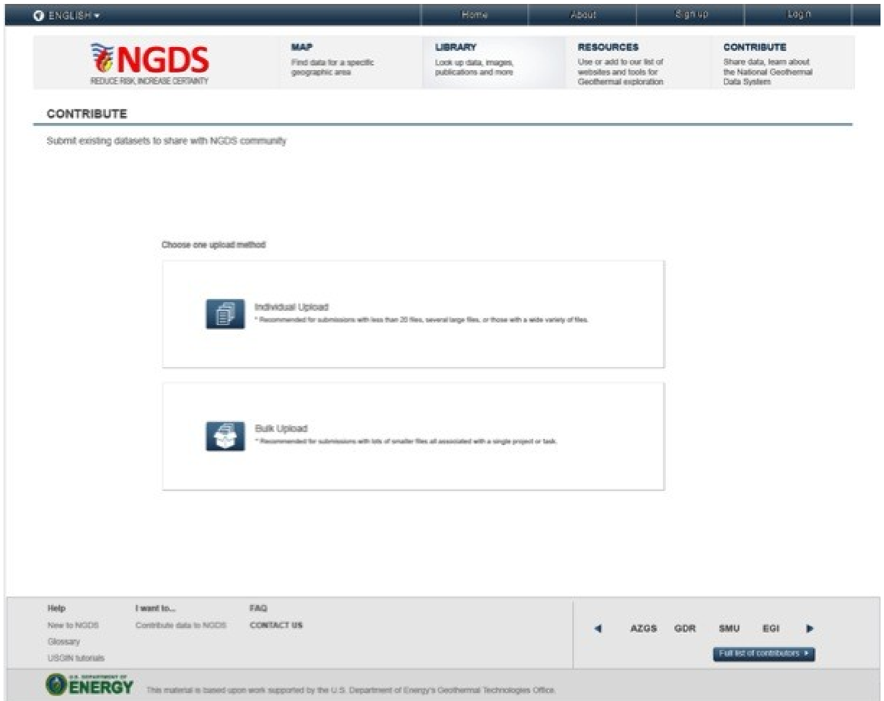


Resource page

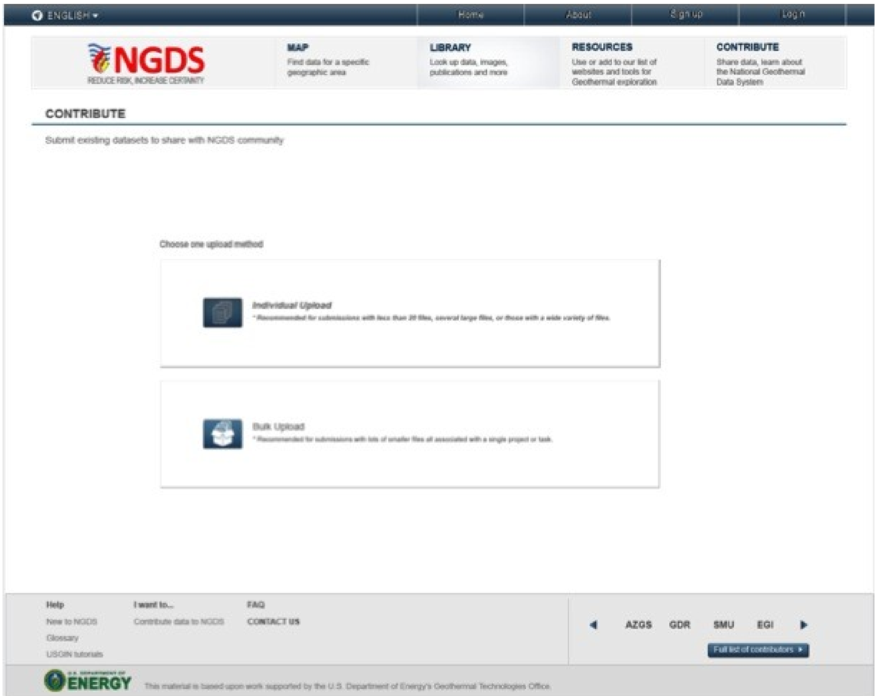
2/2



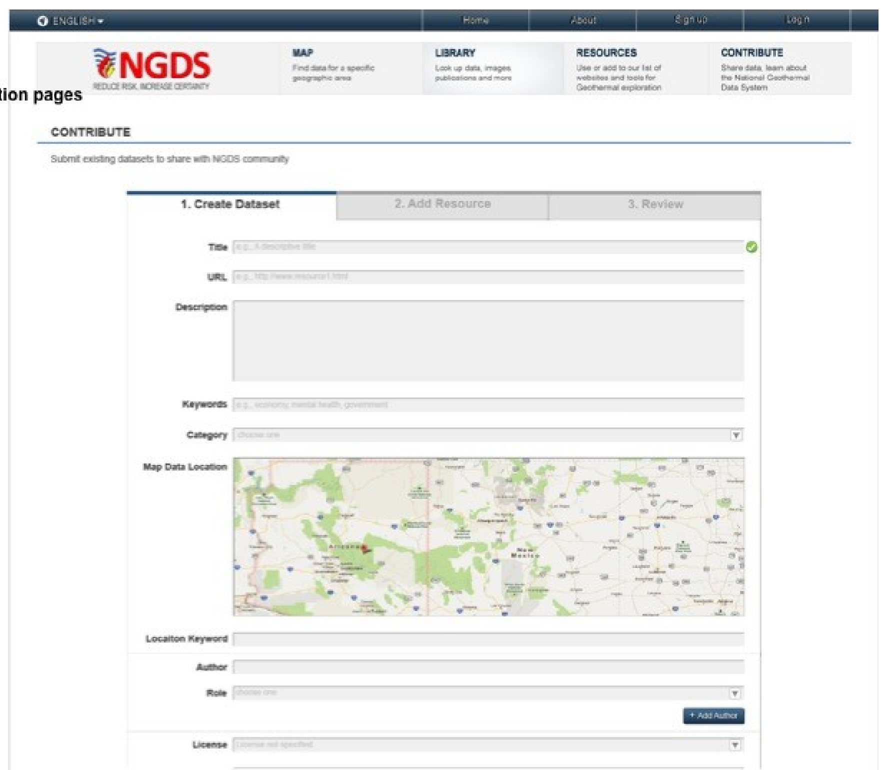
Contribute landing page



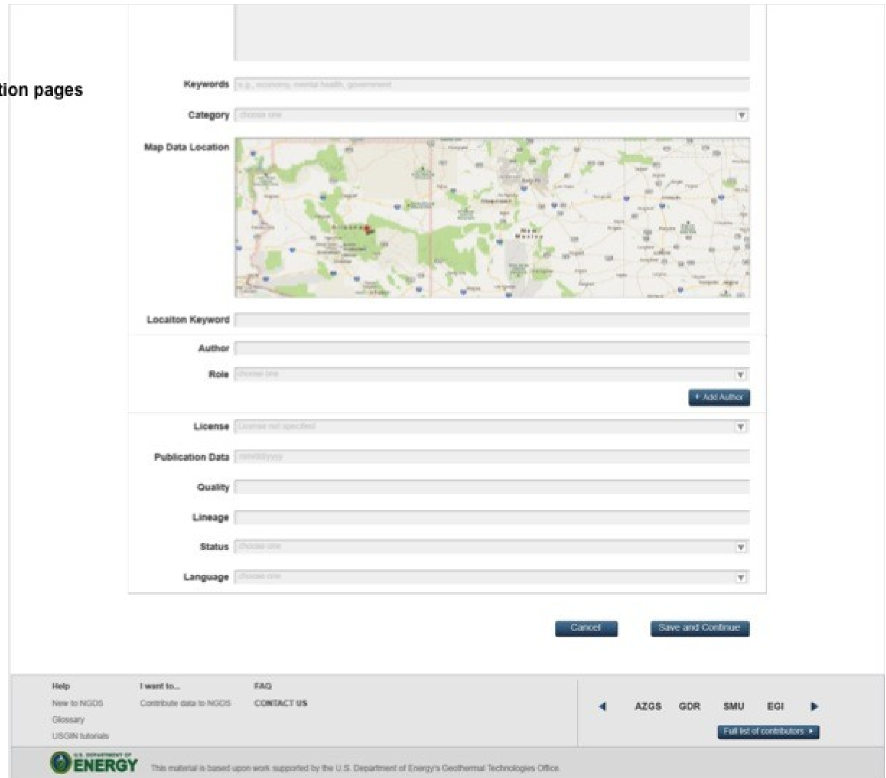
Contribute
landing page



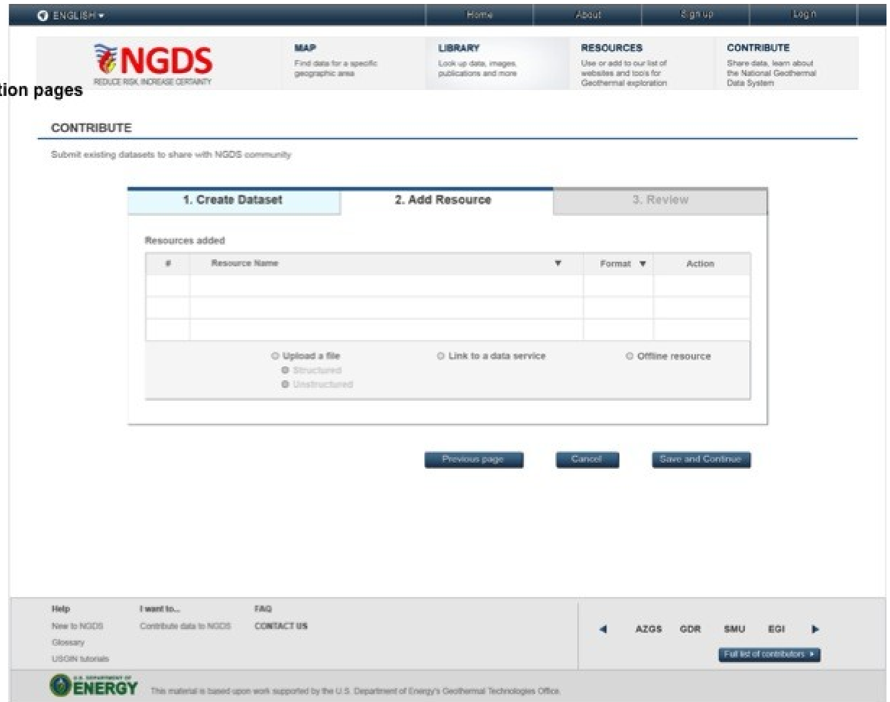
Contribute Section pages
1/10



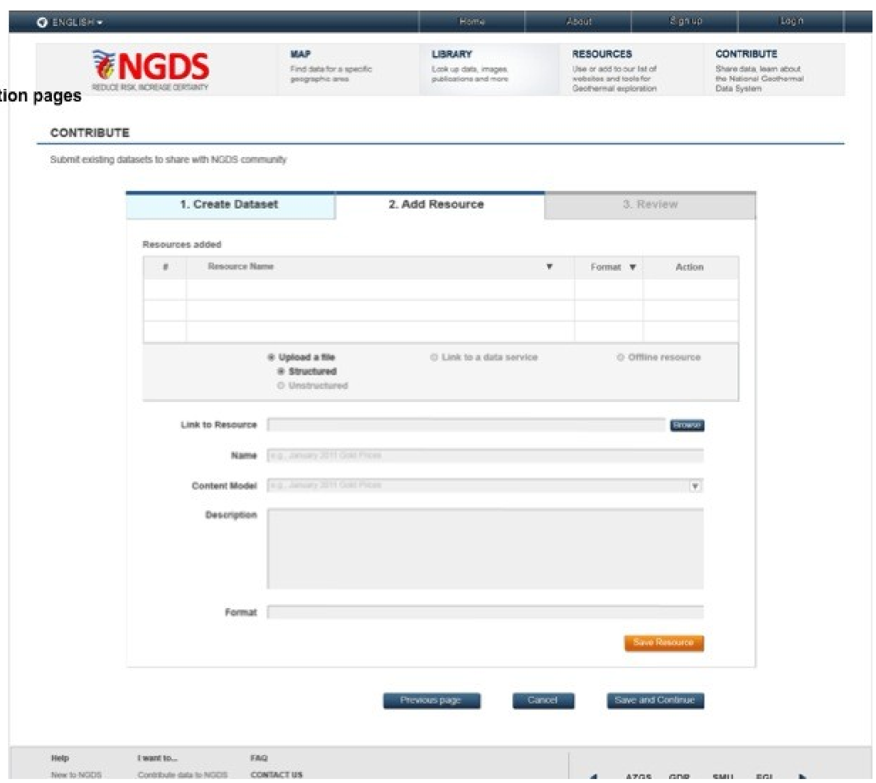
Contribute Section pages 2/10



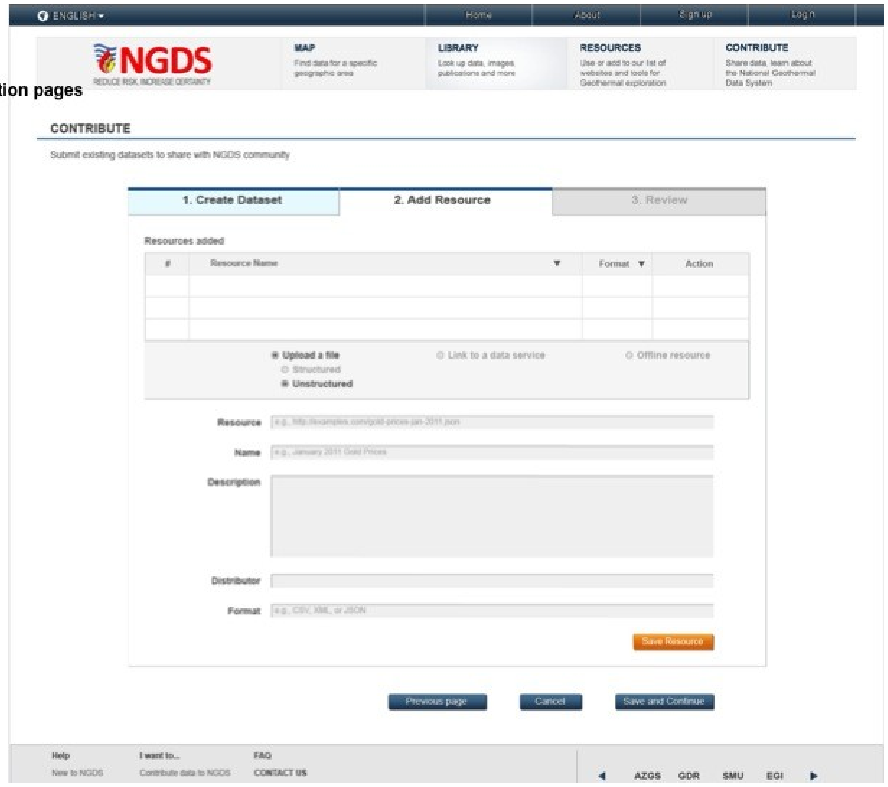
Contribute Section pages 3/10



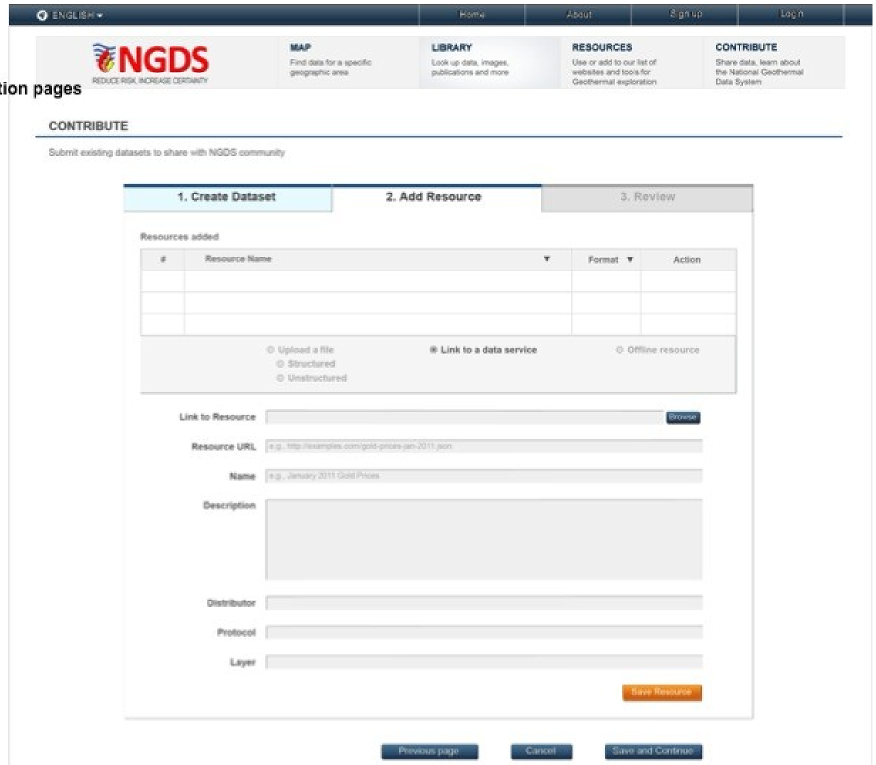
Contribute Section pages 4/10



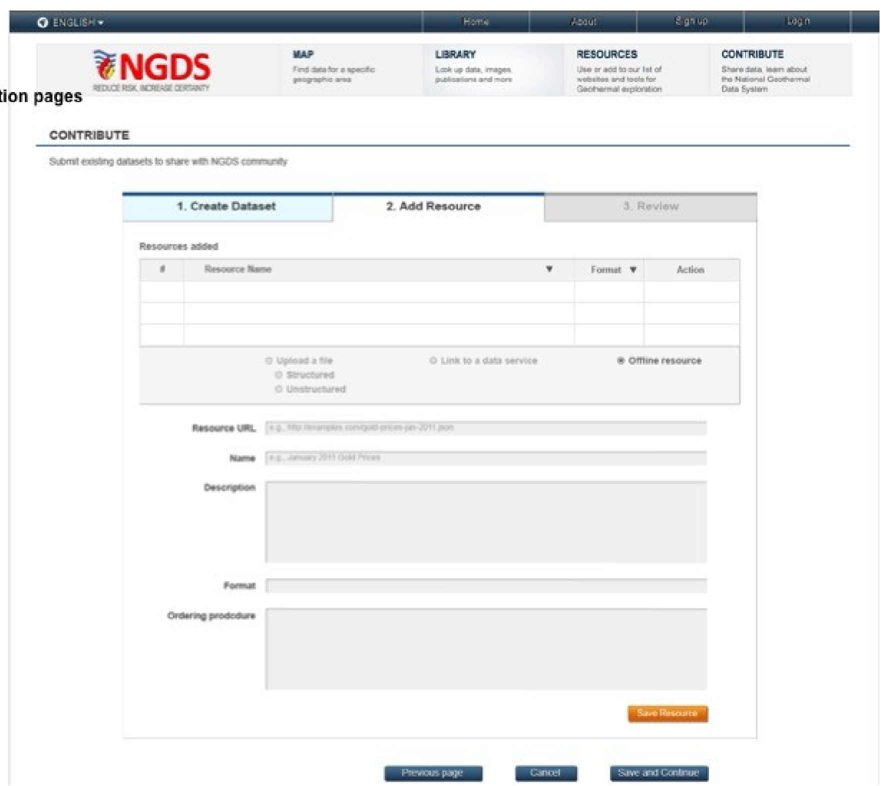
Contribute Section pages 5/10



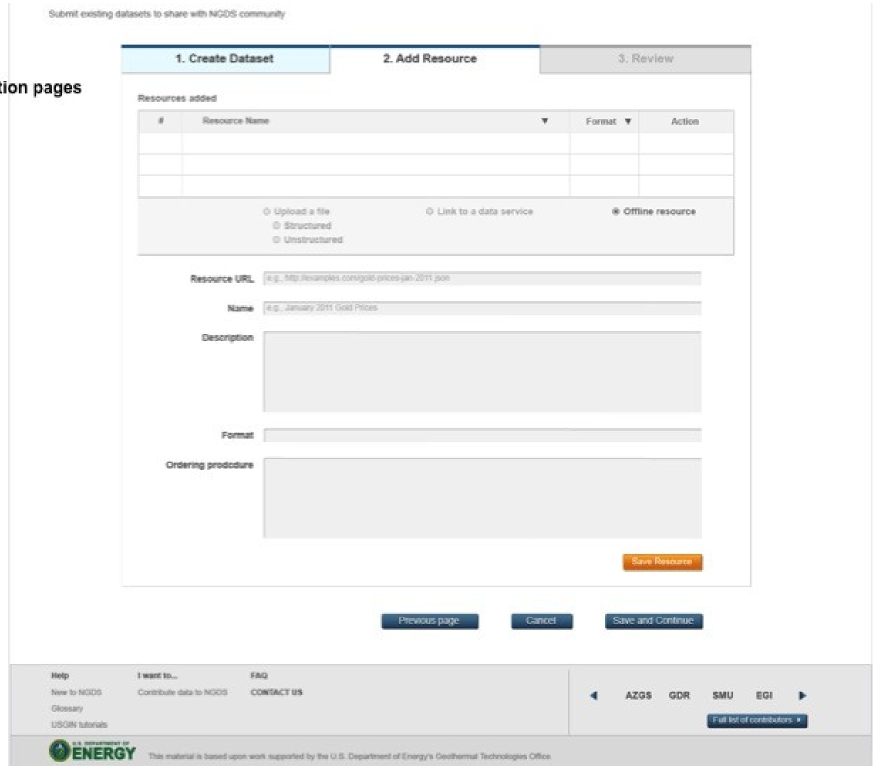
Contribute Section pages 6/10



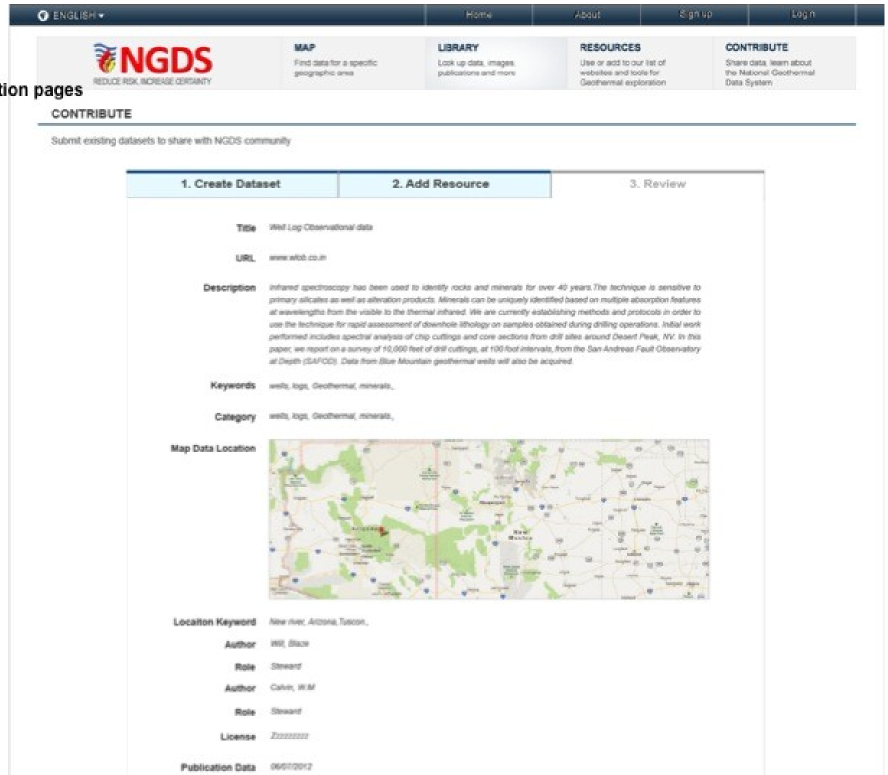
Contribute Section pages 7/10



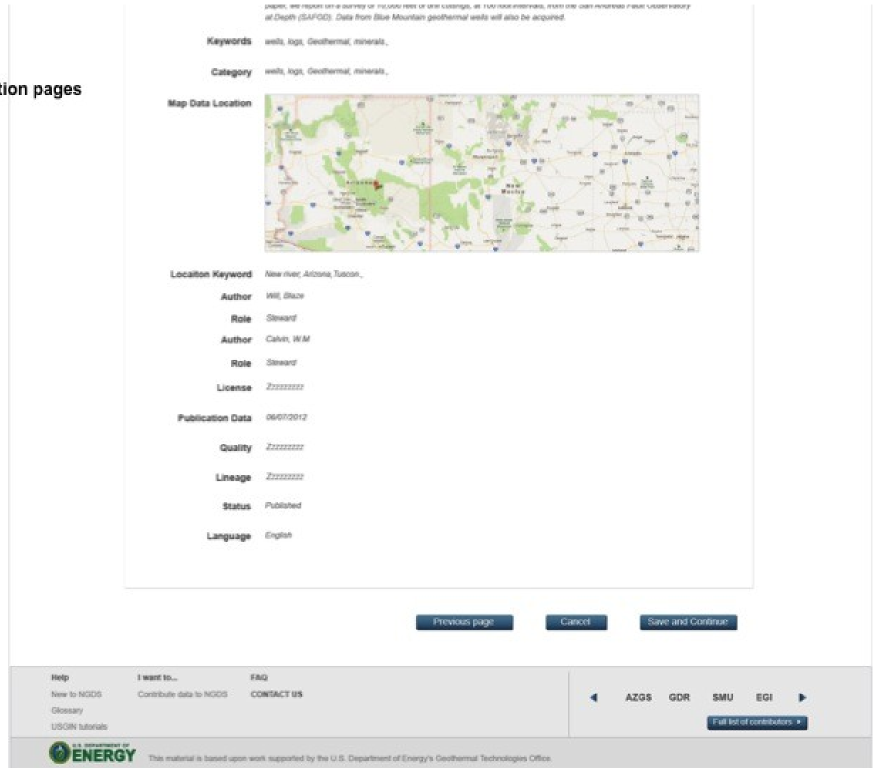
Contribute Section pages 8/10



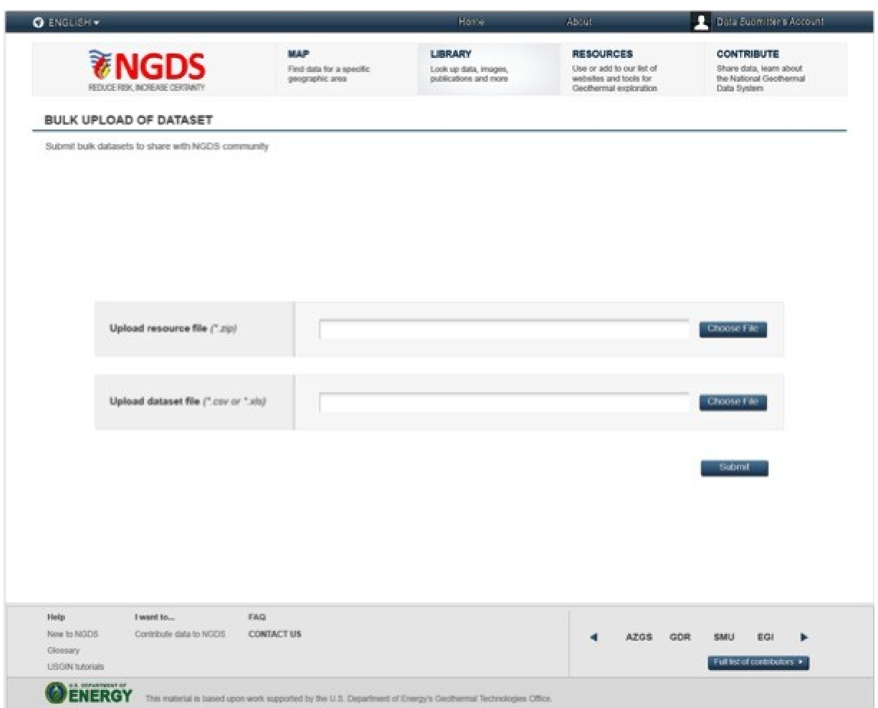
Contribute Section pages 9/10



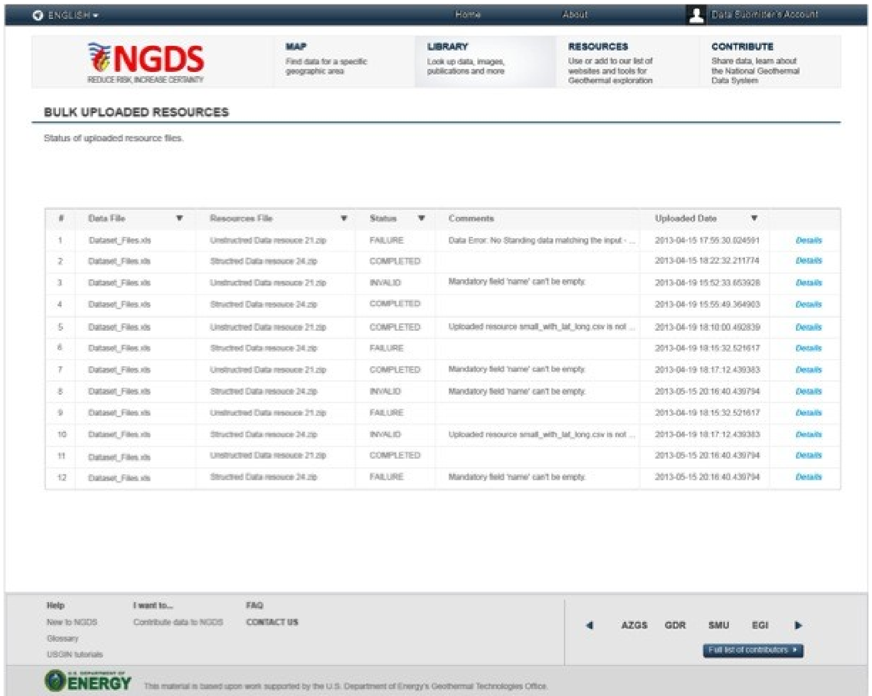
Contribute Section pages 10/10



Bulk Upload 1/2

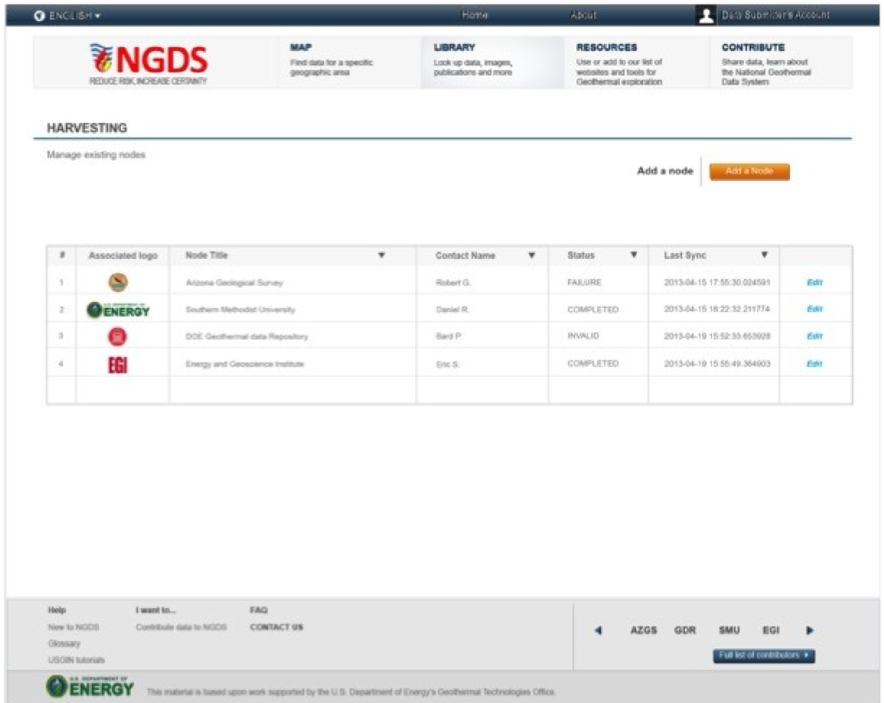


Bulk Upload 2/2

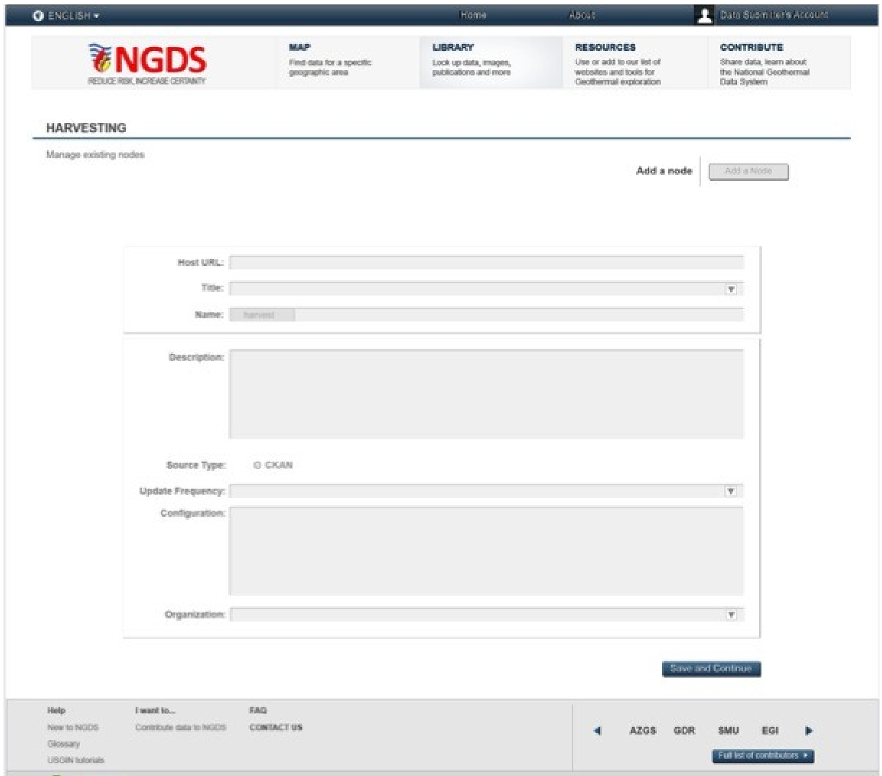


#	Data File	Resources File	Status	Comments	Uploaded Date	
1	Dataset_Files.xls	Unstructed Data resouce 21.zip	FAILURE	Data Error: No Standing data matching the input - ...	2013-04-15 17:55:30.024591	Details
2	Dataset_Files.xls	Structred Data resouce 24.zip	COMPLETED		2013-04-15 18:22:32.211774	Details
3	Dataset_Files.xls	Unstructred Data resouce 21.zip	INVALID	Mandatory field 'name' can't be empty.	2013-04-19 15:52:33.653928	Details
4	Dataset_Files.xls	Structred Data resouce 24.zip	COMPLETED		2013-04-19 15:55:49.364903	Details
5	Dataset_Files.xls	Unstructred Data resouce 21.zip	COMPLETED	Uploaded resource small_with_lat_long.csv is not ...	2013-04-19 18:10:00.492839	Details
6	Dataset_Files.xls	Structred Data resouce 24.zip	FAILURE		2013-04-19 18:15:32.521617	Details
7	Dataset_Files.xls	Unstructred Data resouce 21.zip	COMPLETED	Mandatory field 'name' can't be empty.	2013-04-19 18:17:12.439383	Details
8	Dataset_Files.xls	Structred Data resouce 24.zip	INVALID	Mandatory field 'name' can't be empty.	2013-05-15 20:16:40.439794	Details
9	Dataset_Files.xls	Unstructred Data resouce 21.zip	FAILURE		2013-04-19 18:15:32.521617	Details
10	Dataset_Files.xls	Structred Data resouce 24.zip	INVALID	Uploaded resource small_with_lat_long.csv is not ...	2013-04-19 18:17:12.439383	Details
11	Dataset_Files.xls	Unstructred Data resouce 21.zip	COMPLETED		2013-05-15 20:16:40.439794	Details
12	Dataset_Files.xls	Structred Data resouce 24.zip	FAILURE	Mandatory field 'name' can't be empty.	2013-05-15 20:16:40.439794	Details

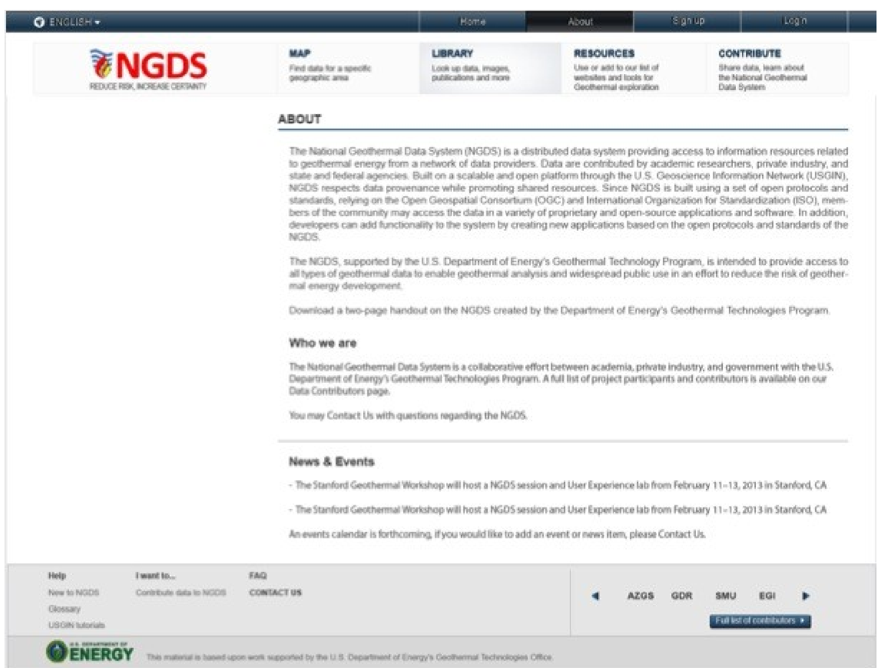
Harvest Home Page



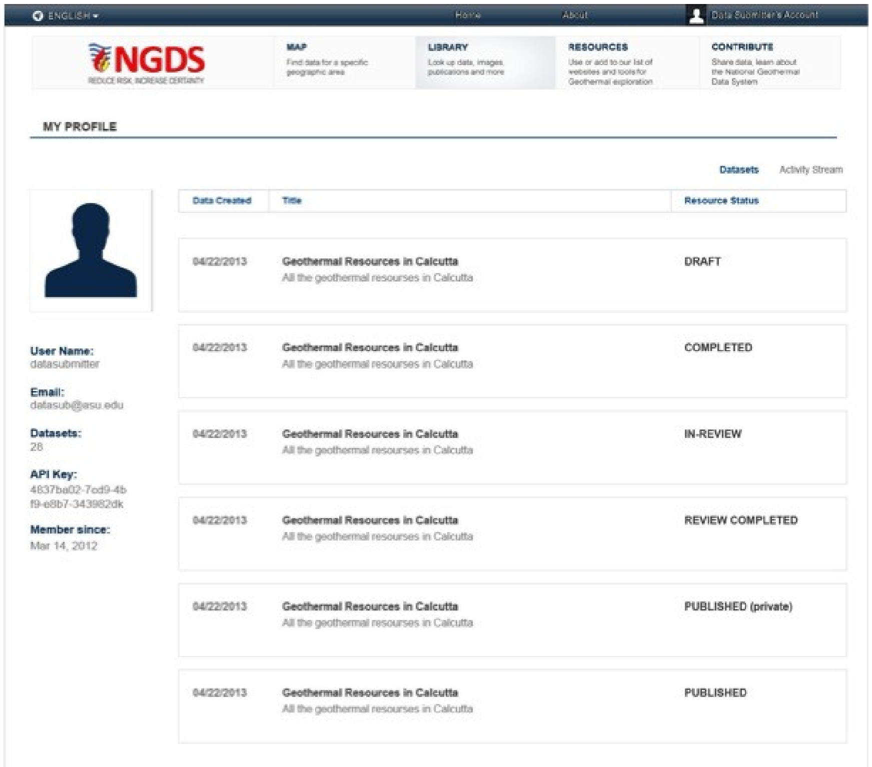
Harvest : Add Node Page



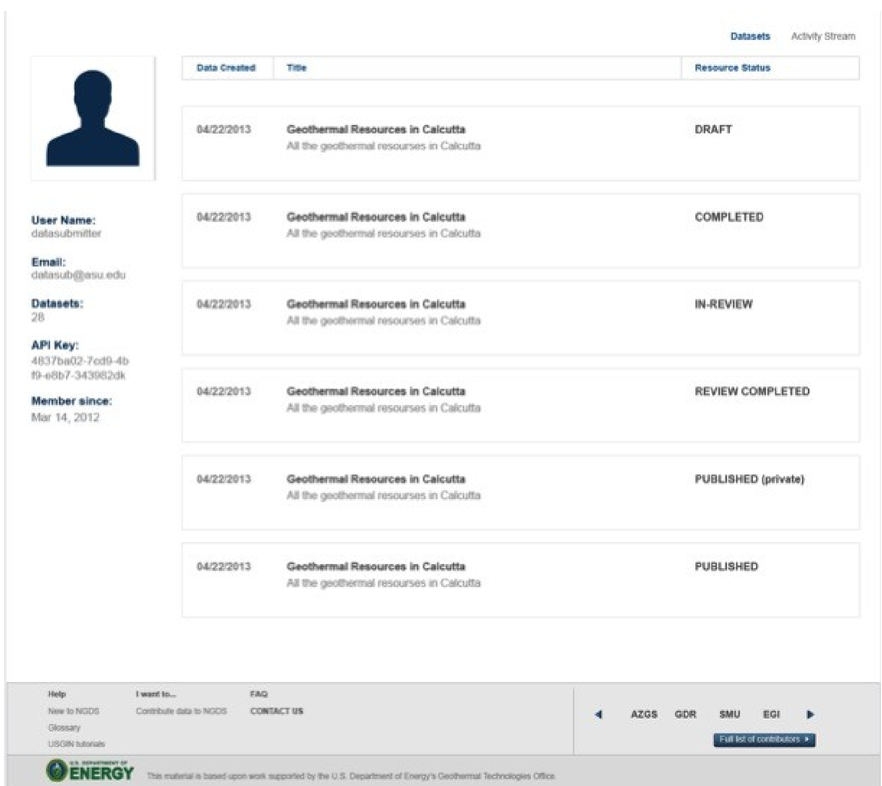
About Page



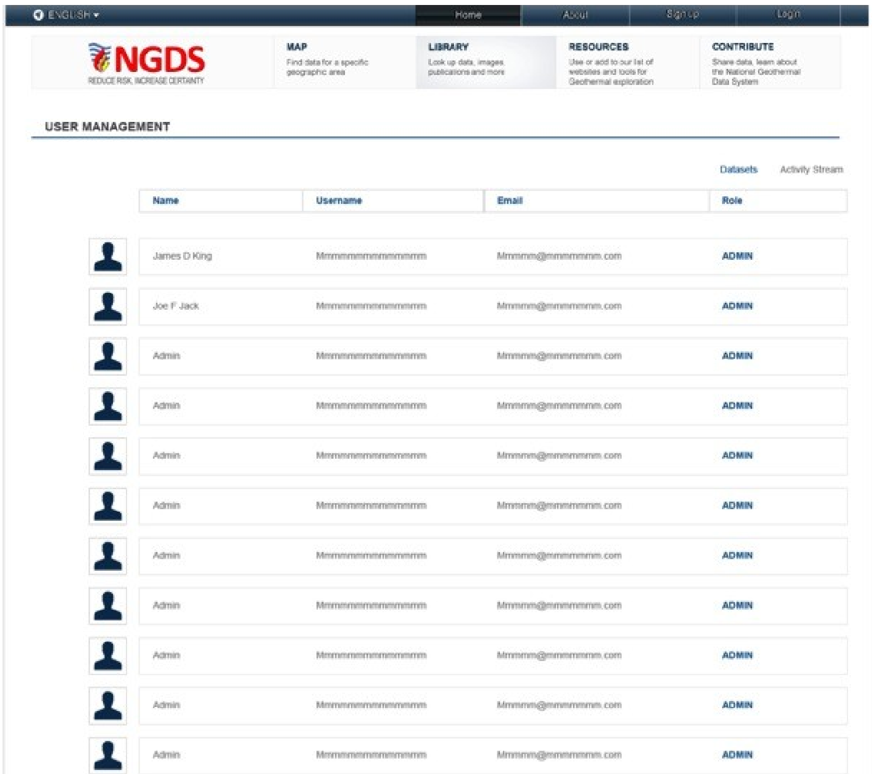
User Profile
Page
1/2



User Profile
Page
2/2

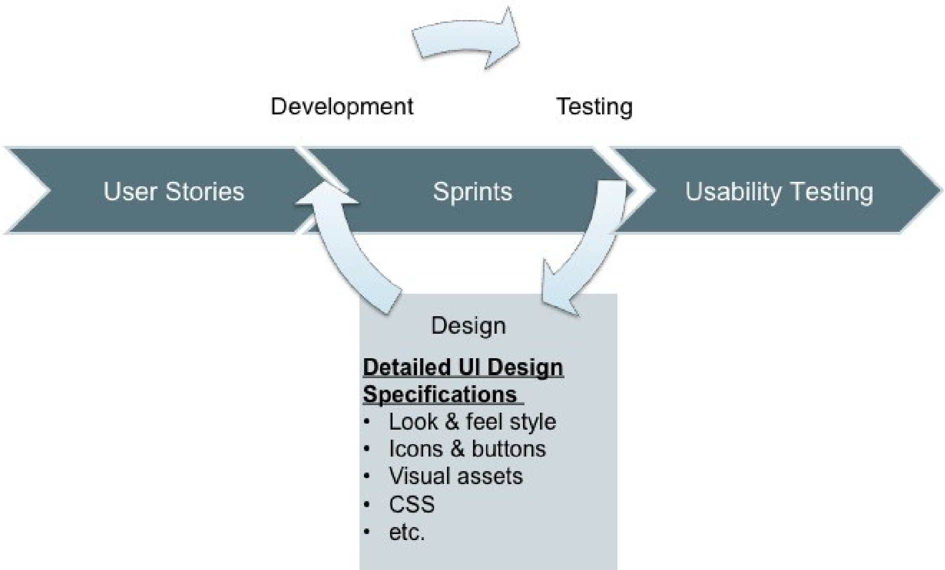


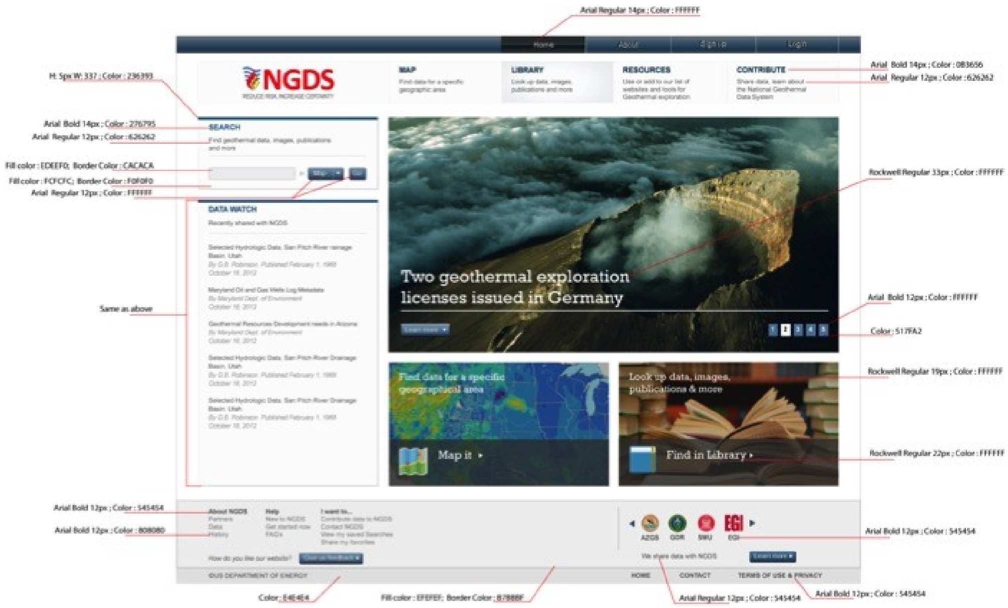
Admin user Management Page

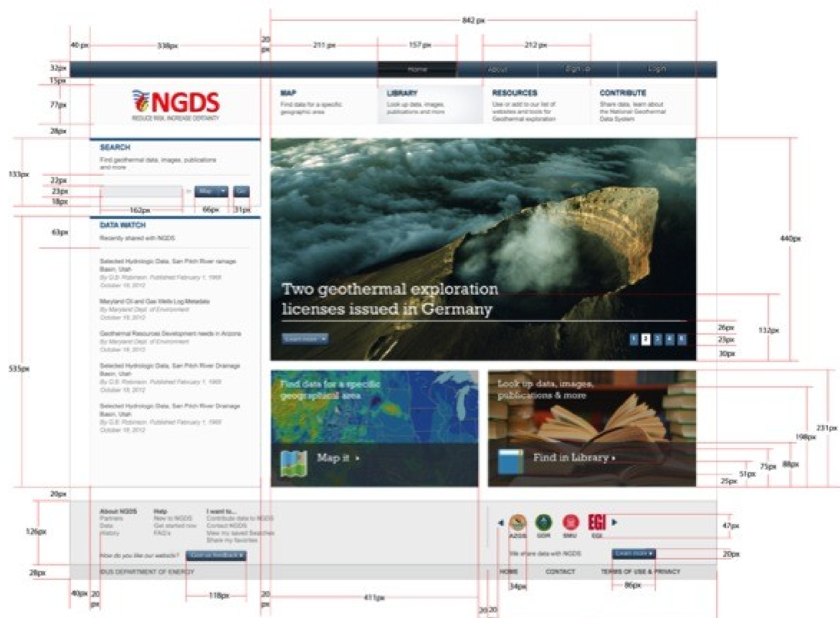


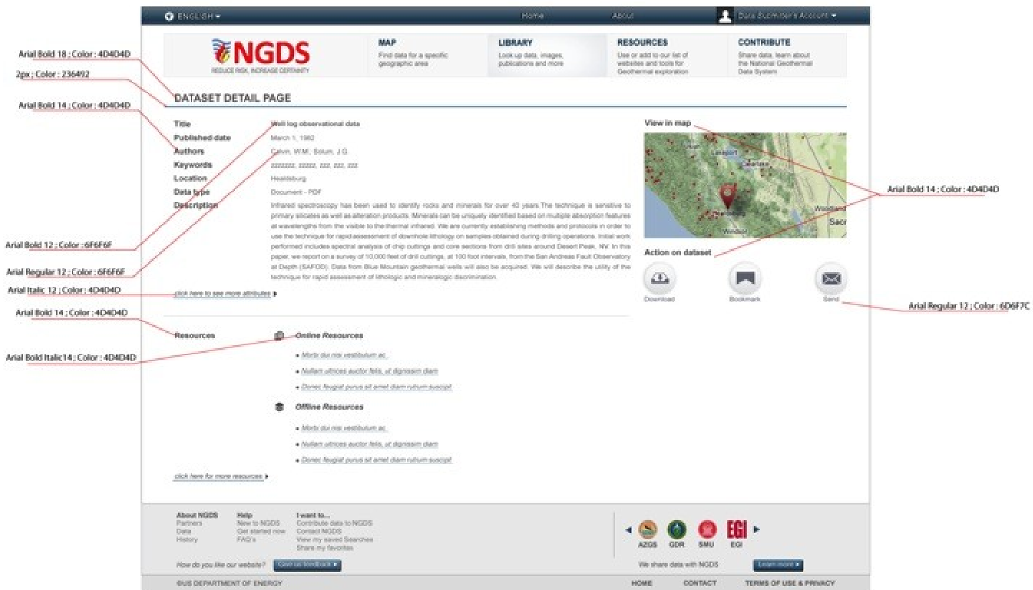
SIEMENS

User-Centered Design Process (Cont.)




Arial Regular 14px ; Color : FFFFFF
Home
About
Sign Up
Login
NGDS
REDUCE RISK, INCREASE CERTAINTY
MAP
LIBRARY
RESOURCES
CONTRIBUTE
H: 5px W: 337 ; Color : 236393
Arial Bold 14px ; Color : 0B3656
Arial Regular 12px ; Color : 626262
SEARCH
Arial Bold 14px ; Color : 276795
Arial Regular 12px ; Color : 626262
Fill color : EDEEF0; Border Color : CACACA
Fill color : FCFCFC; Border Color : F0F0F0
Arial Regular 12px ; Color : FFFFFF
DATA WATCH
Same as above
Two geothermal exploration licenses issued in Germany
Rockwell Regular 33px ; Color : FFFFFF
Arial Bold 12px ; Color : FFFFFF
Color : 517FA2
Find data for a specific geographical area
Map it
Look up data, images, publications & more
Find in Library
Rockwell Regular 19px ; Color : FFFFFF
Rockwell Regular 22px ; Color : FFFFFF
Arial Bold 12px ; Color : 545454
Arial Bold 12px ; Color : 808080
AZGS
GDR
SMU
EGI
Arial Bold 12px ; Color : 545454
HOME
CONTACT
TERMS OF USE & PRIVACY
Color : E4E4E4
Fill color : EFEFEF; Border Color : B7B8BF
Arial Regular 12px ; Color : 545454
Arial Bold 12px ; Color : 545454


842 px
40 px
338px
20 px
211 px
157 px
212 px
32px
15px
77px
28px
133px
22px
23px
18px
162px
66px
31px
63px
440px
26px
23px
30px
132px
535px
231px
198px
88px
75px
51px
25px
20px
126px
28px
47px
20px
40px
118px
411px
34px
86px
433px



Arial Bold 18 ; Color : 4D4D4D
2px ; Color : 236492
Arial Bold 14 ; Color : 4D4D4D
Arial Bold 12 ; Color : 6F6F6F
Arial Regular 12 ; Color : 6F6F6F
Arial Italic 12 ; Color : 4D4D4D
Arial Bold 14 ; Color : 4D4D4D
Arial Bold Italic14 ; Color : 4D4D4D
Arial Bold 14 ; Color : 4D4D4D
Arial Regular 12 ; Color : 6D6F7C
NGDS
REDUCE RISK, INCREASE CERTAINTY
MAP
LIBRARY
RESOURCES
CONTRIBUTE
DATASET DETAIL PAGE
Title
Published date
Authors
Keywords
Location
Data type
Description
Well log observational data
View in map
Action on dataset
Download
Bookmark
Send
Resources
Online Resources
Offline Resources

SOFTWARE REQUIREMENTS SPECIFICATION
NGDS

Version 2.7

11/05/2012

Version History

Version	Author	Date	Reason
0.1	SCR	08/24/2012	Initial Draft Created
1.0	SCR	09/15/2012	Initial Requirements Specifications
1.1	SMR/USGIN	9/18/2012	Review and comment (track changes on)
1.2	SCR	10/2/2012	Address Feedback
2.0	SCR	10/11/2012	Incorporated Steve Richard's changes
2.1	SCR	10/16/2012	Incorporated AnthroTech end-user research input
2.2	SCR	10/18/2012	Incorporate Matt MacKenzie's revisions
2.3	SCR	10/19/2012	Incorporate Ryan Clark's revisions
2.4	SCR	10/19/2012	Duane Nickull
2.5.3	SCR	10/26/2012	Restructured document to incorporate Steve Richard's proposal.
2.6	SCR	10/31/2012	Incorporate Matt MacKenzie's revisions
2.7	SCR	11/05/2012	Incorporate Duane Nickull

Executive Summary

It is the U.S. Department of Energy (or DOE) vision to discover and exploit geothermal energy sources (s. Section 1.1). The DOE and other organizations are funding a variety of research activities around that vision. So far these research activities focus on either **collecting** geothermal data or **presenting and analyzing** such data.

Existing Data Collection Activities:

Data collection activities focused on the creation of various repositories of geothermal data. For example, DOE-GDR – Geothermal Data Repository (https://gdr.openei.org), SMU – Southern Methodist University (http://geothermal.smu.edu), EGI – Energy and Geosciences Institute (http://egi.utah.edu/), USGS – U.S. Geological Survey (http://energy.usgs.gov/OtherEnergy/Geothermal.aspx), and AASG – Association of American State Geologists (http://www.stategeothermaldata.org/). These repositories collect structured data (e.g. well headers or heat flow data expressed in well formed tables) as well as unstructured data (mainly publications as well as other documents – which have varied content as text, photos and had written text, that are usually not represented in well known tabular formats). Due to these initiatives, a large amount of data has been made digitally available: Structured data is aggregated in datasets which are exposed as Web Feature Services [WFS], Web Map Services [WMS] and Web Coverage Services [WCS] while unstructured data is made available for download. These repositories are all based on individual software systems which all comply with standardized protocols (WFS, WMS, and WCS).

Moreover each repository exposes a catalog that allows for searching within the data of the repository. The catalog has a metadata entry for each structured dataset (i.e. aggregated structured data). The metadata entry describes the content of the dataset, such as the type of data (e.g. heat flow, well log, etc.), its origin, and the geographic region covered by the dataset. The details of the metadata entry depend, at least partially, on the type of data. In the case of unstructured data, the metadata entry comprises the results of a keyword indexing service.

The catalog is made accessible via the Catalog Service for the Web (CSW) for all users without any password protection. The publication of datasets in NGDS, however, is restricted to authorized users.

Existing Presentation and Analysis of Data Activities:

Data analysis activities focused on the development of tools (such as the geothermal prospector [URL: http://maps.nrel.gov/gt_prospector]) that allow for analyzing and visualizing data. There are also commercial tools for data analysis available.

How is NGDS going to contribute to this landscape of research projects?

NGDS (National Geothermal Data System) shall fill the gap between data collection activities and data analysis activities in existing DOE funded projects.

The most important new feature will be its ability to harvest the catalogs of all existing geothermal repositories (specifically the SMU repository as well as the AASG repository) provided that they comply with the standardized CSW and make it freely available.

But NGDS shall provide more than an aggregating CSW catalog: It shall also provide the future default solution for geothermal repositories and help the DOE to build a grid of geothermal data repositories. The content of all repositories in this grid shall be made searchable via a federated search mechanism that will give the user the possibility to execute facetted searches across all repositories, evaluate the detected datasets and make them available for download[1]. The federation will make search transparent to end users/ data consumers that will have access to all data in the NGDS network of repositories through any node of the system.

The target audience of NGDS, here called End Users/Data Consumers is a variety of users including legislators, federal and state agencies, financial investors, researchers, educators and students, interested public in general and industry representatives.

In the following we list the five basic needs that need to be realized by NGDS:

1. NGDS shall enable data collectors to create and administrate a repository for geothermal data.
2. NGDS shall enable end users/data consumers to search geothermal data across a multitude of repositories
3. NGDS shall enable end users/data consumers to evaluate discovered data
4. NGDS shall enable end users/data consumers to acquire (i.e. download) selected data
5. NGDS shall enable the end users/data consumers to analyze selected data

These five basic needs will be further explained as follows.

Need for a Standard Data Repository

1 Datasets are made available for download if they are freely accessible.

As outlined above, the currently funded activities lead to a variety of software systems for archiving geothermal data. However, the DOE requires a default software system to be used for future geothermal data collection projects. We call this software system a “node-in-a-box”.

This node-in-a-box must have a simple way for setting up such a repository. The repository must be simple to administrate, flexible with respect to configuration and adaptation by the data collector, and must rely on standard technologies. Most importantly, it must allow for federating its content thus allowing the DOE to realize a grid of repositories for geothermal data.

Besides enabling data collectors to store their geothermal data, it must also provide a minimum set of housekeeping features such as system monitoring, user management, logging of activities, support for backups, and basic security. The system must also realize a number of basic features to adjust access rights to data within the system, distinguishing between data consumers (readers that can access and search for data published in the system) and providers (writers that can provide new content and modify existing data and metadata).

Also, the repository must support some basic business process for uploading, evaluating and publishing datasets. This process may involve multiple users, e.g. one user responsible for uploading data, and a second user responsible for reviewing the uploaded data and making it publicly available (or rejecting it).

Search across Multiple Repositories

NGDS must allow for searching through the catalog of datasets within the grid of federated NGDS nodes. This catalog, which indexes the data across all nodes, must be accessible by each participating node.

End users/data consumers must be provided with a user interface that allows for executing faceted searches that combine many different search filters such as geographic region, type of data, data provider as well as keywords. The found results shall be visualized in an appropriate user interface. The most important aspect of this user interface is that it shall visualize the found datasets in a map in an appropriate way (if found results can be geo-located and displayed appropriately). The map must allow for the usual map features such as panning and zooming. Also, the user interface must display metadata for the found search results in an appropriate way. It must be possible to sort the tables and to export them in an appropriate file format (typically CSVs).

Data Evaluation

Data analysts need to have the possibility to quickly review and evaluate found data sets. They may do this, for example, by relying on peer ratings of these datasets (ratings given by other users). Hence, data analysts need to be enabled to rate a dataset as well. They may also want to give ratings to the origin of data. Analysts may also want to be able to triangulate regions of interest by overlaying information from multiple datasets, utilizing the geo-location associated to each dataset.

Data Acquisition

Metadata must contain a URL for accessing the complete datasets. In case of unstructured data, the URL may point to resources such as a PDF, TIF, JPEG or other type of file. In case of structured data, links for the supported services (WFS, WMS, WCS) must be provided and it must be possible to download the complete dataset from the originating server. The NGDS portal must provide a user interface that allows for downloading datasets in the appropriate form.

Data Analysis

Data Analysis is the strength of existing commercial and open source tools. The NGDS user interface may provide some very basic features for analyzing data. However, it is not the goal to develop a full fledged data analysis tool. Therefore, data analysis will be limited to some very basic features such as display data in tables, and layering of different types of data as a way to improve data discovery. All other data analysis will be performed by third party tools that will download the dataset for their particular use.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

This Software Requirements Specification (SRS) collects, organizes and describes requirements for the NGDS software system captured through use-case models, and from natural language requirements statements from a sample of system users. These include functional requirements, non-functional requirements (NFR), design constraints, and other factors needed to provide a comprehensive picture of the software's operation.

Expected readers of this document are:

- *Project management*
- *Project partners*
- *Software architects and designers*
- *User interface designers*
- *Software developers*
- *Software testers*

This document was prepared by Siemens Corporate Research (SCR) to document requirements for the National Geothermal Data System (NGDS).

1.1 VISION & GOAL

The ultimate goal of the National Geothermal Data System (NGDS) is to support the discovery of geothermal sources of energy. The NGDS will provide online access to important geothermal-related data from a network of data providers in order to:

I. *Increase the efficiency of exploration, development and usage of geothermal energy by providing a basis for financial risk analysis of potential sites*
II. *Assist state and federal agencies in making land and resource management assessments*
III. *Foster the discovery of new geothermal resources by supporting ongoing and future geothermal-related research*
IV. *Increase public awareness of geothermal energy*

1.2 LANDSCAPE OF GEOTHERMAL DATA TOOLS

The NGDS will enable the discovery of geothermal data covering a wide range of topics, from well logs and drilling data to temperature, geochemical, and geophysical measurements. Standardized data access to important datasets will facilitate utilization of these information resources.

A key component of the system is the catalog service through which data providers will register the availability of geothermal resources, and through which users will discover, evaluate and access these resources. A resource will be considered part of the system when it is published. i.e. its metadata is available in the NGDS catalog, allowing it to be discovered. The metadata record shall describe, among other things, the geo-location of the resource and how the resource can be accessed. Data providers will maintain nodes in the network, connected through the use of predetermined web-service protocols and standardized data and metadata representations for exchanging information. These standards will be developed in conjunction with the US Geoscience Information Network (USGIN), thereby providing interoperability with a wider range of geo-scientific information.

Figure 2 illustrates the overarching NGDS vision, showing its role in the landscape of Geothermal Tools and Data Repositories: NGDS will facilitate the integration of geothermal applications (front-ends) with a diverse set of data repositories ("back ends"). A core feature of NGDS is a **catalog service** that allows users to locate data repositories based on their metadata information (the catalog will include geographic location information thus allowing the representation of data on a map). Once the data is located, using the catalog services, user applications can directly access and import that data from a registered NGDS or third party repository, through the use of standard protocols (such as WCS and WFS).

Note that NGDS is not only going to provide the NGDS catalog service. It is in addition going to provide the **NGDS repository** that will provide simple-to-use means for data providers for uploading, validating and publishing their data to NGDS.

Note also that NGDS is going to provide a user application called **NGDS WebApp**. This Web-based frontend shall serve as a user interface for end users who are consumers of geothermal data. The WebApp shall allow for efficient search mechanisms, powerful means to represent discovered data, validate and access it. The WebApp shall also include simple mechanisms for analyzing the data, however its focus Is placed on searching and representing on a map-centric user interface. Specifically the WebApp shall be able to overlay various datasets on a map thus allowing for locating regions that are of potential interest for the end user (e.g. regions that have a high potential to be used for geothermal power exploitation.

For a deeper statistical analysis of individual datasets the user shall be enabled to download the data from the originating data repository.

The combination of the WebApp, with the NGDS repository, and a local node of NGDS catalog represent a node-in-a-box distribution, an application that, once installed, will allow users to participate in the NGDS network. Hence, there may be many instances of node-in-a-box components. i.e. there may be many NGDS repositories, many NGDS WebApps and many NGDS catalogs (that are federated into a distributed catalog services layer).

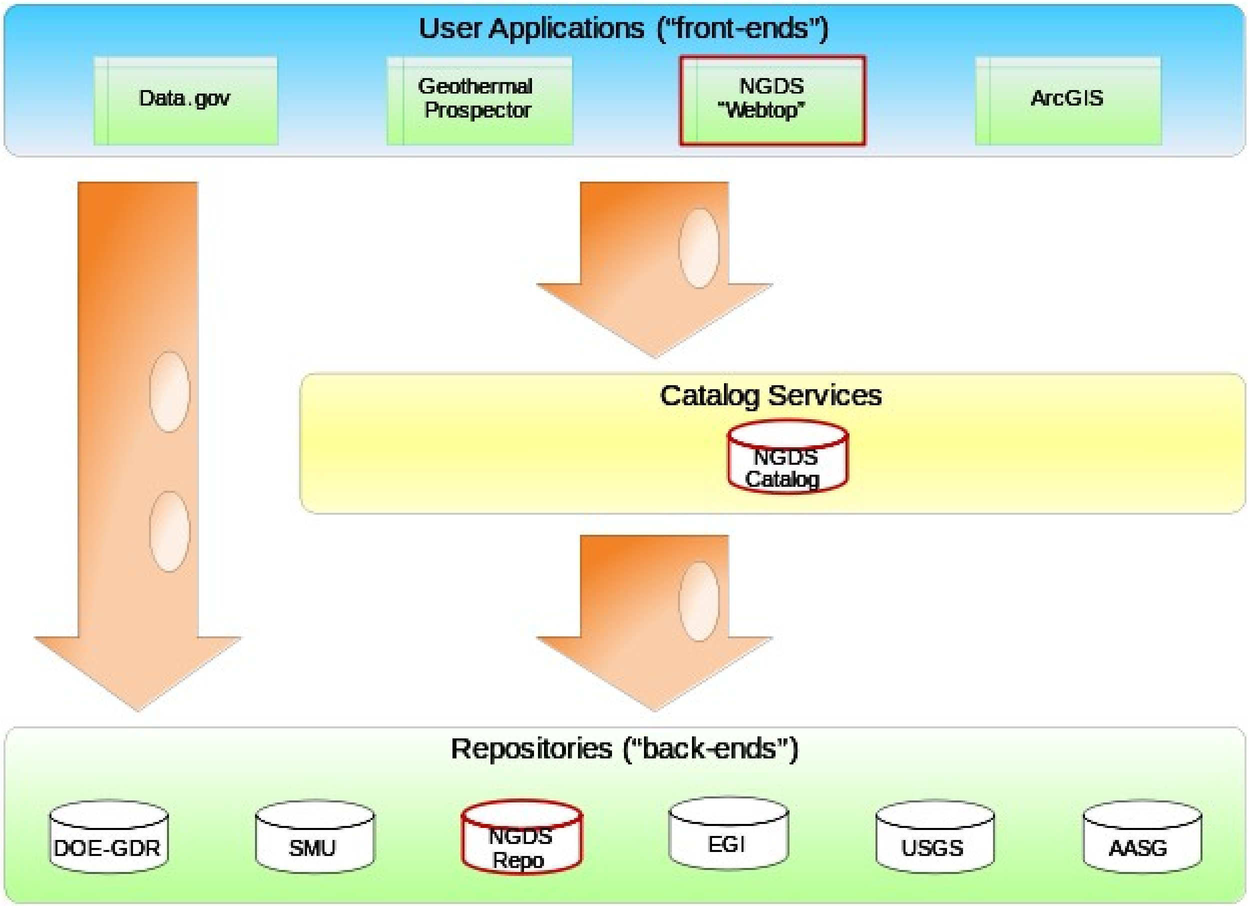


Figure 2 Geothermal Landscape of Tools and Data Repositories of NGDS.
Note that red-framed elements represent sub-components of NGDS node-in-a-box.

1.3 End User Survey Results

Anthro-tech (http://www.anthro-tech.com/) has executed a survey in order to collect information about end expectations to NGDS (**see /P05 document in References section**). Among other information Anthro-tech collected data about the types of data end users are interested in comparing to geothermal data stored in NGDS. This survey made it clear that users are not only interested in strictly geo-thermal data. They also need access to data that is indirectly related to the geothermal research field, such as location and type of power lines, roads, or locations of national preserves. Table VII lists the results some of these terms from the survey:

Table VII types of data users are interested in comparing to geothermal data

Active faults BLM boundaries	Powerlines Production data

Brown fields	Proppants
Case studies	Proximity to transmission lines
Commercial volume	Radioactivity
County borders	Remote sensing data of soil chemistry
Drill hole temperature	Roads
Electrical resistivity	Rock density
Enhanced geothermal systems	Rock type
Financial investments in geothermal exploration	Sage grouse habitat
Fluid composition of spring	Satellite images
Fluid composition of well water	Scaling
Fluid inclusion data	Seismic activity
Fractures	Seismogenic folds
Fumeroles	Soil gas chemistry
Gamma ray	Soil samples
Gas sample	State borders
Geothermal permit data	Surface expressions
Geothermometers	Surface geological map
Gravity	Surface manifestations
Heat capacity of rock	Temperatue of surface manifestation
Heat flow	Temperature gradient map
Hot spring locations	Thin-section analyses
Hydrothermal alteration	Tracers
Hydrothermal eruptions	Transmission lines
Leases	Vegetation
Magnetic survey	Volcanic activity

Micro earthquakes/ seismicity	Volcanoes
Mineral exploration wells	Water availability
Mineral rights	Water rights
MT surveys	Water wells
Mud pots	Well casing
National parks	Well logs
Permits issued	Who owns the land
Permits requested	Wilderness area
Power plants	Zoning information

The survey made it also clear that "Overlaying" is a crucial feature for the NGDS user interface: It must be possible to overlay datasets on top of a geographic map thus allowing the end user to "triangulate" regions that are of specific interest. In the following we list data that the NGDS WebApp must be able to overlay during map-based searches:

- **Geographical –** coordinates for the area.
- **Geological** – rock type, mineral, seismic activity, surface expressions, heat capacity of rock, rock density
- **Geochemical** – soil samples,
- **Geophysical** – gravity, mineral mapping, electrical resistivity, geophysical surveys.
- **Well / Drilling data** – wells, well log temperature, lithology, resistivity, radio activity, gamma ray, who drilled, when, how deep, results, current status, logs, samples, related documents, publications
- **Land status** – who owns the land, zoning, mineral and water rights, permitting, near transmission lines proximity, water availability

2 User Communities and Roles

A variety of user groups can potentially interact and use the NGDS system. In addition to the three main target user groups (defined in the System Vision document, see /P01/), we also include Administrators who are responsible for maintenance of the searchable catalog and entry-point web-application, as well coordinating management of system-wide standards and protocols. Various user groups or roles and their relationships are illustrated in Figure 3.

In Figure 3 we also outline the three main target user communities – Data Provider, Software Developer, and End User. These communities are discussed in general terms, with more detailed descriptions of their respective use cases outlined in the next section.

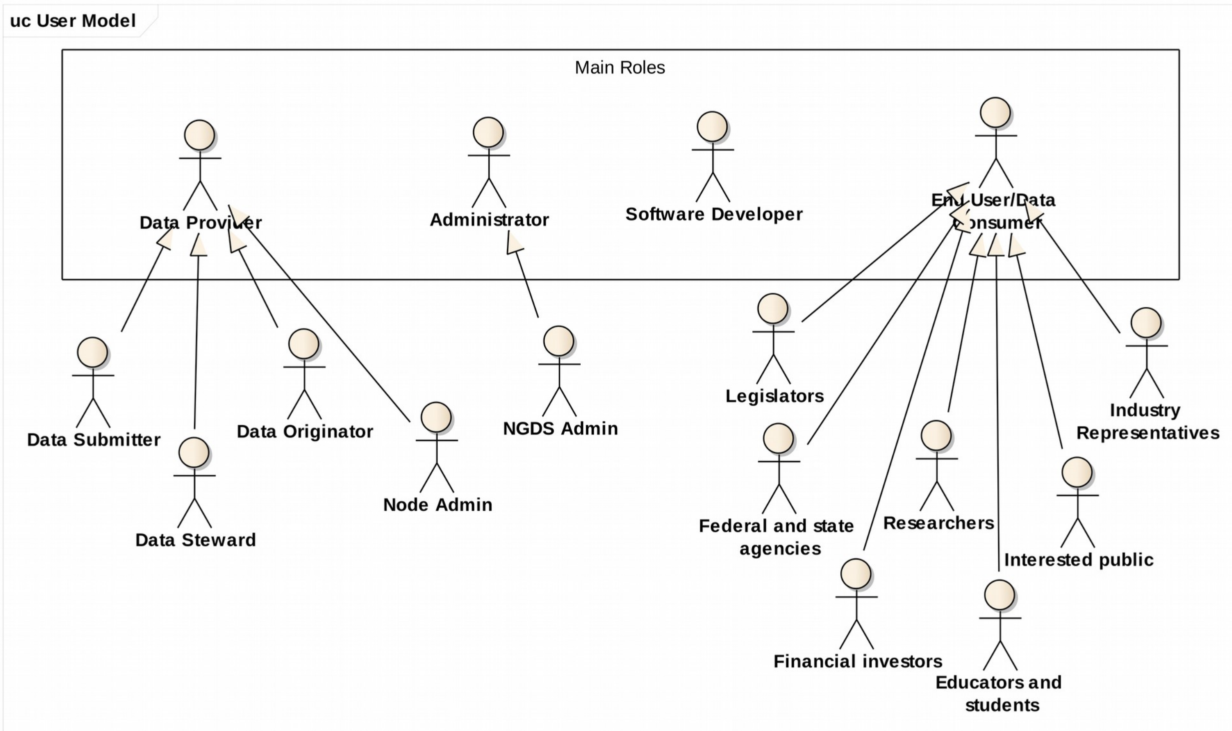


Figure 3 Main user roles and their relationships

The National Geothermal Data System will be a network consisting of four linked communities:

- **Data providers** who will expose information to the system through standardized, internet-accessible interfaces and interchange formats
- **End-users/Data consumers** who will utilize the software and information provided by the system in order to understand and develop geothermal resources.
- **Administrators** who are responsible for administrating and monitoring the system. Typical tasks are installation and configuration, user management, node monitoring, or system backup.
- **Software developers** who will build applications that utilize the data in the system, and make it easier for end-users to interact with the system.

2.1 Data Provider Community

Data providers represent the collection of users who will work together to publicize information to NGDS through standardized, internet-accessible interfaces using one of the supported interchange formats. The publication consists on making data available through a NGDS compatible repository, possibly the NGDS repository from the node-in-the-box distribution, and the publication of the metadata in the NGDS catalog, either via a WebApp component, or programmatically, using NGDS protocols.

The data provider community will generally consist of groups of individuals, often representing a single organization, who work together to maintain a repository of geothermal data that is accessible to NGDS users. This community of users can be further broken down into three distinct roles.

- **Data Submitter**: the user who uses the NGDS protocols and services to publish a piece of data
- **Data Steward**: the user who maintains the quality of a piece of published data
- **Data Originator:** the person that created a piece of data, e.g. a publication. The data originator is not an active actor in any of the use cases involved in the NGDS: as soon as the originator wishes to contribute to NGDS, they take on the role of Data Submitter
- **Node-in-a-Box (or simply Node) Administrator**: the user who is responsible for operating and maintaining an organization's data repository and insures that the data is provided according to NGDS standards and protocols

In fact, data providers play a fundamental role in the success of the NGDS because only when a critical mass of information has been published into the system will it become a useful tool for end-users / data consumers.

2.1.1 Data Submitter Role

Data submitters will publish geothermal related data to NGDS through one of several methods described below.

- Providing metadata to individual resource hosted in an existing repository, e.g. a URL to a file in a website, and registering it to the NGDS catalog service (through one of the federated NGDS catalog components).
- Performing a bulk registration of a collection of resources that are hosted in a Web repository. This is performed by uploading metadata in a template table in which each record describes a single resource (e.g. a URL to a file or spreadsheet) to the NGDS catalog service (again, using one of the federated nodes of the catalog).
- Registering a harvesting endpoint with the NGDS catalog that supports system protocols (CSW, WFS) for sharing metadata according to **NGDS standards**
- Uploading and processing data files in **supported formats** to a NGDS repository component, in order to deploy data services according to system protocols and interchange formats supported by NGDS, as well as constructing standard metadata for the new data and publishing it to the NGDS catalog service.

The data submitter user role will also be able to:

- Validate input files to verify that they conform to supported formats for automated service deployment
- View and browse services deployed from data files
- Validate metadata records

- Verify that access information in metadata records successfully locate and retrieve resources
- View logs created during data publication, processing and maintenance
- Replace existing metadata records or datasets with newer versions.

2.1.2 DATA STEWARD ROLE

The data steward user role will verify and maintain the quality of published data. The data steward will have write access to data under his or her responsibility.

The data steward will perform quality assurance tasks, such as:

- Viewing quality assurance reports
- Flagging a piece of data according to its quality, e.g. giving it a 1-5 star rating
- Performing error correction and data updates

2.1.3 NODE ADMINISTRATOR ROLE

The node administrator is responsible for the administration of one of the data repositories, or nodes, in the system. This user will perform maintenance tasks including:

- Add and delete users
- Assign user roles (data submitter, data steward)
- Administer logs
- Monitor system usage
- Backing up resources

If the node is implemented in a proprietary fashion, i.e. the node-in-a-box components are not utilized to host the data, this user is also responsible for insuring that the data in their repository or node is available to the NGDS in supported data formats and through appropriate system-wide protocols. If, otherwise, the NGDS node-in-a-box application is utilized, this will not be a concern since the repository, catalog and web app will comply with NGDS standards.

2.2 END USER COMMUNITY

End user/data consumer will utilize the software and information provided by the system in order to understand and develop geothermal resources. We use End-user and Data Consumer interchangeably in this document.

Specifically, End Users/Data Consumer will:

- Search for the entire NGDS system utilizing the federated NGDS catalog service either via the NGDS WebApp application, or potentially any other catalog search application that operates with NGDS search protocols and metadata interchange formats.
- Visualize and explore data in map, text, or other graphic presentations (as prioritized by user research findings).
- Select and acquire data via service interfaces using NGDS interchange formats
- Access files in NGDS data repositories
- Save and re-load search queries
- Set filters to be notified as new data of interest is available via the system

In general End Users / Data Consumers can access the NGDS data without authentication (i.e. anonymously). Any metadata in the catalog is freely available. Authenticated End Users / Data Consumers, however, gain the right to provide ratings for the datasets they are accessing, saving and re-loading searches, and posting subscriptions to new data. They, however, do not have the right to modify the data itself. Only Data Stewards and Data Providers can modify data they are directly responsible for.

2.3 System Administrators

The system administrator is responsible for the administration of system components. Node-in-a-box administrators are responsible for their own node (repository, WebApp and catalog), while the NGDS administrator is responsible for the whole NGDS set of nodes, and the federated catalog service. Administrator can perform different tasks including monitoring and registering nodes in the system as well as maintenance of the NGDS catalog and entry-point web-application. This role includes tasks such as:

- Supervise the addition and deletion of nodes to NGDS system
- Monitor usage and system performance
- Respond to inquiries from end-users, active and potential data-providers
- Manage user accounts related to the entry-point web-application
- Security audits, system backup, setting policies, etc.

The system administrator is also responsible for coordinating review and adoption of system protocols and interchange formats and for the registration of new interchange formats and associated schema.

2.4 Software Developer Community

Software developers may build applications that access NGDS resources using the protocols and standards outlined as part of the system architecture. These applications may serve a variety of purposes, from assisting in search and discovery of data available in the system, to visualization of datasets, through more complex analysis of the data. By utilizing existing, community-driven web-service protocols, the NGDS sets itself up to function using a number of already existing pieces of software, such as Data.gov, Geothermal Prospector, ArcGIS, OpenLayers and many others.

3 Use Case Models

This section provides an overview of use case models representing interaction with the system by various user groups as defined above. These use case models are used to identify system requirements related to user interaction. Each of the use case models is represented by a use case table, which in turn delineates concrete functional requirements of the system. Use Case tables are identified by blue headers and UC_XXX identifiers, while functional requirements are listed as bullet points in the functions row of the table. Whenever applicable, we also present derived non-functional requirements (or quality attributes) that are labeled with SRSXXX identifiers.

The use cases will later be linked to system components, explicitly verifying that the system design satisfies all requirements. See the **Software Design Description (SDD)** /P04/, for system design details and the linkage between components and requirements. The following sections describe use-case views of the System with a brief description of each use case.

3.1 High-level Use Cases View

Figure 4 shows a view of the NGDS system from the users' perspective, illustrating different types of users that may interact with the NGDS system by performing different tasks. 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.

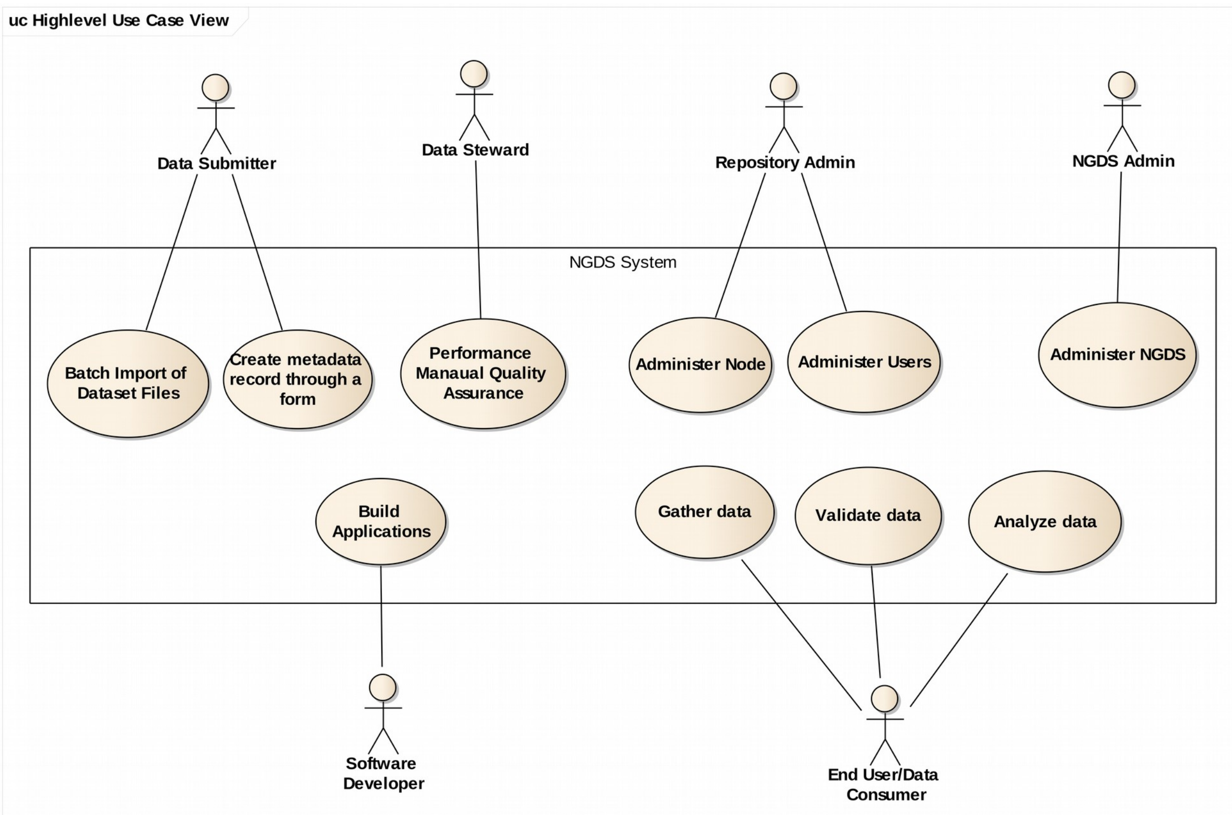


Figure 4: High-level Use Case View

3.2 Data Provider Use Cases

Data provider community's use cases are inherently correlated to the data repository in use by that subsection of the community. As such, these use cases and requirements are scoped to individual nodes or data repositories that make up the NGDS. The focus of these use cases is to create and publish a "resource", which is defined here as a single, coherent dataset which may exist in a number of different representations (e.g. uploaded .xls or .csv files, as well as scanned images, web-services or other content already online and hosted elsewhere). Each resource has a corresponding metadata record that describes the work as a whole, and clearly indicates the access mechanisms available to view that work in any of its available representations.

Use Cases here are primarily broken down into categories according the user role responsible for performing them. However, data submitters, stewards and administrators will have to first log into their node to perform data publishing and administration tasks.

1. **Tier1: Unstructured data** - Manually creating database records through a forms interface, uploading files to a local file repository if they are not already online.
2. **Tier2: structured, but not standardized data** - Bulk loading metadata to the catalog from a metadata input CSV text file for resources that are already accessible online.
3. **Tier3: Structured standardized** data supported natively by NGDS.

In cases 1 and 3, if NGDS node-in-a-box repository is used to host the data resource, a metadata record will need to be created using a form interface in the NGDS WebApp UI. In Case 3, the NGDS WebApp UI quality control application will parse uploaded files, according to NGDS supported format, to determine if they conform to a known Content Model Template. After that, the data will be imported into the NGDS Data repository, having its metadata publicized to the NGDS catalog.

Alternatively, data providers can utilize command-line scripts and programs to automatically import data into the system. In those situations, no UI is provided, but users will still have to publicize the metadata to the NGDS catalog service utilizing NGDS protocols.

In Cases 1 and 3, uploaded files will be stored in a file repository (either the ones provided by NGDS nodes-in-boxes, or in a third-party repository supporting NGDS protocols, as previously discussed). The files are public, however, they can only be modified by their data stewards, submitters and administrators. Creation of metadata will be assisted by node-in-a-box components that parse and extract category and attribute information from uploaded files when possible (based on file type and content). Third-party repositories that utilize alternative software are also required to provide standard NGDS metadata including such category and attribute information.

For Tier 3 access, data will be mapped into prioritized content models determined by AZGS (Arizona State University) and encoded in the adopted interchange formats for publication via NGDS services. Tier 2 data is incompatible with NGDS data model and cannot be automatically indexed by NGDS catalog, however, they can be handled as Tier1 data. Tier 1 data (an image or PDF file for example) will be referenced by a URI pointing to its data location, will have a geo-location tag, and will be indexed based on metadata (manually) provided by the publisher. **See content model documentation /P02/**, for details of these categories. The standard content models will be used to structure a browse tree interface and data browsing capabilities with filtering and display functions. These will be posted in the NGDS Website.

Once the catalog entries have been made, it will be possible to modify them manually, by adding data items as necessary. This will enable the update and addition of metadata to the catalog entries, and will support third party data repositories to publish metadata about their site.

Additionally, end users can annotate metadata records with comments, and can provide star ratings (from 1 to 5). These ratings and comments will potentially help other users in the classification discovery, filtering and gathering of data.

Only Data Submitters, Stewards and System Administrators will have Read/Write access to the metadata posted in the catalog. All Data Consumers (Geothermal Analysts/End User) will have Read only access to the data. The NGDS repository will keep a record of all import operations performed. This will be provided as an import log when requested by Administrators, Submitters and Stewards.

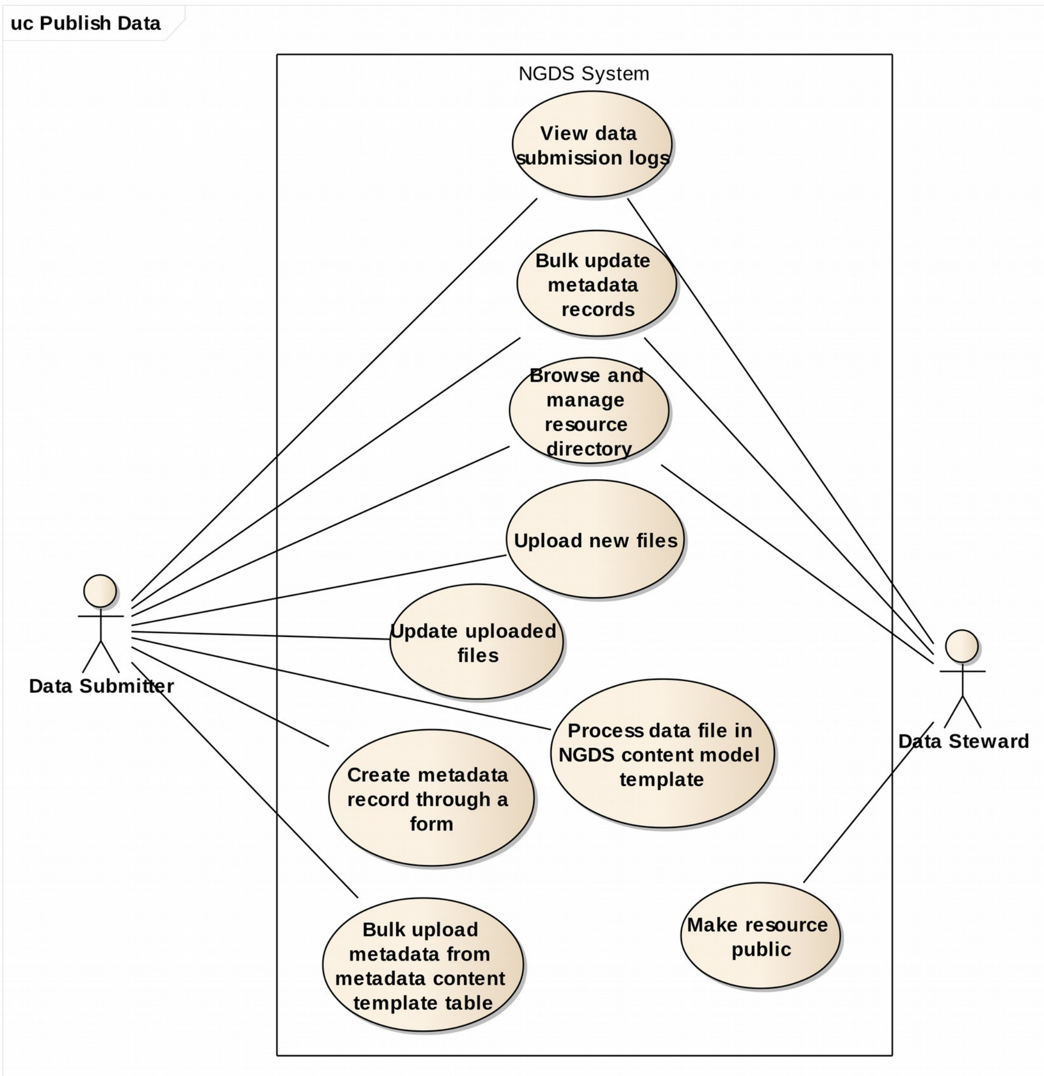


Figure 5 : Data Entry and Submission Use Case View

Figure 5 above illustrates the functionality of the data provider node application that facilitates the submission of data to the NGDS. Once a piece of data is imported into the repository by the Data Submitter, it is handed over to the Data Steward who reviews it. If the Data Steward is satisfied with the content and quality of a dataset and its metadata, he or she can publish the data, making it available throughout the NGDS network by publishing its metadata to the NGDS catalog.

If the published data conforms to NGDS content model, the NGDS node-in-a-box provides functionality to upload a template file and deploy NGDS services based on the file.

Note that, from the point of view of End Users / Data Consumers, all data and metadata publicized through NGDS catalog & NGDS data repository will be freely available, with read only access. Data submitters, Administrators and Data Stewards will have write access to the data and metadata they are responsible for.

Here follows a task-based description of each of the use-cases.

As one of NGDS's main purpose is to support the location of publicly available data, End Users/Data Consumers may freely access public catalog information and data. In these situations, they will be automatically assigned 'guest' access permissions. Data Submitters, data stewards and Administrators, however, will have to first log into the system to perform data publishing and administration tasks. It is assumed that Login and Logout use cases are included in each of the functional areas that follow.

In general, the use cases for NGDS system are organized by user category as follows.

- Data Provider Use Cases
- End-user/Data Consumer Use cases
- System administration Use Cases
- Software Developer Use Cases

When certain tasks or workflows require multiple users to collaborate, they will be represented as workflow use cases, such as Quality Assurance Workflow.

3.2.1 Data Submitter

Data submitters publish geothermal data at their local node repository. Their primary objectives are to make files available online, set up data-services where appropriate, and to generate metadata describing a resource.

3.2.1.1 File Uploads

The first step in making a data resource available in the NGDS network is to store it in a web repository. This process is captured by the file upload use case. The data submitter should also be able to upload multiple files in one operation, if those multiple files pertain to a single resource. Generally such files are not expected to exceed 2GB in size.

Use Case ID	**UC_001a**
Use Case Name	**Upload new files**
Short Description	The goal of this use case is to allow data submitters to upload one or more files to be stored in the NGDS data repository. After the upload, the submitter will also update the metadata

		record of that file, thus allowing it to be cataloged. We assume the file is opaque, i.e. stored as is, with no further content parsing.
Actors		Data submitter
Pre-Conditions		Data submitter is properly authenticated;
Success End Conditions		The files are successfully uploaded and stored in the NGDS repository The metadata record for the provided file is successfully created The metadata remains "private", waiting to be made public by a data steward
Data		Metadata attributes for the specific data type as input to the form Files to be uploaded Geographical location of files
Functions		• Upload files • Form-based metadata input for specific data type • Auto-complete of user contact information • metadata validation • URI creation • Metadata duplicate detection • Tagging of metadata with geographical information • Converting non-standard location coordinates into latitude/longitude and shapes. • Log changes to metadata log file
Main Sequence		
Step	**Actor**	**Description**
1	Data Submitter	Navigates to the files upload screen from NGDS System
2	NGDS System	Presents upload files screen to Data Submitter
3	Data Submitter	Selects the list of files to be uploaded Selects the metadata record type to be created for the files Inputs geographical location
4	NGDS System	Presents the appropriate metadata import form to the metadata type the data submitter selected. Presents a form for selecting and uploading files
5	NGDS System	Uploads selected files Associates uploaded files to the data submitter account
6	Data Submitter	Fills out the form with metadata and finalized data input
7a	NGDS System	Performs validation of metadata form based on content completeness Creates a data location URI and updates metadata form Performs duplicate detection Validates, normalizes, converts geo-location attached to the data Log changes to metadata log file Provides a file upload success message.
8	Data Submitter	Acknowledges the success/failure completion of the operation
9	NGDS System	Makes metadata "private" for discovery, waiting to be made public by a Data Steward Send data steward an e-mail notification about the new data or If the user is both a Data Submitter and a Data Steward, makes the data "public", notifying

		the user of the successful operation.
Variants		
Step	**Actor**	**Description**
7b	NGDS System	In case of duplicates, or incomplete information in the form, the system will provide a failure notification message, indicating the type of error. The user-provided metadata form will be presented to the Data Submitter for correction
7c	Data Submitter	Will correct the form data and resubmit for validation or will quit the import procedure.
Exceptions		
Step	**Actor**	**Description**
	NGDS System	In case of internal file upload/metadata record creation failure, the system will roll back all existing transactions, returning to its previous state.
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	Can't it be assumed that the submitter is working in cahoots with the data steward? Can the approval step be short circuited by assigning both submitter and steward roles to an individual? Yes, it is possible.	
2	IN step 9, perhaps the data steward should be notified that there is a new record to be approved? This could be done by the NGDS system once the data submitter confirms the upload is completed?	

Use Case ID	**UC_001b**
Use Case Name	**Update uploaded file**
Short Description	The goal of this use case is to allow data submitters to update a file that has been uploaded. This process includes the update of the metadata record of the file. We assume the file is opaque, i.e. stored as is, with no further content parsing.
Actors	Data submitter
Pre-Conditions	Data submitter is properly authenticated;
Success End Conditions	The file is successfully updated and new content stored in the NGDS repository The metadata record for the provided file is successfully updated The entry for the new file is also propagated & updated in the NGDS catalog.
Data	Possible new metadata record changes Files to be uploaded
Functions	• Upload files • Auto-complete of user contact information • metadata validation • Metadata duplicate detection • Tagging of metadata with geographical information • Converting non-standard location coordinates into latitude/longitude and shapes. • Log changes to metadata log file

Main Sequence		
Step	**Actor**	**Description**
1	Data Submitter	Searches for a file in the database includes <<gather data>> use cases Selects an option to update file record
2	NGDS System	Presents a update file metadata record screen

3	Data Submitter	Selects new file to be uploaded Updates metadata record for the file Selects update command
4a	NGDS System	Updates file and/or updates metadata record Performs validation of metadata form based on content completeness Updates a data location URI and updates metadata form Performs duplicate detection Provides a file upload success message. Validates, normalizes, converts geo-location attached to the data Log changes to metadata log file
5	Data Submitter	Acknowledges the success/failure completion of the operation
6	NGDS System	Makes metadata "private" for discovery, waiting to be made public by a Data Steward Send data steward an e-mail notification about the new data or If the user is both a Data Submitter and a Data Steward, makes the data "public", notifying the user of the successful operation.
Variants		
Step	**Actor**	**Description**
4b	NGDS System	In case of duplicates, or incomplete information in the form, the system will provide a failure notification message, indicating the type of error. The user-provided metadata form will be presented to the Data Submitter for correction
4c	Data Submitter	Will correct the form data and resubmit for validation or will quit the import procedure.
Exceptions		
Step	**Actor**	**Description**
	NGDS System	In case of internal file upload/metadata record update failure, the system will roll back all existing transactions, returning to a valid state.
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	What happens to the old file? Is it deleted in the database or DOEs it remain there? Is a history of all metadata changes kept? DN: This is a good question and one that probably has no right answer for everyone. Does the new item get a new UUID or is it a revision (version X++) of the existing item? Earlier in the document there is a statement that says "all" node interactions are logged. This might require that the old version stay there. Probably best to ask domain experts.	

3.2.1.2 Data Processing

In some cases, the file to be uploaded may conform to an NGDS content model and be provided in a valid file-format, allowing the information from the file to be processed in order to create a compliant NGDS web-service. Processing will include validation of the data's schema, loading of data from the uploaded file into the repository. The file will then be available for publication by the data steward.

Publication consists in making the metadata of the file available in the NGDS catalog and having its contents publicized via supported protocols (i.e. WMS and WFS).

<table>
<tr><td colspan="2">Use Case ID</td><td>UC_006</td></tr>
<tr><td colspan="2">Use Case Name</td><td>Process data file in NGDS content model template</td></tr>
<tr><td colspan="2">Short Description</td><td>This allows data submitter to process (upload/parse) a file to NGDS data provider nodes.
The difference between the upload/update file use cases is that the file here is formatted according to an existing template, and therefore, can be parsed and checked for correctness. Processing will include validation of data schema, loading data into a data store on the provider node.</td></tr>
<tr><td colspan="2">Actors</td><td>Data Submitter</td></tr>
<tr><td colspan="2">Pre-Conditions</td><td>Requires authentication, access permission to edit metadata records.
File must be formatted according to one of the NGDS content model templates (See /P02/ data specification from Arizona State University, for supported file formats and content models</td></tr>
<tr><td colspan="2">Success End Conditions</td><td>File is submitted to a repository, being accessible through a valid URI</td></tr>
<tr><td colspan="2">Data</td><td>Files properly formatted according to NGDS supported formats</td></tr>
<tr><td colspan="2">Functions</td><td>• import data files
• Validate data file content & formats
• Log changes to metadata log file</td></tr>
<tr><td colspan="3">Main Sequence</td></tr>
<tr><td>Step</td><td>Actor</td><td>Description</td></tr>
<tr><td>1</td><td>Data Submitter</td><td>Navigates to the page that allows upload of files to NGDS data repository</td></tr>
<tr><td>2</td><td>Data Submitter</td><td>Provides file path & name to the system
Provides a data type from the NGDS supported content models and file formats
Send file for upload</td></tr>
<tr><td>3</td><td>NGDS System</td><td>Validates file data format based on its provided type
Informs the user about possible errors in the format
Perform file content validation
Performs duplicate detection

Informs the user about errors and fails, without importing the file;
Or provides a success message

Makes metadata “private” for discovery, waiting to be made public by a Data Steward
Send data steward an e-mail notification about the new data
Or
If the user is both a Data Submitter and a Data Steward, makes the data “public”, notifying the user of the successful operation.</td></tr>
<tr><td>4</td><td>Data Submitter</td><td>Verifies the operation status. If a file has errors, the user will have to correct them without the help of the system.</td></tr>
</table>

Variants		
Step	**Actor**	**Description**
3b	Data Submitter	Is notified of suboptimal data content Chooses to submit the data anyways
3c	NGDS system	Accepts the file, flags the problems with content or metadata Hand file over to the data steward
Exceptions		
Step	**Actor**	**Description**
	NGDS System	In case of internal import failure, the system will roll back the existing transaction, returning to a valid previous state.
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

3.2.1.3 Metadata Generation

The creation of metadata follows three distinct use cases. First is to generate a metadata for a single resource through a form-based interface. The purpose of such an interface is to make it as easy as possible for the data submitter to create high-quality, NGDS standards-compliant metadata describing a single resource.

Use Case ID	**UC_003**
Use Case Name	**Create metadata record through a form**
Short Description	The goal of this use case is to allow data submitters to create a metadata record describing a resource (tier1, tier2 data) by input of information manually through a form interface for inclusion in the NGDS catalog.
Actors	Data submitter
Pre-Conditions	Data submitter is properly authenticated; Data is available through one of the NGDS data repositories, and is identified through a valid URI. Metadata includes the geological location of the geological feature associated to the data.
Success End Conditions	The meta-data for the provided geological feature is successfully imported into the NGDS catalog The data remains “private”, waiting to be made public by a data steward
Data	Metadata attributes for the specific data type as input to the form
Functions	• Form-based metadata input for specific data type • metadata validation • URI validation • Metadata duplicate detection • Tagging of metadata with geographical information • Converting non-standard location coordinates into latitude/longitude and shapes. • Log changes to metadata log file

Main Sequence		
Step	**Actor**	**Description**
1	Data Submitter	Navigates to the metadata input screen from NGDS System
2	NGDS System	Presents metadata input screen to Data Submitter
3	Data Submitter	Selects the meta-data type to be imported
4	NGDS System	Presents the appropriate metadata import form to the metadata type the data submitter selected.
5	NGDS System	Automatically fills out form with data submitter contact information
6	Data Submitter	Fills out the form with metadata and finalized data input
7a	NGDS System	Performs validation of metadata form based on content completeness Validates data location URI Performs duplicate detection Provides an import success message. Validates, normalizes, converts geo-location attached to the data Log changes to metadata log file
8	Data Submitter	Acknowledges the success/failure completion of the operation
9	NGDS System	Makes metadata "private" for discovery, waiting to be made public by a Data Steward Send data steward an e-mail notification about the new data or If the user is both a Data Submitter and a Data Steward, makes the data "public", notifying the user of the successful operation.
Variants		
Step	**Actor**	**Description**
7b	NGDS System	In case of duplicates, invalid URIs or incomplete information in the form, the system will provide a failure notification message, indicating the type of error. The original metadata form will be presented to the Data Submitter for correction
7c	Data Submitter	Will correct the form data and resubmit for validation or will quit the import procedure.
Exceptions		
Step	**Actor**	**Description**
	NGDS System	In case of internal import failure, the system will roll back all existing transactions, returning to a valid state.
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	Can you define content completeness? What's the criteria to be used here? Are there required fields and optional fields for example? ND: Believe this will be set forth in the metadata cardinality rules (example 0..1)	
2	In case of duplicates, which one is the most important the new instance of the data or the existing version of it? How to resolve these conflicts? DN: The domain experts said if records are overlapping, they would want both. If they are truly identical, it doesn't really matter does it (from a pure logic perspective anyways).	
3	What if the user DOEs not provide a geo-location to the data? DN: Given the UCD feedback, a geospatial reference is mandatory for every record for map based searching.	
4	The existence of a make public use case performed by a Data Stewart implies that after an import, the data is made	

	private, waiting for being publicized by the steward. Is it a correct assumption? DN: I believe so. This is reflected elsewhere in the document.

Where a data submitter wishes to create metadata for a larger collection of resources for which files are already available online, such metadata information can be uploaded to the node through the use of a metadata template table. In such a table, each row represents a single resource, and the columns are translated into appropriate parts of a standard NGDS metadata record.

Use Case ID		**UC_004a**
Use Case Name		**Bulk upload metadata from metadata content template table**
Short Description		The goal of this use case is to allow data submitters to import a CSV file containing metadata into the NGDS catalog. New records will be marked ‘submitted’; Data Steward will then have option to review metadata through forms interface; records will become visible to public when marked ‘published’. Metadata will be validated for content completeness, URLs checked for http 200 responses, and new metadata record will be run through a duplicate-detection process to identify existing metadata that may already describe the resource.
Actors		Data Submitter
Pre-Conditions		Data submitter is properly authenticated; Individual data items, referenced to in the CSV file are available through one of the NGDS data repositories, and is identified through a URI. Metadata includes the geological location of the feature associated to the data. CSV file containing metadata in table form according to the NGDS Compilation template Metadata Excel workbook
Success End Conditions		The meta-data for the provided geological feature is successfully imported into the NGDS catalog. The imported data remains private, waiting for the Data Steward to make it public.
Data		CSV file containing resources metadata, formatted according to the NGDS compilation template metadata excel workbook.
Functions		• Form-based metadata input for specific data type • metadata validation • URI validation • Metadata duplicate detection • Tagging of metadata with geographical information • Converting non-standard location coordinates into latitude/longitude and shapes. • Log changes to metadata change log file
Main Sequence		
Step	**Actor**	**Description**
1	Data Submitter	Navigates to the bulk metadata input screen from NGDS System
2	NGDS System	Presents bulk metadata input screen to Data Submitter
3	Data Submitter	Selects the metadata CSV file type to be bulk imported
5	NGDS System	Automatically fills out form with data submitter contact information
4	Data Submitter	Inputs CSV file name and path into the metadata input screen
6	Data Submitter	Starts import process

7	NGDS System	Validate metadata for content completeness Validates URI Detects metadata duplication Marks problematic entries for revision Marks individual entries as “submitted” Normalizes geo-location coordinates to the internal system representation (UML or lat/long) Log changes to metadata change log file
8	NGDS System	Presents import data report to user in a form
9	Data Steward	Reviews “submitted” data for errors. The forms allow users to modify/correct the individual metadata records
10	Data Steward	After reviewing the data and correcting errors, send metadata for publication
11	NGDS System	Repeats step 7 for a final check
12	NGDS System	Makes metadata “private” for discovery, waiting to be made public by a Data Steward Send data steward an e-mail notification about the new data OR If the user is both a Data Submitter and a Data Steward, makes the data “public”, notifying the user of the successful operation.
Variants		
Step	**Actor**	**Description**
11b	NGDS System	Rejects data based on invalid records
Exceptions		
Step	**Actor**	**Description**
	NGDS System	In case of internal import failure, the system will roll back the transaction, returning to a previous valid state.
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	See: “**Create metadata record through a form”** use case. The same issues are present here	
2	Only CSV files are supported? Prescriptive?	

Finally, a data submitter may wish to update metadata records for existing resources, perhaps one at a time, or in bulk. In either case, the data submitter must first be able to locate the resources that s/he wishes to see updated.

Use Case ID	**UC_008**
Use Case Name	**Browse and manage resource directory**
Short Description	Allows the User to view & manage all of the resources (or datasets) under their stewardship (data steward) or that they have submitted (Data submitter), based on metadata describing the resources. The resource listing will be presented to the User in a tree view directory structure. Users can define collections (folders, subdirectories) to organize the listing according their needs. A resource may be assigned to multiple collections. Access control may be assigned at the collection level. The display should indicate clearly any resources that have quality issue flags attached

Actors	Data submitter, Data Steward
Pre-Conditions	Requires authentication, access permission to view & edit individual user metadata records & datasets.
Success End Conditions	Users can adequately manage (create, rename, delete) resources Users can assign resources to collections (updating metadata) Users can delete existing resources (and their associated metadata) Metadata is kept in synchrony with changes in the collection
Data	Resources: metadata records or files
Functions	• Visualize flagged resources with quality issues • Create, delete, reorganize resource collections • delete resources • Visualize resources & collections

Main Sequence

Step	Actor	Description
1	Users	Navigate to the resources management page
2	NGDS System	Provides a tree view of existing resources and their collection structure Allow users to navigate through the structure Allow users to perform management operations (move, delete, add, rename)
3	Users	Perform collection management operations (add, delete, rename, move) Perform resource operations (delete, move, rename)
4	NGDS System	Responds to user operations, providing an up-to-date view of the resources available for the user. DOEs not allow the users to see resources that are not under their responsibility. Apply selected management operations on the resources

Variants

Step	Actor	Description

Exceptions

Step	Actor	Description

Open Issues (Please use this field to indicate questions/comments on the use case)

ID	Issue Description
1	It seems to me that, in this view, the only operation that users can do with resources is delete. (Importing and modification of resources must be done in other views). Is it correct?
2	Missing requirements: tree-view portrayal of resource hierarchy. More clarity required as to who can create collections and sub-collections at what levels in the hierarchy. [Data steward assigns permissions to create public collections; perhaps allow authenticate users to define personal collections that they persist in a user workspace to use between sessions]. How do permissions work if metadata can belong to more than one collection? [Metadata record has only one owner; do we need a use case for reassigning ownership of a metadata record?] If a record is part of one collection and I have edit privileges to that collection, but it is also included in another collection to which I do not have privilege, what are my options? [Looks like assigning access control at collection level won't work; work access control through users and group membership, with permissions at group level to share edit capabilities]
3	DN: I found this to be very prescriptive (are all resources always hierarchic?) Is it possible they have nodes that are traversed in other ways?

Once the data submitter has located the record or records that need to be updated, they are either adjusted using the form-based interface described above for single resources, or through a bulk updating workflow for the case of multiple resources.

Use Case ID		**UC_005**
Use Case Name		**Bulk update metadata records**
Short Description		Metadata records may be selected based on a content-based filter query based on fields as contact information or linkage URIs. Once selected, those records can be replaced using regular expressions. User selects metadata content item to update, value to replace, and new value to use.
Actors		Data submitter, Data steward, System administrator
Pre-Conditions		Requires authentication, access permission to edit metadata records.
Success End Conditions		All relevant metadata record attributes that match the search criteria are replaced with the new value defined by the user
Data		Specific fields of metadata records: contact information and URIs
Functions		• Search & replace of metadata record attributes based on simple regular expressions • Log changes to metadata log file
		Main Sequence
Step	**Actor**	**Description**
1	Users	Navigates to the search & replace metadata record feature page
2	Users	Input search criteria using simple regular expression Input replace data Select metadata record attributes to be replaced
3	NGDS System	Performs search & replace routine using user input parameters on the specified metadata record attributes Presents a report of affected records and their suggested changes Asks users for confirmation Logs changes in metadata change log file
4	Users	Accept or reject bulk change based on report produced by the system
5	NGDS System	Commit metadata record changes; or roll back if users reject changes
6		
		Variants
Step	**Actor**	**Description**
		Exceptions
Step	**Actor**	**Description**
	NGDS System	In case of internal import failure, the system will roll back all existing transactions, returning to a valid state.

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	How interactive should be the process of search & replace?

Once the data submitter has completed the process of creating a resource, that resource will be marked as submitted. As a submitted resource, access to the metadata, data and files associated with the resources is restricted to users with permissions appropriate to edit, and the metadata is not included in any public harvesting endpoints (i.e. CSW interface).

3.2.2 Data Steward

It is the job of the data steward to verify and maintain the quality of published data. Users in this role will perform quality assurance tasks, as well as maintain appropriate permissions over a subset of resources available on a node. These permissions are likely to be applied to groups of users on distinct collections of resources.

3.2.2.1 Activity Logs

In order to assist the data steward in the analysis of resource quality, activity logs will be generated during any data or metadata creation, update or QA procedure. These logs will include information such as:

- time of activity
- actions taken
- user responsible
- user comments

These activity logs will be accessible to data submitters and data stewards with appropriate permissions for a particular resource.

Use Case ID	UC_007 & UC_046
Use Case Name	**View resource submission and update logs**
Short Description	Allows authorized users to view the logs for metadata record creation, file uploads to file repository, and processing of NGDS files that conform to valid content model templates. These logs are created during any data submission or update activity.
Actors	Data submitter, Data Steward, System administrator
Pre-Conditions	Requires authentication, access permission view metadata records import logs. These logs are created during any data submission or update activity, so at least one data submission must have occurred.
Success End Conditions	Successful and correct visualization of data submission logs. No missing logs
Data	Data submission logs containing details such as time of activity, actions taken, data submitter, Data Steward, size of data, submitter comments, etc

Functions		• Submission log capture • Submission log visualization	
Main Sequence			
Step	**Actor**	**Description**	
1	Users	Navigate to the data submission log page Select among existing logs by date or by name of data upload operation or metadata creation operation	
2	NGDS System	Searches for the log given user search criteria Succeeds presenting the log Or fails, providing a error message if no data record is found	
Variants			
Step	**Actor**	**Description**	
Exceptions			
Step	**Actor**	**Description**	
Open Issues (Please use this field to indicate questions/comments on the use case)			
ID	**Issue Description**		
1	Missing requirements detailing the logs and their creation. What exactly is logged? Also, how is the log portrayed to the user, or is that too much in the realm of design?		

3.2.2.2 Validation and Quality Assurance Logs

When data or metadata are processed in NGDS content model templates, these logs will include a report on validation of the processed information and adherence to the content model (syntax, completeness, cardinality, data types, URL checking and de-duplication). A validation report will be created which will list any recognized quality issues.

The data steward will review these quality assurance reports, and perform appropriate actions based on their assessment of the report.

Figure 6 illustrates the functionality supported by the data provider node application to facilitate quality assurance workflow. Functionality includes automated quality assurance and reporting functionality where that is practical. When data or metadata are processed in NGDS content model templates, the application will validate adherence to the content model (syntax, cardinality, data types, other tests that may be automated), and will generate a report listing recognized quality issues. Part of the validation process will be a function to compare the new metadata to existing metadata in the local catalog to identify existing records that may describe the same resource. Data Submitter will need to manually inspect the flagged records in the submitted data or metadata to correct issues. They can modify the record, or indicate that the record is correct and add a note explaining their reasoning.

In addition, authorized End Users can flag a data or metadata record for quality issues if they consider that the content is not accurate. The System will handle these flags (annotations) by notifying the appropriate Data Submitter to take action to correct the issue. The data submitter than sends it back to the steward for approval and publication.

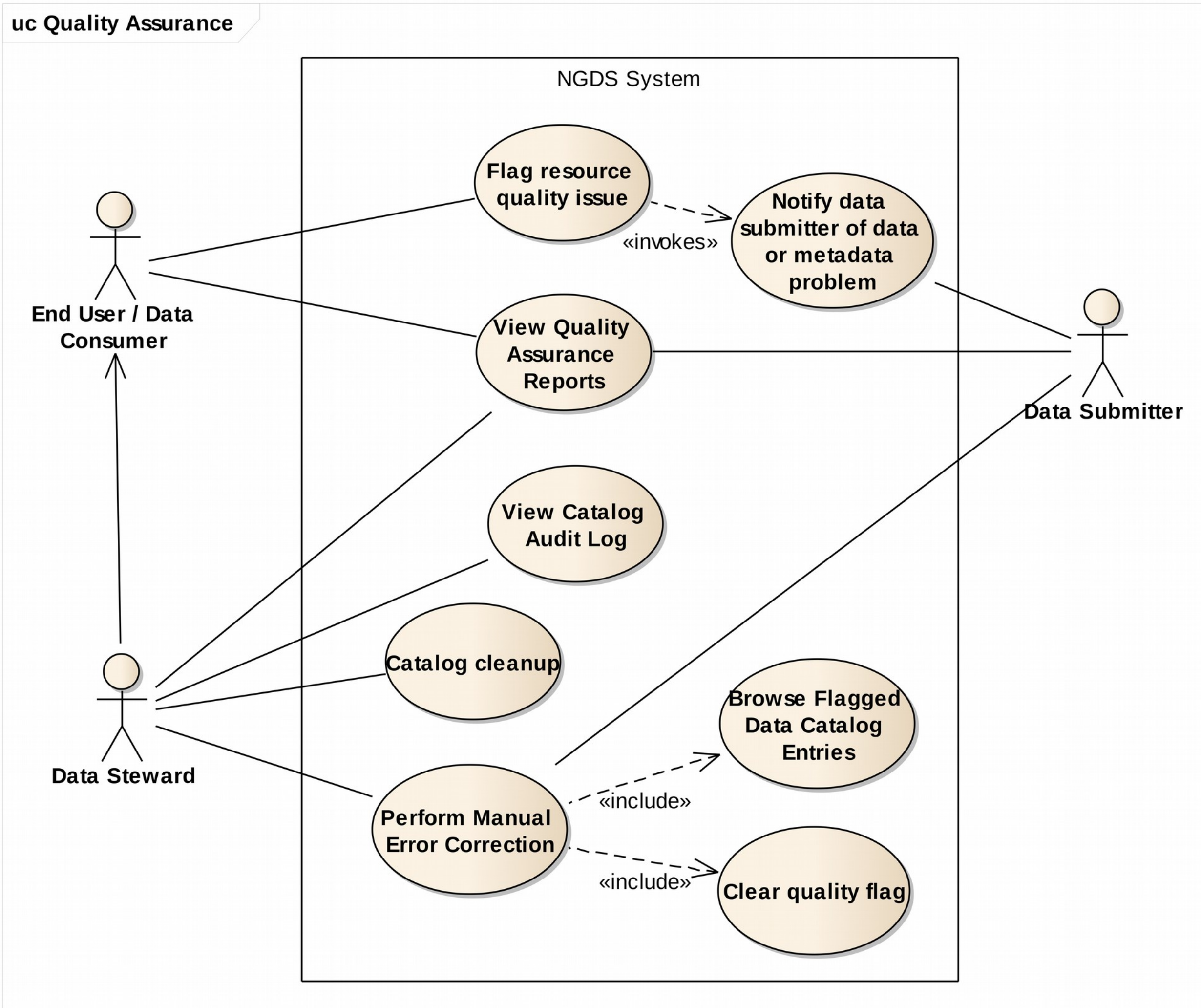


Figure 6 Quality Assurance Use Case View

Here follows a brief description of each of the use-cases.

Use Case ID	UC_046
Use Case Name	**View catalog audit log**
Short Description	At every change the system creates a log for a given catalog item. This use case allows authorized users to view the log of all changes made to a data or metadata record, changes made both manually and automatically.

Actors		Administrator, Data Steward
Pre-Conditions		An audit log of catalog changes has been created and actively updated by the system
Success End Conditions		Users are able to visualize the change logs for a given record
Data		Catalog audit log
Functions		• View catalog audit log
Main Sequence		
Step	**Actor**	**Description**
1	Users	Navigate to quality assurance page Select a metadata record Visualize the log for that record
2	NGDS System	Responds by loading and displaying the data record log.
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	Missing requirements detailing the validation and generation of the validation log. Here are two, but where are ones about URL checking? Syntax? Is the validation log attached to the activity log for a particular resource?	
2	DN: I am also confused if this is happening at the node or NGDS system level? If another node change must be recorded in every other node, this could get quite messy fast (math suggests that it would be equal to ((N * (N^2 – 1) * R) * MR) where N is the number of nodes, R is the number of records and MR is the number of metadata records and that is assuming a 1:1 relationship between instances or R and MR.	

Use Case ID	**UC_042**
Use Case Name	**View quality assurance reports**
Short Description	Allows the Data submitter or Data Steward to view quality assurance reports for resources they submit or maintain. Quality assurance can be automatic or manual. As an automated process, it is performed by NGDs during import, flagging possible typos and simple input errors in the metadata being managed by the system As a manual process, it involves users that see possible issues and flag them in the data records.
Actors	Data Submitter, Data Steward, End User/Data Consumer
Pre-Conditions	The metadata record has been imported into NGDS catalog and a set of quality assurance functions were automatically ran in those records. As a result, the metadata records were flagged for possible errors. Users have write access to the data i.e. they are either submitters or stewards of that data
Success End Conditions	Users are able to identify the flagged errors in the metadata they provide or maintain
Data	Metadata records, Quality Report
Functions	• Visualize quality assurance report

Main Sequence		
Step	**Actor**	**Description**
1	Users	Navigate to quality assurance page Select view quality assurance reports
2	NGDS System	Responds by displaying a list of metadata records that were flagged as having possible quality assurance problems
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	Should consumers be able to see QA reports? If they are not addressed, it seems like it would serve as a good "warning" to analysts. The Quality Report is for steward and submitter consumption, with lots of details that may not be relevant to End Users, e.g. wrong measures, duplicated fields. We assume the steward will fix them before making the data public. However, some stewards may want to make that data public anyway, in that case, consumers could benefit from this information.

Use Case ID	**UC_043**
Use Case Name	**Flag resource quality issue**
Short Description	Allows authorized & authenticated End-User/Data Consumer to create a flag, indicating that some issue exists with the quality of a data or metadata record. The process creates an annotation record documenting the dataset, URI for the resource in question, identity of the user raising the flag, timestamp. The annotation record should also include notes on resolution process, who, when, what. Flagged resources should be clearly marked
Actors	Data Steward. End User/Data Consumer
Pre-Conditions	The metadata record has been imported into NGDS catalog. Metadata QA records can be modified by any user in the system
Success End Conditions	Users are able visualize metadata records, and to provide quality assurance feedback to these records as they see fit.
Data	Metadata records (read only) Metadata QA (quality attribute) records (read/write) User information
Functions	• Manually flag resource quality issues

Main Sequence		
Step	**Actor**	**Description**
1	End User/Data Consumer	Navigate to quality assurance page Visualize metadata records Edit record quality attributes The system will automatically fill in user information based on her credentials
2	NGDS System	Responds by displaying a list of metadata records that were flagged as having possible quality assurance problems Include use case <<Notify data submitter of data or metadata problem>>
Variants		
Step	**Actor**	**Description**
1b	End User/Data Consumer	End user can flag data records directly on the metadata visualization page, without going to the quality assurance page.
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID	**UC_046c**
Use Case Name	**Browse flagged data catalog entries**
Short Description	Allows Users to navigate through the catalog of resources that have had their quality flagged for correction. A special view of the resource browse list, filtered for records that have quality flags raised from those that do not have them
Actors	Data Steward, Data Submitter
Pre-Conditions	a list of metadata resources is available, allowing filtering by quality attribute flag
Success End Conditions	Users are able to identify flagged entries in the metadata catalog
Data	Metadata record
Functions	• Browse data catalog entries • Filter catalog entries by flagged attribute

Main Sequence		
Step	**Actor**	**Description**
1	Users	Navigate to quality assurance page Visualize metadata record Filter out un-flagged metadata records
2	NGDS System	Responds by displaying filtered metadata records

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Finally, the data steward is capable of resolving quality issues with resources which they are authorized to edit. Updates to data and metadata are performed similarly to as described for data submitters above. As described above, activity logs are generated for all of these actions, including indication of the user who made the corrections and user-provided notes indicating what was adjusted.

Use Case ID	**UC_044**
Use Case Name	**Perform manual error correction**
Short Description	Allows a Data submitter or Data Steward to manually address issues present in a data or metadata record that has been flagged for having quality issues. Note that this process may require resubmission of data files. Quality check and error editing will be built into the data submission process, but will require the steward to access the data in the appropriate environment for quality issues recognized after a submission is complete. The process will require the reviewer to make a note that is recorded as annotation on the record, along with the identity of the reviewer and a time stamp.
Actors	Data Submitter, Data Steward, Administrator
Pre-Conditions	The metadata record has been imported into NGDS catalog.
Success End Conditions	Users are able visualize metadata records, and to provide quality assurance feedback to these records as they see fit.
Data	Metadata records (read/write) Metadata quality attribute records (read/write) User information
Functions	• Edit metadata record • Edit data • Log metadata record change

Main Sequence		
Step	**Actor**	**Description**
1	Users	Include use case <<browse flagged data catalog entries>> Edit record quality attributes Edit metadata record Edit data Include use case <<clear quality flag>> The system will automatically fill in user information based on his/her credentials
2	NGDS System	Responds by displaying a list of metadata records, and by allowing users to edit metadata and data. Data will be typically edited by a third party too, e.g. excel spreadsheet, and will have to be re-imported by the user after it is modified.
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID	UC_045
Use Case Name	**Clear quality flag**
Short Description	Allows authorized User to clear a quality flag on a data or metadata record. This will occur when the User believes that the flagged issue is actually valid in the context of that record. The process will require the reviewer to make a note that is recorded with the flag, along with the identity of the reviewer and a time stamp.
Actors	Administrator, Data Steward
Pre-Conditions	Metadata records having quality flag attributes are available in the system The users have corrected the quality attribute issues, especially those that are automatically detected.
Success End Conditions	Users are able to remove data quality attribute flags
Data	Metadata quality attribute records (read/write)
Functions	• Remove quality issue flag • Log metadata record change

Main Sequence		
Step	**Actor**	**Description**
1	Users	Include use case << browse flagged data catalog entries>> Select data with quality issue flag Turn off quality issue flag
2	NGDS System	Responds by : Automatically validates data once more to check for possible remaining quality issues

		turning off the quality attribute flag in case of no quality issue found logging operations
Variants		
Step	**Actor**	**Description**
3	NGDS System	If some quality attribute is found, the metadata record flag is turned back on and the user is notified of the issue.
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	DN: Who has ultimate authority to determine the true data in the event of a dispute?	

Use Case ID	**UC_004b Catalog cleanup**
Use Case Name	**Catalog cleanup**
Short Description	The NGDS catalog will periodically and automatically verify the liveliness of the data sources referenced in its metadata catalog. That information will be used to mark the “unreachable”, or “non-public” data items in the metadata catalog. The data steward can use that information to delete metadata records that are invalid, or to fix those records with valid URIs.
Actors	Data Steward
Pre-Conditions	Requires authentication, access permission to edit metadata records. The catalog has metadata records for which URIs that will be checked for liveliness There is also a predetermined threshold time period above which the metadata record will be marked as having “broken links” There is also a period of time parameter used by the NGDS catalog to periodically sweep the catalog, looking for those links
Success End Conditions	The NGDS catalog is cleared up from metadata records that have invalid URIs or those records have their URIs fixed.
Data	All metadata records of the NGDS catalog
Functions	• Automatic verification of broken links • Automatic verification of “private links” from third party repositories • Automatic tagging of broken metadata records • Deletion of catalog metadata records based on broken link attributes

Main Sequence		
Step	**Actor**	**Description**
1	NGDS System	Runs period data catalog verification routine (There should be a very reasonable timeout period on this like 30 days where the record is marked private to allow data stewards ample time to fix things if on vacation or away from their premise.) Mark records as "unreachable" when URIs are inaccessible for a period more than the threshold Mark records as "private" in case authentication is required. Keeps record of unreachable links Keeps record of private links.
2	Data Steward	Request "unreachable" metadata records report/screen
3	Data Steward	Corrects URIs making them "alive" again, possibly by contacting the data submitter to ask rather than embark on a search unless there is some other data that would aid them.
4	Data Steward	Instructs the system to remove selected "unreachable" metadata records
5	Data Steward	Select remaining "unreachable" metadata for deletion
6	NGDS System	Deletes selected "unreachable" metadata records after
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
	NGDS System	In case of internal import failure, the system will roll back all existing transactions, returning to a valid state.
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	Is there any undo capability? All deletions are final? DN: I would defer this question to the domain experts. IN some cases, I would imagine that knowing there once was a record might be useful if it can be tracked down via other means (or if they want to talk to the data submitter directly). Good question	
2	Before any data can be deleted or removed (potentially via the duplicate detection process) the node must ensure that there are no metadata records existing for it.	

The data steward may choose to delegate the responsibility of correcting flagged resources to the data submitter responsible for the record. In these cases, the data steward can send notifications to data submitters indicating resources flagged for quality that need to be addressed.

Use Case ID	**UC_004c**
Use Case Name	**Notify data submitter of data or metadata problem**
Short Description	The goal of this use case is to allow the system to notify data submitters whenever other users other than the submitter herself, detect problems in the data or metadata.
Actors	Data Submitter, Data Steward, End User
Pre-Conditions	The metadata record has been imported into NGDS catalog.

Success End Conditions		The data submitter receives e-mail notifications with quality issues detected by other users.	
Data		Metadata records (read/write) Metadata QA (quality attribute) records (read only) Data submitter information: e-mail	
Functions		• Send e-mail to data submitter with metadata and data issues	
Main Sequence			
Step	**Actor**	**Description**	
1	Users	Inspect data or metadata record Find issues with the data or data record, for example, typos, wrong information Choose to create an e-mail describing the issue Provide e-mail for further contact Sends e-mail	
2	NGDS System	Responds by automatically filling in the e-mail of the data submitter and the reviewer and producing an e-mail with the collected comments from other users.	
Variants			
Step	**Actor**	**Description**	
Exceptions			
Step	**Actor**	**Description**	
	NGDS System	Bounding of e-mail may occur if the data submitter e-mail is invalid	
Open Issues (Please use this field to indicate questions/comments on the use case)			
ID	**Issue Description**		
1			

3.2.2.3 Resource Publication

Once the data steward is confident that a resource's quality is acceptable, the user can mark the resource metadata as "published". This allows un-restricted, read-only access to the resource's metadata, and includes the resource's metadata in any public-facing harvest interface (i.e. CSW).

Use Case ID	**UC_009**
Use Case Name	**Make resource public**
Short Description	This allows Data Steward to indicate that a resource is available for public discovery and access once they are satisfied with the data quality.
Actors	Data Steward
Pre-Conditions	Metadata record is properly created in the catalog, waiting to be made public by a data steward Data is properly uploaded to a data repository, waiting to me made public.
Success End Conditions	The metadata record is publicized AND the data is made available through a public URI.
Data	Metadata records in the catalog or files in the repository

Functions		• Make data public though a URI • Make metadata record public
Main Sequence		
Step	**Actor**	**Description**
1	Data Steward	Include use case <<Browse and manage resource directory>>
2	Data Steward	Select data or metadata for publication
3	NGDS system	Make data or metadata public (as a valid URI or in the catalog respectively)
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

3.2.3 *Node Administrator*

The primary purpose of the node administrator is to perform administrative actions on the node, manage user accounts related to their node, to insure the node's performance and that it communicates relevant information according to NGDS standards and protocols. Figure 7 illustrates the node administration use cases.

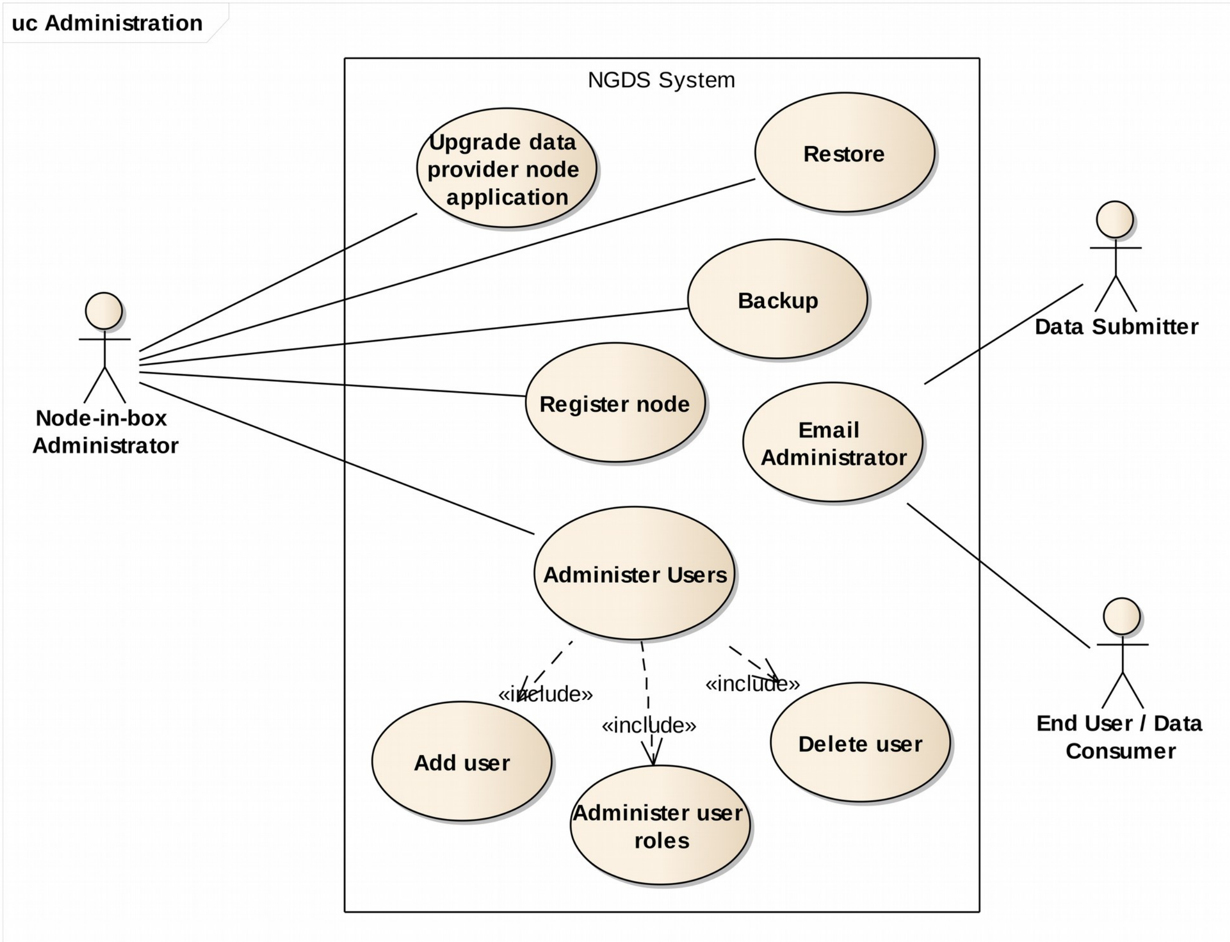


Figure 7 Administration Use Case View

Here follows a brief description of each of the use-cases.

3.2.3.1 User Account Management

The node administrator must be able to manage user accounts. This includes the ability to add and remove users, assign permissions to users or groups, add and remove users from groups.

Use Case ID	**UC_029a**
Use Case Name	**Administer users**
Short Description	Allows the system administrator to manage data provider node users. This will allow the system administrator to add and remove users on the administered node, and assign user roles and group membership.
Actors	Node-in-box administrator
Pre-Conditions	Node-in-the box is properly installed and configured
Success End Conditions	The administrator is able to perform the main administration operations
Data	User records

Functions		• Add user • Delete user • Modify user permissions and roles
	Main Sequence	
Step	**Actor**	**Description**
1	User	Include use cases <<add user>>, <<administer user roles>>, <<delete user>>
2	NGDS System	Responds to administration operations, enforcing role and users policies.
	Variants	
Step	**Actor**	**Description**
	Exceptions	
Step	**Actor**	**Description**
	Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**	
1	DN: Within this requirement, it will be necessary to perform a check to ensure that a user being deleted does not have custody of any records otherwise we may end up with orphaned records (unless they default to the node administrator). Just a though that came to mind.	
2	DN: A user should only be deleted once all their records (data and metadata) is transferred to another user. Otherwise there will be orphaned data and metadata.	

Use Case ID		**UC_029b**
Use Case Name		**Enrolment**
Short Description		Allow users to self enroll. i.e. to create their account in the NGDS system for the purpose of supporting NGDS data import/export and exploration activities.
Actors		End User/Data Consumer, Data Submitter
Pre-Conditions		Node-in-the box is properly installed and configured
Success End Conditions		The users have their accounts created, and their user data and credentials accessible throughout the system
Data		User records
Functions		• Enroll user • Un-enroll user
	Main Sequence	
Step	**Actor**	**Description**
1	User	Navigates to the enrollment screen of the system Inputs enrolment data including: password, login, e-mail, and some preferences Includes <<add user use case>>
2	NGDS System	Responds by: Validating enrollment data, for example, checking for repeated logins Creating a user account

Variants		
Step	**Actor**	**Description**
1b	User	Navigates to the enrollment screen of the system Requests deletion of her profile -> un-enrolment
2b	NGDS System	Responds by: Deleting the user record and its credentials Includes <<delete user>> use case
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	Do we really want to allow users to self enroll as data providers? DN: My answer would be yes since the system has been designed with the data steward as a failsafe against publishing bad data. The less human involvement the better.	
2	DN: Same caveat here WRT orphaned data and metadata. Do not allow a user to remove themselves if they have data submitted until the ownership of the data is re-assigned or the data is removed.	

Use Case ID		**UC_030**
Use Case Name		**Add users**
Short Description		Creates a new user account in the system, allowing the storage of important user information such as e-mail, name, login, password, address, and enabling features such as subscription and search saving, and auto complete.
Actors		Node-in-box administrator, Users
Pre-Conditions		Node-in-the box is properly installed and configured
Success End Conditions		A new user account is created
Data		User records
Functions		• Add user • Add users via invitation
Main Sequence		
Step	**Actor**	**Description**
1	Administrator	Navigates to user management screen Starts new user creation Inputs user information
2	NGDS System	Responds by checking for repeated user credentials and if positive, creating a new account for a user

Variants		
Step	**Actor**	**Description**
1b	Administrator	Navigates to user management screen Starts new user creation Inputs user information Send invitation to user
2b	NGDS System	Responds by checking for repeated user credentials and if positive, send invitation to user
3	User	Responds by accepting or rejecting invitation
4	NGDS System	Creates user account or does nothing if the invitation was rejected
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID	**UC_031**
Use Case Name	**Delete users**
Short Description	Removes a user, its credentials and associated information from the system
Actors	Node-in-box administrator
Pre-Conditions	Node-in-the box is properly installed and configured
Success End Conditions	A new user account is deleted, together with its data including subscriptions, saved searches, history, etc.
Data	User records
Functions	• remove user

Main Sequence		
Step	**Actor**	**Description**
1	Administrator	Navigates to user management screen Selects the user to be removed Selects user removal option
2	NGDS System	Responds by removing user profile with its associated information including subscriptions, saved searches and history. Notify administrator of unpublished records by a data steward, and change permissions of those data records to the administrator.

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	What happens if a user that is a data steward is deleted? Are her metadata records unpublished, assigned to a system pseudo user?	
2	DN: QUESTION: Should the system remove the associated information or should certain info persist (logs, metadata etc)?	

Use Case ID	**UC_032**	
Use Case Name	**Administer user roles**	
Short Description	The goal of this use case is to allow the administrator to assign different roles to users. These roles control the users' abilities to publish data in the system or to administer system functions.	
Actors	Node-in-box administrator	
Pre-Conditions	Node-in-the box is properly installed and configured	
Success End Conditions	User role assignment is updated according to administrator needs	
Data	User records	
Functions	• assign role to user • remove use role	
Main Sequence		
Step	**Actor**	**Description**
1	Administrator	Navigates to user management screen Selects the user to be managed Selects user roles
2	NGDS System	Responds by assigning selected roles to user.
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	Are users federated across nodes? RSSF: Stewards and submitters are local users, end user/consumers are global users.	
2	DN: DERIVED REQUIREMENT: Every Node must have at least one administrator, one steward correct?	

Use Case ID		**UC_032b**
Use Case Name		**Backup**
Short Description		The goal of this use case is to allow node administrators to backup the data and meta-data being stored in the system.
Actors		Node-in-box administrator
Pre-Conditions		Node-in-the box is properly installed and configured
Success End Conditions		The data, metadata and indexes of a node-in-a-box is successfully backed up
Data		User records Metadata records Data indexes
Functions		• manual backup NGDS node • automatic backup of NGDS node
Main Sequence		
Step	**Actor**	**Description**
1	Administrator	Navigates to user management screen Selects backup option Provides destination folder
2	NGDS System	Responds by backing up node data to the assigned repository
Variants		
Step	**Actor**	**Description**
1	Administrator	Navigates to user management screen Selects backup option Set ups automatic backup option by providing a destination and a day of week/time
2	NGDS System	Responds by backing up node data to the assigned repository at the given day of week/time
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID		UC_032b
Use Case Name		**Restore**
Short Description		The goal of this use case is to allow node administrators to restore the backed up data and meta-data being stored in the system.
Actors		Node-in-box administrator
Pre-Conditions		Node-in-the box is properly installed and configured
Success End Conditions		The data, metadata and indexes of a node-in-a-box is successfully backed up
Data		Backed up blob with important node data
Functions		• restore NGDS node
Main Sequence		
Step	**Actor**	**Description**
1	Administrator	Navigates to user management screen Selects restore option Provides source folder and file
2	NGDS System	Responds by restoring up node data to the assigned repository
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

3.2.3.2 System Management (Meet NGDS System Requirements)

If a third party repository is used to store data, it is the responsibility of the node administrator to ensure that their node provides information in accordance with NGDS standards and protocols. This set of requirements must be filled by any system that wishes to play the role of a node or data repository within the NGDS. Alternatively, the NGDS node-in-a-box, described in this document, will simplify the node administrator's workload by providing an easily installed and configured software package that satisfies these requirements.

The following are requirements that the node administrator must insure are satisfied. They are outlined as requirements since they are targeted functionality for software development.

- The node shall provide access to tier 1 data and file-based representations of tier 2 and 3 data via standard HTTP protocols

- The node shall provide access to tier 2 and 3 data through OGC web-services (WFS, WMS, WCS) where appropriate

- The node shall provide metadata that conforms to the USGIN metadata profile (http://repository.usgin.org/sites/default/files/dlio/files/2011/u11/usgin_iso_metadata_1.1.3.pdf)

- The node shall provide access to metadata through a CSW 2.0.2 discovery service

- The node shall create de-referenceable URIs for all applicable resources, including data sets, metadata records, files, and features within data services

Once the node administrator is satisfied that their node meets the demands of standards and protocols required to participate in the NGDS, the node administrator may request registration of their node with the system's aggregating catalog, thus publishing their node's data into the system.

Use Case ID		**UC_034**
Use Case Name		**Register node**
Short Description		The goal of this use case is to allow new nodes to be added to the NGDS grid. These nodes can be of different types included "node-in-a-box" installations, or third party data provider repositories.
Actors		Node-in-a-box Administrator, NGDS Administrator
Pre-Conditions		Node-in-the box properly installed as a NGDS node management hub Client node-in the box properly installed but not yet registered
Success End Conditions		A new node is registered in the NGDS network, and the data it provides becomes available to the other nodes, and searchable through the by the NGDS catalog
Data		e-mails, node credentials
Functions		• add new node • index new node
Main Sequence		
Step	**Actor**	**Description**
1	Node administrator	Get NGDS administrator information Include <<e-mail node-in-a-box administrator>> to NGDS administrator
2	NGDS Administrator	Include << register new node into NGDS>>
3	NGDS System	Responds by: Locating the NGDS administrator user/node automatically filling in the user e-mail address; sending an e-mail to the NGDS administrator

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	Who manages the grid? Is the grid configured centrally? NGDS Administrator in a special user in a node elected to be the main node.
2	DN: DESIGN QUESTION: Will the NGDS system use a one master – many slaves model or a truly federated model where there are only peers? RSSF: We will decide this after a trade study.

3.2.3.3 Routine Maintenance

The node administrator must also perform routine maintenance tasks pertaining to their node. These include upgrading software and responding to user questions and requests.

Use Case ID		**UC_033**
Use Case Name		**e-mail (NGDS or node-in-a-box) administrator**
Short Description		The goal of this use case is to allow users to send e-mail to administrators of the system to handle matters such as granting of especial access rights, or to register new nodes in the NGDS network, or other issues. There are two types of administrators: node-in-a-box administrators, and NGDS administrators.
Actors		All users, Administrators
Pre-Conditions		Node-in-the box is properly installed and configured Administrator has registered her e-mail information
Success End Conditions		Users can communicate their needs/issues with the node-in-a-box administrator
Data		e-mails
Functions		• send email to administrator
Main Sequence		
Step	**Actor**	**Description**
1	All users	Navigates to e-mail node-in-a-box administrator page Select send e-mail to administrator Type –mail message Selects send e-mail, confirms operation
2	NGDS System	Responds by: automatically filling in the user e-mail address; sending an e-mail to the administrator

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID		**UC_035**	
Use Case Name		**Upgrade node-in-the-box application**	
Short Description		The goal of this use case is to upgrade the software that implements the node-in-the-box application. This can potentially include the data repository, the catalog service, and the WebApp application	
Actors		Node-in-a-box administrator	
Pre-Conditions		Node-in-the box is properly installed	
Success End Conditions		The software components(s) of the node are updated without data loss and will minimum impact on other nodes of the NGDS network	
Data		Possibly all the data stored in the data repository and index	
Functions		• update software components • shutdown node • restart node	
Main Sequence			
Step	**Actor**	**Description**	
1	Node Administrator	Gracefully disconnects the node from the network Shuts down the server component (s) Update software Restart service(s)	
2	NGDS System	Shuts down gracefully & restarts without data loss	
Variants			
Step	**Actor**	**Description**	
Exceptions			
Step	**Actor**	**Description**	
Open Issues (Please use this field to indicate questions/comments on the use case)			
ID	**Issue Description**		
1			

3.3 Software Developer Use Cases

Software developers are users who build applications that access NGDS resources using the system's public, service-oriented APIs. These applications may be built in order to satisfy any number of functions, including data discovery, accessibility, visualization and analysis.

Figure 8 illustrates use cases for software developers, who will build applications that access NGDS resources using the public web services API of the system for their own applications, and the public API of the system will mainly support the standard web services WMS, WFS, and CSW. The current system will allow users to publish links to their application websites, but will not provide any advanced catalog feature to manage those applications such as what is provided by app stores.

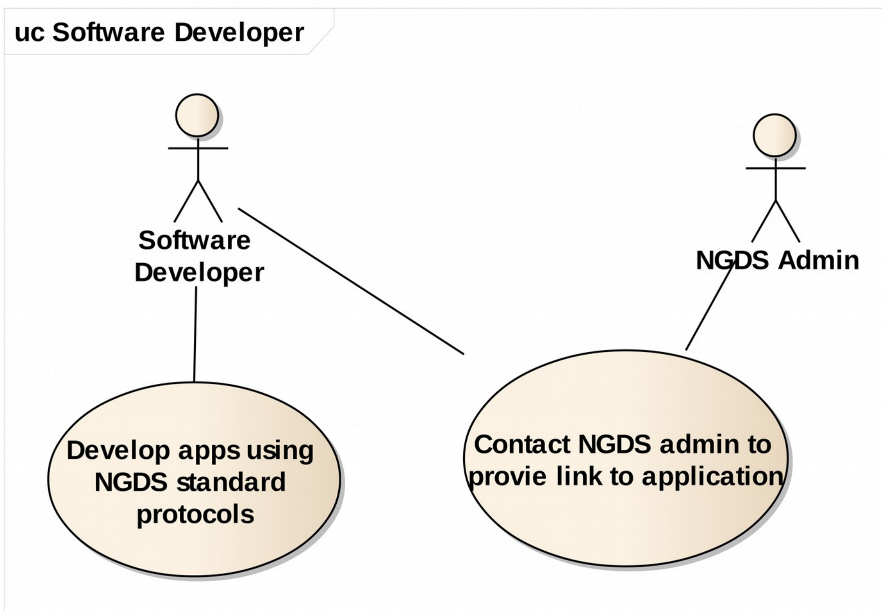


Figure 8 Software developer use cases

3.3.1 Develop Applications

Use Case ID	UC_040
Use Case Name	Develop apps using NGDS standard protocols
Short Description	Through the use of web protocols, in particular: CSW, WFS and HTTP, software developers can build applications that utilize the data and meta-data stored in NGDS. They can also use NGDS to locate referenced data (data that is not stored in NGDS components but is referenced by the NGDS catalog). They can, for example, develop applications that display that information on maps, perform data analysis and discover geological information of geothermal sites within US
Actors	Software developer

Pre-Conditions		NGDS data repositories and catalogs are available and accessible through standard internet protocols. Software developers may need an account to interact with the system if their application involves the updating of information in repositories and catalog, or if they utilize services such as posting of reviews.	
Success End Conditions		Software developers can successfully build applications that utilize the NGDS system resources	
Data		Data models, metadata schemas, key system URIs, e.g. catalog service main URI, protocol and data model documentation	
Functions		• Standard protocols • Ability to interact with NGDS via APIs • Logging of API usage statistics and accesses	
Main Sequence			
Step	**Actor**	**Description**	
1	Software Developer	Build applications that utilize standard protocols and data models from NGDS	
2	NGDS System	Responds to these protocols interactions via well known APIs	
Variants			
Step	**Actor**	**Description**	
Exceptions			
Step	**Actor**	**Description**	
Open Issues (Please use this field to indicate questions/comments on the use case)			
ID	**Issue Description**		
1	DN: It might be a good idea to issue API keys to control the number of queries and use of data or at least have a metric to understand who is using it and for what. DN: Discussion about developer API keys. Not critical but something to keep in mind. Given there are not concrete plans on how to pay for maintenance after the system is up and running, providing unlimited API access might not be something that can be afforded.		

3.3.2 Register New Application with NGDS

Furthermore, the software developer will wish to register their new application with the NGDS in order to promote its use throughout the system.

Use Case ID	**UC_041**
Use Case Name	**Contact NGDS admin to provide link to application**
Short Description	The goal of this use case is to capture the need for a way to software developers request the inclusion of application links in the NGDS web site. They will do so via a public e-mail address or via a form that allow them to submit this information.
Actors	Software developer, NGDS administrator
Pre-Conditions	NGDS website is available and its URI is know by the software developer

Success End Conditions		NGDS software developers can provide a link to their applications and this information is incorporated in the list of applications hosted in the NGDS website
Data		URIs from software developers e-mail or URI to website location where developers can post those requests
Functions		• Publish new application links via the NGDS system
Main Sequence		
Step	**Actor**	**Description**
1	Software Developer	Post links to software via the NGDS website
2	NGDS System	Captures the request and routes it to the NGDS administrator for approval
3	NGDS Administrator	Responds by approving or rejecting a new website link and description
4	NGDS System	If approved, automatically updates the website.
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

3.4 END-USER USE CASES

Once datasets are published into the NGDS from a node, it will be possible for end-users to discover and access those datasets through a variety of different applications.

End-user use cases are organized around the basic workflows that the NGDS must support. These are: discover and gather, validate and evaluate, and analyze and visualize data. As shown in Figure 9 in each one of these steps, users can perform different activities.

Note that the use cases discussed in this section summarize the results described in the End-User Research Summary provided by Anthro-Tech (**see document /P05**)

Once datasets have been published to the system by registering them in the NGDS catalog and making the resource accessible through standard web service protocols and data formats, it will be possible for End Users to discover and browse the dataset using the NGDS WebApp application. They can browse resource descriptions returned by searching the NGDS catalog, and view the resource described by a particular catalog entry. Browsing capabilities will depend on the type of resource. Documents can be accessed through standard web linking and browser display for file types like PDF, TIFF, TXT. Tabular data in represented in NGDS supported formats will be displayed in a standard table type grid display. Geo-referenced resource locations will be visualized through a map interface that provides data exploration capability. This map-based data exploration utilizes the information provided in the metadata hosted in the catalog, as well as third party data layers to help users triangulate the data with that external information. Map-based search can be combined with other types of searches such as text-based, and can be further narrowed down by data type and metadata content filters.

For file-based resources, End Users will be able to download the associated files using links from catalog entries. The WebApp application will provide data filtering and download capabilities for data sets that accessible through NGDS services in standard NGDS interchange formats. End users will be able to enter filter criteria, and subset datasets to meet user-defined criteria, and last, end users will be able to export these results to their local file system.

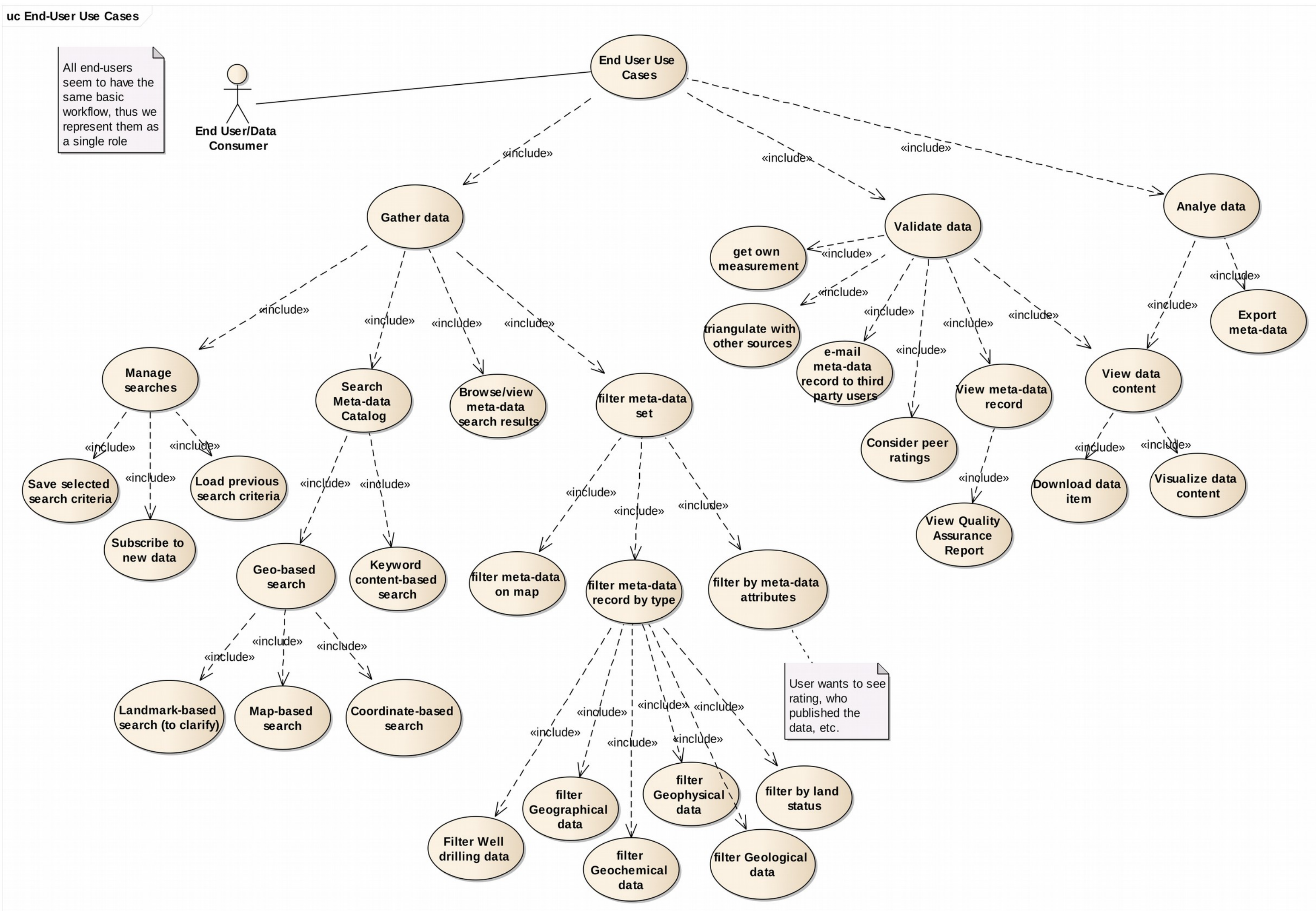


Figure 9 End-user use cases overview

In the following sections, we analyze these 3 different steps individually.

3.4.1 DISCOVER AND GATHER DATA

The first and most important use of the NGDS to end-users is the ability to discover and gather geothermal data. Gathering data involves searching the NGDS metadata catalog for different types of geothermal information, casually browsing the search results, and further filtering those results based on the user's own criteria of relevance. The complete tree of Use Cases is illustrated in Figure 10. We will describe only the most important of those use cases in the following.

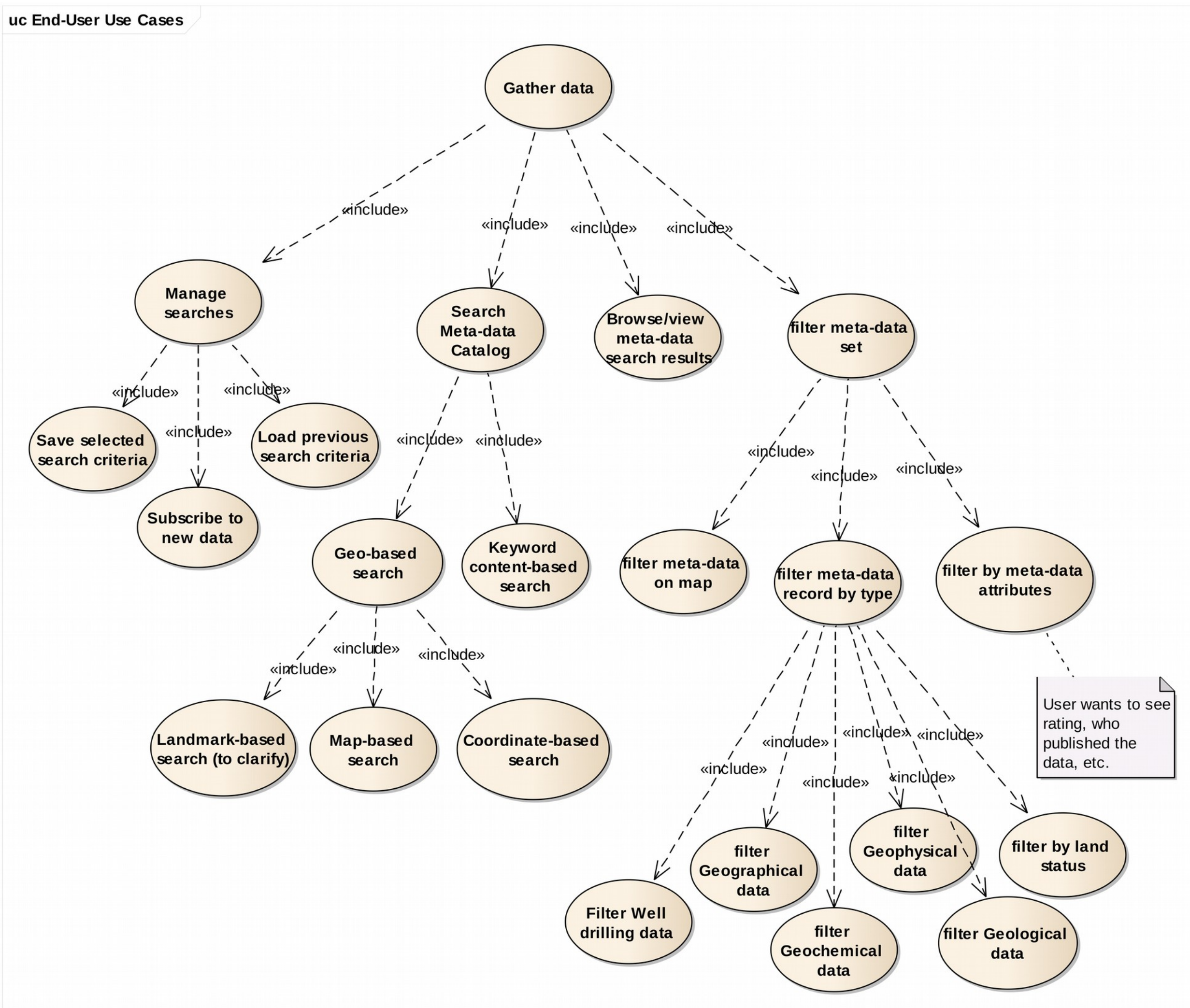


Figure 10 Data gathering supporting use cases

The "gather data" use case, shown in Figure 10, can be broken down into sub use cases. It consists of searching the metadata catalog for different types of geological information, doing a casual browsing, and further filtering that information based on the user's own criteria of relevance. Users can also save existing searches, load them in a later occasion, or use those searches as subscriptions. This publish/subscribe model allow users to be notified whenever new data matching a certain criteria is input into the system catalog.

The NGDS catalog can also be searched by data geo-location. The geo-location search can be based on landmarks, on areas in the map, or can include more precise, coordinate-based searches.

Once a search is performed, it is usually the case that information must be further prioritized and filtered. The filtering of information allows users to select a relevant subset from the totality of data returned by a general search, and to rank the information according to the user's needs. Filtering can be performed by data type, e.g. by selecting or discarding certain data types, including well drilling, geographical, geochemical, geophysical, geological and land data, as well as publications; by geo-location, i.e. filtering results that fit within an area in the map, or by other metadata attribute, including the document content if this information can be extracted from the dataset (e.g. Tier3 dataset types). These filtering capabilities are not mutually exclusive, and are done in a combined and interactive way.

Here follows a brief description of each of the use-cases.

3.4.1.1 Map-Based Search

In searching for data, geographic location provides the first mechanism by which end-users may use to filter results. To that end, users expect a map-based search interface in which they can quickly filter results by panning and zooming the map, or moving directly to an area-of-interest by specifying a landmark name or coordinates.

<table>
<tr><th colspan="2">Use Case ID</th><th>UC_014</th></tr>
<tr><th colspan="2">Use Case Name</th><th>Map-based search</th></tr>
<tr><td colspan="2">Short Description</td><td>The goal of this use case is to support users discovery of metadata by using maps. Map-based search consist in zooming, panning and selecting a region in a map. This search method can be used together with other search methods, to narrow down the data in a search.</td></tr>
<tr><td colspan="2">Actors</td><td>End User/Data Consumer</td></tr>
<tr><td colspan="2">Pre-Conditions</td><td>System is correctly installed and operational</td></tr>
<tr><td colspan="2">Success End Conditions</td><td>The user can visualize a result of a search as layers in a map.</td></tr>
<tr><td colspan="2">Data</td><td>Metadata features shown as layers (WFS protocol), metadata summary of selected data items on map (when users click on a data point in the map)</td></tr>
<tr><td colspan="2">Functions</td><td>• Map navigation capability: panning, zooming, selecting sub-regions
• Retrieve metadata record for selected elements in a map</td></tr>
<tr><th colspan="3">Main Sequence</th></tr>
<tr><td>Step</td><td>Actor</td><td>Description</td></tr>
<tr><td>1</td><td>User</td><td>Navigates to the map screen
Utilizes the map functions (panning, zooming) to navigate to a geographical location
Visualizes data points in a map
Include <<filter metadata on map>> use case</td></tr>
<tr><td>2</td><td>NGDS System</td><td>Responds to user commands by updating maps and data layers</td></tr>
</table>

Variants		
Step	**Actor**	**Description**
1a		Include other types of search as landmark-based, coordinate-based, content-based, etc.
Exceptions		
Step	**Actor**	**Description**
	NGDS System	In case of failure to display layers due to network errors, for example, the system should notify user of the situation
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID		**UC_013**
Use Case Name		**Landmark-based search**
Short Description		The goal of this use case is to allow users to utilize landmarks (state, city, county, district, known geothermal sites) to narrow down the search in the catalog. This search method can be used together with other search methods, to narrow down the data in a search.
Actors		End User/ Data Consumer
Pre-Conditions		A database of landmarks and their geo-locations must be available for search
Success End Conditions		The user finds information based on landmarks
Data		Landmark name
Functions		• Find landmark • Retrieve landmark geo-location
Main Sequence		
Step	**Actor**	**Description**
1	User	Uses landmark name as a search criteria
2	NGDS System	Consider a list of metadata items around a landmark as a filter in the current search

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	As noted by one of the monitors this needs to be clarified with Antro-tech. what is a landmark? What are examples of landmarks? How difficult would it be to support landmark-based search in our approach? Can we reuse functionality from search engines as Google map services to do this? DN: Agree. I think pushing that off to the map vendor is a good approach as funding is limited within NGDS. Still, a request from Anthro-tech for clarification would be good.

Use Case ID	**UC_015**	
Use Case Name	**Coordinate-based search**	
Short Description	The goal of this use case is to allow users to utilize geographical coordinates to narrow down the search in the catalog.	
Actors	End User/Data Consumer	
Pre-Conditions		
Success End Conditions	The user finds information based on geographical coordinates	
Data	Geographical coordinates	
Functions	• Validate coordinates • Retrieve data from the system based on proximity or containment within geographical coordinates	
Main Sequence		
Step	**Actor**	**Description**
1	User	Uses coordinates as a search criteria by possibly typing them in
2	NGDS System	Validates input from the user Consider a list of metadata items around or within the provided coordinates as an additional filter in the current search

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	As noted by one of the monitors this needs to be clarified with Antro-tech. do the users need to type coordinates? Is there a UI-based way to support this search without requiring users to type those coordinates, for example, by drawing a box in a map? I think the onus should be on the end user to figure out where the landmark is and locate it on the map. It is very unclear what the landmark means.

3.4.1.2 Keyword-Based Search

End-users wish to be able to specify a keyword or set of keywords that restrict results to resources which utilize those keywords. This should work in tandem with map-based search, allowing users to filter simultaneously based on keyword and location.

Use Case ID		**UC_016**
Use Case Name		**Keyword content-based search**
Short Description		The goal of this use case is to allow users to search data by its metadata content registered in the catalog. If data comes in tier 3 format, it includes its indexed content; if it comes in Tier1 and Tier2 formats, the search is based on whatever could be extracted/converted/indexed into data or meta-data.
Actors		End User/Data Consumer
Pre-Conditions		There is metadata in the catalog
Success End Conditions		The user finds information based on metadata content
Data		All metadata stored in the NGDS catalog
Functions		• Content-based search of metadata records • Content-based search of data content
Main Sequence		
Step	**Actor**	**Description**
1	User	Navigates to a content-based search panel Types in search criteria
2	NGDS System	Validates input from the user Retrieves metadata records based on user typed content Shows results as text and, when possible, as layers in the map

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	DN: There may be a need for a basic thesaurus but hard to make with a limited budget.

3.4.1.3 Refining Results and Faceted Search

Once the end-user has generated a set of search results, usually through a combination of the methods outlined above, that user will wish to continue to refine their search results. This can occur through further map-based and keyword-based search, or may be accomplished by selecting from a set of search facets, or fields.

Use Case ID		**UC_019**
Use Case Name		**Filter results by type**
Short Description		The user can also narrow down its search results by specifying certain data types of interest, thus ruling out all other data that do not belong to these types from the returned list of search results.
Actors		End User/Data Consumer
Pre-Conditions		A search was performed and a subset of the metadata from the catalog was retrieved by the NGDS catalog. The search result is displayed as a list of metadata records
Success End Conditions		The user can narrow down the search results
Data		A subset of metadata obtained by a search A list of metadata types present in the subset of metadata under consideration
Functions		• Filter metadata set by type
Main Sequence		
Step	**Actor**	**Description**
1	User	Selects one or more data types from a list of types displayed for a search result
2	NGDS System	Responds by displaying only the metadata list for the types the user selected The map may also update in response to this search

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID	**UC_020**	
Use Case Name	**Filter results by metadata attributes**	
Short Description	Different content models prescribe different attributes to different types of data, these attributes can be used to further refine the search result, for example, excluding data points for which their metadata record do not have certain attribute content values.	
Actors	End User/Data Consumer	
Pre-Conditions	A search was performed and a subset of the metadata from the catalog was retrieved by the NGDS catalog. The search result is displayed as points in a map and as a list of metadata records	
Success End Conditions	The user can narrow down the search results	
Data	A subset of metadata obtained by a search A list of metadata types present in the subset of metadata under consideration	
Functions	• Filter metadata set by attribute content	
Main Sequence		
Step	**Actor**	**Description**
1	User	Types in a filtering criteria based on supported metadata attributes
2	NGDS System	Responds by displaying only the metadata list matching user filter criteria The map may also update in response to this search

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	Missing Filters: by location terms, geothermal thematic terms, source organization, publication date, popularity and user ratings. Requirements need to clearly indicate that these vocabularies will exist on which such faceted filtering can be performed. ND: This should be a requirement for the metadata team

End-users also expect to be able to draw a specific area of interest on the map, and have the search results only reflect data from that specific area.

Use Case ID		**UC_018**
Use Case Name		**Filter results on map**
Short Description		Once a search is made and search results is displayed as points on a map, the user can narrow down its search by selecting a sub-area in the map, thus filtering out all data points that are outside that geographical region.
Actors		End User/Data Consumer
Pre-Conditions		A search was performed and a subset of the metadata from the catalog was retrieved by the NGDS catalog. The search result is displayed as points in a map
Success End Conditions		The user can narrow down the search results
Data		A subset of metadata obtained by a search
Functions		• Select region in a map • Filter search results based on a bounding box on a map
Main Sequence		
Step	**Actor**	**Description**
1	User	Select a subset of elements in a map by defining a region. i.e. creating a bounding box in a map
2	NGDS System	Respond to user input by zooming the map to the selected region and filtering out results that are outside the bounding box. The list of metadata result displayed in a textual form may need to be updated.

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

3.4.2 Validate and Evaluate Data

Once an end-user has located resources of interest, they next investigate to learn what they can about the dataset or file content to determine if it is valid for their intended purpose. The first step in the process is adequate visualization of the search results themselves. The complete tree of use cases is illustrated in Figure 11.

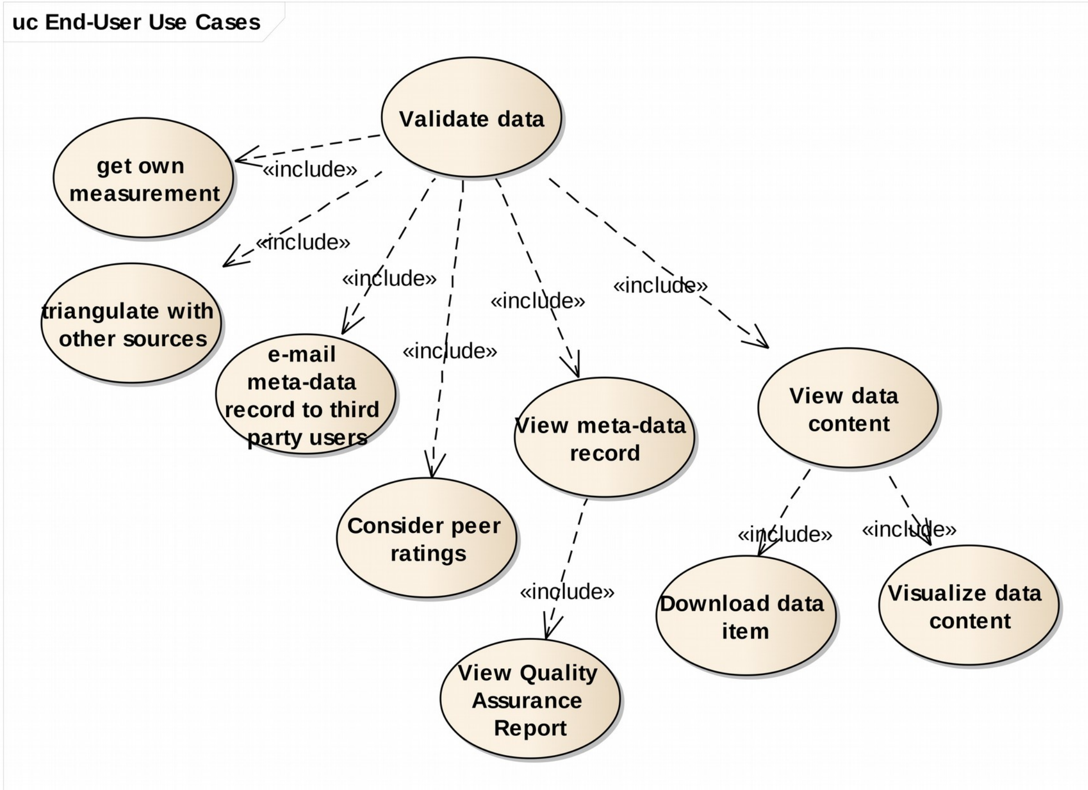


Figure 11 Data validation supporting use cases

In the following we list the most important use cases of this category.

<table>
<tr><td colspan="2">Use Case ID</td><td>UC_017</td></tr>
<tr><td colspan="2">Use Case Name</td><td>Browse/view metadata search results</td></tr>
<tr><td colspan="2">Short Description</td><td>The goal of this use case is to allow users to visualize the results of a search and inspect its content.
This visualization is supported by metadata lists and maps.</td></tr>
<tr><td colspan="2">Actors</td><td>End User/Data Consumer</td></tr>
<tr><td colspan="2">Pre-Conditions</td><td>A search was performed and a subset of the metadata from the catalog was retrieved by the NGDS catalog.</td></tr>
<tr><td colspan="2">Success End Conditions</td><td>The user can find what he/she is looking for</td></tr>
<tr><td colspan="2">Data</td><td>A subset of metadata obtained by a search</td></tr>
<tr><td colspan="2">Functions</td><td>• Browse search results
• Inspect elements in a map</td></tr>
<tr><td colspan="3">Main Sequence</td></tr>
<tr><td>Step</td><td>Actor</td><td>Description</td></tr>
<tr><td>1</td><td>User</td><td>Include <<search metadata catalog>>, <<filter metadata set>></td></tr>
<tr><td>2</td><td>NGDS System</td><td>Shows results as text and, when possible, as layers in the map
Shows metadata record when user clicks on a map or on metadata lists</td></tr>
</table>

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

3.4.2.1 Metadata Evaluation

The user first wishes to assess information available in a resource's metadata such as the originator's description of the dataset, the resource's accessibility, its provenance and peer-reviews of the resource. In order to do so requires that the user be able to access the resource's metadata in its entirety.

The evaluation itself is outside the scope of the system. The system just provides the metadata (and data as per the above use case). The end user's evaluation will probably happen once they get the data they are seeking.

Use Case ID		**UC_026**
Use Case Name		**View metadata record**
Short Description		After locating a piece of metadata in the catalog, the user inspects the metadata record in more detail, for example, to decide if it refers to the data she is looking for.
Actors		End User/Data Consumer
Pre-Conditions		A search was performed and a metadata record was selected for further inspection
Success End Conditions		The user is able to access and view the contents of the metadata
Data		Documents and structured records stored in GTDA repository or third party repositories. The data is located through a URI
Functions		• Retrieve metadata record • Visualize metadata record
Main Sequence		
Step	**Actor**	**Description**
1	User	Select metadata record Opens metadata record in a visualization mode
2	NGDS System	Responds by showing the contents of all the content of the metadata record
3	User	Views the metadata record content

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
	User	In case the URI pointed by the metadata record becomes unavailable during the execution of this procedure, the system must provide an error message. The metadata record may be marked as invalid.
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	Missing requirements for other “views” of metadata, for example as human-readable HTML, or as an ATOM entry. These are just low-hanging fruit.	

Use Case ID	**UC_024**	
Use Case Name	**Provide peer ratings**	
Short Description	By inspecting data reviews posted by other users, a user can gauge the accuracy and validity of data. Peer ratings can include textual description, star ratings or both. This UC captures the fact that the system must provide support for peer ratings. As part of this use case, users can also post peer ratings. There is no restriction of who can post those ratings. The user log-in information is used to identify the peer rating poster	
Actors	End User/Data Consumer	
Pre-Conditions	The existence of ratings posted by other peers for the case of review None if the user will be the first to post a rate	
Success End Conditions	The user is able to view peer ratings text and star ratings in the metadata posted in the NGDS catalog.	
Data	Peer ratings and their textual description that are attached to metadata in features of the map and the search results list	
Functions	• Provide peer ratings star rating and text for a given metadata record • Post a peer rating • Store peer reviews	
Main Sequence		
Step	**Actor**	**Description**
1	User	Include family of use cases <<gather data>>
2	NGDS System	Shows user a result of a data gathering activity Displays peer ratings together with returned data
3	User	Opens peer ratings of selected items read peer ratings star rate and their textual description

Variants		
Step	**Actor**	**Description**
2b	User	Include use case <<view metadata content>>, <<view document content>> Post a peer rating
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	This feature leads me to suggest that optional user profile fields indicating who a reviewer is (e.g Steve Richard of AZGS rates this 3 stars)	

3.4.2.2 Data Comparison

Often end-users wish to look more carefully at the data's values and compare it to other datasets to which they are familiar. This may involve visualization and data analysis within the search interface, or in an external environment. See the discussion of data visualization and analysis below for more information.

Use Case ID		**UC_022**
Use Case Name		**Triangulate with other sources**
Short Description		Users may compare the metadata returned by the system with external data layers, provided by external WFS sources such as demographics, topological, weather, and so on, thus helping them to make inference on the quality of the information. These use cases can also be performed by third party applications that consume the data published in NGDS
Actors		End User/Data Consumer
Pre-Conditions		A search was performed and a subset of the metadata from the catalog was retrieved by the NGDS catalog.
Success End Conditions		The user compares the returned metadata with third party data sources and comes to a conclusion
Data		External data sources as WFS providers for map layers Metadata from the system
Functions		• Download data • Export data via standard protocols to third party applications
Main Sequence		
Step	**Actor**	**Description**
1	User	Include family of use cases <<gather data>>
2	NGDS System	Shows user a result of a data gathering activity
3	User	Selects download report
4	NGDS System	Creates a metadata report
5	User	Compares the data manually with their own, or with the help of a CAD tool.

Variants		
Step	**Actor**	**Description**
3b	User	Uses CAD tool, via WFS to read NGDS repository data
4b	NGDS System	Exports data via WFS for third party CAD tool.
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	It is not clear if the scope of the system includes the mashing up of information with external data sources. This seems to be an advanced feature that may be better achieved by utilizing third party CAD tools. DN: Agree!	
	RC: The general idea is that you compare the data in a particular dataset to some other known data. This should be fleshed out as requirements for accessibility and download of data.	
	MM: There is an opportunity here to enable crowd sourced information about resources via this use case.	

3.4.2.3 Storing and Sharing Search Results

The process of data evaluation can take place over a long period of time, and may involve multiple end-users. In order to facilitate this, end-user search interfaces must support the capability of a user to save and share a set of filtered search results with other users.

Use Case ID		**UC_025**
Use Case Name		**View data content**
Short Description		After locating a piece of data, the user inspects the data content by URI element that the metadata refers to.
Actors		End User/Data Consumer
Pre-Conditions		A search was performed and a metadata record was selected for further inspection
Success End Conditions		The user is able to access the data pointed by the metadata record
Data		Documents and structured records stored in NGDS repository or third party repositories. The data is located through a URI
Functions		• Retrieve URI document • Open and display document content to user
Main Sequence		
Step	**Actor**	**Description**
1	User	Select document view
2	NGDS System	Responds by retrieving the document or data record, and opening it with an appropriate viewer, that will allow user to inspect the document or data record content A tabular viewer for a structured data record would be useful.
3	User	Views the data content

Variants		
Step	**Actor**	**Description**

Exceptions		
Step	**Actor**	**Description**
	User	In case the URI pointed by the metadata record becomes unavailable during the execution of this procedure, the system must provide an error message. The metadata record may be marked as invalid.

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	DN: Long URLs can be aliased via a URL shortener. This may be better as it does not require the system to preserve the state of a specific search yet allows the search to be shared. A discussion to have. Rather than save the search criteria, maybe it should just provide a URL encoded string that can be used to represent the same state. This is less expensive from a systems perspective. Example - https://www.google.ca/search?q=NGDS&oq=ngds&sugexp=chrome,mod=0&sourceid=chrome&ie=UTF-8

Use Case ID	**UC_010**
Use Case Name	**Save selected search criteria**
Short Description	The goal of this use case is to allow users to save searches, to be reused in a later time, and for setting up subscriptions to content changes. In this use case, after the user performs a search, she saves that search parameters for further use. This search then can be used to subscribe to new data, and to continue a previous discovery activity. Searches are saved on the end-user accounts, for their private use. In the future they may be shared among other users. When saving a search, users can opt to make search public so others can reuse it.
Actors	End User/Data Consumer
Pre-Conditions	In order to allow saving and retrieval, the user must be identifiable; hence, there is a need for users (in particular the end user) to be logged in using their unique account.
Success End Conditions	The search criteria is properly validated and saved into the system under a given name.
Data	Search criteria
Functions	• Save search criteria • Validate search criteria • Record search parameters by monitoring user input • Make search public to other users

Main Sequence		
Step	**Actor**	**Description**
1	Users	Include use cases <<Search meta-data catalog >>, <<filter meta-data set>>
2	NGDS System	Performs search according to user discovery workflow Record save parameters by monitoring user input
3	Users	Select save search criteria option from the UI Give the search a name Select option to make search public/private

4	NGDS System	Saves search criteria Make search criteria public if user selected so.
Variants		
Step	**Actor**	**Description**
1b	Users	Input search criteria in a separate content-based search form, instead of using the included use cases
2b	NGDS System	Uses form-based search as criteria
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	Should the system support form-based search only or should it record data as shown above? Which option is better? DN: I would adhere to the architectural principles known as REST. A URI represents the state of a specific resource etc.	
2	Should the saved searches from one user be visible to other users as Antro-tech indicated that searching is a collaborative process? Maybe searches are not saved under a use id but treated as a resource. They could then be saved under the resource name (URI) and the (anonymous) user could create a bookmark to access the search later. This bookmark/URL can also be shared with other users. DN: Agree.	

Use Case ID		**UC_012**
Use Case Name		**Load previous search criteria**
Short Description		The goal of this use case is to support users in loading previously saved search criteria. They do so by browsing through their list of saved searches.
Actors		End User/Data Consumer
Pre-Conditions		In order to allow saving and retrieval, the user must be identifiable; hence, there is a need for users (in particular the end user) to be logged in using their unique account.
Success End Conditions		The loaded search criteria is loaded and executed, displaying results in the system
Data		Saved search criteria
Functions		• Load saved search criteria
Main Sequence		
Step	**Actor**	**Description**
1	User	Navigates to the saved search screen Browses through existing saved searches (if any) Selects an existing saved search Instructs the system to load the search
2	NGDS System	Loads a saved search Uses the search criteria to search the catalog

		Display search results to user
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	DN: Again, about the need to authentication in order to save a search, this is possibly not true. A saved search can be represented by a URI. Who made it is not relevant.	

Use Case ID		**UC_023**
Use Case Name		**e-mail metadata record URI to third party users**
Short Description		After a search, users can also choose to e-mail the metadata set URI to other users in order to collect opinions on the quality of the metadata and possibly the data also.
Actors		End User/Data Consumer
Pre-Conditions		A search was performed and a subset of the metadata from the catalog was retrieved by the NGDS catalog.
Success End Conditions		An e-mail is sent out to a recipient with a URI to the report on the current metadata set obtained through the system
Data		URI to Metadata from the system Search information
Functions		• E-mail metadata report • Build metadata report
Main Sequence		
Step	**Actor**	**Description**
1	User	Include family of use cases <<gather data>>
2	NGDS System	Shows user a result of a data gathering activity
3	User	Selects e-mail metadata results functionality Provides e-mail recipient information
4	NGDS System	Creates a report with the metadata Sends e-mail to the address provided by user

Variants		
Step	**Actor**	**Description**
1b	NGDS System	Runs periodic subscription query Goes to step 4
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	This ought to include not only email, but RSS, Facebook, Twitter, Google+ and maybe Reddit and/or Yammer?	
2	DN: DERIVED REQUIREMENT: If you do this, you must also add in mechanisms to prevent this system from being used to spam people or abuse it in other ways.	

The end-user should also be notified if new data is published to the NGDS that satisfies a particular saved search criteria that they have elected to follows.

Use Case ID		**UC_011**
Use Case Name		**Subscribe to new data**
Short Description		The goal of this use case is to allow users to utilize saved search criteria as subscriptions to new content published in the catalog that matches specific criteria. Users will be notified via e-mail, when new data that has been input in the system, matching that subscription criteria was published
Actors		End User/Data Consumer
Pre-Conditions		In order to allow subscriptions and notifications to occur, the user must be identifiable; hence, there is a need for users (in particular the end user) to be logged in using their unique account. The search criteria used in the subscription is properly validated and saved into the system under a given name
Success End Conditions		A subscription is successfully performed
Data		saved search/subscription criteria, user e-mail obtained from user profile subscription name
Functions		• Load search criteria • Subscribe to search criteria • System notification service that periodically notifies users of matched searches
Main Sequence		
Step	**Actor**	**Description**
1	Users	Include use cases <<Load previous search criteria >> Select use search criteria as a subscription option
2	NGDS System	Fills in user e-mail information using her profile information or Saves search criteria as subscription

Variants		
Step	**Actor**	**Description**

Exceptions		
Step	**Actor**	**Description**

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	DN: Is it possible this can be done without using saved search criteria?

3.4.3 ANALYZE AND VISUALIZE DATA

Finally, once the data set is ready, it can be exported for further analysis. As shown in Figure 12, the user can export metadata to CSV file, or can further gather the data by following the URIs provided by the catalog.

Note that, data export may be the perimeter where the NGDS node stops and anything that happens beyond this boundary is out of scope of the system.

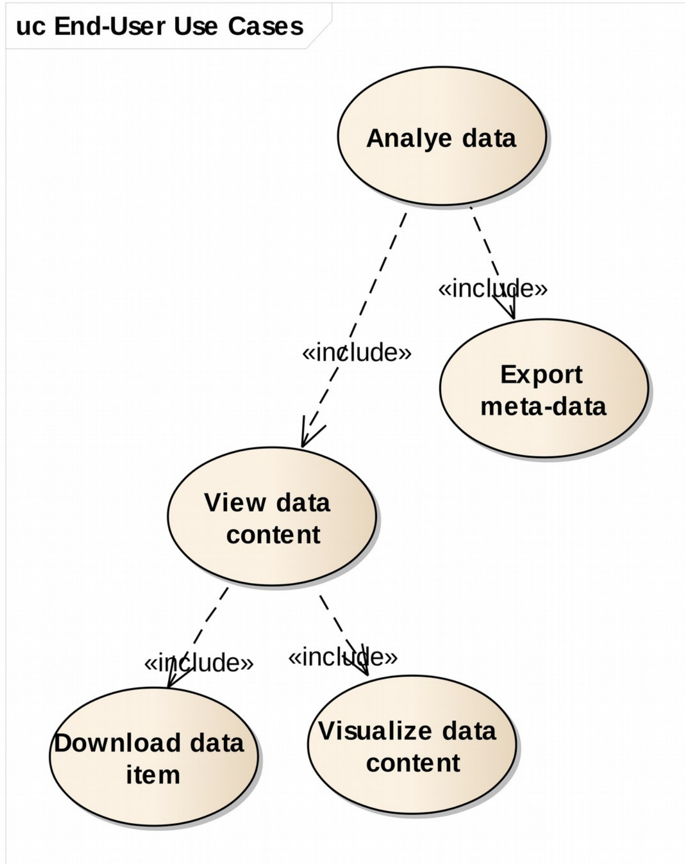


Figure 12 Data analysis

Use Case ID	UC_027
Use Case Name	**Export metadata**
Short Description	Users can export metadata records for different purposes, for example, to integrate them into their reports and spreadsheets, to further analyze these records with a CAD system, etc.
Actors	End User/Data Consumer
Pre-Conditions	Metadata was gathered and filtered
Success End Conditions	The user is able to save the metadata records
Data	Metadata records managed by GTDA
Functions	• Retrieve metadata records • Visualize metadata records

Main Sequence		
Step	**Actor**	**Description**
1	User	Selects metadata set (Include use case <<gather data>>, Include use case <<filter data>>) Indicates the type of format to export the records (we assume only CSV is supported) Initiates report generation
2	NGDS System	Responds by generating a CSV file(s) with the metadata records
3	User	Download generated metadata record files
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	If there is more than one metadata record type, how the user would like to have the report? One CSV file per metadata type?	
2	It is beyond the scope of this project the development of tools or providing support for detailed analysis of data. Hence, the need for exporting the data for further analysis.	
3	DN: Is CSV the only option here? Would it be possible that some users want JSON or XML? Would suggest yes. JSON is probably the most useful but it depends on the data models (content models). I would hate to try and represent complex binary data as CSV.	

3.5 System Administrator Use Cases

As shown in Figure 13, the NGDS administrator will supervise the operation of the whole NGDS system, including its node-in-the-box and third party repositories. The administrator will be able to register/unregister nodes in the NGDS system, manage accounts of participants, including their publication rights. They can also communicate with node-in-box administrators.

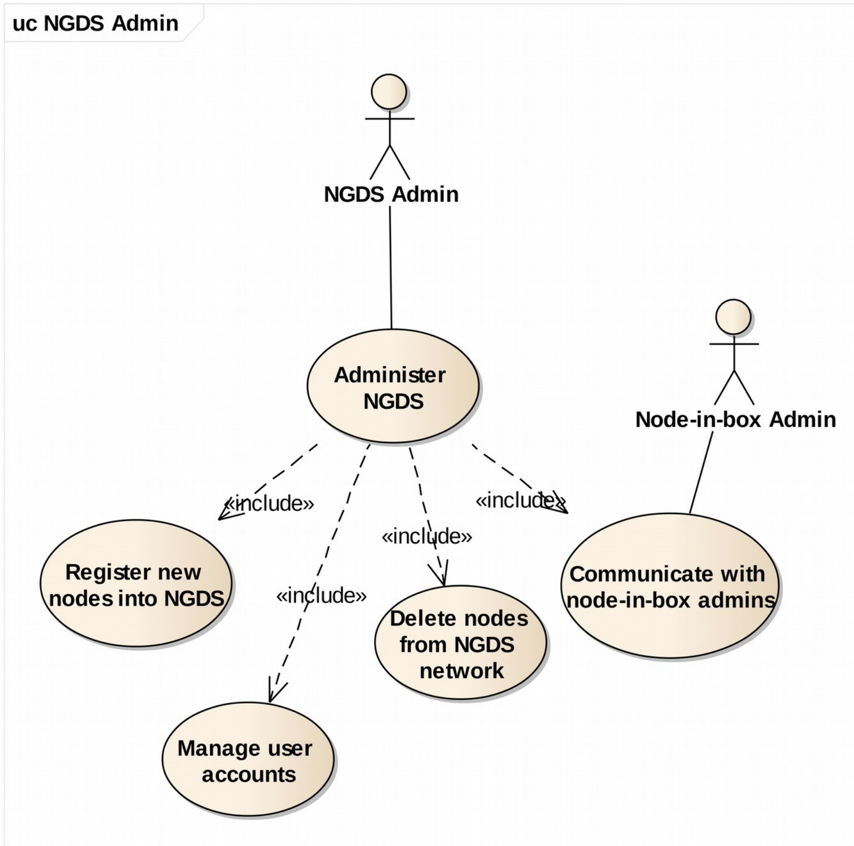


Figure 13 NGDS Administration use cases

Use Case ID	**UC_036**
Use Case Name	**Register new nodes into NGDS**
Short Description	The goal of this use case is to allow NGDS administrators to respond to new node requests. The administrator should evaluate the validity of the request, accepting, or rejecting it.
Actors	NGDS administrator
Pre-Conditions	Node-in-the box properly installed as a NGDS node management hub Client node-in the box properly installed but not yet registered The new node must have at least one unique data or metadata record not currently in the system.
Success End Conditions	A new node is registered in the network, and the data it provides becomes searchable in the by the NGDS catalog
Data	e-mails, NGDS nodes registry
Functions	• add new node • index new node

Main Sequence		
Step	**Actor**	**Description**
1	Node administrator	Include <<e-mail node-in-a-box administrator>> to NGDS administrator
2	NGDS Administrator	Receives a request for new node addition Evaluates the request for its validity If valid request, update node registry
3	NGDS System	Responds by: automatically scanning the new node and creating a metadata index for that node. Utilizes the standard communication protocols CSW.
Variants		
Step	**Actor**	**Description**
2b	NGDS Administrator	Rejects request based on external criteria
Exceptions		
Step	**Actor**	**Description**
1	NGDS System	Raise exception in case of incompatible/invalid protocols. Unless they support the common standards and protocols that they cannot join.
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID	**UC_038**
Use Case Name	**Delete nodes from NGDS network**
Short Description	The goal of this use case is to allow NGDS administrators to respond to remove previously registered nodes from the system.
Actors	NGDS administrator
Pre-Conditions	Node-in-the box properly installed as a NGDS node management hub The node to be removed is currently registered
Success End Conditions	A currently registered node is removed from the network, and the metadata records referencing this node are removed from the NGDS catalog
Data	NGDS nodes registry, NGDS catalog data
Functions	• Remove existing node • Remove index for node

Main Sequence		
Step	**Actor**	**Description**
1	NGDS administrator	Navigates to NGDS node administration page Selects node to be removed Removes selected node
2	NGDS System	Responds by: automatically removing the metadata for the removed node from NGDS index, and by removing the node registration record.
Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
1		
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1		

Use Case ID		**UC_039**
Use Case Name		**Communicate with the Node-in-a-box admin**
Short Description		The goal of this use case is to allow node-in-a-box admins and NGDS admins to communicate in the handling of administration issues such as request node removal, check node registration information, etc.
Actors		NGDS administrator, node-in-a-box administrator
Pre-Conditions		Node-in-the box is properly installed and configured Administrator has registered her e-mail information
Success End Conditions		Administrators can communicate with each other
Data		e-mails
Functions		• send email to administrator
Main Sequence		
Step	**Actor**	**Description**
1	All users	Navigates to e-mail node-in-a-box administrator page Select send e-mail to administrator Type –mail message Selects send e-mail, confirms operation
2	NGDS System	Responds by: automatically filling in the user e-mail address; sending an e-mail to the administrator

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	DN: I would rank the implementation of this use case as a low priority. This can be done by simply placing a mailto:xxx@xxx.com link on a page and letting the users own email client handle it. This seems a bit much to build into the system.	

Use Case ID		**UC_037**
Use Case Name		**Manage NGDS user accounts**
Short Description		Allows the system administrator to manage NGDS users. This will allow the system administrator to add and remove users on the administered node, and assign user roles and group membership.
Actors		NGDS administrator
Pre-Conditions		NGDS network is properly installed and configured
Success End Conditions		The administrator is able to perform the main administration operations
Data		User records
Functions		• Add user • Delete user • Modify user permissions and roles
Main Sequence		
Step	**Actor**	**Description**
1	User	Include use cases <<add user>>, <<administer user roles>>, <<delete user>>
2	NGDS System	Responds to administration operations, enforcing role and users policies.

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**

Open Issues (Please use this field to indicate questions/comments on the use case)	
ID	**Issue Description**
1	DN: Not sure if this is a use case I agree with. I think it would be better to delegate this to node administrators. If there is a user that the NGDS super admin wants removed, he or she can communicate that to the node admin perhaps? The worry is that the super NGDS admin would have to understand the metadata and data that the user is associated with before deleting to ensure no data or metadata is left orphaned. The decision is not mine but I encouraged discussion on this point.

3.6 USE CASES COMMON TO ALL USERS

Some of the use cases of the system do not belong to any of the functional groups in particular, but represent functionality that is used throughout the system.

In particular, users must login and logout the system in different situations as described in the use cases. This allows the system to track the user actions, and to automatically fill in contact information as name, e-mail, etc. these use cases are described as follows.

Use Case ID	**UC_001**
Use Case Name	**Login**
Priority [1-3] higher is more important	3
Reason for priority	User authentication is a basic activity that is a pre-condition for many other use cases. Failing to implement it is a show-stopper.
Short Description	The goal of this use case is to uniquely identify and authenticate a user, allowing the system to enforce access policies, and to use the user information to automatically fill in forms data, save searches and subscriptions, identify comments, etc.
Actors	Data Submitter, End User, Data Steward, System Administrator
Pre-Conditions	The user is logged out of the system
Success End Conditions	The user is logged in and authenticated with the system
Data	User login and password, or credentials collected in a third party authentication service
Functions	• Authenticate user using system credentials • Authenticate user using third party services, for example: Facebook, Gmail, and others

Main Sequence		
Step	**Actor**	**Description**
1	User	Navigates to the system login screen Types in login information Or utilizes one of the existing third party authentication services, eg. Gmail, Facebook.
2	NGDS System	Responds by letting user login, granting access according to her role credentials Or by denying access to the user
Variants		
Step	**Actor**	**Description**
		1. User forgets password but remembers username 2. User forgets username but remembers password 3. User forgets both username and password 4. User forgets username/password and the email they used to register.
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	DN: Password recovery tools could be added. Depending on the level of sophistication, enforcing a inimum set of standards for password might also be prudent.	

Use Case ID	**UC_002**
Use Case Name	**Logout**
Priority [1-3] higher is more important	3
Reason for priority	User authentication is a basic activity that is a pre-condition for many other use cases. Failing to implement it is a show-stopper.
Short Description	This allows a logged-in user to gracefully end and close their session
Actors	Data Submitter, End User, Data Steward, System Administrator
Pre-Conditions	The user is logged in the system
Success End Conditions	The user is logged out with no negative side effects to the system
Data	Current user section and credentials
Functions	• Logout user

Main Sequence		
Step	**Actor**	**Description**
1	User	Indicates logout action to the system or Leaves the website, forcing a timeout logout
2	NGDS System	Responds by finalizing current user session, and logging out the user

Variants		
Step	**Actor**	**Description**
Exceptions		
Step	**Actor**	**Description**
Open Issues (Please use this field to indicate questions/comments on the use case)		
ID	**Issue Description**	
1	DN: This implies session management. If that is the case, it is not noted as a requirement. Otherwise the user would not need to end a session. They could just close the browser window and walk away.	

4 Overall system quality attributes

We captured the system functional requirements in the form of use cases. For each use case, we listed a set of system functions exercised. In this section, we discuss overall system quality attributes that further describe the conditions on which the system will function.

4.1 Maintenance

In this application maintenance plays the major role among the non-functional requirements. The application will be designed such that the software stack can be changed.

In order to improve maintainability the system shall be optimized for a lean code base. This ensures that developers quickly gain an overview. In order to minimize the code base the system shall make use of third party components wherever and whenever possible. In other words "buying" is better than "making". "Buying" in our case also includes making use of open source/ freeware.

There is a trade-off between buying and making. Powerful frameworks may require complex configuration. It may be more costly to configure an off-the-shelf framework than to develop the required functionality. Well-maintained (active) framework projects undergo frequent updates that must be applied to a deployed system in order to stay up-to-date or maintain security. Documenting the configuration of the components – an important aspect with respect to maintainability – may be difficult. Therefore, for each component it must be decided if the maintainability is higher when taking an off-the-shelf product or using self-developed components that are tailored to the specific use case.

For those components that are designed within the project we require unit tests to be available that test at least the minimum functionality. In addition system integration tests are recommended but due to the high costs of designing a system test toolkit this might not be possible with the given time and resources.

NFR001 <u>All project-developed source code shall have comments at least on a per-class level.</u>

Supported Use Cases: ALL

NFR002 The System's architecture shall be documented.

Supported Use Cases: ALL

NFR003 The System's configuration parameters shall be documented.

Supported Use Cases: ALL

NFR004 The System's source code shall be covered by unit tests to at least 50% of coverage. Regression tests will be run as part of the software process.

Supported Use Cases: ALL

4.2 Usability & Accessibility

Here follows the System requirements that affect its usability.

NFR005a The System shall provide a reasonably simple to use installation tool: The tool shall install all required components (potentially with the exception for Java and/ or Python) and guide the administrator through the initial configuration steps.

NFR005 The system shall be cloud-ready: It must be available as an archived Virtual Machine (or VM) that is ready to use after a few configuration steps. Our main target is EX2(from Amazon) and other providers that are able to boot a VM image)

Supported Use Cases: ALL

NFR006 The system shall include detailed instructions that guide the user through the process of installation of one node and joining a grid.

Supported Use Cases: ALL

NFR008 The project-developed graphical user interfaces shall use a uniform look-and-feel for web applications, defined by the UX team. Minor customizations will be possible by adjusting Cascading Style Sheets, for example.

Supported Use Cases: ALL

NFR009 The project-developed applications shall provide online help explaining how to perform user-related functions.

Supported Use Cases: ALL

NFR010 The project-developed applications shall present the user with clear, understandable and accurate information explaining each task that can be performed using the software.

Supported Use Cases: ALL

NFR011 The project-developed applications shall present the user with human understandable error messages explaining the errors that occur during user interactions.

Supported Use Cases: ALL

NFR012 The key data import operations should be transactional. i.e. The user shall be able to abort operations before completion, without any negative consequences.

Supported Use Cases:

NFR013 The project-developed applications shall provide a status indicator showing the progress towards completion of user triggered processing, search queries, exports and downloads.

Supported Use Cases:

NFR013b The system shall comply with the section 508 Amendment to the Rehabilitation Act of 1973 section related to Web-based Intranet and Internet Information and Applications.

Supported Use Cases:

NFR013c The system shall comply with the ISO/TS 16071, "Ergonomics of human-system interaction – Guidance on accessibility for human-computer interfaces.

Supported Use Cases:

4.3 Performance and Scalability

Here follows the performance requirements of the System. We assume the system will be designed to support universities and data providers within USA, with a maximum of 1000 concurrent users.

Our estimates are based on the size of the Geothermal community. It is unlikely, for example, that Geothermal data will be of interest for the average citizen of the US. Besides, from the relatively small community of academic geothermal experts there will be a number of non-academic users, from either government or industry. However, the number of organizations able to exploit this data is limited.

Regarding hardware sizing guidelines: as the project goes along we will collect experience and then can also provide guiding guidelines. This will be addressed in our project plan.

NFR014 Each data provider node must be capable to maintain a list of at least 100 other NGDS nodes for harvest or distributed search.

Supported Use Cases: ALL

NFR015 Each data provider node shall indicate it has taken action in response to all user operations within 2 (two) seconds.

Supported Use Cases: ALL

NFR016 Each data provider node shall be capable of supporting up to 50 simultaneous authenticated, logged-in users.

Supported Use Cases:

NFR017 Each data provider node shall be capable of handling at least 50 (fifty) HTTP requests every 1 (one) minute.

Supported Use Cases:

NFR018 Each data provider node shall respond to every request from the NGDS in no more than 10 (ten) seconds.

Supported Use Cases:

NFR019 The System shall be able to handle the import of data files up to 2GB in size.

Supported Use Cases:

NFR020 The System shall be able to handle the import of up to 1000 data files in any one import operation.

Supported Use Cases:

NFR021 The System shall support the storing of up to 100000 data files in the import directory of each data provider.

Supported Use Cases:

NFR022 The System shall support the storing of up to 500GBs of data files in the import directory of each data provider.

Supported Use Cases:

4.4 Security

We assume that this system is mainly providing a portal to public information. Therefore, the security features of the system will focus on trustworthiness of the data being provided by through it.

We cannot discount, however, nefarious use of the email functions, spam comments being added if annotations are allowed and other typical bad behavior, for example.

The system will include the usual security aspects such as authorization, authentication, access and auditing (the 4 A's of security). The two aspects "authentication", and "access" can be easily integrated as there are very well-defined protocols and technologies available (namely HTTPSs and secure XML for access, and form-based or basic-authentication for authentication). Hence these two aspects have a binary character, i.e. they are either fulfilled or not fulfilled. Authorization and auditing can be implemented with more variability and hence they are not binary. In our case auditing is most likely of less importance. Access control must allow for distinguishing between Data Stewards, Submitters and basic users. Access controls must protect the system's administration interface against unauthorized access. Data stewards need assurance of data integrity against arbitrary modification.

Data Stewards and submitters must have access only to appropriate record collections, requiring a business process for managing access rights for certain data.

In fact security can be one of the main obstacles of the architecture since it is a vertical aspect impacting the complete architecture stack. Due to its "open nature" we therefore intend to reduce security needs to the bare minimum knowing that this will make adding security enhancements in the future more difficult.

Security also impacts the deployment process as the system administrator has to set up some initial accounts, define some user groups and adjust access rights for those user groups. We also must prevent that "default passwords" (e.g. for accessing the database) linger beyond the installation phase.

Here follows the System's requirements for security and accessibility.

NFR024 The System shall embody a security plan and process to ensure that unauthorized users are denied access.

Supported Use Cases: Log

NFR025 Valid login authentication is required for all data submitter, steward, and administrator functions.

Supported Use Cases: ALL that perform changes in data/metadata

NFR026 The System shall only allow users access to write data they have permissions to write

Supported Use Cases: ALL with write accesses

NFR027 The System shall only allow users access to download data files they have permissions to download

Supported Use Cases: ALL downloads

NFR028 The System shall only allow the data steward for a resource permission to delete it.

Supported Use Cases: UC_037

NFR029 The communication between end-users and the services of the system will be encrypted using HTTPS protocol.

Supported Use Cases: ALL communication

NFR030 Data communications between the External Systems and NGDS applications shall be secured by message authentication where applicable/necessary.

Supported Use Cases: ALL communication

NFR031 A data provider node shall only transmit data to clients when the data is published by an authenticated data submitter or steward for access by that client's permission group.

Supported Use Cases: ALL requiring authentication

NFR032 The data provider node shall maintain the integrity and availability of all data stored in its local data store.

Supported Use Cases:

NFR033 The data provider node shall maintain the integrity of all files stored in the node's local file repository.

Supported Use Cases:

NFR034 The data provider node shall maintain a log of activities for auditing purposes.

Supported Use Cases:

NFR035 NGDS applications shall be developed considering good security coding practices, thus minimizing vulnerability to attacks. In particular, it should comply with FIPS (Federal Information Processing Standards)

Supported Use Cases:

4.5 SUPPORTABILITY

Here follows the requirements that enhance the supportability and maintainability of NGDS project-developed software.

NFR036 NGDS software shall be written using the standard coding style for the used programming languages. For example, Oracle Java Coding Styles, if Java turns out to be the selected language.

Supported Use Cases: ALL

NFR038 NGDS software shall be designed utilizing the concept of encapsulation. Components shall be created that encapsulate related functionality within them, and nothing else.

Supported Use Cases: ALL

NFR039 All software shall be modular to minimize the time and complexity involved in maintaining and extending the platform and application.

Supported Use Cases: ALL

NFR040 NGDS software shall not contain any statically detectable dead code.

Supported Use Cases: ALL

4.6 DATA REQUIREMENTS

The Data Assessment Team will inventory additional data to be submitted by project subcontractors, and project management will prioritize datasets for delivery based on recommendations from the Domain Steering Committee and User-Centered design team. A prioritized list of important geothermal data types will be ascertained and used as the basis for populating NGDS data assets. See the Data Requirements Specification /P02/, for a description of data categories and attributes that will be stored in the System’s database.

Specification of metadata for any data asset provides the information necessary to enable discovery and evaluation of that asset through various search mechanisms, as well as access to the asset by people or automated processes. For Geothermal and other Geosciences data assets, the NGDS will use the US Geosciences Information Network (USGIN)[2] Metadata Recommendations for Geoscience Resources for guidance on metadata content, and the USGIN ISO metadata profile and implementation guidelines for encoding metadata content for interchange between catalog search and client applications. An NGDS metadata compilation template is available for tabular metadata compilation or setting up existing metadata for bulk loading.

A number of content models are available defining data interchange content for important geothermal data. These content models are implemented in XML as Geography Markup Language (GML). Simple features for interchange using OGC Web Feature services to enable interoperability among computing systems. Content models include a description of data attributes for geothermal types such as:

- Active Fault/Quaternary Fault
- Aqueous Chemistry
- Borehole Temperature Observation Feature
- Direct Use Feature
- Drill Stem Test Observations
- Earthquake Hypocenter
- Fault Feature
- Geologic Contact Feature
- Geologic Unit Feature
- Geothermal Area
- Geothermal Fluid Production
- Geothermal Power Plant
- Heat Flow
- Heat Pump Facility
- Lithology Interval Log Feature
- Thermal/Hot Spring Feature
- Volcanic Vents
- Well Header
- Well Log Data Compilation

These content models are intended to specify interchange formats, not database tables for data management. The models are denormalized to facilitate queries via the OGC service interfaces, and to minimize joins and data lookups by data consumers. The objective is to make data access and utilization simple. The data provider node will support upload of data in a CSV encoded table structure implementing these content models. Such uploaded files will be validated for conformance

2 http://usgin.org

with the model, then the data will be transferred to a GIS feature class and NGDS web services deployed using the data.

4.7 Design Constraints

The follow requirements are derived from design decisions that represent constraints that are mandated and must be adhered to.

NFR041 The NGDS components shall interface to NGDS data provider nodes via NGDS web services.

Supported Use Cases: ALL

NFR042 The NGDS components shall use the API provided by NGDS web services for data functions, for example WCS, WFS.

Supported Use Cases:

NFR043 The NGDS participants shall use a data abstraction layer for access to databases used for metadata management and management of data in NGDS content models.

Supported Use Cases:

NFR044 The system components shall use web services for communication with NGDS client software and other NGDS nodes.

Supported Use Cases:

NFR045 The NGDS components shall use web services to send data to NGDS client applications, including the NGDS portal application.

Supported Use Cases:

NFR046 The NGDS components shall use web services for the querying of data from NGDS nodes.

Supported Use Cases:

NFR047 The system shall provide requested metadata to the NGDS as XML files

Supported Use Cases:

NFR048 NGDS metadata catalog services shall provide metadata search results using the USIGN ISO XML encoding profile in response to CSW 2.0.2 protocol.

Supported Use Cases:

NFR049 The software shall be designed with Linux as its main target platform. However, as much as possible, it shall utilize portable technologies such as Java, that will facilitate its porting to other operating systems and platforms.

Supported Use Cases: ALL

NFR050 Under no condition should the failure of one node be capable of crippling or rendering the entire NGDS system useless. The system must be capable of adding or removing nodes while maintaining normal operations.

Supported Use Cases: ALL

4.8 Licensing Requirements

The software system being developed as part of the project shall have an open source license variant. The details of the license are being developed in coordination with DOE and project partners. The license will include provisions to allow users to copy, distribute and transmit the software, to adapt the software for other applications, and to make commercial use of the software, under the condition that the following attribution for the source of the software is included in any copy or derived work:

In order to preserve the original NGDS licensing terms, the use of third party libraries and application servers that violate these terms will be vetted.

The data provided by the data providers will not be made available to the public domain until it is released and published through the "publish" feature of the System. Other access control constraints may be applied by individual nodes at their discretion.

4.9 Applicable Standards

See the WSS **/P05/** for the details of the web services standards to be applied.

4.10 Installation and Deployment

The NGDS website will be a web application hosted by a server to be identified by project management. See the Deployment Specifications Document (DSD) **/P03/** for more details. It contains details of hardware components, operating systems, licensing agreements, etc.

The SCR development team will deliver the NGDS node-in-a-box software stack (WepAPP, NGDS Repository and NGDS catalog) to BSU project management team for deployment, testing and acceptance on a server of their choice, in accordance to the project's schedule. The NGDS node-in-a-box software will be delivered as an installable application to BSU project management for deployment, testing and acceptance on a server of their choice, in accordance to the project's schedule.

The software package will include an installation program in the form of a shell script of an apt package for Ubuntu Linux.

Appendix

5 Node-in-a-box Software Package

A redistributable, node-in-a-box, free/open-source software package will be developed (composed of an NGDS repository, an NGDS catalog and an NGDS Desktop as shown in Figure 14). This node-in-a-box software application will give data providers a simple way to register data sources, load data and expose those data as a node in the NGDS network. The software will support batch import and upload of shared datasets in supported formats adhering to standard content models. As part of the WebApp, a user interface will be provided to help users upload data to the system. The use of the NGDS data repository or the WebApp, however, is not required in order to participate as a node in the NGDS network; data providers may use whatever tools they wish to expose their data, as long as they utilize data interchange formats and web-service protocols conforming to NGDS specifications, and register themselves, as data providers, with the NDGS Catalog.

In this arrangement, different node-in-a-box instances can co-exist in the system. This requires their catalogs to be synchronized via a federation service or some sort of aggregating catalog. The mechanisms of federation, either centralized, hierarchical or peer-to-peer, for example, will be later decided, in the design stage of the project (see /P04/). A centralized aggregating catalog service is an option.

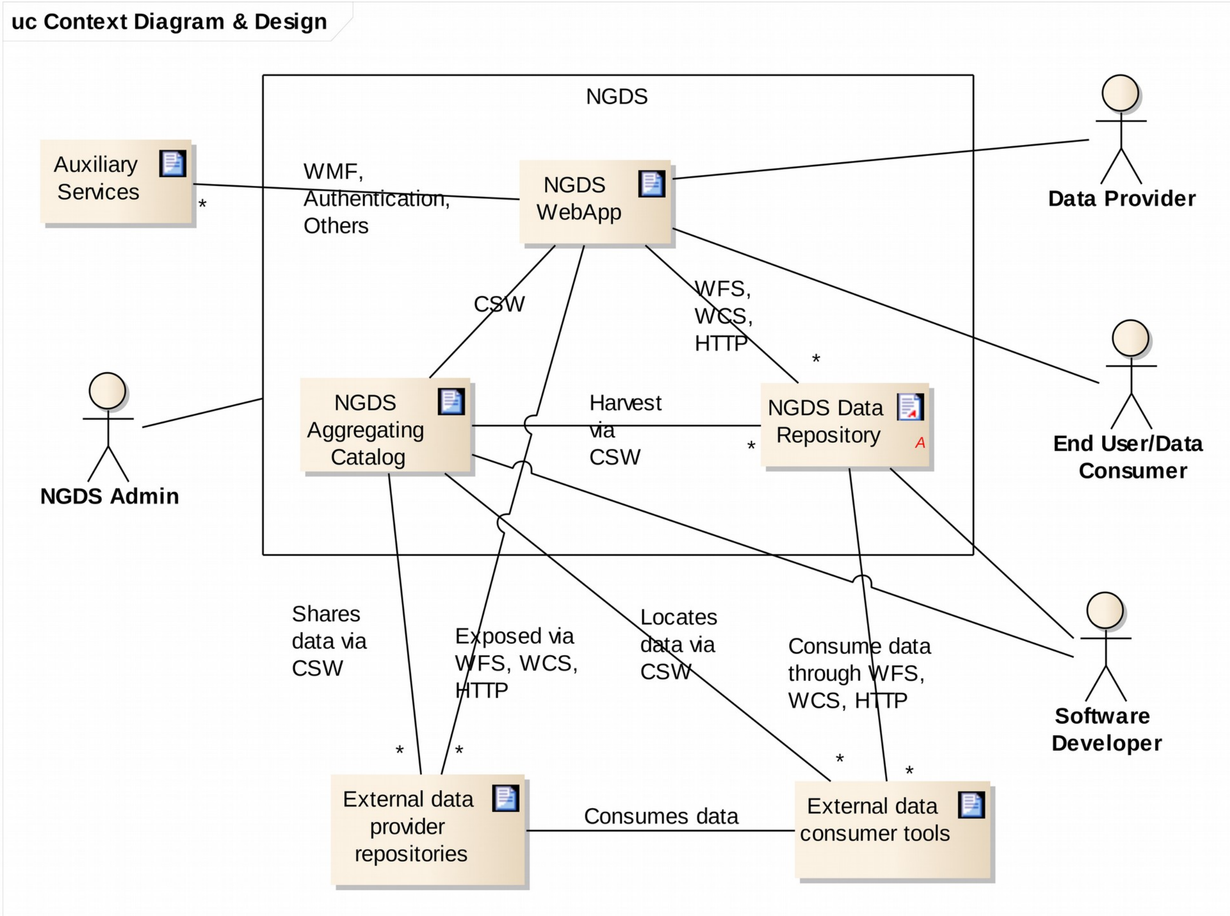


Figure 14 NGDS Data Provider Software Package

Specifically, Figure 14 shows that the NGDS will facilitate publication, visualization and discovery of geothermal data using services and applications. The diagram depicts the various major interfaces for NGDS node. The NGDS will provide a catalog, a web top UI and a data repository application. End users/Data consumers participate in the network by browsing through metadata and consuming data. Data Providers can publish metadata to NGDS catalogue service, and can use NGDS repository to store their data. Optionally, they can provide their own data sources, as long as they publicize metadata to the system in the Catalog and make data available through standard protocols and data exchange formats. NGDS System Administrator can install and manage data providers, granting them the ability to publish data to the system. Software developers produce applications that consume the data and metadata published in the system. NGDS also relies on functionality from external Web services such as authentication, and maps (via WMS protocol).

6 End-User/Data Consumer Software

End-users may interact with the system through a variety of entry points, but the project will implement two primary access points: a NGDS website and a NGDS WebApp application. As much as possible, these will be integrated in order to appear to the end-user as a single web-based experience. The primary access point to the system will be the NGDS WebApp application.

6.1 NGDS WebSite

The website will be designed to provide information about all the NGDS participants. It will serve as gateway to the system, allowing users to discover data and applications that utilize NGDS resources. The site will include information on the project's progress, NGDS specifications, the access to the map-centric search application, other software applications utilizing NGDS services, NGDS presentations, documentation and tutorials, a link to the catalog of NGDS nodes, and any other results as they become available. Note that Arizona State University is in charge of developing this Website.

6.2 NGDS WebApp

A user-friendly, web-based application will be created in order to support finding, visualizing, mapping, and acquisition of data by end-users/data consumers. This application will provide a user interface that allows end-users/data consumers to discover and access resources made available across all NGDS nodes, and to search for data across the system based on topic, location, time or other criteria. Standardized metadata describing each dataset will provide the user with the information necessary to determine the utility of that dataset for their purposes. Geographic datasets will be visualized through a map interface that will also allow users to inspect the details of individual data points (e.g. wells, temperature measurements, etc.) from properly formatted datasets. Note that the map UI is not intended for analysis and comparison of different data layers. It works only as a graphical way for searching data in the map and inspecting individual elements metadata. The WebApp will also provide a user interface for node-in-a-box data providers to publish data to their NGDS data repositories.

7 References

7.1 Project References

The following table identifies all of the references to project documents applicable to the development of this SRS document.

Reference	**Status**	**Document Name**

/P01/	Approved	System Vision
/P02/	Planned	Data Specification (to be provided by AZGS)
/P03/	Planned	Deployment Specification
/P04/	Planned	Software Design Description
/P05/	Planned	Web Services Specification
/P05/	Reviewed	Anthro-Tech End User Research Summary Report

Table 8: Project Reference Documents

Note that a number of the above referenced documents may be under development at the time of release of this document.

7.2 External References

The following table identifies all of the references to documents external to the project that are applicable to the development of this SRS document.

Reference	Document Name
/E01/	

Table 9: External Reference Documents

8 Acronyms, and Abbreviations

The following table lists the abbreviations used in this document, in order to promote their unique and unambiguous usage throughout the document and the Project.

Abbreviations	Definition
DOE	Department of Energy
NGDS	National Geothermal Data System
SDD	Software Design Description
DIS	Data Import Schema
OGC	Open Geospatial Consortium
WSS	Web Services Specification
ORM	Object-Relational Mapping
WMS	Web Map Service

WFS	Web Feature Service
CSW	Catalogue Service for the Web
WCS	Web Coverage Service
NetCDF	Network Common Data Form
API	Application Programming Interface
CSV	Comma-Separated file format
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
OAI-PMH	Open Archives Initiative – Protocol for Metadata Harvesting
UTM	Universal Transverse Mercator coordinate system

Table 10: Abbreviations

9 Glossary of Geological Terms

The following table lists the terms used in this document, in order to promote their unique and unambiguous usage throughout the document and the Project.

Terms	Definition
Geological feature	Any type of geological phenomena such as: • Faults • Fluid inclusions • Fluid rock interaction [Use for non-water fluids only, usually carbon dioxide] • Hydrothermal alteration [Same as rock-water interaction] • Fractures • Magmatism [Includes magma and magmatic processes; use more specific term “Volcanism” if applicable, although some articles address both; see also “Magma energy”] • Volcanism [Includes volcanoes and volcanic processes] • Permeability [Do not overuse; ability of fluid to flow through rock] • Pore pressure [Includes capillary pressure] • Porosity [Do not overuse; amount of pore space within rock] • Rock mechanics [Response of rocks to external forces, especially related to stress] • Seismic attenuation [Ability of rocks to absorb seismic waves] • Seismicity [Same as earthquakes; see also “Induced seismicity”] • Microseismicity [Same as microearthquakes] • Seismic velocity

	• Stress fields
	• Borehole breakouts [Bulging of wells, related to stress fields]
	• Tectonics [On a regional scale]
	• Any rock formations treated in detail
	  o Breccia
	  o Granite
	  o Greywacke [also spelled graywacke]
	  o Sedimentary basins

Data	Documents, Computer database, pictures, spreadsheets about geological features
Metadata	Summary of important characteristics of a piece of data used for the purpose of searching and discovery by users of a geological feature database.
Dataset	A group of geological data, typically related to one geological feature.
Geological survey	Gathering data using geological methods, or descriptions of regional geology; includes structure
Resources	A generic name to include data, metadata and other artifacts that are under the responsibility of a user.
Geophisical survey	Different types of descriptions of physical characteristics of regional characteristics including electromagnetic surveys, gravity surveys, magnetic, sysmic, shallow temperature surveys, etc.
Heat flow	Energy coming to the surface in a given area, often measured in watts/m^2 and used in resource assessment; derived by multiplying temperature gradients and thermal conductivity; essentially equivalent to heat flux, which is an engineering term
Temperature gradients	Change in temperature with depth, e.g., °C/m; includes temperature depth profiles
Thermal conductivity	Ability to conduct heat; e.g., salt domes have high conductivity, while sedimentary basins have low conductivity
Well logging	Well measurements and descriptions, including borehole geology
Metadata template	A predetermined set of metadata attributes used to describe data about geological features. E.g. document meta-data include a document bibliographical citation, its geo-location, ownership, and URI (for the actual document)
Landmark	Locations and regions in the globe that are geo-referenced: city, district, state, county, known geothermal sites, etc. A land mark is implemented as a dictionary, a map from names into geographical coordinates.

Table 11: Terms

10 NGDS DATA ACCESS PROTOCOLS

Software developers and the applications they develop will utilize the system through various interfaces. The OGC Catalog Service for the Web (CSW 2.0.2) will be used to enable catalog search via a Web API. Data services will be implemented using OGC WMS, WFS, NetCDF services, as well as other services adopted by the technical and steering committee as the system evolves. File-based resources will be accessed using standard HTTP GET requests.

11 DATA MODEL

In order to be made available in the NGDS system, data must be provided in predetermined formats, and must be made available in the Web through standardized protocols. The publication of data in NGDS is supported by two types of data repositories: **NGDS Data Repositories**, and **External Data Sources.**

- **NGDS Data Repositories** (or node-in-a-box) are repositories which utilize software designed as part of this project to implement standard NGDS sharing protocols. These repositories are installed and maintained by individual data providers. NGDS data repositories support the WFS and WCS and CSW protocols, supporting the catalog and discovery of data.
- **External Data Sources** are repositories maintained by third party entities which utilize any software of their choosing. Integration with the NGDS is contingent on the publication of data in the supported protocols (CSW, WFS), the utilization of supported data representation, and the listing of their data in the NGDS Aggregating Catalog.

Users of the NGDS system may want to publish data in one of three major data formats:

- **Tier 1: unstructured** — represent file based resources, unstructured data in text, image, etc; requires human to extract data for analysis.
- **Tier 2: structured, but not standardized** — represent data structured in proprietary formats that are not compatible with standard NGDS content model. Data in this tier will need to be translated and adapted to NGDS standard representations, by their owners, before they can be published and accessible in the NGDS network.
- **Tier 3: structured, standardized** -- data published in the NGDS standardized protocols and interchange formats supported by NGDS content model.

Note that NGDS, and in particular, the NGDS data repository, will only support Tier1 and Tier 3 data, following Arizona State University data formats. Tier2 data will have to be first converted to standardized data models by their respective owners, before it can be made available via NGDS.

NGDS will provide links to those data templates, at Arizona state university, or DOE websites, and will provide forms to help users input that information in the system.

Appendix 3b

Software Architecture

Siemens AG

National Geothermal Data System
Software Architecture Document (SAD)

Version: V1.1

Author: Christoph Kuhmuench

Version:	Date:	Author:	Details
0.1	05/20/2013	Kuhmuench	Creation of Template
0.3	07/16/2013	Kuhmuench/ Bruschi	Release for Review by P. Bruschi
0.4	07/22/2013	Bruschi	Review contents.
0.5	07/25/2013	Kuhmuench	Integrated comments
0.6	07/30/2013	Kuhmuench	Completing several sections and tables.
0.7	07/30/2013	Bruschi	Reviewed complete document
1.0	08/20/2013	Kuhmuench	Finishing Touches

Table of Contents

List of Figures

List of Tables

1 PREFACE

This Software Architecture Document (SAD) documents the software architecture of the National Geothermal Data System (NGDS). The document's outline is inspired by Kruchten's 4+1 View Model for Software Architecture [KRUCHTEN].

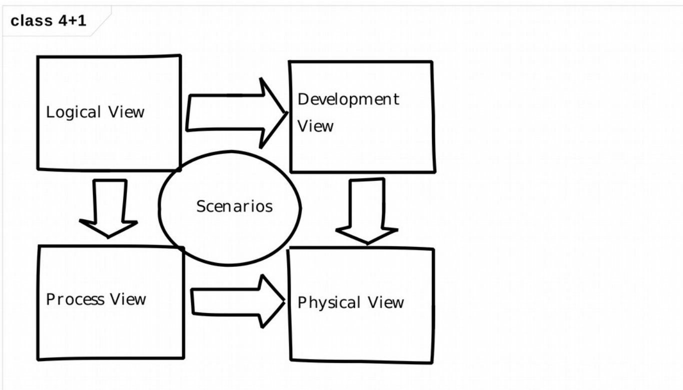


Figure 15: 4+1 View Model for Software Architectures

Kruchten's model is chosen because it is widely accepted and allows a very fine-grained tailoring to the specific needs of a project. Furthermore the model is reasonably lightweight.

The description of the NGDS architecture utilizes the document template provided by the Software Engineering Institute (SEI) [SEI]. However, this template is very complex and goes beyond the needs of a project of the size of NGDS. Also, it is hard to read because it contains many cross-references between the sections of the document and a deep hierarchy level.

Therefore the two approaches are melded together by thinning out the SEI document template and replaced sections with ideas from the 4+1 View Model.

The theme of this architecture document is to strike a balance between lightweight documentation and structured and detailed insight.

1.1 PURPOSE AND AUDIENCE

This document is intended for a technical audience who need to understand the concepts and the reasoning behind the NGDS architecture. Targeted audience includes:

- Software Architects
- Software Developers
- DoE Monitors

The purpose of the document is to clarify for the audience how NGDS works, which base components are used and what the main technical decisions were made during the design and implementation of NGDS. This helps to fulfill one of the main goals of NGDS, which is to provide a basis for a sustainable open source software project that is attractive for an open source team to maintain. With this documentation future developers will be able to quickly understand the system and become productive. The NGDS documentation also provides the basis for a future open source organization to take over architectural oversight of the software.

1.2 Responsibilities (Stakeholders)

Figure 16 depicts the stakeholder organizations of NGDS considered in the development of the architecture described by this SAD. This section lists the concerns that each stakeholder organization has that can be addressed by the information in this SAD.

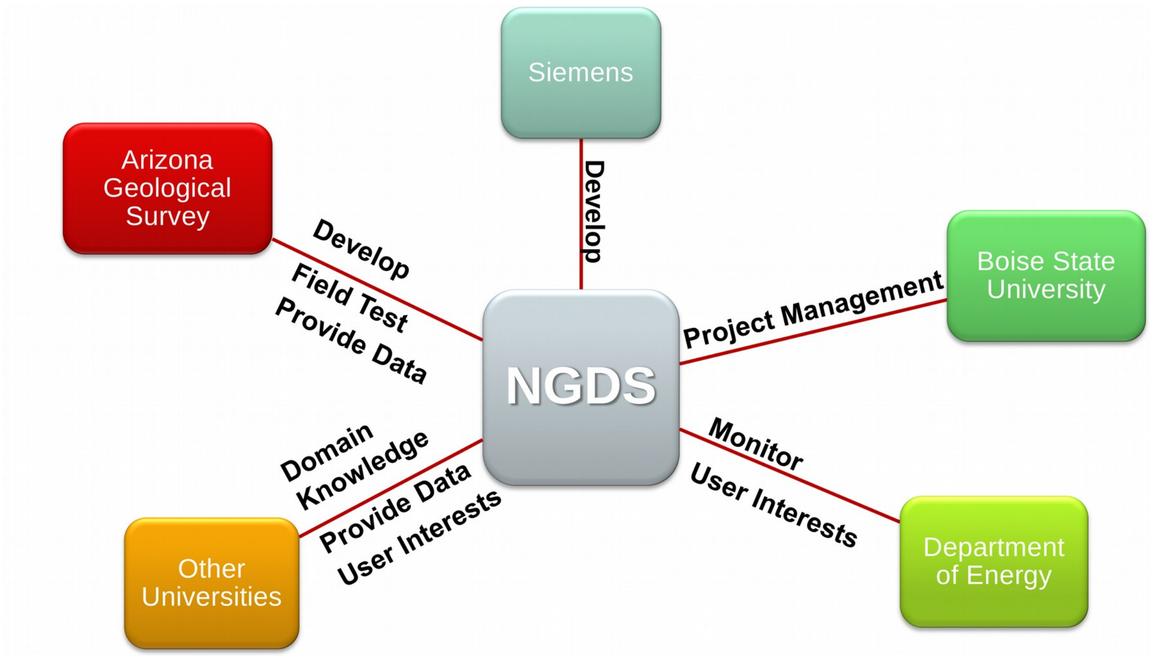


Figure 16: NGDS Stakeholders

The Department of Energy is the project sponsor of NGDS. The DOE represents the interests of all NGDS end users including commercial/ industrial/ financial end users. For the DOE it is important to have a selection of data in the system that is of interest to the various user groups and that it is made available in a simple and easy to understand way. It is also important to the DOE to ensure that the product is sustainable and lives on after the project funding is exhausted.

Boise State University has the overall project lead and is responsible to ensure that the deliverables are created and that all project partners are cooperating harmoniously. Boise's main interest is that the project creates a useful and attractive output that will be useful to the NGDS end users.

Siemens' role is to understand the requirements of NGDS, implement the system according to the requirements, and provide the documentation and installation tools. Siemens interest is to provide a functional and valuable solution that provides the expected features. Siemens' role is to research powerful technical solutions and provide an effective and efficient system that meets the customer's needs.

Arizona Geological Survey (AZGS) also contributes to the implementation of the system, and provides Siemens with domain knowledge, executes the field tests, and is a source of some of the NGDS data. AZGS is a candidate to provide NGDS as both a Node-in-a-Box and possibly act as a Central Node. Thus, AZGS is interested in the relevant installation and data upload features. AZGS wants to profit from being a core host of geothermal data. AZGS is also a candidate to play a major role in long-term maintenance of NGDS.

Other universities are providing domain knowhow, data and end user requirements. Universities will make use of the NGDS software as a Node-in-a-Box and participate in the NGDS grid of shared geothermal information. Their interest is to increase the quality and quantity of their research activities by sharing their NGDS data with other universities and organizations.

1.3 Document Roadmap

This document outlines the architecture of NGDS and is structured in the following way:

- Section 2 provides a brief overview of the system scope and background.
- Section 3 describes the most important use cases and scenarios.
- Section 4 highlights the key architectural decisions that were made during the implementation of NGDS.
- Section 5 contains five views of the NGDS Design (logical view, development view, process view, and physical view).

2 SYSTEM SCOPE AND BACKGROUND

The requirements of NGDS are discussed in detail in the requirements document [REQ2.7]. This section discusses the most important aspects of those requirements that are important for understanding what the NGDS does and its architecture.

2.1 PROBLEM BACKGROUND

Creating (electrical) energy from geothermal resources is a business with a high level of risk. Investments before exploitation can take place are high and it takes many years before they become profitable; however, geothermal energy is a renewable form of energy and therefore interesting as a long term energy resource.

It is the U.S. Department of Energy (or DOE) vision to discover and enable exploitation geothermal energy sources. The DOE and other organizations are funding a variety of research activities around this vision. So far these research activities focus on either collecting geothermal data or presenting and analyzing such data.

Existing Data Collection Activities:

Prior data collection activities focused on the creation of various repositories of geothermal data. For example:

- DOE-GDR – Geothermal Data Repository (https://gdr.openei.org),
- SMU – Southern Methodist University (http://geothermal.smu.edu),
- EGI – Energy and Geosciences Institute (http://egi.utah.edu/),
- USGS – U.S. Geological Survey (http://energy.usgs.gov/OtherEnergy/Geothermal.aspx),
- and AASG – Association of American State Geologists (http://www.stategeothermaldata.org/).

These repositories are used to collect structured data (e.g. well headers or heat flow data expressed in well formed tables) and unstructured data (mainly publications as well as other documents – which have content as varied as text, photos and hand-written text, that are usually not represented in well-known tabular formats). Due to these initiatives, a large amount of data has been made available digitally including structured data aggregated in datasets which are exposed as Web Feature Services [WFS], Web Map Services [WMS] and Web Coverage Services [WCS] while unstructured data is made available for download. These repositories are all based on individual software systems that all comply with standardized protocols (WFS, WMS, and WCS).

Moreover, each repository exposes a catalog that allows for searching within the data of the repository. The catalog has a metadata entry for each structured dataset (i.e. aggregated structured data). The metadata entry describes the content of the dataset, such as the type of data (e.g. heat flow, well log, etc.), its origin, and the geographic region covered by the dataset. The details of the metadata entry depend, at least partially, on the type of data. In the case of unstructured data, the metadata entry comprises the results of a keyword indexing service.

The catalog is made accessible via the Catalog Service for the Web (CSW) for all users without any password protection. The publication of datasets in NGDS, however, is restricted to authorized users.

Existing Presentation and Analysis of Data Activities:

Data analysis activities focus on the development of tools (such as the Geothermal Prospector http://maps.nrel.gov/gt_prospector) that allow for analyzing and visualizing data. There are also commercial tools for data analysis available.

How is NGDS going to contribute to this landscape of research project results?

NGDS fills the gap between data collection activities and data analysis activities in existing DOE funded projects. The most important new feature is NGDS's ability to harvest the catalogs of all existing geothermal repositories (specifically the SMU repository as well as the AASG repository) provided that these repositories comply with the standardized CSW and make it freely available.

Still, NGDS provides more capabilities than an aggregating CSW catalog. It also provides the future default solution for geothermal repositories and helps the DOE to build a grid of geothermal data repositories. The content of all repositories in this grid is searchable via a federated search mechanism that gives the user the possibility to execute faceted searches across all repositories, evaluate the detected datasets, and make them available for download[3]. The federation makes searches transparent to End Users/Data Consumers that have access to all data in the NGDS grid of repositories through any node of the system.

The NGDS End Users/Data Consumers are a variety of users that include legislators, federal and state agencies, financial investors, researchers, educators and students, interested public in general and industry representatives.

2.1.1 Goals and Context

The following are the basic goals of NGDS:

1. NGDS enables data providers the ability to create and administrate a repository of geothermal data.

3 Datasets are made available for download only if they are freely accessible.

2. NGDS enables End Users/Data Consumers to easily search geothermal data across a multiple of repositories of geothermal data.
3. NGDS enables End Users/Data Consumers to evaluate discovered data.
4. NGDS enables End Users/Data Consumers to select data to download.
5. NGDS in conjunction with other NGDS compatible tools enables End Users/Data Consumers to analyze downloaded data.
6. NGDS enables Software developers to extend NGDS functionality with new features.

Figure 17 depicts the basic goals of NGDS.

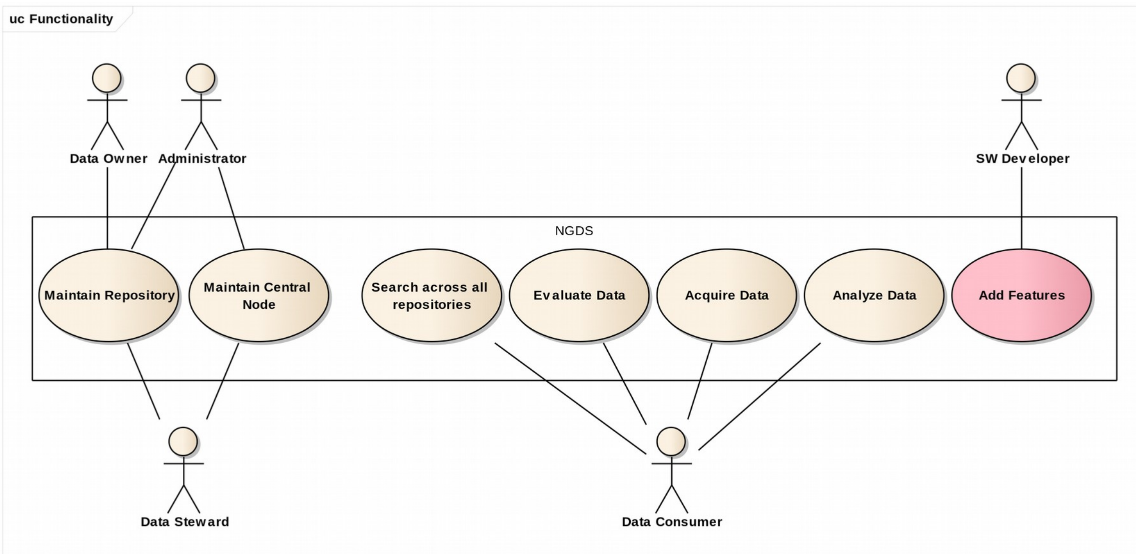


Figure 17: Basic Goals of NGDS

Need for a Standard Data Repository

As outlined above, the activities lead to a variety of software systems for archiving geothermal data. However, the DOE requires a default software system to be used for future geothermal data collection projects. This software system is referred to as a "Node-in-a-Box".

A Node-in-a-Box is a simple way for setting up a geothermal repository. The repository is simple to administrate, flexible with respect to configuration and adaptation by the data collector, and relies on standard technologies. Most importantly, it allows for federating its content as part of a grid of repositories for geothermal data.

Besides enabling data collectors to store their geothermal data, the NGDS provides a minimum set of housekeeping features for system monitoring, user management, logging of activities, support for backups, and basic security. The system has a number of basic features to adjust access rights to data within the system, distinguishing between End Users/Data Consumers (readers that can access and search for data published in the system) and data providers (writers that can provide new content and modify existing data and metadata).

Also, the NGDS Node-in-a-Box supports basic business process for uploading, evaluating and publishing datasets. This process may involve multiple users, e.g. one user responsible for uploading data, and a second user responsible for reviewing the uploaded data and making it publicly available (or rejecting it).

Search across Multiple Repositories

The NGDS allows for searching through the central catalog of datasets within the grid of federated NGDS nodes (see Figure 18). This catalog, which indexes the data across all nodes, is accessible by each participating node.

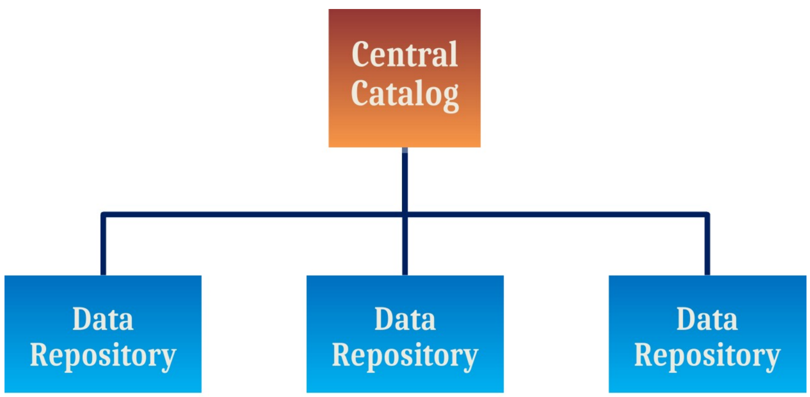


Figure 18: NGDS is a grid of repositories

End Users/Data Consumers utilize the NGDS user interface for executing faceted searches that combine many different search filters such as geographic region, type of data, data provider as well as keywords. The found results are visualized in an appropriate user interface screen. The most important aspect of the NGDS user interface is that it visualizes the found datasets in a map in an appropriate way (if found results can be geo-located and displayed appropriately). The map allows for the usual map features such as panning and zooming. Also, the user interface displays metadata for the found search results in an appropriate way and be able to sort the tables and to export them in an appropriate file format (typically CSVs).

Data Evaluation

Data analysts have the possibility to quickly review and evaluate found data sets. They may do this, for example, by relying on peer ratings of these datasets (ratings given by other users). Hence, data analysts are enabled to rate a dataset as well. They may also want to give ratings to the origin of data. Analysts are also able to triangulate regions of interest by overlaying information from multiple datasets, utilizing the geo-location associated to each dataset.

Data Acquisition

The metadata contains a URL for accessing the complete datasets. In case of unstructured data, the URL points to resources such as a PDF, TIF, JPEG or other type of file. In case of structured data, links for the supported services (WFS, WMS, and WCS) are provided allowing the complete dataset to be downloaded from the originating server. The NGDS user interface provides for downloading datasets in the appropriate form.

Data Analysis

Data Analysis is the strength of existing commercial and open source tools. The NGDS user interface provides some very basic features for analyzing data. However, it is not the goal to develop a full-fledged data analysis tool. Therefore, data analysis is limited to some very basic features such as display data in tables, and layering of different types of data as a way to improve data discovery. All other data analysis will be performed by third party tools that utilize the downloaded dataset for their particular use.

2.1.2 Significant Driving Requirements

The following table shows the architecturally significant functional requirements:

Requirement	*Description*	*Impact on Architecture*
Decentralized Repositories	Support for independent repositories that store data. Support for a centralized catalog aggregating the catalogs of the distributed repositories.	The NGDS SW architecture needs to be configurable to provide functionality for a Node-in-a-Box repository or a Central Node repository.
CSW Support	Support for importing catalogs from any repository that exports its catalog via CSW (harvesting).	The framework chosen has support for CSW import/ export. CKAN supports import while export is supported via a third party extension.
Metadata Support	Support for all metadata fields that are collected about datasets and resources. Support shall include harvesting of metadata and search of the metadata.	The framework chosen has support for configuration of metadata. CKAN allows for changing the default metadata fields.
Data Storage	Support for storing resources (files) of arbitrary type. The files are available for download.	The framework chosen has support for upload and download of arbitrary files and intelligent storage algorithm of files on the server. CKAN provides various algorithms out of the box.

OGC Support	The content of structured data files is available as WFS and WMS.	Integration of a solution for OGC services is required: The NGDS architecture uses Geoserver because it provides the reference implementation for OGC services.
Faceted Search	The UI allows for using facets to filter search results.	CKAN brings support for facets but it does not provide the configurability of facets out-of-the-box. NGDS implements this functionality as a UI extension.
Map Search	The UI allows for using a map to search for search data (geographic search).	CKAN does not support a map out-of-the box but there are various adaptations of CKAN that integrate a map. For the architecture, this only impacts the UI. Leaflet is used as a map widget.

Table 2-12: Architecturally significant functional features

The following table outlines the most architecturally significant quality features:

Requirement	*Description*	*Impact on Architecture*
Sustainability	It is expected that the product will have a life span that is much longer than this project lasts. Therefore it is important to find a way how the product can be maintained after the project funding is used up.	The system must be easy to take over and maintain by an open source organization. The architecture promotes this by focusing on popular open source frameworks and libraries as a basis. The architecture also includes nose tests as a test framework. This is popular among python projects.
Implement ability	Funding and development time are both very limited. It is therefore required to find a way how the multitude of features can be realized with maximum development efficiency. The solution should make use of many existing frameworks in order to speed up development. The choice of these frameworks and platforms depend upon the team's experience with platforms as well as feature completeness play an	Since CKAN is the base platform that provides a high percentage of the functionality. Other functionality is integrated by combining other existing open source projects such as Geoserver rather than developing these functionalities from scratch.

	important role.	
Performance	The system is able to efficiently search through hundreds of thousands of datasets. Key performance bottleneck is the ability to execute a geographic search rapidly. Another factor is the ability of the system to harvest from nodes with reasonable execution speed. NGDS is not expected to have many users in parallel (between 10 – 20 parallel users). Therefore scaling up to many users will not be an issue.	NGDS relies on PostGIS built-in capability for geographic search. This is most likely going to lead to bottlenecks. Therefore an algorithim might be integrate that uses a bucket mechanism to reduce the number of datasets that need to be searched. This mechanism might not be completed due to budget restrictions.

Table 2-13: Architecturally significant quality features

2.2 Further Reading on Requirements and Functionality

The NGDS team has created several documents that provide an insight into the required functionality.

First of all, the requirements are documented in a requirements engineering document [REQ-2.7]. In addition the team maintains an Excel Spreadsheet that outlines the high-level Features of the system [FEATURES]. It groups the features into 7 feature groups that are listed in the following table.

Feature Group	*Description*	*# Features*
Contribute Datasets	This feature group contains all features related to making datasets available via a NGDS Node-in-a-Box. It covers uploading of different types of resources either individually or in bulk mode and making data available as OGC services.	6
User Administration	This covers the standard management features for a Web application and specifically user management.	3
Harvesting	This covers the require functionality to harvest from arbitrary nodes using CSW.	2
Library Search	This covers various search features that do not rely on the map.	5
Map Search	This covers the remaining search features that do rely on the map.	10
Explore Dataset Details	This covers various features that allow to explore the content of an individual dataset	5
Configuration	This covers the features required to adapt NGDS to specific usage scenarios and updates of standards.	3
General UI	This covers features that have to do with the functionality of the UI.	3
Total		37

Table 2-14: NGDS Feature Groups

The development of the system is executed in an agile development process and this process is driven by use case stories. Several stories areaggregated into a so-called epic story. The development process and the epic stories are tracked in an Excel Spreadsheet [EPICS].

2.3 Solution Background

This section describes the rational of the chosen architecture.

The solution space for the NGDS architecture is limited by two key quality requirements:

1. Implementability
2. Sustainability

Budget and time restrictions (Implementability) alone prohibit an implementation from scratch. Moreover, a very quick study of the field proves that there are various existing frameworks or solutions available that provide a large subset of the functionality. It was therefore clear from the beginning on that a green-field approach would not be efficient. The second factor (Sustainability) requires choosing a solution that has good chances to be picked up by open source communities. Of course, chances are higher if NGDS would piggyback on an existing and successful open source project that already has a large community.

Finally, the customer's (DoE) preference for a certain framework (CKAN) had a high impact. The reason for the customer to prefer this framework is that it has been selected as the customer's preferred platform for other projects besides from NGDS.

2.3.1 *Architectural Approaches*

As mentioned above the customer has a preference for the framework CKAN for strategic reasons. However in order to ensure that the strategic requirement does not collide with the specific requirements of NGDS, CKAN and various other framework candidates were verified against NGDS key functional and non-functional requirements. This was done in a two-step approach. A list of candidate solutions and frameworks was researched and determined. Afterwards the candidates were reviewed by a team of four stakeholders (Steve Richard, Ryan Clark, Paul Bruschi, and Christoph Kuhmuench). The results of the review were collected in an Excel Spreadsheet and proved that CKAN was indeed the best fitting framework.

2.3.2 *Analysis Results*

In the following candidate frameworks and their rating with respect to functional completeness and non-functional rating are listed:

#	*Framework*	*Functional Completeness*	*Non Functional Rating*
1	CKAN	80%	75%
2	GTDA	76%	68%
3	AASG	72%	55%
4	Geonode	54%	Not rated
5	DSpace	62%	Not rated
6	DataONE	<50%	Not rated
7	Cuashi	<50%	Not rated
8	SISSweb	<50%	Not rated

Table 2-15: Rating of Framework Candidates

Although the two frameworks GTDA and AASG have a high rating, CKAN was chosen because it was more complete regarding features, and is the framework preferred by the customer. Note that the frameworks do cover large percentage of the functionality required for NGDS which is the main reason to base further development on one of those frameworks.

2.4 Product Line Reuse Considerations

NGDS is tailored towards managing arbitrary unstructured and structured data. NGDS does provide some functionality specific to the geothermal domain. However, NGDS is designed in such a way that it could be tailored for other domains that require storage and access of distributed data.

3 USE CASE SCENARIOS

This section contains several use case scenarios that are used to show the interaction between the components of the system. Use case scenarios are a good way to demonstration how requirements are fulfilled by the chosen architecture. Use case scenarios further help developers and architects to understand the dynamics of the system.

Note that this document employs use case scenarios and not use cases. In order to prevent confusion between ***use cases*** and ***use case scenarios*** our definition of both follow:

Definition Use Case:

A use case represents the actions that are required to enable or abandon a goal. A use case has multiple "paths" that can be taken by any user at any one time.

Definition Use Case Scenario:

A use case scenario is a single path through the use case.

The seven most important use cases scenarios (simply called scenarios for the remainder of this section) are described. The chosen scenarios touch many of the NGDS requirements and reflect the most frequently used functionality and the most important ones from the customer's perspective.

The scenarios cover all seven main functionality fields [FEATURES] of NGDS. The scenarios depict successful paths of a use case (e.g. only successful upload creation of a dataset) because the non-successful path is usually very simple because it only requires discarding any uploaded information.

#	*Scenario*	*Description*
1	Successful Upload of unstructured Dataset Resource	In this story the user creates a new dataset, enters all relevant metadata, adds one resource with unstructured content to the dataset, adds the remaining required metadata and publishes it. This story assumes that the user has the role Data Steward and thus has the right to create and publish the data.
2	Successful Upload Structured Dataset Resource	In this story the user creates a new dataset, enters all relevant metadata, adds one resource with structured content that follows a certain content model to the dataset, adds the remaining required metadata and publishes it. This story assumes that the resource validates and that the user has the role Data Steward and thus has the right to create and publish the data.
3	Successful Publish structured Resource as OGC Services	In this story the user selects a dataset that has at least one structured resource. The user publishes one of the structured resources as OGC services by clicking on the “Publish as OGC” Button. This story assumes that no errors occur while publishing.
4	Map Search	In this story the user marks an area on the map, enters some search keywords and clicks on the search button (or presses the “Return” key). This story assumes that datasets matching the query are found. They are displayed on the map and in the results list.
5	Keyword Library Search	In this story the user enters some keywords into the search field and presses the search button (or presses the “Return” key). This story assumes that several (more than one) results are found. The user then activates a facet in the facet UI Widget to filter the result.
6	Successful Upload Shapefile	In this story the user creates a new dataset, enters all relevant metadata, adds one shapefile as a resource that follows a certain content model to the dataset, adds the remaining required metadata and publishes it. This story assumes that that the shapefile validates and that the user has the role Data Steward and thus has the right to create and publish the data.
7	Successful Harvest Metadata	This story assumes that the user has already used the UI of the Central Node to schedule harvesting from a certain node. Therefore in this story the Linux CRON damon represents the “user”. Upon the scheduled time the cron damon triggers that the Central Node contacts the node and requests an update of its content catalog using CSW and integrates the

	returned catalog update into its own catalog.

Table 3-16: Use Case Scenarios

4 MAIN DECISIONS

4.1 ONE SOFTWARE FOR BOTH NODE-IN-A-BOX AND CENTRAL NODE

NGDS developed only one software stack that provides the functionality for both central node and Node-in-a-Box. The reason for this decision is that so many features of both node types are the same that it is easier to develop one software stack and configure the behavior of the system with a configuration file. In order to do so a new parameter to CKAN's configuration file ("development.ini") that defines the behavior of a node during startup was added. The node can either be configured as Central Node or Node-in-a-Box. As a Central Node the NGDS software provides the harvesting but no uploading capability while as a Node-in-a-Box the NGDS software provides content uploading but no harvesting capability.

4.2 OPERATING SYSTEM SUPPORT

Ubuntu 12.04 LTS was chosen as the NGDS reference platform. Due to the nature of NGDS, most users will most likely want to run it on a Linux Operating System. Ubuntu is a well-known and well-documented Linux OS. Also, CKAN is optimized for Ubuntu or other Debian-based Linux distributions.

NGDS has been developed with Ubuntu Version 12.04 LTS because it has a long support cycle. When the next LTS version becomes available the NGDS development environment will be upgraded. Installation files, etc. are written in such a way that they will likely work on any Debian-based Linux version. Therefore, porting to other Linux platforms is a minor effort. NGDS is also frequently tested and installed on the Mac OS X platform but NGDS is not test installed the system on Windows. Most likely it is possible to start the system on Windows but to fully support that would drain too many development resources and it is questionable if this form of testing adds much value.

4.3 BASE BACKEND TECHNOLOGIES

NGDS relies heavily on Python and CKAN. This defines the architecture of NGDS to a great extend because CKAN is a framework with well-defined extension points where NGDS functionality has been added. Further, CKAN is used as-is without modifying the CKAN core. Bugs are found in the CKAN core are reported and tracked through the Open Knowledge Foundation's support team and their tools.

Since CKAN is targeted towards Postgres, NGDS's reference database is also Postgres. The PostGIS extension of Postgres is used for geographic features.

In order to serve OGC services NGDS uses Geoserver which runs in its default setting on jetty.

For indexing of Metadata and full-text indexing NGDS uses SOLR (on jetty). SOLR is configured according to the CKAN recommendations.

4.4 Base Frontend Technologies

NGDS uses HTML5 and CSS3 as the base technology for the frontend. HTML5 and CSS3 are now supported by all major browsers (even later versions of Internet Explorer). In addition NGDS uses various JavaScript Libraries. Specifically, NGDS uses JQuery and various libraries based on JQuery. Furthermore, NGDS uses CSSless in order to reduce CSS complexity. At the time of writing of this document these are the state-of-the-art technologies for developing Web applications.

For the production of the HTML pages, NGDS uses the Jinja2 templating system that is built into CKAN. The templating system can be compared to PHP and allows to bring backend-information (made available via Python) into the frontend HTML content. Jinja2 is very flexible and works well in the CKAN environment. The reason for using Jinja2 is that it is the reference solution for CKAN. Building other templating systems into CKAN would be extra effort and causes problems regarding maintainability.

In many cases it is possible to create the page by either using Jinja2 templating or JavaScript. Jinja2 is preferred rather than JavaScript whenever possible because Python code and Jinja2 code is in general easier to maintain than JavaScript code.

The Leaflet Map Widget is used to present maps (http://leafletjs.com/). This widget is currently popular and has an active developer community. OpenLayers (http://openlayers.org/) was considered but due to higher experience with Leaflet among the NGDS developers, it was decided to go with Leaflet.

Naturally it is not possible to simply replace Leaflet with another map widget solution. However to the extent possible, the Leaflet-specific code is encapsulated in separate classes and modules so that the classes can be replaced in the future in case that the map widget needs to be exchanged.

5 HIGH-LEVEL CONTEXT VIEWS

This section contains the core technical information about how NGDS is designed and developed. As described in the document preface, the Kruchten's 4+1 View Model and added a set of System Context Views are used. The system context view depicts the external components that NGDS has to cooperate with and the protocols used.

5.1 SYSTEM CONTEXT VIEWS

NGDS is a grid of data repositories containing datasets of geothermal information. Each of these repositories maintains a catalog listing all of its (publicly) available datasets. This catalog must be made available via the standardized protocol "Catalog Service for the Web – CSW". The catalog must provide metadata information about each dataset according to the ISO metadata standard ISO 19115.

NGDS provides a reference implementation for such a data repository (the so called "Node in a Box") supporting CSW. However, institutes can rely on other solutions as long as they provide the catalog using CSW and according to ISO 19115.

There is at least one node in the system playing the role of the central access point that aggregates the catalogs of the data repositories. Note that it only aggregates the metadata provided in the catalog. It does not duplicate the datasets. These remain on the repositories.

The Central Node also provides a Web-based UI that allows user to search, preview, and download data. The Web-based UI can also access structured data via standardized protocols WFS and WMS and must provide viewers for these protocols.

Figure 19 depicts the interaction between Central Node and repositories.

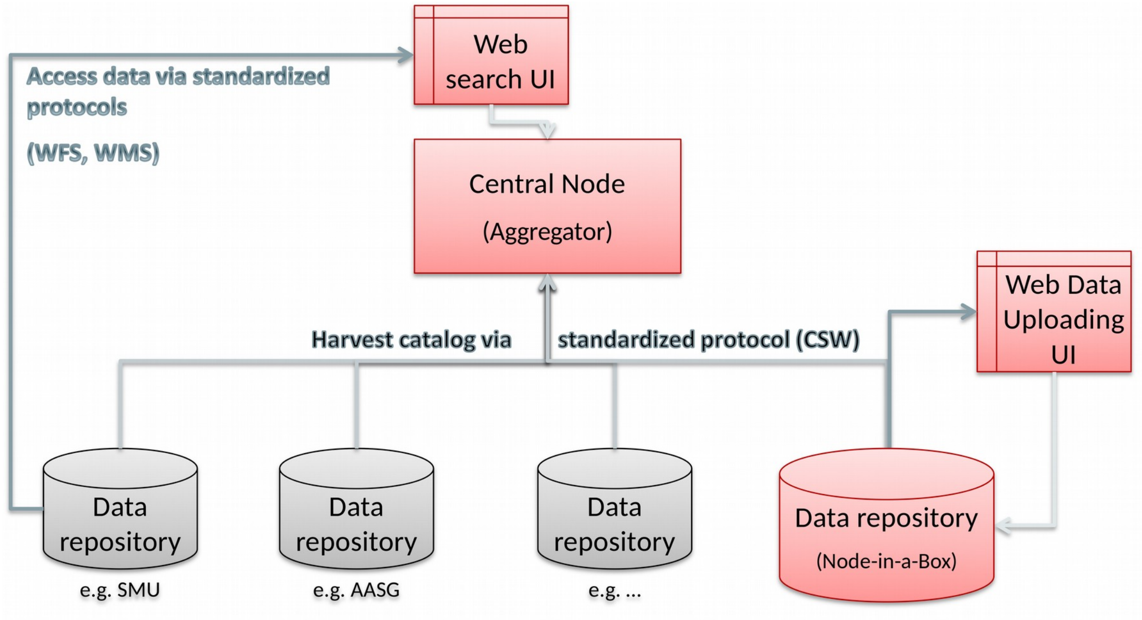


Figure 19: NGDS System Context View

5.2 Logical Views

In the NGDS documentation the logical view package contains various view angles. A domain model represents an abstract view of the NGDS domain.

The solution domain shows the high-level building blocks of the architecture. First, a static view of the high-level building blocks and then a dynamic view describe several interactions between the components according to user scenarios.

5.2.1 Domain Model

The Domain Model of NGDS can be represented as a class diagram (see Figure 20).

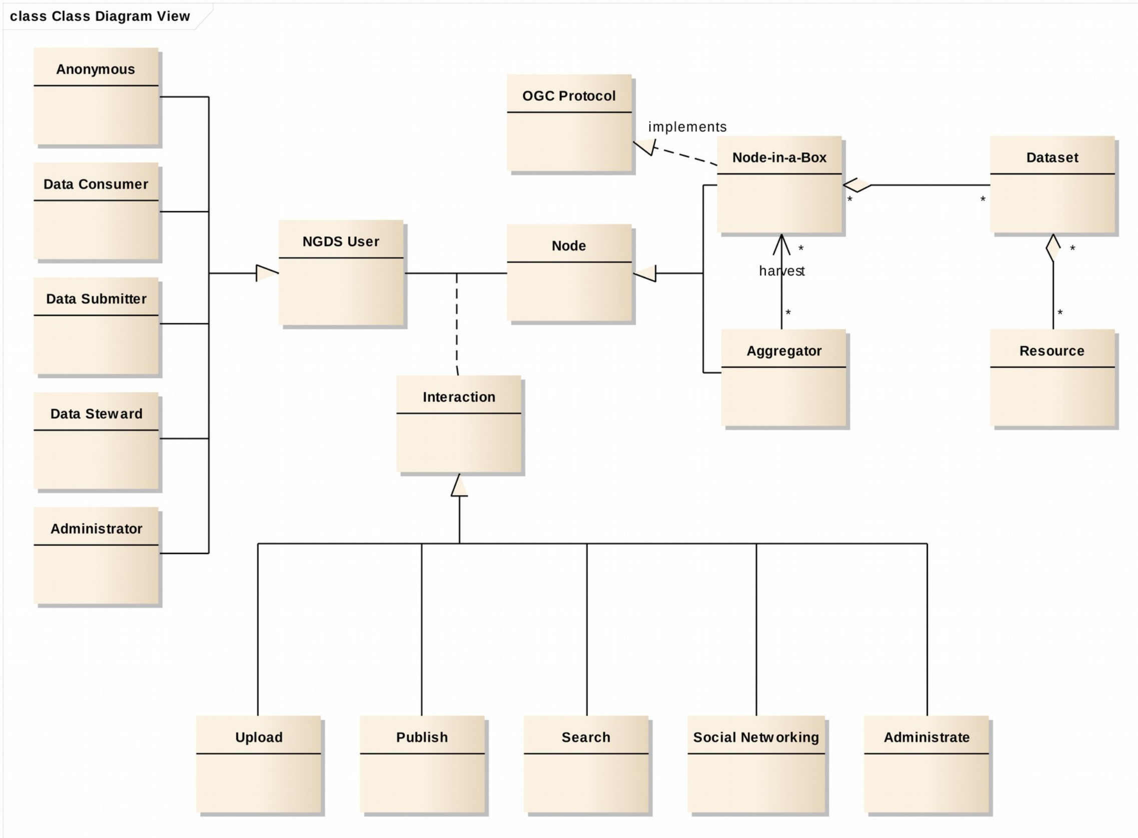


Figure 20: NGDS Domain Model as a Class Diagram

NGDS distinguishes between five user roles interacting with the system. In a similar fashion the five types of interaction between the user and the system are represented by different nodes.

Anonymous users will have limited access to NGDS use the system to search for datasets. An anonymous user has no account and therefore cannot log into NGDS and cannot be identified.

End Users/Data Consumers will make use of the search interactions and the social networking features. A Data Consumer has logged into the system and can therefore be recognized. This is required in order to allow access to the social networking features. Note that any user can sign up to NGDS and gain the privileges of this role.

Data Submitters will make use of the search, social networking, and upload interactions. A Data Submitter has the extra privilege to upload data. Note that uploaded data stays invisible to NGDS End Users/Data Consumers until a Data Steward makes them “public”. This process shall give Data Stewards the opportunity to ensure the quality of the system.

Data Stewards will make use of the search, social networking, uploading, and publishing interactions. As stated above the Data Steward is responsible for data quality.

Administrators will make use of all interactions including the system administration features.

As already outlined, the system distinguishes between two node types: repository nodes – also called “Node-in-a-Box” – store datasets and aggregating nodes – also called “Central Node” – allow for searching across all repositories.

A dataset represents a unit of geothermal information. The core of this information is stored in the resources contained in this dataset. In addition a dataset maintains metadata about this core information (e.g. location on the map, author, title of the resource, an abstract of the contained data, etc.). Essentially, this metadata shall help users to identify datasets that are interesting for their problem domain.

NGDS supports various types of resources as outlined in the following Figure 21.

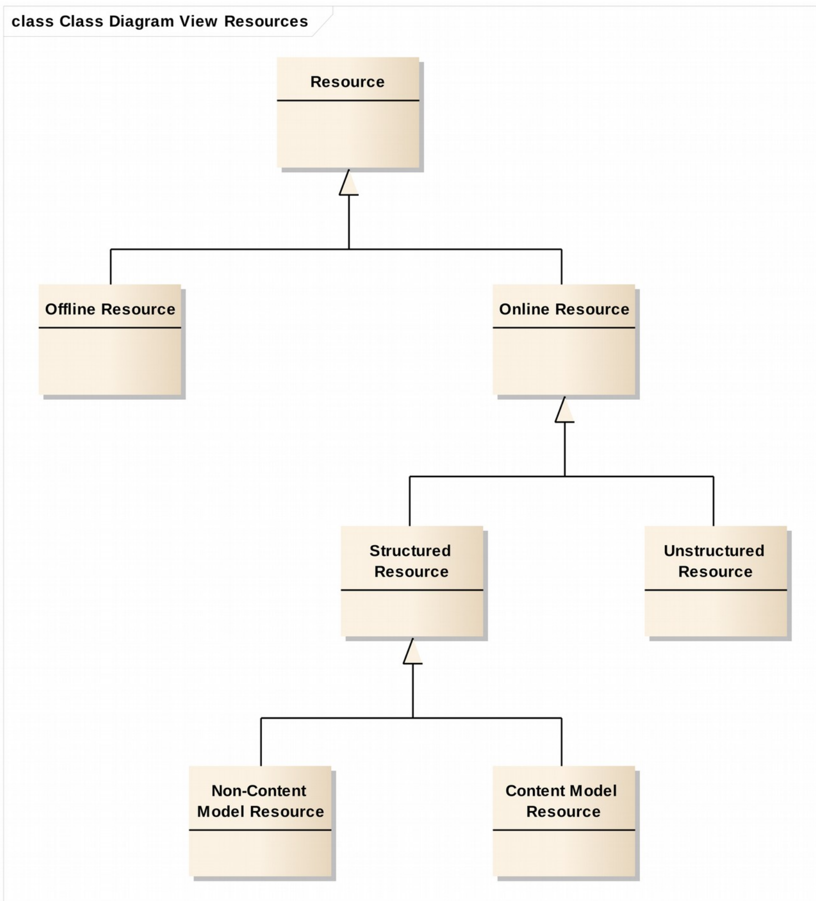


Figure 21: NGDS Domain Model of Resources as a Class Diagram

Offline Resources are entries that are non-virtual. Examples are cores, non-scanned maps and other tangible artifacts. Online resources can be in general downloaded as files or accessed as services. Structured and unstructured resources are distingushed between each other. Further, the structured resources are split into non-content model resources and content model resources. The latter ones follow a well-defined template and NGDS can validate the resource against that content model.

5.2.2 HIGH-LEVEL COMPONENTS

As discussed before NGDS is based on CKAN and therefore the system architecture is largely defined by this framework plus additional tools and components. Figure 22 outlines the high-level components. Note that not all used CKAN extensions are shown in this figure. A complete list of all extensions can be found in section 5.3.1 on page 32.

Note that components that are colored in green are existing components that may be configured for NGDS purposes. The green-colored components were not modified by the NGDS development team other than modifying configuration files to adapt them. Components that are colored in orange are new CKAN extension components developed by the NGDS team for NGDS. All components developed by the NGDS development team are realized as CKAN extensions.

Components colored in blue are base software that is not modified for NGDS.

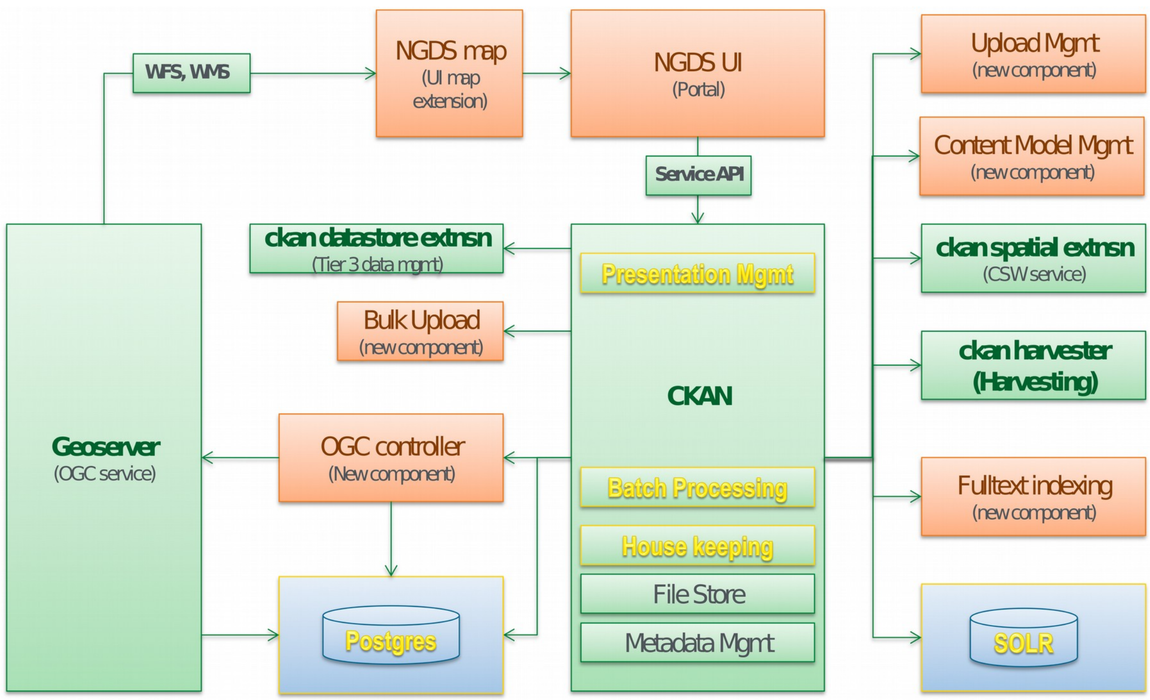


Figure 22: NGDS High-level Components

The following sections discuss the components in more detail.

5.2.2.1 CKAN

CKAN is the core framework of NGDS and provides a large percentage of the functionality. CKAN is a Python based data management system that provides a plug-in mechanism allowing for extending and tailoring it to specific needs. Besides from the core features File Storage and Metadata Management it also provides a multitude of features that are only indirectly features. Examples for those components are batch processing, housekeeping, Presentation Management and Security.

5.2.2.2 SOLR

SOLR is an indexing framework developed and maintained by Apache. Essentially it can index data and provides a search API to find data that contains that keyword. SOLR is a crucial element of CKAN. CKAN uses SOLR to index the metadata of each dataset. Essentially all search requests are forwarded from CKAN to SOLR which returns a list of the unique IDs of all datasets that have matching metadata.

SOLR is also used to do full-text indexing of uploaded text files.

5.2.2.3 Geoserver

Geoserver is the OGC's reference implementation for the standardized protocols WCS, WFS, and WMS. It is based on Java and requires a servlet container to run. NGDS is currently based on Jetty.

Geoserver essentially accesses a SQL database, and can represent arbitrary tables in this database as WCS, WFS, and WMS, provided that they provide geo-information. NGDS uses Postgres/ PostGIS as relational database with geo-coding support. Geoserver provides a UI to control which tables are made available as a service but it also provides a RESTful API. NGDS uses this RESTful API to control Geoserver from a CKAN plug-in.

5.2.2.4 Postgres/ PostGIS

Postgres plays the role of the relational database server. PostGIS is a Postgres extension that provides support for geographic features. Postgres/ PostGIS could be replaced with another relational database system provided it supports geographic features. But since Postgres/ PostGIS is the reference database system for CKAN, Postgres/ PostGIS is the reference solution for NGDS.

Postgres is used to store all Metadata. It also hosts dynamically created tables that represent the content of structured data resources (i.e. CSV files, XLS files, XML files). Essentially each column in the structured files is translated into a column in the dynamically created table.

The Postgres database also provides the input for Geoserver which computes the OGC services (WFS, WMS, and WCS) from the dynamically created tables that represent the content of the structured files.

5.2.2.5 CKAN Data-store and Data-storer Extension

The CKAN Data-store extension is an already existing extension of CKAN developed by the Open Knowledge Foundation. In short, the Data-store is the place to store structured data. It is paired with a second extension called CKAN Data-storer Extension. The Data-storer is capable of parsing arbitrary XLS, CSV, and XML files and stores their content in the Data-store. In essence it creates a new table for each table in the uploaded structured file. The table is named after the unique id of the file.

The CKAN Data-store provides a RESTful API to access the data in those dynamically created tables. Various CKAN UI extensions provide previews of the data, e.g. as searchable and sortable tables (in HTML), Graphs and Charts, or as an overlay on a map. These CKAN UI extensions make use of the API provided by the Data-store.

5.2.2.6 OGC Controller Extension

This component is developed by the NGDS team. It is realized as a CKAN extension. It is responsible for connecting the CKAN data-store with the Geoserver and thus allows a Node-in-a-Box to export structured resources as OGC services.

5.2.2.7 Upload Management

This component is developed by the NGDS team. It is realized as a CKAN extension. It is responsible for managing the additional metadata fields that NGDS requires on top of the metadata already managed by NGDS. It also provides a UI that guides the user through the process of uploading one individual file.

5.2.2.8 Bulk Upload Management

This component is developed by the NGDS team. It is realized as a CKAN extension. It is responsible for managing the upload datasets in a bulk format, i.e. it allows a user to upload multiple datasets in one step. In principle the user has to generate an Excel file containing the metadata for each resource that shall be uploaded. The all resources to be uploaded are stored in a ZIP file. A simple UI allows the user to provide both Excel and ZIP file. The system processes the uploaded content in a batch process and provides the user with information about the progress and if required error situations (e.g. incomplete metadata).

5.2.2.9 Content Model Management

This component is developed by the NGDS team. It is realized as a CKAN extension. NGDS supports geothermal-specific content models. Essentially a content model provides a template for certain data resources (in general this is an Excel file).

Content Model Management verifies the data of a resource. If the uploaded resource does not comply with the content model it will reject it and provide the user with an error message.

5.2.2.10 CKAN-Spatial Extension

The CKAN-Spatial extension is an already existing extension of CKAN developed by the Open Knowledge Foundation. The NGDS team modified the configuration of the extension so that it supports ISO metadata.

In general this extension allows for exporting the content of the NGDS metadata as a CSW service. This functionality is only required on a Node-in-a-Box.

5.2.2.11 CKAN Harvesting Extension

The CKAN-Harvesting extension is an already existing extension of CKAN developed by the Open Knowledge Foundation. The NGDS team modified the configuration of the extension so that it supports ISO metadata.

In general this extension allows for reading metadata catalogs that are either provided in the CKAN native protocol format or via CSW.

This functionality is only required on a Central Node.

5.2.2.12 CKAN Full-text Indexing Extension

This component is developed by the NGDS team. It is realized as a CKAN extension. The component provides full-text indexing functionality. It is hooked into the CKAN file upload mechanism. When a file is uploaded that is of one of the supported file formats (at the moment only text and PDF) the full-text indexing component will create a full-text index of the file content using SOLR. The content is linked the ID of the dataset so that search results will include hits into the full-text index.

5.2.2.13 NGDS UI

The NGDS UI is also hooked into the system as a CKAN extension. It makes use of the templating system (Jinja2) to create a UI according to the specifications of the NGDS design team (Siemens UX team and Anthro-Tec). The NGDS UI simply overrides the CKAN default UI by using CKAN's UI override mechanisms. In general overriding makes use of two mechanisms: First of all it overrides the NGDS default look & feel by overriding the CKAN Cascading Style Sheets and header and footer templates. In addition it also overrides the main content of those pages for which the design team has created a Look and Feel that is significantly different from the CKAN default look & feel.

5.2.2.14 CKAN Map Extension

The CKAN Map Extension is part of the NGDS UI component. It is also hooked into the system as a Jinja2 template. However, due to its dynamic nature it is heavily JavaScript oriented. The corner piece of the component is the Leaflet Map Widget (http://leafletjs.com/). It is a very intensively developed widget and provides many features. It provides the capability to display map base layers and display WMS maps on top.

5.2.3 Layered View

A layered view provides next level of detail insight to NGDS. Note that a layered view is still a high level of abstraction that may lead to misunderstandings. Figure 23 shows NGDS as a layered block diagram. If a block is placed on top of another block this means that the upper block calls the API of the lower block. Neighboring blocks may use each-others API to communicate. There should not be any communication “upwards”, with the exception of callback functions and methods as this would create circular dependencies.

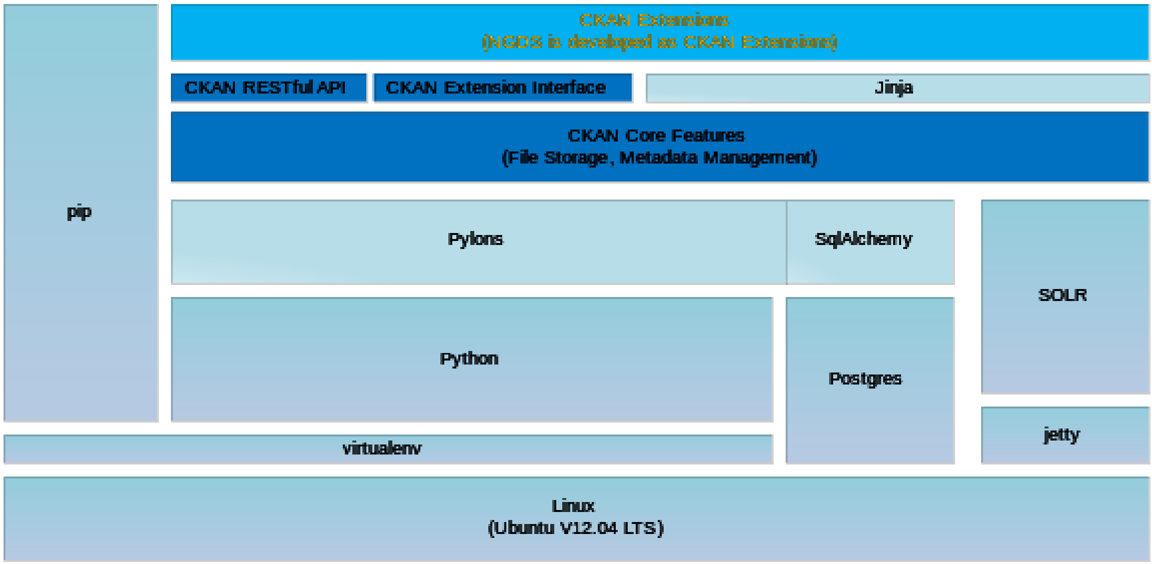


Figure 23: NGDS Layered View

The following components are added to the component diagram shown in the previous subsection:

5.2.3.1 Pylons

Pylons is the base framework of CKAN. It provides a multitude of functionality for Web Applications. Details about the framework can be found here: [http://www.pylonsproject.org/]

5.2.3.2 SqlAlchemy

SQLAlchemy is CKAN’s Object Relational Mapping Framework (ORM). It is a framework that is used in many Python projects. Details about SQLAlchemy can be found here: [http://www.sqlalchemy.org/]

5.2.3.3 Jinja2

Jinja2 is the templating system used by CKAN. It provides features that are similar to PHP and it is well-integrated with Python, i.e. it allows the transfer of data between the Python environment and the templating system. Details about Jinja2 can be found here: [http://jinja.pocoo.org/]

5.2.3.4 Virtualenv

Virtualenv allows creating a virtual Python environment. With virtualenv one can create a Python Base Directory in Linux. All files required to run the Python interpreter are copied from the default directories of the operating system into that folder. When additional libraries are installed from the internet (e.g. via PIP) they are installed into that directory and not in the default directory of the operating system. Details can be found here:

[https://pypi.python.org/pypi/virtualenv]

5.2.3.5 PIP

PIP is a package management system used to install and manage software packages written in Python. It is the default tool for managing Python libraries. Details can be found here:

[https://pypi.python.org/pypi/pip/]

5.2.3.6 CKAN, CKAN RESTful API, and CKAN Extension Interface

The components CKAN Core, CKAN RESTful API, and CKAN Extension Interface comprise the CKAN system. The CKAN RESTful API and the Jinja2 templating System can be used to develop a new API on top of CKAN. The CKAN Extension Interface on the other hand provides the means to develop extensions in Python that extend the functionality of CKAN. It also allows for hooking into the existing CKAN API, e.g. by adding additional code into the CKAN file upload mechanism and other core features.

Details can be found here:

[http://CKAN.org]

5.2.3.7 CKAN Extensions

The CKAN extensions make use of the three mechanisms CKAN RESTful API, CKAN Extension Interface, and Jinja2, in order to provide additional functionality. There are already many extensions available that are maintained by either the OKFN or other organizations. Some extensions are installed with the CKAN core package; others need to be installed separately. NGDS makes use of the following CKAN extensions:

Extension	*Description*	*URL*
CKAN-ext-data-store (data-store)	The data-store allows for storing and accessing structured data via a RESTful API. It is part of the CKAN installation package	http://CKAN.org
CKAN-ext-data-store r (data-storer)	The data-storer stores structured data in the data-store. It reads structured files (CSV, XLS) and stores them in the data-store.	https://github.com/okfn/CKANext-data-storer
Statistics (stats)	CKAN's stats extension analyzes your CKAN database and displays several	http://docs.CKAN.org/en/latest/stats.html

	tables and graphs with statistics about your site, including: • Total number of datasets • Dataset revisions per week • Top-rated datasets • Most-edited Datasets • Largest groups • Top tags • Users owning most datasets	
Data Viewer (json_preview, resource_proxy, recline_preview)	CKAN provides an API to create previewers for data resources. It also provides a number of plug-ins for structured data. PDF, text and images are simply presented within an iFrame. For structured data it provides a previewer based on recline.js. The plug-ins recline_preview, resource_preview, json_preview, and resource_proxy are part of the Data Viewer.	http://docs.CKAN.org/en/769-docs-reorg/data-viewer.html
Geospatial Capabilities (spatial_metadata, spatial_query)	CKAN offers a set of geospatial features that allow adding spatial information to your datasets. • Spatial_metadata: Associate geographic data with datasets, with automatic geo-indexing of datasets. • spatial_query: Search for datasets using spatial queries. • Wms_preview:	http://docs.CKAN.org/en/CKAN-1.7.2/geospatial.html
CKANext-spatial (wms_preview)	This extension contains plug-ins that add geospatial capabilities to CKAN. It contains a large number of plug-ins of which only a small subset are used. • Wms_preview: provides a previewer for WMS resources.	https://github.com/okfn/CKANext-spatial

Table 5-17: Existing CKAN Extensions used in NGDS

The following table lists the extensions developed by the NGDS team. The last column maps the extension to the component described in the logical view. In general one component from the logical view in Figure 22 comprises of one or more CKAN extensions.

Extension	*Description*	*Contributes to Components*
OGC-controller (Geoserver)	Manages the connection to Geoserver and makes WMS, WFS, and WCS services available.	OGC-controller
NGDS Metadata (metadata)	Handles the NGDS-specific Metadata	Upload Mgmt
NGDS Harvesting (ngdsharvest)	Handles the harvesting.	CKAN Harvester
ngdsui	Serves the UI. Currently also includes other features (planned to be refactored)	Ngds map, ngds ui, upload mgmt., content model mgmt., full-text indexing

Table 5-18: New CKAN Extensions developed by the NGDS team

5.3 Development Views

5.3.1 Component View of CKAN

This section describes some of features and components of CKAN. The API of the system is well-documented and can be found here: http://CKAN.org.

CKAN is based on a multitude of Python frameworks and libraries. The glue between all components is provided by Pylons (http://www.pylonsproject.org/) which by itself again consists of a multitude of frameworks and libraries. Essentially Pylons provides the base functionality to create a Web application by developing extension, hooked into the core system. It makes use of SQL Alchemist for Object-Relational Mapping (ORM). It also provides all base functionality to create a Web Application (Session Management, Security, Forms Management, Templating Language, etc.).

Figure 24 outlines the most important Pylons libraries. Please refer to [PYLONS] for more details about the structure and components of Pylons.

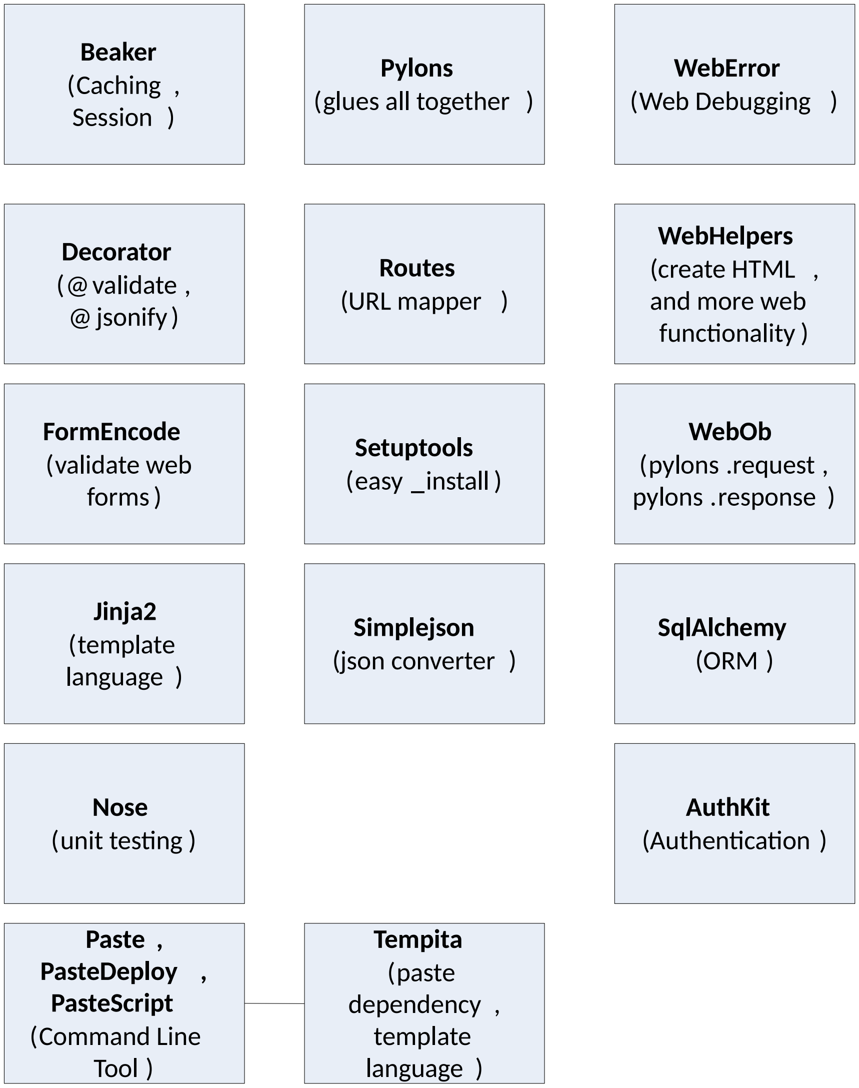


Figure 24: Pylons Base Libraries

Figure 25 outlines the most important components of CKAN.

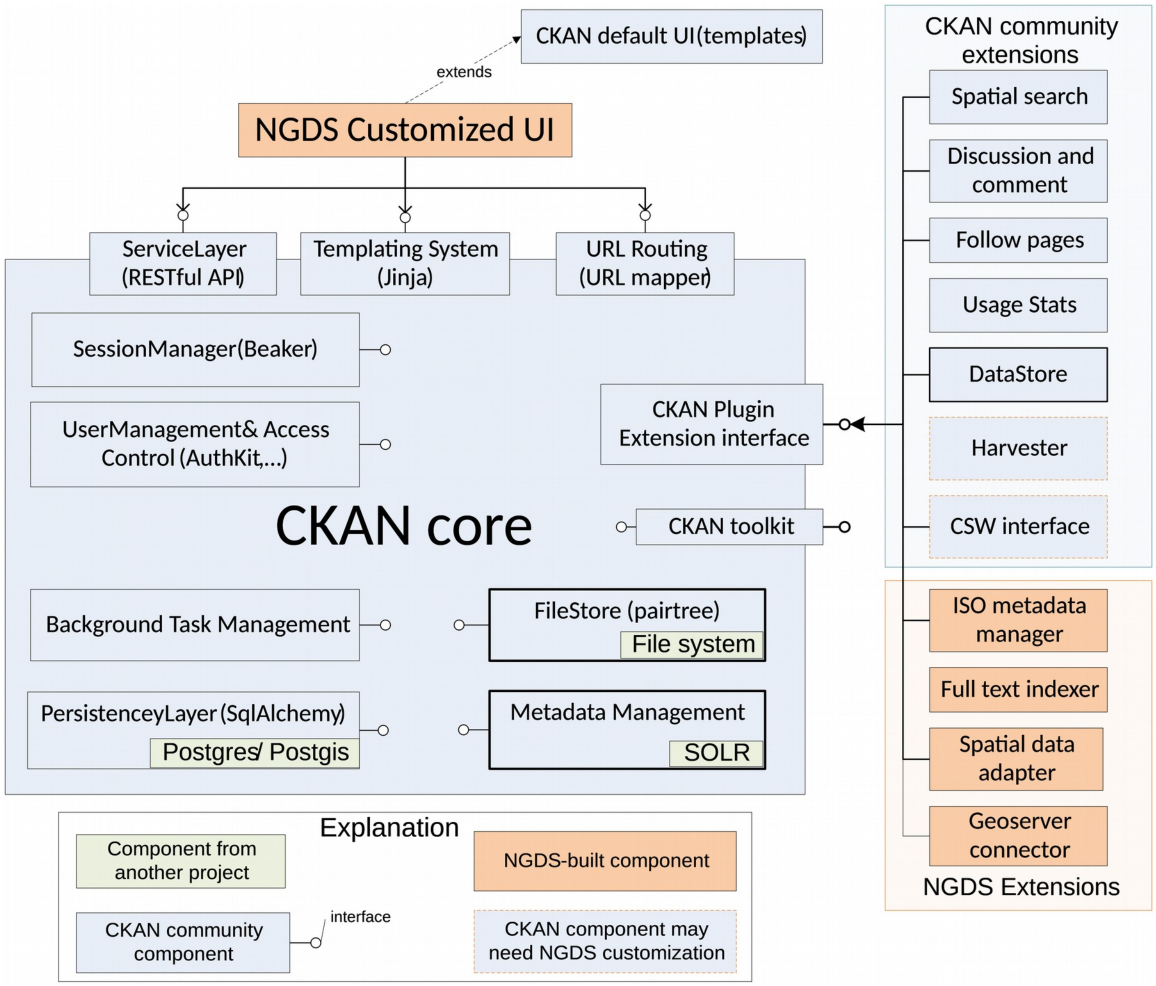


Figure 25: CKAN Component View

Please refer to the CKAN documentation [http://www.CKAN.org] for more information about CKAN.

5.3.2 GITHUB CODE OVERVIEW

NGDS maintains it software in github.com. In the following the structure of the NGDS github presence is outlined:

Repository	*Description*
ckanext-ngds	Most important of all NGDS repositories on github. Contains the source code for the NGDS extensions.
dev-info	Contains information for developer. Also contains the installation scripts.
documents	Contains development documents such as sprint plans, requirements documents, and design documents.
ckan-dev	Copy of the CKAN repository. It is intended to capture any NGDS modifications of CKAN. Currently, there are no modifications.
ckanext-spatial-dev	Cop of the CKANext-spatial repository. Is intended to capture any NGDS modifications to this plug-in. Currently, there are no modifications.
ckan	Contains a fork of the CKAN project.
ckanext-spatial	Contains a fork of the CKAN spatial extension project.

Table 5-19: NGDS Repositories on http://github.com

The NGDS source code can be found in two of the repositories mentioned above:

3 *ckanext-ngds:* This repository contains all code written by the NGDS development team. All code is written in a single CKAN extension that provides various plug-ins (see Table 5 -18).

4 *ckan-dev:* This repository contains a copy of the CKAN source code. It is designed to also hold any modifications that NGDS team needs to apply to CKAN. However, up to now there are no modifications and the repository contains the code of CKAN version 2.0.

5 *ckanext-spatial:* This repository contains a copy of the CKAN spatial extension source code. It is designed to also hold any modifications that NGDS team needs to apply to CKAN. However, up to now there are no modifications.

The most interesting of the repositories to understand is the one in ckanext-ngds. The screenshot shown in Figure 26 outlines the structure of the repository.

The repository follows the standard approach for CKAN extensions. The majority of the code (i.e. the ngds plug-ins can be found in the directory CKANext. The directory i18n contains code and data for internationalization. The installation directory simply contains the word file that describes the installation procedure for NGDS. In the directory sample-data a number of geothermal resources can be found that can be used to test NGDS content model functionality. The scripts directory contains a number of configuration scripts. Among the files in the directory setup.py shall be highlighted. It follows the standard approach for CKAN extensions and provides information about all plug-ins provided with this CKAN extension. The two test files are used to execute tests on the command line.

i18n	Updated for internationalization	5 months ago
installation	Adding a version of the installation document that is still in progress.	2 days ago
sample-data	sample data	9 days ago
scripts	update config script	2 days ago
.gitignore	testing package	a month ago
MANIFEST.in	Updated for internationalization	5 months ago
__init__.py	refactor to share base model, base controller, and a single database …	6 months ago
ckanclient.cfg	updated for fixes	2 months ago
facet-config.json	updated for faceted search	2 months ago
readme.md	Updated readme.md	5 months ago
setup.cfg	Updated for internationalization	5 months ago
setup.py	Some reformatting.	4 months ago
test-core.ini	Updated readme.md	5 months ago
test.sh	Update test.sh	5 months ago

Figure 26: Content of the base directory of the repository ckanext-ngds

5.4 System Dynamics Views

This section describes the major NGDS use cases and their dynamic aspects. The description uses message sequence charts to show which components are involved and how they interact.

5.4.1 *Use Case: "Upload unstructured Dataset Resource"*

In this use case the user creates a new dataset and adds an unstructured resource to this dataset. The task of the NGDS is to store the resource and associate it with the metadata. It also needs to update the search index so that the dataset can be found using the search mechanisms (Map Search, Keyword Search, and Faceted Search). Figure 27 shows the interaction between the components.

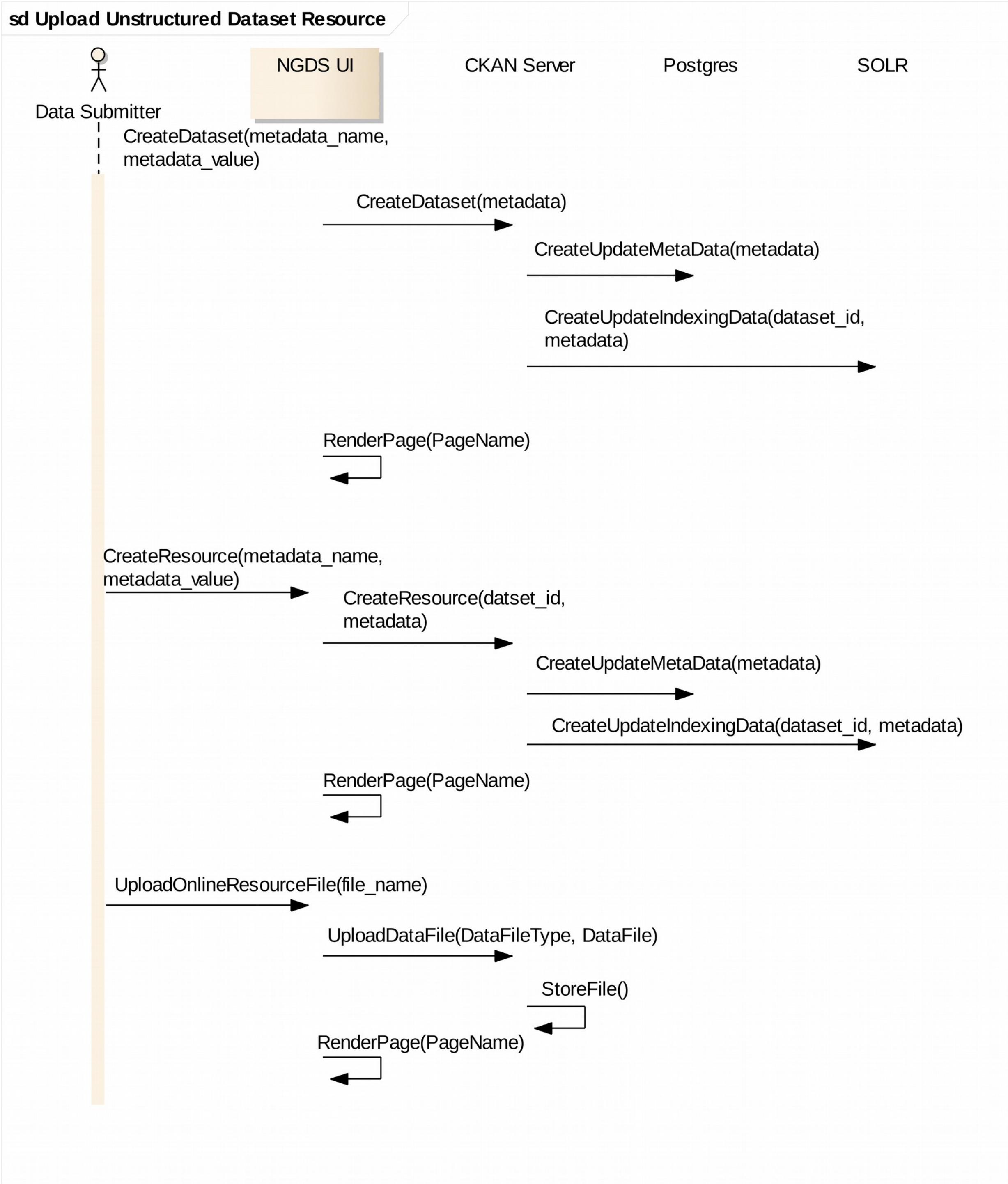


Figure 27: MSC for Use Case "Upload Dataset with Unstructured Resource"

The user enters the relevant metadata into the CKAN UI which is then transferred to the CKAN core system. The CKAN core creates a unique ID for the dataset and stores the metadata information in the Postgres SQL database. Then it forwards the Metadata to SOLR for indexing. SOLR creates an index and assigns it to the dataset ID. When the user uploads the unstructured dataset it will be stored in the Linux file system. Note that the metadata for the resource is also stored in Postgres and indexed by SOLR. This step is omitted in the figure above to simplify visualization.

5.4.2 Use Case: "Upload Structured Dataset Resource"

In this use case the user creates a new dataset and adds a structured resource to this dataset. The task of the NGDS is to store the resource and associate it with the metadata. It also needs to update the search index so that the dataset can be found using the search mechanisms (Map Search, Keyword Search, and Faceted Search). Furthermore it needs to verify the resource against its content model template and store the data in the CKAN data-store for retrieval and statistical analysis.

Figure 28 shows the interaction between the components.

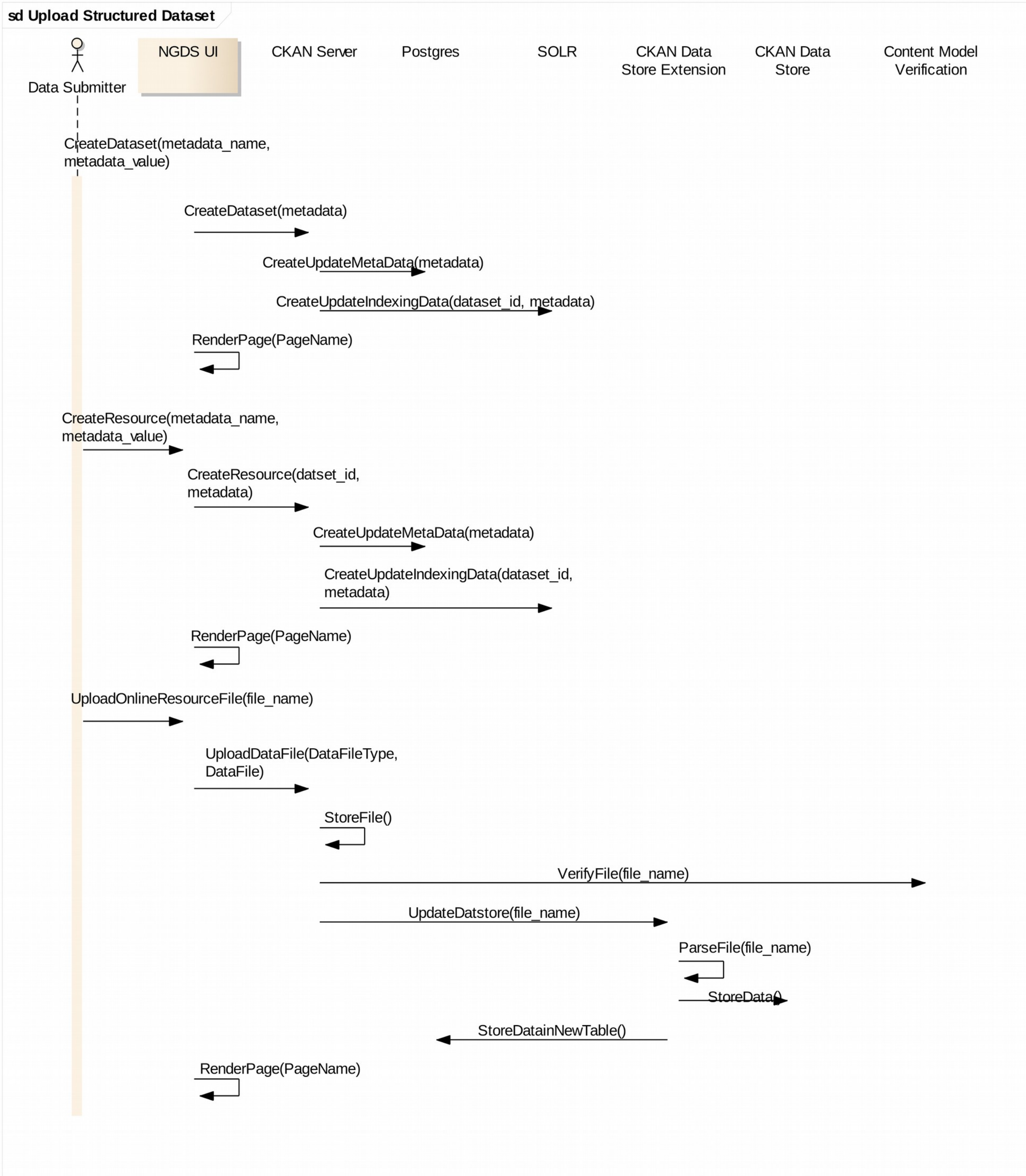


Figure 28: MSC for Use Case "Upload Dataset with Structured Resource"

In many aspects the use case causes the same interactions as uploading unstructured resources. The difference here is that Content Model Verification is involved to check the consistency of the uploaded file. Furthermore the CKAN data-store is involved to create a new table containing the structured data. The reason for adding the data to the CKAN data-store is that it can be statistically analyzed and also previewed using CKAN's built-in previewers for structured data. Also, uploading the data to the CKAN data-store is a prerequisite to make the resource available as OGC services.

5.4.3 *Use Case: "Publish Structured Resource as OGC Services"*

In this use case the user wishes to make a resource of a dataset available as OGC services (i.e. as WFS, WMS, and WCS services). Since it is a structured resource it is stored in the CKAN data-store, i.e. a table containing the content of the resource exists. It is the task of the NGDS software to add a new column of type PostGIS/GeoShape to that table and calculate its value for each row. It then needs to inform Geoserver about this table so that Geoserver starts serving it as WFS, WMS, and WCS.

Figure 29 shows the interaction between the components.

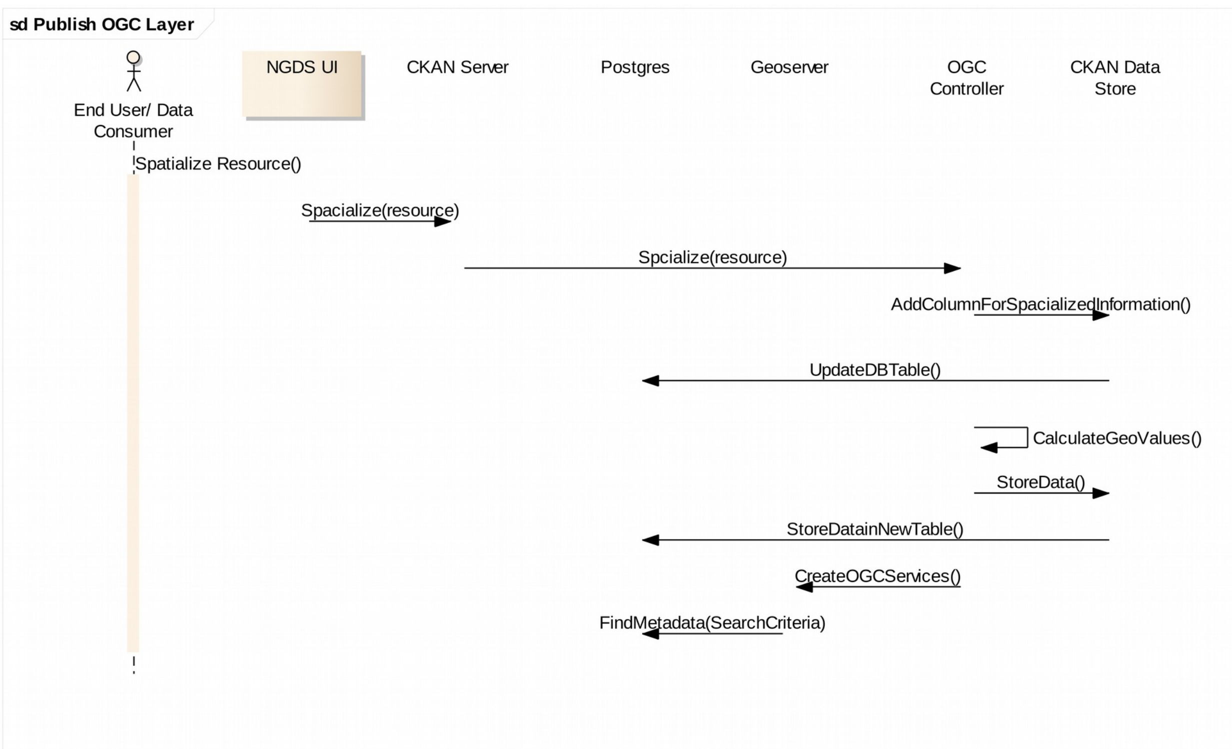


Figure 29: MSC for Use Case " Publish structured resource as OGC Services"

The user triggers the process by pressing the "Spatialize" button in the NGDS UI on the dataset details page. The UI calls a RESTful interface which CKAN routes to the OGC Controller Plug-in. The plug-in then makes use of the CKAN data-store API which in turn adds the new column to the table associated with the resource. Then the OGC controller plug-in calculates the values for the new column and updates the table again indirectly by making use of the CKAN data-store API.

In the next step the OGC Controller plug-in calls Geoserver's RESTul API. Geoserver then publishes the relevant OGC services using the table associated with the resource. The newly created column is used for geo-referencing.

5.4.4 Use Case: "Map Search"

In this use case the user wishes to search for datasets using the map. After marking an area on the map and entering search keywords the user expects to receive a list of matching datasets and a representation of these datasets on the map.

It is the task of the NGDS software to execute the geographic and the keyword search and render the information.

Figure 30 shows the interaction between the components.

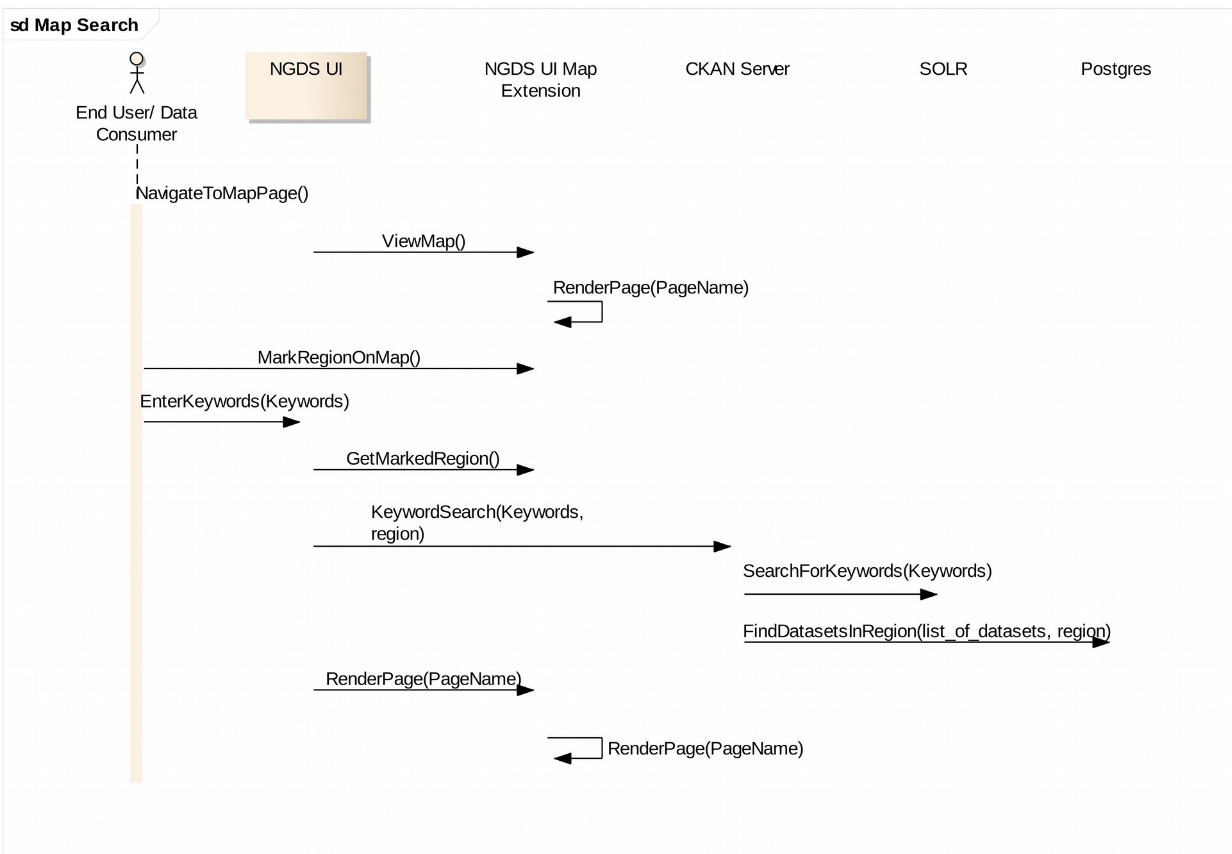


Figure 30: MSC for Use Case "Map Search"

The user initiates map search by navigating to the map search. The NGDS UI will trigger that the map is being rendered by the NGDS UI Map Extension. The user then triggers the map search by marking the search area on the map and by entering keywords into the search bar. When the user presses the search button the NGDS UI retrieves the marked area from the NGDS UI Map Extension and sends the information together with the search keywords to CKAN using CKAN's default search API. CKAN then uses SOLR to retrieve a list of IDs of all datasets matching the search criteria. It then contact Postgres and uses the PostGIS extension for Postgres to execute a geographic search for all datasets that are overlapping with the marked area. The metadata of all datasets fulfilling this criterion are returned to the UI for rendering buy the NGDS UI Map Extension.

5.4.5 USE CASE: "KEYWORD LIBRARY SEARCH"

In this use case the user wishes to search for datasets using keywords and facets. After entering the keywords and activating the relevant facets the user expects to see a list of all datasets fitting the search criteria.

It is the task of the NGDS software to execute the keyword search, filter according to the facets and render the information.

Figure 31 shows the interaction between the components.

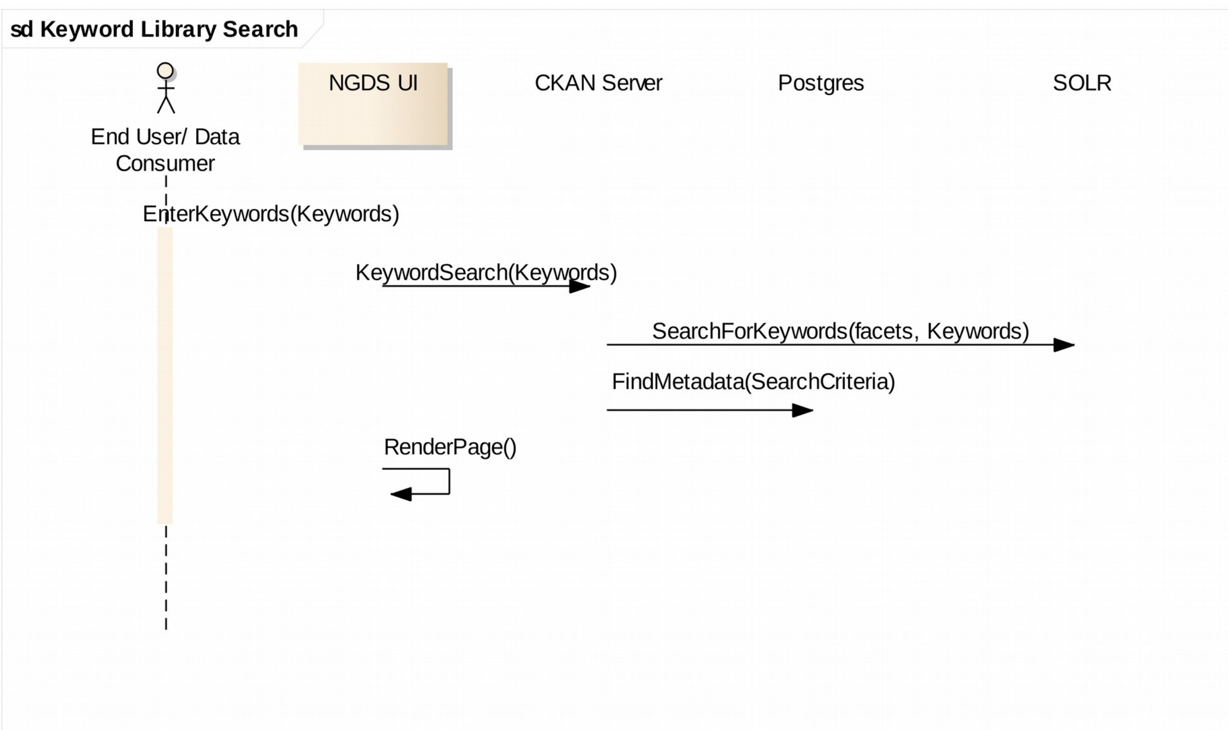


Figure 31: MSC for Use Case "Keyword Search"

The Faceted Search dynamics simply require that the CKAN server is using SOLR to find the IDs of all datasets that fit the search criteria. It then retrieves the metadata of the fitting datasets from Postgres and returns the information to the NGDS UI for rendering.

5.4.6 Use Case: "Upload Shapefile"

In this use case the user creates a new dataset and adds a Shapefile resource to this dataset. The task of the NGDS is to store the resource and associate it with the metadata. It also needs to update the search index so that the dataset can be found using the search mechanisms (Map Search, Keyword Search, Faceted Search). Before it can verify the Shapefile against a content model and add it to the CKAN data-store it first needs to unpack it so that the individual files within the Shapefile can be processed.

Figure 32 shows the interaction between the components.

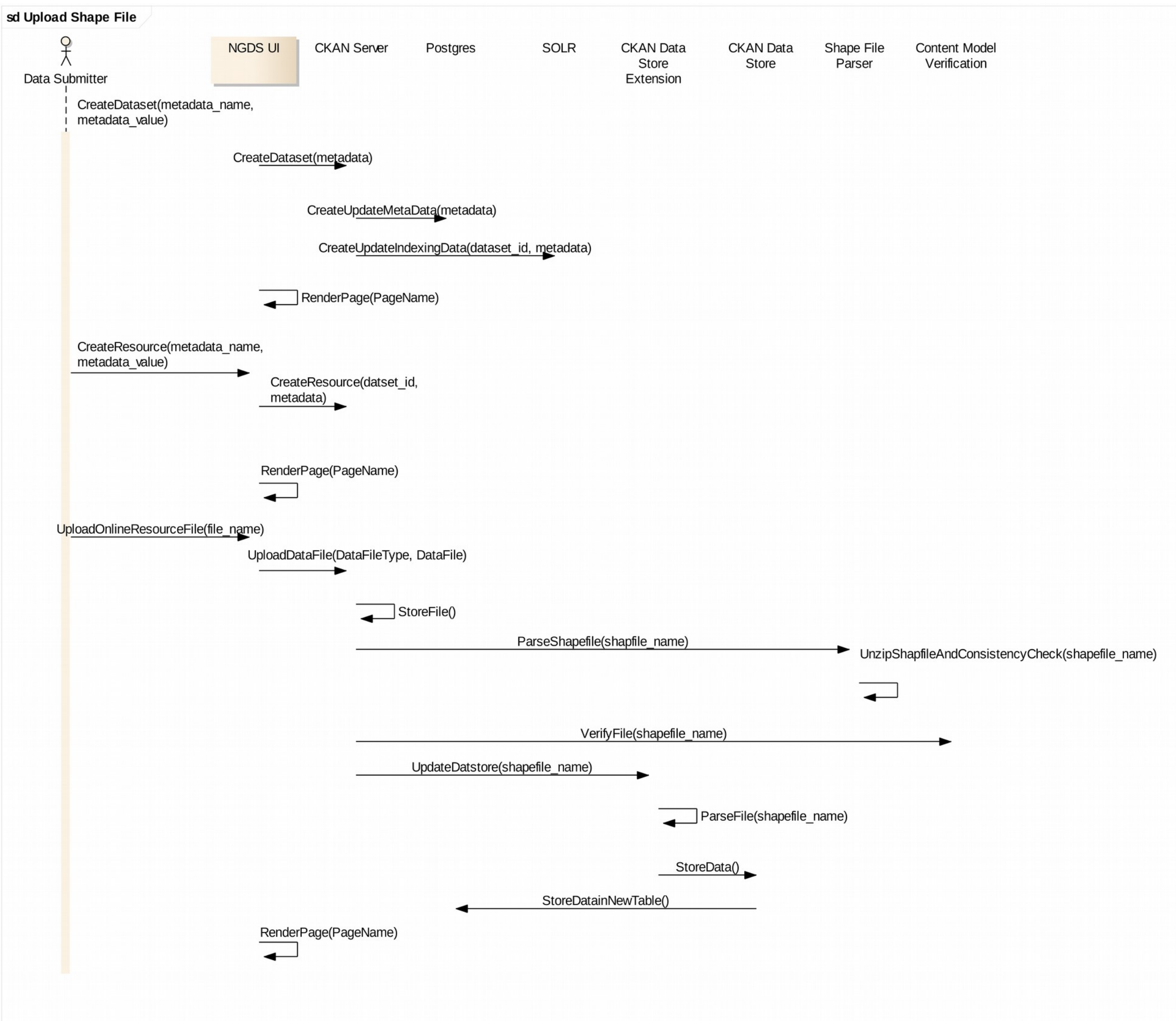


Figure 32: MSC for Use Case "Shapefile Upload"

In many aspects the use case causes the same interactions as uploading structured resources. The difference here is that Shapefile needs to be unpacked before it can be verified and added to the data-store.

5.4.7 Use Case: "Harvest Metadata"

In this use case no human user is directly involved. Rather the harvesting process is triggered as a Linux CRON job. At harvesting time the CRON job triggers a process that causes the CKAN harvesting plug-in to start harvesting from a repository. Figure 33 assumes that the repository is a NGDS Node-in-a-Box. It is expected that at the end of the harvesting process all changes within the repository that happened since the last harvesting time are applied to the aggregated catalog that is maintained by the central harvesting node. In other words, a user must be able to find the datasets in that repository by using the search interface of the central harvesting node.

The task of the NGDS software is to contact the repository using the CSW protocol and to request for a list of all changes since the last harvesting. The Node-in-a-Box must then compile the change list and send it via CSW to the central harvesting node. The list can be very long. The central harvesting node must store the changes in its local catalog.

Figure 33 shows the interaction between the components.

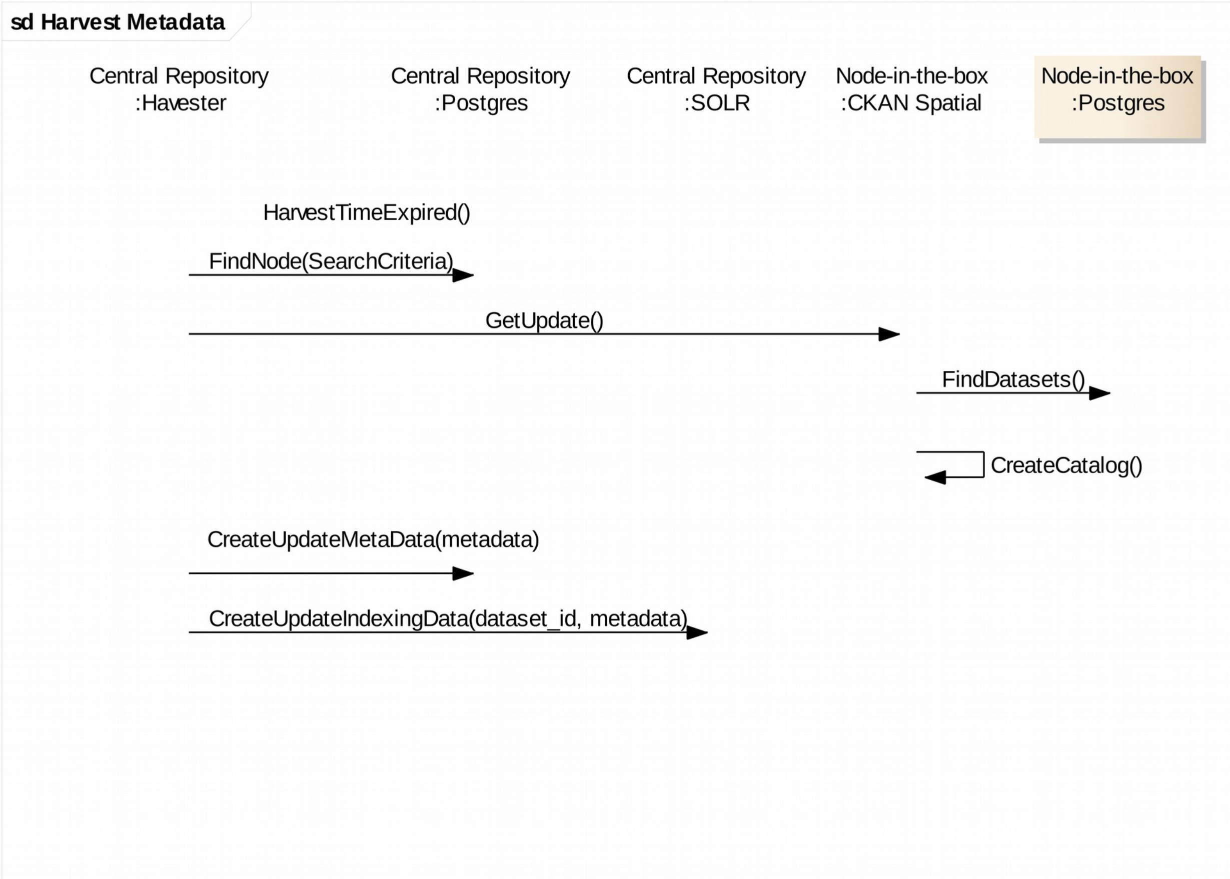


Figure 33: MSC for Use Case "Harvest Metadata"

As can be seen the CKAN Plug-in Central Repository Harvester and its counterpart CKAN Plug-in Spatial Extension do the heavy lifting in this use case. Essentially the harvester contacts the spatial extension over the network. The Spatial Extension compiles the list and returns it to the harvester which applies the retrieved information to both the Postgres database and the SOLR indexer.

5.4.8 Use Case "Preview WFS WCS Resources"

In this use case the user wants to preview a resource of type WFS or WCS. The task for the NGDS System is to visually present the WFS/ WCS resource. CKAN come with recline.js support which is able to present structured resources as a table, graph, or on a map (if the structured resource provides geographic information). However, recline does not directly WFS or WCS. However, it does support JSON representations of tables. Therefore NGDS UI does not access the WFS/ WCS resource directly. It rather accesses it indirectly via translation service provided as a CKAN extension. Figure 34 shows the interaction between the components is presented as a message sequence chart.

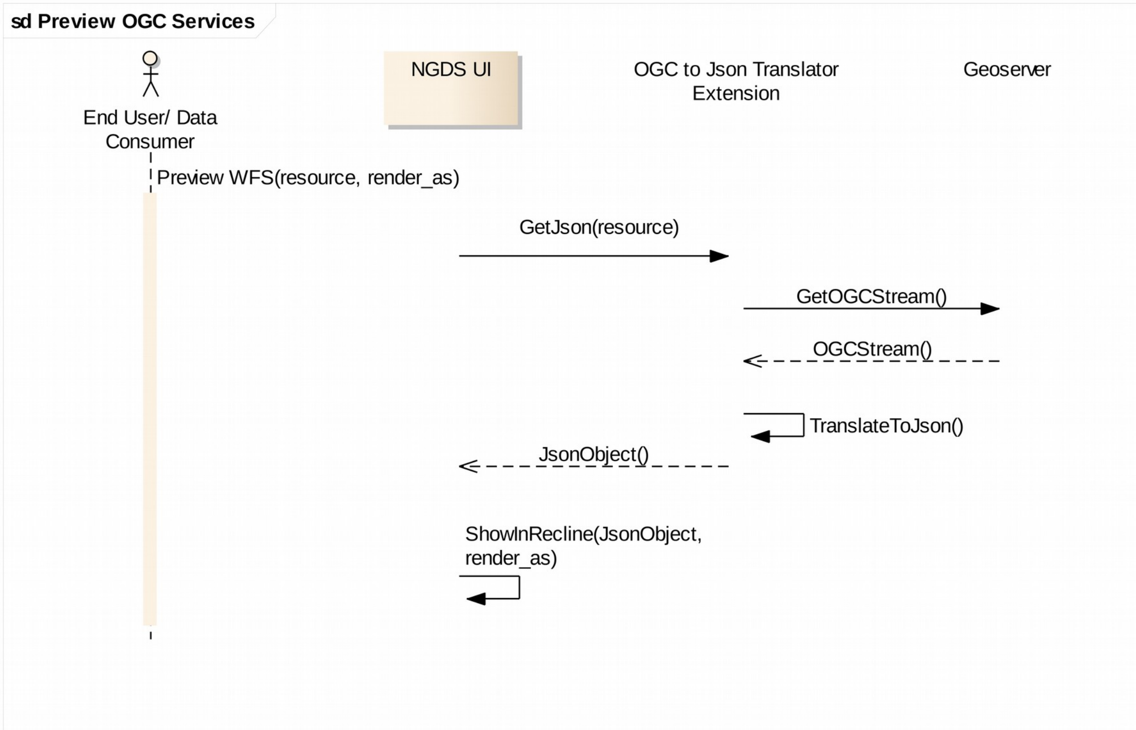


Figure 34: Use Case "Preview WFS WCS Resources"

Essentially the user initiates the interaction by clicking on the "preview resource" button of a WFS or WCS resource. The UI then calls the RESTful API of the "OGC to JSON Translator Extension". This extension contacts the OGC server to receive the WFS/ WCS stream. In Figure 34, the OGC server is represented as a Geoserver. However, any server that provides WFS or WCS can be contacted. The data stream from the server is translated into a JSON object which in turn is returned to the UI. The UI then uses its recline.js library to present the WFS/ WCS resource.

5.5 Process Views

The following table provides an overview of all processes triggered by NGDS and how these processes are started.

Process	*Description*	*Service Type*
NGDS-CKAN core process	Runs the NGDS Web Application. Started with: `paster server configuration.ini`	Started as a Linux service. During development strated from the command line.
CKAN celery damon	Runs the celery damon which acts as a batch processor for CKAN activities. It runs them in an asynchronous process. When the queue of batch tasks is empty the damon idles. `paster celeryd`	Started as a Linux service. During development started from the command line.
NGDS harvester	Runs the NGDS harvesting process. Started with: `paster harvest ...`	Started as a Linux service only on the Central Node. During development started from the command line.
NGDS full-text indexer	Runs the NGDS full-text indexing service. Started with: `paster fulltextindex`	Started with a script. The process dies when there is nothing to index. It is restarted by a cronjob 5 minutes after it dies last time. During development started manually from the command line.
Postgres	Postgres Database	Started as a Linux service.
Geoserver	OGC Server. Started with: `service jetty start`	Started as a Linux service.
SOLR	Indexing service. Started with: `service jetty start`	Started as a Linux service

Table 5-20: NGDS Processes

5.6 Physical Views

NGDS is a Web Application and physically runs on PC server hardware. The system can be deployed completely on a single machine running all required applications. During development virtual machines (virtual box) with 1 Intel I-7 core, ~1GB RAM and ~20GB disk space are used. For production purposes larger settings are highly recommended. For installation on a single machine the system should provide at least 4 Intel Xeon cores, 8GB of RAM, and several hundred MB of disk space.

For performance reasons the system may also be scaled to run on multiple machines. Scaling both horizontally and vertically is possible as is described in the following two Figures.

Note that the NGDS development team does not test any of the scaling options.

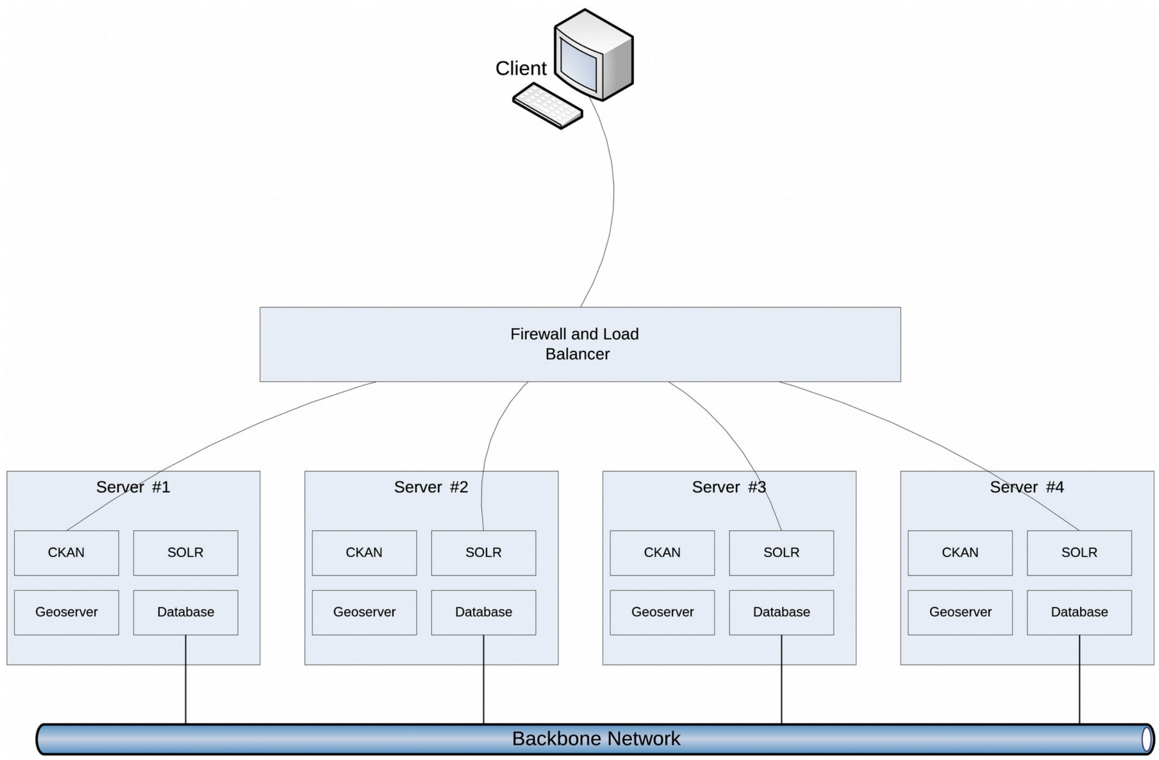


Figure 35: NGDS scaled horizontally

When horizontal scaling is used, the complete system is duplicated on multiple server machines. Each server provides all required instances. The database servers underneath are synchronized using built-in synchronization mechanisms of the database management system. A Load Balancer is required to distribute requests across the machines. NGDS does not have a single-sign-on mechanism so the Load Balancer has to be configured to use sticky sessions.

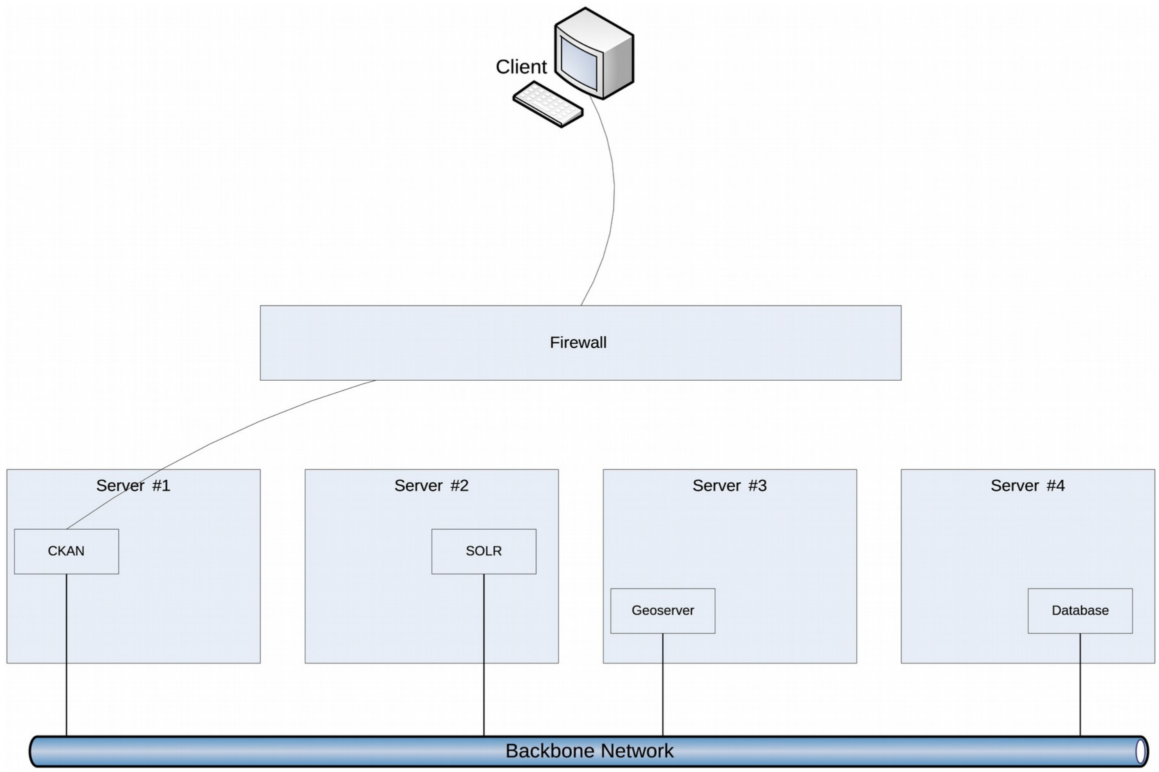


Figure 36: NGDS scaled vertically

Vertical scaling requires running each main component of NGDS on a separate server. This option is likely to be implemented easiest. Some of those components could be duplicated as necessary. Note that vertical scaling does not increase the system reliability since each node in the system is a single point of failure.

6 Referenced Materials

KRUCHTEN	Kruchten, Philippe (1995, November). Architectural Blueprints — The "4+1" View Model of Software Architecture. IEEE Software 12 (6), pp. 42-50.
Barbacci 2003	Barbacci, M.; Ellison, R.; Lattanze, A.; Stafford, J.; Weinstock, C.; & Wood, W. *Quality Attribute Workshops (QAWs)*, Third Edition (CMU/SEI-2003-TR-016). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2003. <http://www.sei.cmu.edu/publications/documents/03.reports/03tr016.html>.
Bass 2003	Bass, Clements, Kazman, *Software Architecture in Practice*, second edition, Addison Wesley Longman, 2003.
Clements 2001	Clements, Kazman, Klein, *Evaluating Software Architectures: Methods and Case Studies*, Addison Wesley Longman, 2001.
Clements 2002	Clements, Bachmann, Bass, Garlan, Ivers, Little, Nord, Stafford, *Documenting Software Architectures: Views and Beyond*, Addison Wesley Longman, 2002.
IEEE 1471	ANSI/IEEE-1471-2000, *IEEE Recommended Practice for Architectural Description of Software-Intensive Systems*, 21 September 2000.
REQ-2.7	Kuhmuench et. al.: NGDS Software Requirements Specification. V2.7. http://gdsdpwg.net/NGDSDesign/Shared%20Documents/Requirements/SCR_NGDS_Software_Requirements_Specification_v2.7.docx
FEATURES	Bruschi et.al.: NGDS Feature List. http://gdsdpwg.net/NGDSDesign/Shared%20Documents/Architecture/Features.xlsx
EPICS	Bruschi, Tanikella et.al.: Prioritization of Epics. http://gdsdpwg.net/NGDSDesign/Shared%20Documents/Sprints/Prioritization%20of%20Epics.xlsx
PYLONS	James Gardner: The Definitive Guide to Pylons. 2009. ISBN 1-59059-934-9

7 DIRECTORY

7.1 GLOSSARY

Term	*Definition*
software architecture	The structure or structures of that system, which comprise software elements, the externally visible properties of those elements, and the relationships among them [Bass 2003]. "Externally visible" properties refer to those assumptions other elements can make of an element, such as its provided services, performance characteristics, fault handling, shared resource usage, and so on.
view	A representation of a whole system from the perspective of a related set of concerns [IEEE 1471]. A representation of a particular type of software architectural elements that occur in a system, their properties, and the relations among them. A view conforms to a defining viewpoint.
view packet	The smallest package of architectural documentation that could usefully be given to a stakeholder. The documentation of a view is composed of one or more view packets.
viewpoint	A specification of the conventions for constructing and using a view; a pattern or template from which to develop individual views by establishing the purposes and audience for a view, and the techniques for its creation and analysis [IEEE 1471]. Identifies the set of concerns to be addressed, and identifies the modeling techniques, evaluation techniques, consistency checking techniques, etc., used by any conforming view.

7.2 Acronym List

Acronym	*Definition*
API	Application Programming Interface; Application Program Interface; Application Programmer Interface
ATAM	Architecture Tradeoff Analysis Method
CMM	Capability Maturity Model.
CMMI	Capability Maturity Model Integration
CORBA	Common object request broker architecture
COTS	Commercial-Off-The-Shelf
EPIC	Evolutionary Process for Integrating COTS-Based Systems
IEEE	Institute of Electrical and Electronics Engineers
KPA	Key Process Area
OO	Object Oriented
ORB	Object Request Broker
OS	Operating System
QAW	Quality Attribute Workshop
RUP	Rational Unified Process
SAD	Software Architecture Document
SDE	Software Development Environment
SEE	Software Engineering Environment
SEI	Software Engineering Institute, Systems Engineering & Integration, Software End Item
SEPG	Software Engineering Process Group
SLOC	Source Lines of Code
SW-CMM	Capability Maturity Model for Software
CMMI-SW	Capability Maturity Model Integrated - includes Software Engineering

UML	Unified Modeling Language
NGDS	National Geothermal Data System
CKAN	Comprehensive Knowledge Archive Network
DOE	Department of Energy
GDR	Geothermal Data Repository
AZGS	Arizona Geological Survey
WFS	Web Feature Service
WMS	Web Map Service
WCS	Web Coverage Service
CSW	Catalog Service For the Web
OGC	Open Geographical Consortium
OKFN	Open Knowledge Foundation
SAD	Software Architecture Document

Appendix 4a

Developer Guide

Siemens AG

National Geothermal Data System Software Developer's Guide

Version: V1.1

Editor: Christoph Kuhmuench

Version:	**Date:**	**Author:**	**Details**
0.1	Roberto Silva Filho	05/28/2013	Initial Draft Created
0.2	Monica McKenna	06/11/2013	Minor updates
0.3	Monica McKenna	06/24/2013	Combining comments from a few people
0.4	Monica McKenna	06/25/2013	Added appendix with summary of development.ini changes
0.5	Monica McKenna	07/22/2013	A little re-organization, more hints, and added gdal
0.6	Monica McKenna	07/24/2013	Updating with feedback
0.7	Christoph Kuhmuench	12/26/2013	Updating to latest installer.

Table of Contents

List of Figures

Preface

NGDS or National Geothermal Data System is a government funded project. The system supports the storage and search of geothermal data from different organizations. It is a hub and spoke network of data providers through its many node-in-a-box distributions, with the hub being a centralized node that supports the harvesting and search of this distributed data.

Purpose and Audience

This document is a step by step tutorial to guide new developers and users in setup both node-in-the-box and harvesting catalogs instances of NGDS.

This document is intended for a technical audience who need to understand the concepts and the reasoning of the installation process. Targeted audience includes:

NGDS System Administrator
Software Architects
Software Developers
DoE Monitors

The purpose of the document is to help software developers to become productive quickly. We assume that the developer has browsed through the requirements document [REQ2.7], and has intensively studied the software architecture document [ARC1.0]. The architecture document is a key input for the developer and should help to understand the grand picture of NGDS. Finally the developer must also read the installation instructions and execute them carefully and precisely.

This helps to fulfill one of the main goals of NGDS, which is to provide a basis for a sustainable open source software project that is attractive for an open source team to maintain. With this documentation future **NGDS System Administrators** will be able to quickly understand the system and become productive. The NGDS documentation also provides the basis for a future open source organization to take over architectural oversight of the software.

Document Roadmap

This document outlines the architecture of NGDS and is structured in the following way:

- Configure the production version for development
- Relevant frameworks and links to their description

- Configure Eclipse to be used with CKAN
- Debugging with Eclipse
- Highlevel guide line through the code

Document Scope and Background

The requirements of NGDS are discussed in detail in the requirements document [REQ2.7]. The software Architecture is discussed in detail in the software architecture Document [ARC1.0]. The installation is discussed in detail in the installaion guide [INS1.0].

As outlined in the architecture document the NGDS software stack is the same for a node installation and a central installation (see Figure 18). The install process includes a configuration parameter that allows to choose between installation as a Node-in-a-Box and Central Harvesting Node.

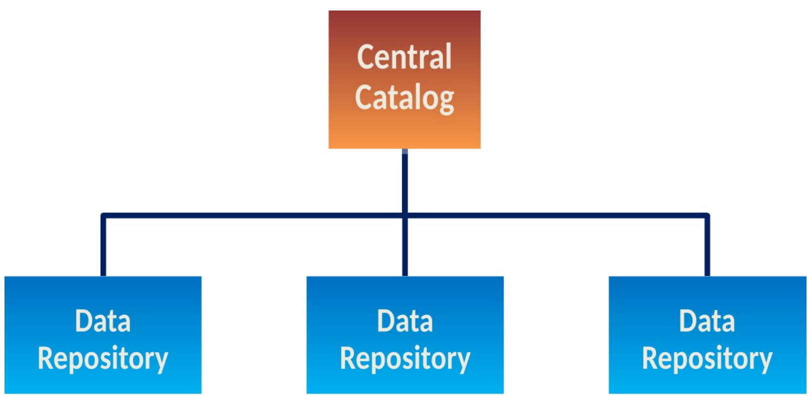


Figure 37: NGDS is a grid of repositories

Configure System for Development

First of all, please also refer to our WIKI:

https://github.com/ngds/dev-info/wiki/_pages

In this section we outline how to modify the system so that it can be used in development mode. We assume that the installation has been completed as it was outlined in the Installation Guide [INS1.0].

The following figure outlines the main components of NGDS.

The developer could very well use the system in production setup. In this setup CKAN is served indirectly via apache2 web server. However, as a developer it is preferable to run the system directly via the command line. Also the developer probably want to turn on the development mode in order to get more feedback. The development mode also has the advantage that it serves the source JavaScript files and not the compressed JavaScript files.

The installer script also configures the celeryd demon to run as a service. Most likely this is ok from the developers perspective but we also show how to shutdown the service in the following section

Stopping Apache2

In order to run the system in debugging mode we first shut down apache2:

```
% sudo service apache2 stop
```

This command shuts the apache server down. Note that the server will come up again when you restart the system. De-Installation of apache2 via apt is overkill. An easy way to deactivate a service is to execute the following command:

```
% sudo sh -c "echo 'manual' > /etc/init/apache2.override"
```

Stopping the Celeryd (Optional)

As outlined above for most developers it is actually convenient to run the celeryd as a service. If there is a need to deactivate and run it manually this is how it is done:

```
% sudo service ngds-celeryd stop
```

As before apache2 the service will restart when the system is rebooted. This can be prevented with:

```
% sudo sh -c "echo 'manual' > /etc/init/ngds-celeryd.override"
```

Configure CKAN for development mode

In order to run CKAN in development mode you want to execute it from the command line via a paster command. We assume that you have executed the step 2.1 above and stopped apache2. You can leave the celery demon running. All NGDS source code can be found under:

```
/opt/ngds/bin/default
```

So in order to run the system from the command line do the following:

```
% cd /opt/ngds/bin/default
% . ./bin/activate
% cd ckan
% cp /opt/ngds/etc/ckan/default/production.ini ./development.ini
% paster serve development.ini
```

After this command is executed NGDS runs from the command line. The service can be reached under:
http://127.0.0.1:5000/

In the next steps you want to modify some entries in the development.ini file:
[line017] Debug= true

This is flag causes two things:
Cause CKAN to create Debugging output
Cause CKAN to serve the uncondensed JavaScript resources

Especially the second aspect is important for developers. If the flag is not set you have to compile the JavaScript resources every time you make a change to one of them. This can be done with the command:

```
% paster --plugin=ckanext-ngds ngdsapi compile_client_scripts
```

Everything else can stay unchanged.

Relevant Frameworks

Please refer to the architecture document to get an overview of all frameworks.

Understanding CKAN Routes

This is a list of URLs that CKAN listens to. For each URL, I've printed out the "Controller" and "Action" that is performed when a request comes in to that URL. This provides you with a starting point for understanding what CKAN does at any particular page or API call. "Controller" provides a pointer to a class, and "Action" is the name of a function inside that class.

If the Controller is not fully qualified, then they exist at ckan.controllers.{whatever}, for example "home" really means "ckan.controllers.home"

```
/error/%(action)s
    Controller: error
    Action: None

/error/%(action)s/%(id)s
    Controller: error
    Action: None

%(url)s
    Controller: home
    Action: cors_options

/api/2/search/%(register)s/geo
    Controller: ckanext.spatial.controllers.api:ApiController
    Action: spatial_query

/package/%(id)s/map
    Controller: None
    Action: None

/dataset/%(id)s/map
    Controller: ckanext.spatial.controllers.view:ViewController
    Action: wms_preview

/proxy
    Controller: ckanext.spatial.controllers.view:ViewController
    Action: proxy

/hello/world/%(eggs)s
    Controller: ckanext.helloworld.controller:HelloWorldController
    Action: dispatch_get

/
    Controller: home
    Action: index

/about
    Controller: home
    Action: about

/api%(ver)s/action/%(logic_function)s
    Controller: api
```

```
    Action: action

/api%(ver)s
    Controller: api
    Action: get_api

/api%(ver)s/search/%(register)s
    Controller: api
    Action: search

/api%(ver)s/tag_counts
    Controller: api
    Action: tag_counts

/api%(ver)s/rest
    Controller: api
    Action: index

/api%(ver)s/qos/throughput/
    Controller: api
    Action: throughput

/api%(ver)s/rest/%(register)s
    Controller: api
    Action: list

/api%(ver)s/rest/%(register)s
    Controller: api
    Action: create

/api%(ver)s/rest/%(register)s/%(id)s
    Controller: api
    Action: show

/api%(ver)s/rest/%(register)s/%(id)s
    Controller: api
    Action: update

/api%(ver)s/rest/%(register)s/%(id)s
    Controller: api
    Action: update

/api%(ver)s/rest/%(register)s/%(id)s
    Controller: api
    Action: delete

/api%(ver)s/rest/%(register)s/%(id)s/%(subregister)s
    Controller: api
    Action: list

/api%(ver)s/rest/%(register)s/%(id)s/%(subregister)s
    Controller: api
    Action: create

/api%(ver)s/rest/%(register)s/%(id)s/%(subregister)s/%(id2)s
    Controller: api
    Action: create
```

```
/api%(ver)s/rest/%(register)s/%(id)s/%(subregister)s/%(id2)s
    Controller: api
    Action: show

/api%(ver)s/rest/%(register)s/%(id)s/%(subregister)s/%(id2)s
    Controller: api
    Action: update

/api%(ver)s/rest/%(register)s/%(id)s/%(subregister)s/%(id2)s
    Controller: api
    Action: delete

/api%(ver)s/util/user/autocomplete
    Controller: api
    Action: user_autocomplete

/api%(ver)s/util/is_slug_valid
    Controller: api
    Action: is_slug_valid

/api%(ver)s/util/dataset/autocomplete
    Controller: api
    Action: dataset_autocomplete

/api%(ver)s/util/tag/autocomplete
    Controller: api
    Action: tag_autocomplete

/api%(ver)s/util/resource/format_autocomplete
    Controller: api
    Action: format_autocomplete

/api%(ver)s/util/resource/format_icon
    Controller: api
    Action: format_icon

/api%(ver)s/util/group/autocomplete
    Controller: api
    Action: group_autocomplete

/api%(ver)s/util/markdown
    Controller: api
    Action: markdown

/api%(ver)s/util/dataset/munge_name
    Controller: api
    Action: munge_package_name

/api%(ver)s/util/dataset/munge_title_to_name
    Controller: api
    Action: munge_title_to_package_name

/api%(ver)s/util/tag/munge
    Controller: api
    Action: munge_tag
```

```
/api%(ver)s/util/status
    Controller: api
    Action: status

/api%(ver)s/util/snippet/%(snippet_path)s
    Controller: api
    Action: snippet

/api%(ver)s/i18n/%(lang)s
    Controller: api
    Action: i18n_js_translations

/packages
    Controller: None
    Action: None

/packages/%(url)s
    Controller: None
    Action: None

/package
    Controller: None
    Action: None

/package/%(url)s
    Controller: None
    Action: None

/dataset/%(id)s/related/new
    Controller: related
    Action: new

/dataset/%(id)s/related/edit/%(related_id)s
    Controller: related
    Action: edit

/dataset/%(id)s/related/delete/%(related_id)s
    Controller: related
    Action: delete

/dataset/%(id)s/related
    Controller: related
    Action: list

/apps/%(id)s
    Controller: related
    Action: read

/apps
    Controller: related
    Action: dashboard

/dataset
    Controller: package
    Action: search

/dataset/%(action)s
```

```
    Controller: package
    Action: None

/dataset/%(action)s/%(id)s/%(revision)s
    Controller: package
    Action: read_ajax

/dataset/%(action)s/%(id)s
    Controller: package
    Action: None

/dataset/activity/%(id)s/%(offset)s
    Controller: package
    Action: activity

/dataset/%(id)s.%(format)s
    Controller: package
    Action: read

/dataset/%(id)s
    Controller: package
    Action: read

/dataset/%(id)s/resource/%(resource_id)s
    Controller: package
    Action: resource_read

/dataset/%(id)s/resource_delete/%(resource_id)s
    Controller: package
    Action: resource_delete

/dataset/%(id)s/resource_edit/%(resource_id)s
    Controller: package
    Action: resource_edit

/dataset/%(id)s/resource/%(resource_id)s/download
    Controller: package
    Action: resource_download

/dataset/%(id)s/resource/%(resource_id)s/embed
    Controller: package
    Action: resource_embedded_dataviewer

/dataset/%(id)s/resource/%(resource_id)s/viewer
    Controller: package
    Action: resource_embedded_dataviewer

/dataset/%(id)s/resource/%(resource_id)s/preview/%(preview_type)s
    Controller: package
    Action: resource_datapreview

/groups
    Controller: None
    Action: None

/groups/%(url)s
    Controller: None
```

```
    Action: None

/group
    Controller: group
    Action: index

/group/list
    Controller: group
    Action: list

/group/new
    Controller: group
    Action: new

/group/%(action)s/%(id)s
    Controller: group
    Action: None

/group/activity/%(id)s/%(offset)s
    Controller: group
    Action: activity

/group/%(id)s
    Controller: group
    Action: read

/tags
    Controller: None
    Action: None

/tags/%(url)s
    Controller: None
    Action: None

/tag/read/%(url)s
    Controller: None
    Action: None

/tag
    Controller: tag
    Action: index

/tag/%(id)s
    Controller: tag
    Action: read

/users/%(url)s
    Controller: None
    Action: None

/user/
    Controller: None
    Action: None

/user/edit
    Controller: user
    Action: edit
```

```
/user/activity/%(id)s/%(offset)s
    Controller: user
    Action: activity

/user/activity/%(id)s
    Controller: user
    Action: activity

/dashboard/%(offset)s
    Controller: user
    Action: dashboard

/dashboard
    Controller: user
    Action: dashboard

/user/follow/%(id)s
    Controller: user
    Action: follow

/user/unfollow/%(id)s
    Controller: user
    Action: unfollow

/user/followers/%(id)s
    Controller: user
    Action: followers

/user/edit/%(id)s
    Controller: user
    Action: edit

/user/reset/%(id)s
    Controller: user
    Action: perform_reset

/user/register
    Controller: user
    Action: register

/user/login
    Controller: user
    Action: login

/user/_logout
    Controller: user
    Action: logout

/user/logged_in
    Controller: user
    Action: logged_in

/user/logged_out
    Controller: user
    Action: logged_out
```

```
/user/logged_out_redirect
    Controller: user
    Action: logged_out_page

/user/reset
    Controller: user
    Action: request_reset

/user/me
    Controller: user
    Action: me

/user/set_lang/%(lang)s
    Controller: user
    Action: set_lang

/user/%(id)s
    Controller: user
    Action: read

/user
    Controller: user
    Action: index

/revision
    Controller: revision
    Action: index

/revision/edit/%(id)s
    Controller: revision
    Action: edit

/revision/diff/%(id)s
    Controller: revision
    Action: diff

/revision/list
    Controller: revision
    Action: list

/revision/%(id)s
    Controller: revision
    Action: read

/feeds/group/%(id)s.atom
    Controller: feed
    Action: group

/feeds/tag/%(id)s.atom
    Controller: feed
    Action: tag

/feeds/dataset.atom
    Controller: feed
    Action: general

/feeds/custom.atom
```

```
    Controller: feed
    Action: custom

/ckan-admin
    Controller: admin
    Action: index

/ckan-admin/%(action)s
    Controller: admin
    Action: None

/api/storage
    Controller: ckan.controllers.storage:StorageAPIController
    Action: index

/api/storage/metadata/%(label)s
    Controller: ckan.controllers.storage:StorageAPIController
    Action: set_metadata

/api/storage/metadata/%(label)s
    Controller: ckan.controllers.storage:StorageAPIController
    Action: get_metadata

/api/storage/auth/request/%(label)s
    Controller: ckan.controllers.storage:StorageAPIController
    Action: auth_request

/api/storage/auth/form/%(label)s
    Controller: ckan.controllers.storage:StorageAPIController
    Action: auth_form

/storage/upload
    Controller: ckan.controllers.storage:StorageController
    Action: upload

/storage/upload_handle
    Controller: ckan.controllers.storage:StorageController
    Action: upload_handle

/storage/upload/success
    Controller: ckan.controllers.storage:StorageController
    Action: success

/storage/upload/success_empty
    Controller: ckan.controllers.storage:StorageController
    Action: success_empty

/storage/f/%(label)s
    Controller: ckan.controllers.storage:StorageController
    Action: file

/i18n/strings_%(lang)s.js
    Controller: util
    Action: i18n_js_strings

/util/redirect
    Controller: util
```

```
    Action: redirect

/testing/primer
    Controller: util
    Action: primer

/testing/markup
    Controller: util
    Action: markup

/stats
    Controller: ckanext.stats.controller:StatsController
    Action: index

/stats/%(action)s
    Controller: ckanext.stats.controller:StatsController
    Action: None
```

Remote Debugging the node in a box and central deployments on the amazon machine

Prerequisites

- PyCharm Professional Edition
- The NGDS project setup on PyCharm. (Importing the pyenv directory into pycharm is the ideal setup)
- Ability to ssh to the amazon machine
- Either linux, or putty on windows. (I coulnd't get OpenSSH to work well on my windows)

Outline

We're going to create configuration entries for a remote python interpreter (amazon), a remote debug configuration which uses the remote python interpreter to serve our application. We're also going to create a mapping between our local and remote source directories so that we can set breakpoints and so forth and have these reflected on the remote instance.

We'll then create an ssh tunnel and forward a local port to the amazon machine so that we can navigate the UI locally on our browser by setting a socks proxy variable in our network configuration.

We also will have to create a copy of the configuration file, one for each instance we're debugging, and modify it to change the port number that the application will be served on. We'd like to leave everything else the same so that we have all the data and indexes that the actual instance has.

Local Steps

PyCharm

Configuring a remote python interpreter

- Go to Settings, and find the Python Interpreters section.

- Click on the + button and select remote in the dropdown.

The values that go into the dialog should look like this -

```
Host : 1.1.1.1               Port : 22
Username : ubuntu
Auth Type : Keypair (OpenSSH)
Private Key File : /path/to/my/priv_key.txt
Passphrase : abracadabraismypassphrase
// The passphrase you typed in when you generated your private key.

Python Interpreter Path : /home/ubuntu/pyenv/bin/python
// This is the path to the virtual environment's python binary. This
should correspond to the instance you're debugging.
```

- Click on Test Connection to verify that your settings are accurate.
- You'll do this once for every instance you'd like to debug, for example, one for the central instance and one for the node instance.

Creating a debug configuration

- Click on the Run menu item and click on 'Edit Configurations'
- Click on the + button to add a new debug configuration entry.

The values that go in there should somewhat like this

```
Name : Remote Node-in-a-box
Script : bin/paster serve src/ckan/debug.ini --reload
// Assuming you set your working directory to /home/ubuntu/pyenv
Python Interpreter : Remote Python
// Select the interpreter you created in the earlier step
Working directory : /home/ubuntu/pyenv/
Path Mappings : /home/ubuntu/pyenv - /home/mylocalusername/mylocalpyenv
```

- Click Apply.

SSH Port Forwarding

- Open up a terminal, type in ssh -D 7777 ubuntu@1.1.1.1 -v
- Substitute 1.1.1.1 with the IP of the remote machine.
- Open up your browser, I prefer firefox since you can apply the proxy to only firefox and not modify the system's network configuration.
- Navigate to Advanced > Network Settings and select the SOCKS Proxy item and enter in localhost and port 7777 into the Host and Port entries respectively.

Remote Steps

- SSH to the remote machine
- Navigate to the ckan directory of the instance you want to debug.
- Create a copy of development.ini in the same directory and call it debug.ini. cp development.ini debug.ini
- Modify development.ini and in the [server:main] section, modify the port number and set it to 7777 (or some other value that is not being used. You have to make sure this is the same as the port you entered in the SSH port forwarding step)

Testing the setup

- In pycharm, select the debug configuration we just created and click on the debug button (or select debug from the tools menu).
- The console readout will indicate an ssh connection being made, and if it's successful, will tell you which port the application is being served on.
- Now open up a terminal and type in ssh -D 7777 ubuntu@1.1.1.1 -v (Replace 1.1.1.1 with the IP of the machine you're debugging)
- Navigate to localhost:7777 on firefox. If all the previous steps were done correctly, you should see the NGDS UI come up.

FAQs and gotchas

How do I compile the project's less files to css files and minify js resources?

A) Run the paster command

paster --plugin=ckanext-ngds ngdsapi compile_client_scripts from the ckan installation directory.

How do I run a paster command?

A) Paster commands are declared in setup.py files. For this reason, they are always to be run from the directory containing the setup.py file that declares that command.

CKAN commands are run in two ways - either

cd /path/to/src/ckan
paster --plugin=ckanext-ngds ngdsapi doc-index #to run an ngds command.

Or,

cd /path/to/ckanext/ngds
paster ngdsapi doc-index --c /path/to/src/ckan/development.ini

The first command is run from the ckan installation directory where it's development.ini and setup.py file are to be found. The second command is run from the ckanext-ngds/ckanext/ngds directory where it's setup.py file declaring the ngdsapi command is to be found, and the path to ckan's development.ini is provided. The only reason for this is that ckan (and ckan plugin) commands are run against a live ckan instance.

As soon as I complete installing NGDS, when I navigate to the library page, I get a 404 - Group not found. What do I do?

A) You'll need to create an organization called public. Open up development/production.ini and remove ngdsui from the plugin list (it is to be found under ckan.plugins). Restart CKAN and then navigate to the organization page and create an organization with the title set to 'public'. Then add ngdsui back into the plugins list and restart ckan. You should now be able to see the library page with no errors.

System Design Framework for National Geothermal Data System

USGIN development team
version 1.0
May 1, 2014

Contents

1. Introduction

This report presents a framework for National Geothermal Data System (NGDS) architecture focused on the network data publication and data access aspects of the system. The design is based on work done at the Arizona Geological Survey under auspices of DOE award DE-EE0001120, in conjunction with related work on development of the Geoscience Information Network (USGIN) supported by NSF grant EAR-0753154, and a parallel DOE award DE-EE1002850 to compile and publish geothermal data from state geological surveys to integrate with the NGDS.

The design is intended to provide an incremental development framework that utilizes existing and open-source technology wherever possible, builds on a variety of existing standards and specifications, and allows for agile development of the NGDS in the current, rapidly evolving technology environment. To be sustainable, NGDS has provided a framework to promote community engagement and incorporate new technology and ideas as they are developed without disrupting existing practices.

This document includes an introductory section discussing the scope of the system based on the original Department of Energy Funding Opportunity Announcement (FOA) and system requirements which can from NGDS consortium's proposal. Requirements regarding specific data discovery and system architecture needs are discussed in the third section, followed by implementations of that system architecture and system deployment including software developments.

2. Scope and purpose of system

As described in original Department of Energy Funding Opportunity Announcement (FOA):

National Geothermal Database Description
The National Geothermal Database will store critical geothermal site attribute information such as temperature at depth, seismicity/microseismicity, fracture maps, drilling data, permeability data, well logs, geophysical surveys, etc. The database should be inclusive of all types of geothermal resources such as hydrothermal, geopressured, Enhanced Geothermal Systems, geothermal fluids coproduced with oil and/or gas, etc. It should also utilize information from existing USGS geothermal resource assessments and DOE funded R&D projects. This standardized set of geothermal resource data will be made available to the public and serve to focus geothermal exploration activities, thereby mitigating investment risks.

From http://apps1.eere.energy.gov/geothermal/projects/projects.cfm/ProjectID=27:

“The NGDS will be able to handle the full range of geoscience and engineering data pertinent to geothermal resources as well as incorporate data from the full suite of geothermal resource types. It will be able to handle data on geothermal site attributes, power plants, environmental factors, policy and procedure data, and institutional barriers. It will provide resource classification and financial risk assessment tools to help encourage the development of more geothermal resources by industry. It will be an easy to use system that meets the needs of the professional and the public for information on geothermal resources.”

Abstracted from Original Project Proposal from the Geothermal Data Coalition:
Goal: build a state-of-the art data system.

- *reduce social-cultural barriers that could hinder the development of a comprehensive database*
- *Provide access to critical data and data products.*
- *Provide the basis for financial investment risk analysis.*
- *Provide geothermal-resource information to the public and decision-makers*
- *support state and federal agencies with land and resource management missions*
- *support ongoing and future geothermal-related research*
- *contribute to enhancing the education pipeline for careers in the geothermal energy industry*

3. System Technical Design principles

The National Geothermal Data System was tasked with providing online resources to make it easy for users to extract, assess, and synthesize data according to criteria they select. Data are provided by a community of data providers, many of whom maintain their own data management systems. Numerous kinds of existing, “legacy” data in various tables, spreadsheets and databases from state surveys nation-wide have been digitized and made accessible through the system.

Resources (e.g. data, metadata, catalogs, services, tools) are made accessible through the system by creating metadata conforming to a shared content model and sharing them through the main metadata aggregator (‘central node’). The metadata provide information describing resources that can be indexed for discovery by search engines, information about provenance and quality of the resource so users can evaluate the resource for their application, and information describing how to access the resource. The access instructions should be in a format that can be utilized by software clients to automate the access process and minimize the amount of user interaction required to bring the resource to their desktop.

The central node at http://www.geothermaldata.org is a single search client for users to search all resources in the system. Any search client that implements the system catalog service profile should be able to conduct search against any system catalog that also implements the profile. This means that there can be multiple portals and client applications for accessing system resources; it requires that a single client can search different catalogs in the system without the user having to reconfigure the software.

Providing quality information to evaluate system resources requires criteria that can be used to filter data and categorize them according to established and user-defined quality levels. These quality filters will vary depending on the type of data and their targeted use.

Structured data are provided through NGDS services that have published protocol and documented interchange formats. The idea is that multiple data providers can present the same kind of information in the same way, and a client that implements an NGDS service can access that service from any server in the system that offers that service and get data that integrate with minimum operator intervention.

The following bullet points were extracted from the original project proposal and subsequent SOPO:

- Design must be expansive; capture the full physical, geologic, geophysical, and geochemical context of geothermal systems on scales ranging from regional to the individual well bore to the thin section and microscopic scales.
- Information in system must be supported by metadata to document authority and to provide people and projects that compile data the appropriate level of recognition and support
    - All data will credit the original intellectual source and host server of record for that data.
    - Standard measures of "quality" should be available. E.G. variability, bias, systematic error, imprecision, accuracy, precision, reproducibility, etc.
- Able to adapt to evolving requirements, new technologies and standards, and expanded scope as necessary.
- Use existing or emerging standards and technology whenever possible rather than developing new ones
- Open source and open accessibility is preferred to encourage third parties to independently develop software applications that can use the content and services provided by the system
- People who produce data can integrate those data into the data system.
- Provide a means of capturing legacy data
- Distributed data system, connected by the principle of data sharing and interoperability among linked sites
- Two-way system of both data-in and data-out.
- Provide the users with the base data behind data products
- Assign Digital Object Identifiers (DOI) to datasets
- Accessible through multiple browsers
- Easily maintained

Data Access:

- Provide open access to public data
- Contributors can require user consent to license conditions on data (e.g. noncommercial use only)
- Implement access controls and security to limit access to datasets at discretion of provider
- Data owner retains control of access to all data regardless of where it is stored.

3.1 Approach

One of the basic objectives of the NGDS is to make access to data simpler. A major time consuming aspect of bringing disparate datasets together is data integration. This process involves matching field or element names in the schema for various data sets, selecting those that contain the information of interest, and then merging content into a single data set with consistent usage of vocabulary and units of measure in a standardized collection of fields or elements. Data integration may be done by data providers who choose to deliver data in standardized interchange formats, by data consumers who acquire data in heterogeneous formats and schema and figure out how to extract what they need, or data integration may be done by middleware layers that implement transformations between known formats and schema.

Data integration in our current system of scientific information interchange is mostly left to the data consumer. Until recently, the most common approach was has been for an investigator to collect various datasets and integrate them into a single database that was used for some analysis; some small part of the data might get published, and the compiled dataset was subsequently committed to oblivion. Centralized data aggregation schemes have also been developed and deployed, but rarely outlive project funding or are not maintained and rapidly grow stale due to out-of-date data or use of retired technology. A tremendous amount of effort has been made towards developing systems to promote the management of data such that it may be reused without having to repeat the same integration and cleanup processes over and over.

The path adopted for the Geoscience Information Network to simplify data access and promote reuse is to develop standard formats and access protocols used to deliver common data sets (e.g. borehole temperature data, heat flow measurements) to consumers. The onus of data maintenance is shifted towards organizations that are tasked with data management and preservation. By documenting data schema, encoding formats and practices for vocabulary usage, data can be put into the ‘data integration’ format, or ‘information exchanges’ 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.

This requires education of the data providers/publishers on the use of the information exchanges, but results in a larger community of IT personnel who know how to get data into and out of the information exchanges. Mapping data into an interchange format is likely to be done more accurately by those who originate the data working in conjunction with data managers who understand the interchange formats. The net effect is a greater likelihood that the federated information system using the documented interchange formats will outlast any particular researcher, data provider, project, or agency. HTML on HTTP, NetCDF, and XML are examples of data integration formats that have achieved wide usage and long term usefulness.

The use of schema and encoding specifically designed for data integration and interchange means data producers and consumers can continue to use internal data formats that are optimized for their business requirements. Use of the community interchange formats reduces the amount of work required because only one transformation from internal to interchange format has to be engineered for each interchange format in use.

Data integration by providers introduces additional costs into the data delivery process, and this cost dictates that there must be consideration of the benefits obtained. For data that are not provided using documented interchange formats, detailed metadata describing the schema and encoding of the data will be necessary to enable reuse. The NGDS steering committees originally developed information exchanges based on input from experts in the geothermal community, but now that process has moved to the broader community. Policies regarding defining new, relevant information exchanges and determining what data should be presented in and in what formats, as well as what data are specialized to a degree that data integration by the providers is not warranted for the broader system are determined by the users of the system. Criteria for such decisions will likely include how many providers have a particular kind of data, how often that kind of data are known or expected to be used, the cost of obtaining or reproducing the data, and the expected useful lifetime of the data.

3.2 Requirements

3.2.1 Data discovery

The fundamental use case addressed by a distributed system is to find resources of interest via the internet, based on criteria of topic, place, or time, evaluate resources for an intended purpose, and learn how to access those resources. Detailed metadata describing a resource data schema, describing service or application operation, or providing detailed descriptions of analytical techniques and parameter are outside the scope intended for basic search and discovery metadata. Our contention is that this more domain/resource specific type information is better accounted for with linked documents utilizing schema appropriate to those specific resources. Some examples include OGC getCapabilities, WSDL, and ISO 19110 feature catalogs. Along with a basic search capability, several data discovery components were identified, targeted for software development of a web portal:

- An NGDS Node application, or node-in-a-box, which assists a data provider in sharing their data using the appropriate NGDS standards and protocols, as well as in describing their data using the system's metadata standards
- A web-application capable of creating and dereferencing URIs for resources in the system
- A single, aggregating catalog that maintains a registry of data-providing nodes in the system, and provides a single point of search for data from the entire system
- A user-centered, entry-point web-application where data consumers go to find, evaluate, explore and acquire data in the NGDS

3.2.2 System Architecture Requirements

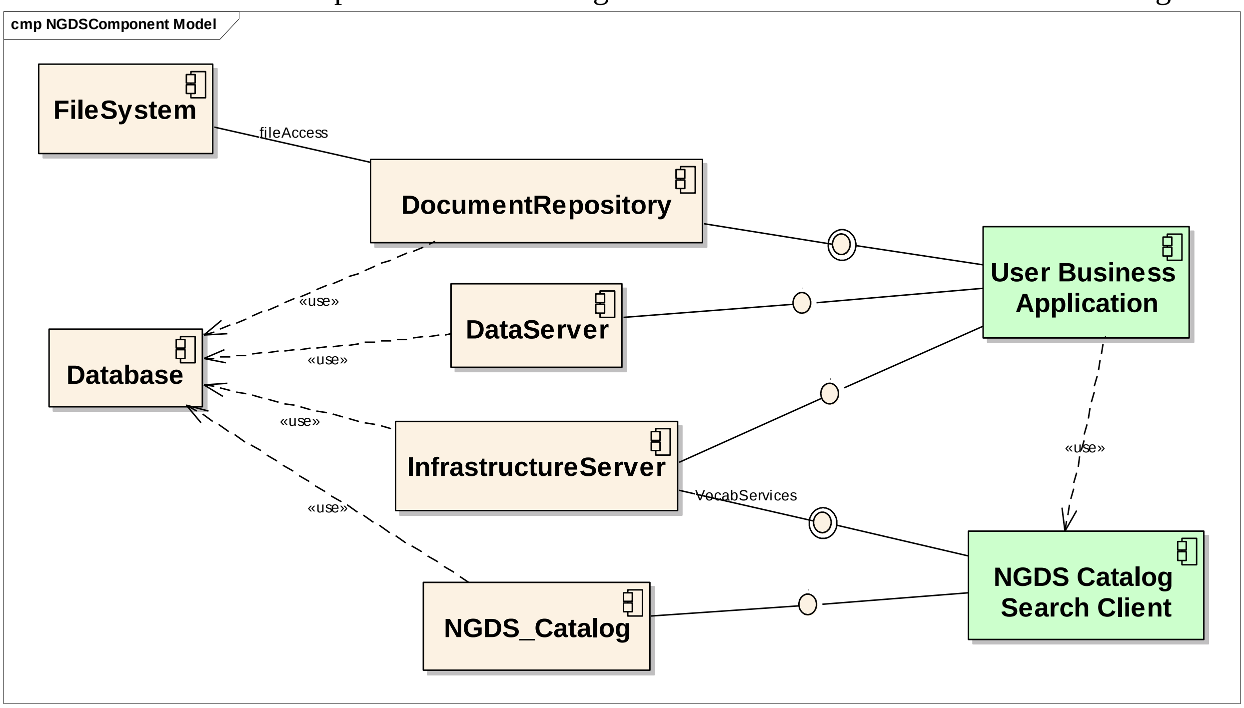
NGDS system architecture and software must support a set of pre-defined functions deemed necessary for the user and as well as for a living system which incorporates technologies and developments without disrupting existing practices. The following is a list of those necessary functions::

- Set of standards, content models and protocols that allow the various components in the system to function interoperable
- Network of distributed nodes that provide geothermal data following the standards and protocols defined by the system
- A variety of applications built to search for, or consume data using the system's standards and protocols
- Searchable catalog that maintains a registry of data provided by all the nodes in the network
- A mechanism for creating and dereferencing unique identifiers (or URIs) for resources in the system. "Resources" include metadata records, datasets, individual rows within a dataset, and physical objects

- Web-application that serves as an single entry point to the system in order to establish and maintain system identity and branding
- Data in the system is categorized in three distinct groups, or tiers
    - Tier 1: unstructured data such as scanned well logs, geologic maps, physical resources (e.g. core, rock or water samples, etc.), or peer-reviewed scientific articles
    - Tier 2: tabular, structured, georeferenced data that does not conform to a content model
    - Tier 3: tabular, structured, georeferenced data that does conform to a content model
- The system defines a set of content models for common geothermal data themes and provides a canonical encoding of these models
- The system defines a standard model for metadata describing data in the system and provides a canonical encoding of that model
- The system defines a set of protocols used to exchange data, metadata and other information between components of the system
- File-based data is transferred via standard HTTP protocols (e.g. GET), or through person-to-person communication in the case of physical or otherwise offline resources
- Tier 2 data can be transferred via standard Open Geospatial Consortium (OGC) web services (e.g. Web Feature Service or WFS, Web Map Service or WMS, Web Coverage Service or WCS)
- Tier 3 data are expected to be transferred via standard Open Geospatial Consortium (OGC) web services (e.g. Web Feature Service or WFS, Web Map Service or WMS, Web Coverage Service or WCS)
- Metadata is transferred via a standard OGC web service, Catalog Service for the Web or CSW

4. System Architecture

The framework for implementing data handling requirements is a community of data providers exposing information through standardized internet-accessible interfaces (services), a community of software developers building applications that will utilize the information resources available to the community, and a community of users taking advantage of the software and information to develop geothermal resources. The service inventory would be focused on entity services that provide information resources. As used here, an entity service is a service that provides a requested resource packaged in some interchange format in response to a request, as opposed to a functional service that takes some input package of information and produces an output response according to some processing logic operating on the input information. A key component is the catalog service —an entity service through which data providers register the availability of resources, and users discover, evaluate, and access resources. The system architecture will be described in terms of the functional components shown in Figure 39. These are discussed in the following



sections.

4.1 Functional components

Figure 38. Functional components of National Geothermal Data System. A variety of implementation choices are available for each of the components. Components on the left are mostly hosted by system servers, and interact with the client components on the right through a collection of interfaces defined by the service profiles.

4.1.1 Catalog

A NGDS catalog component implements one or more protocols for searching a metadata store and returning metadata. At least one of the implemented protocols and interchange formats used for delivering metadata must conform to an NGDS specification. Initial catalog testing and prototypes are using the Open Geospatial Consortium Catalog Service for the Web (CSW), but other protocols such as the Open Archive Initiative Protocol for Metadata Harvest (OAI-PMH) or the OpenSearch protocol may also prove to be useful. The CSW was selected for initial development work because it operates in the same framework as the other Open Geospatial Consortium services being tested for data delivery (the Web Map Service and Web Feature Service), is designed for geospatial data, and has a variety of free, open-source software projects developing clients and servers for the protocol, as well as a variety of commercial products (including ESRI ArcGIS) that are implementing the protocol. The CSW service requires all conformant implementations to return metadata using a simple XML encoding of the Dublin Core Elements and Terms (csw:record), and defines a collection of metadata content elements as core queryable and returnable elements (see OGC 07-006r1). The base CSW specification adds a bounding box as a core queryable requirement for any CSW catalog. Any CSW server must be able to search for criteria based on core queryable elements, and must include the core returnable elements in csw:record XML response documents (although element values may be nil). In addition a CSW service can offer any other xml schema for metadata content, and in the geospatial community, the most widely used profile is for the ISO 19115/19115 metadata. Use of this metadata schema allows richer metadata content that enables greater automation of access to resources. NGDS Catalog instances may be implemented with various software and hardware configurations on any node in the system. To be an NGDS compatible/compliant catalog, the only requirement is that they implement an NGDS catalog service profile, and provide metadata in at least one output format schema and profile that conforms to an NGDS metadata interchange specification.

NGDS metadata content requirements are based on USGIN ISO 19139 and ISO19115/119 metadata profiles for encoding the NGDS metadata content model, outlined at https://github.com/usgin/usginspecs/blob/master/USGIN_ISO_MetadataV1.2_tag.pdf?raw=true. This scheme includes additional metadata attributes and elements for more in depth metadata. Encoding of metadata using the ATOM publishing protocol (http://tools.ietf.org/html/rfc5023) has recently been utilized extensively for describing network resources using a simplified scheme similar to csw:record, but with more structured XML to promote greater interoperability. The intention of the USGIN approach is that a small number of these encoding schemes would be adopted, with mappings allowing lossless conversion of content between schemes, allowing implementation of software metadata clients with advanced functionality to streamline user access to the actual described resources.

4.1.2 Document Repositories

Data in documents will be accessed via URL from document repositories, which are basically web-accessible file systems. In this context, ‘document’ is used in a very general way as a packaged body of intellectual work with an author (or editor, compiler, or similar originating role), a title, and some status with respect to Review/authority/quality. Documents can be packaged in a single file (e.g. a MS Word document) or a group of related, linked digital files (e.g. ESRI Shape file). Documents provide a straightforward path to get data online quickly and easily for the data provider, but if this approach is used for datasets (e.g. Excel spreadsheets, Microsoft Access databases), it requires the data consumer to do all data integration work themselves. In addition, for the datasets to be useful for data consumers, the metadata descriptions must clearly define the entities and attributes (or features and properties) of the datasets such that users can understand their meaning.

Many options are available for implementing document repositories, including DSpace (FOSS, http://www.dspace.org/), OCLC ContentDM (commercial), Fedora (http://fedora-commons.org/), and the Drupal-based document repository developed in collaboration with the USGIN project (http://repository.stategeothermaldata.org). In order to integrate holdings in system document repositories, a system repository must make available metadata for contained resources using a NGDS metadata interchange format that can be inserted into the NGDS central node. This metadata must contain the required minimum content to allow discovery and access to any document in an NGDS repository, including a URL that will retrieve the resource.

4.1.3 Data Servers

A Data Server is any component that implements a service providing data using at least one protocol and interchange format conforming to an NGDS specification. Data service delivery of content differs from the simpler document-based delivery because it requires that the format and content delivered will conform to some know set of rules, allowing software to interact directly with the data server to facilitate user acquisition and integration of data into their work environment.

Data delivery through a service requires the service provider to perform any necessary data integration operations to get content into the schema conforming to the service profile. This requires more work for the data provider than the simpler document deliver approach, and thus will have to be implemented incrementally based on the quantity and significance of various data items. Data types that are deemed suitable for service delivery will have NGDS protocols, interchange formats, and vocabularies defined to enable automated access to those data.

Since many of the data types are associated with geographically located features, the Open Geospatial Consortium Web Feature Service (WFS) is proposed as the starting point for implementation of feature services. This protocol uses GML geometry for location description, and allows feature types to be defined that are characterized by feature specific xml schema.

A number of international efforts are under way to develop specifications for data interchange of geoscience information (GeoSciML), and basic observation and measurement data (ISO19156). These xml schema are very flexible to allow representation of a wide range of content, but are thus correspondingly complex. Currently there are no client applications that can do more that transform complex xml to html for display.

Thus, services are defined using simple xml schema with string and numeric-valued elements. These services can be consumed by existing clients like ArcMap and Quantum GIS. Simple feature schema will be compatible with GeoSciML, ISO specifications, and other complex standard schema to the degree that is practical. As clients are developed for richer-content complex feature services, the NGDS can adopt more complex, information-rich schema. There are also a number of other data formats in use in related communities for geoscience information interchange, including WaterML in use by the CUAHSI project, NetCDF, which is widely used for large numeric data sets in the atmospheric and remote sensing communities, and an xml markup developed for geochemical data by the EarthChem project. NGDS has used such schemas as a basis for constructing current information exchanges so that uniformity and interoperability in the science community are more likely to be achieved.

4.1.4 Metadata and Tiered Data

To make any resources available to NGDS, metadata must be created for that resource, containing files, URLs, and other distributions. To be accepted in the system, a metadata record would be created for a resource and then loaded into a catalog server or web-accessible directory that is then harvested by the NGDS. The metadata record must provide the user not only the fact of the resource's existence, but to give the user adequate information to evaluate and access it. Required metadata is outlined in ISO19115/19139 metadata standards and in the following document: https://github.com/usgin/usginspecs/blob/master/USGIN_ISO_MetadataV1.2_tag.pdf?raw=true.

Individual documents require one metadata record per document. Some document types may consist of a bundle of files, e.g. ESRI shape file. In general these should be bundled into a single file like a zip archive or UNIX tar file. The metadata must include the URL at which the document can be accessed. These documents might be scans of well logs, scanned reports or publications, or data in a spreadsheet, such as an Excel file.

Datasets include internal record level source information, documenting details of observation or measurement procedure and other information specific to a particular data type. This includes information such as location, data and time of observations, and the source of the data. These metadata are delivered with the data, and only summarized in the dataset metadata that are published to the NGDS-compliant catalog.

The actual mechanics of bringing particular datasets online will be dependent of the format of existing data, and the IT resources of the data owner. Some organizations may choose to implement web services on their own servers to expose datasets, others may choose to work with a partner or hub that has better IT support to host services.

4.1.4.a Tier 3 Structured, Standardized Data

The top tier (most desirable) data types in NGDS conform to specified schemas given in information exchanges. These data are then made available as Open Geospatial Consortium (OGC) services, particularly Web Map Service (WMS) and Web Feature Service (WFS), with the exception of data provided using the Metadata information exchange. This information exchange outlines ISO 19115/19139 metadata standards necessary for metadata delivery describing any resource. Bulk uploads of metadata provided in this content model (generally translated from an original database or catalog to the provided Excel content model, then tranformed into XML for catalog management) is thus also considered standardized data.

Data which conform to NGDS interchange formats are made available in user-defined data files described by metadata in the system catalog and placed in web-accessible servers. Standardization of interoperable data services and community interchange formats have been developed for NGDS as an ongoing, living process for the geoscience community. The infrastructure that supports the normative schema locations is at https://github.com/usgin/modelmanager which houses the code base that supports the Django NGDS schemas management site http://schemas.usgin.org/models/. The information exchanges are developed and maintained at https://github.com/usgin-models, where GitHub repositories exist for each given data type. These tagged versions of schemas are then implemented at the aforementioned repositories and sites for use in the system. The community of users are tasked with continuing development and versioning of new or existing information exchanges as needed. A detailed description of the workflow for defining a new information exchange can be found at https://github.com/usgin/usginspecs/wiki/Define-New-Information-Exchange, with a detailed description of NGDS standards for constructing them at https://github.com/usgin/usginspecs/wiki/Content-Model-Guidelines. Some important requirements for creating new information exchanges include:

4. Ensuring interoperability among data sets with members adopting common standards and protocols.
5. Data schema must be vetted with stakeholders.
6. Data schema for interchange formats and instance documents based on these schema must be versioned, such that expanded or modified versions can be introduced without disrupting working systems.

4.1.4.b Tier 2 Structured, Non-Standardized Data

Tier 2 data sets do not have standard interchange protocols or specifically defined file formats, but are structured data usually in tabular file format. These provide a simplified exchange of some kinds of information, and if widely used these would be obvious candidates for system interchange formats. The recommended metadata for tier 2 resources is designed to allow discovery, evaluation of the resource based on text description, and access to the resource via a web link (URL).

4.1.4.c Tier 1 Non-Structured Data

All other data, not in a structured form, is considered tier 1 data. Reports, logs, maps and other documents pertinent to geothermal energy exploration, evaluation, development, and production that have been converted to digital form by scanning to create digital image files contribute to the 'legacy' tier 1data collection in NGDS. Tier 1 map resources should be georeferenced (geoTiff or world file) if possible. Preferred image and document formats are pdf, tif, jpg, or png. File formats that are specific to particular (especially proprietary) software are undesirable, as the system is built on ease of accessibility. OCR processing of text to make Adobe Acrobat files searchable is highly desirable. Georeferenced map images ideally will be published through a Web Map Service (WMS) as well as accessed from document repositories. Digital documents must be publicly available online. Those resources that are not online, such as a core samples facility, are indicated in NGDS as 'offline resources'. These resources are simply identified by location and other information in the metadata record.

4.1.5 Database and File System

Various databases and file systems accessed by server applications will house the actual system resources. For security and simplicity, these are not directly accessible for system users, but are accessed through NGDS and other client-side interfaces. Many user applications (like the Geothermal Prospector) may also have local data stores, in databases or file systems, used to cache resources obtained from the system for offline usage, better performance, and reliability.

A document repository implementing using Drupal software (with a coucbDB, node.js back-end) was created as an online repository for state geothermal data collection (http://repository.stategeothermal.org/). This application also supports production of metadata meeting NGDS requirements. Instructions for deployment of a similar system is available to interested users at http://lab.usgin.org/groups/drupal-development/creating-document-repository-drupal.

4.1.6 Clients

The client applications implement most of the desktop analytical and search functionality required by the system. The user interface developed in CKAN (see section below) ingests OGC CSW endpoints like that of the state geothermal catalog for resource discovery. As the system is built to encourage client-side development, other applications are meant to be built on top of the NGDS frameword and provide in-depth evaluation and manipulation of NGDS data. See Section 4.2.2 for specific client-side development examples.

4.2 System deployment

4.2.1 Nodes

Any server that is internet accessible and implements one or more NGDS services, including document and other data repositories containing files indexed by metadata in a catalog made available to NGDS, is effectively a node in the system (Figure 2). Each node will implement one or more of the abstract components shown in Figure 1, and will need to register public resources available at that node in the system.

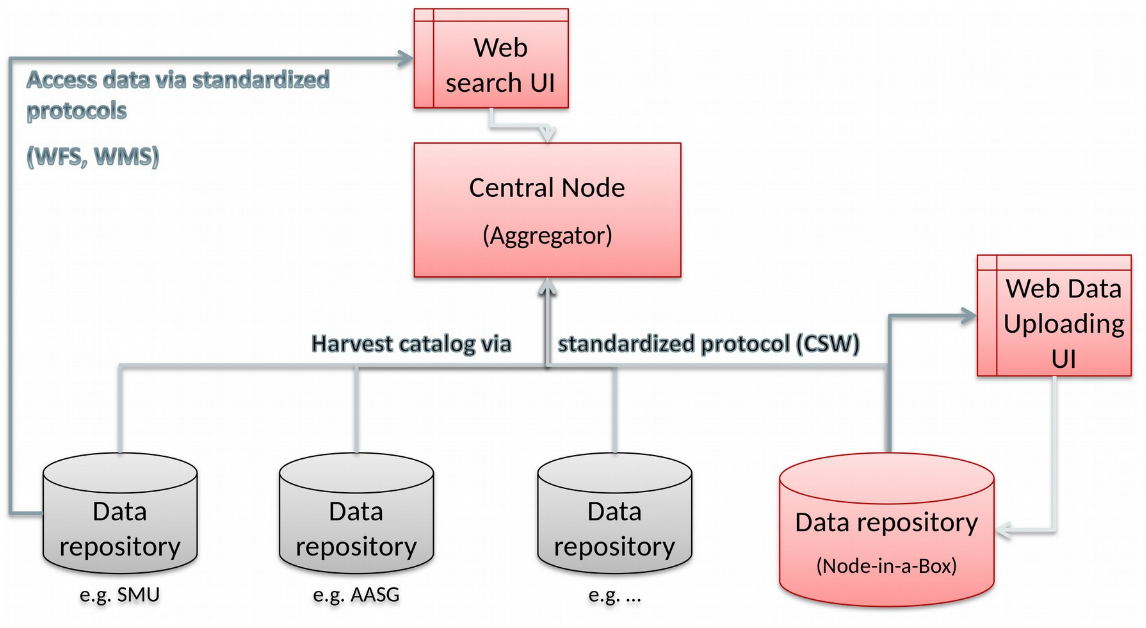


Figure 2. Deployment of nodes in the system. Nodes will implement functions including archives, system specification repositories, and registries of identifiers, as well as standard catalog and data services. Other nodes will implement catalog and data services, and may provide applications that utilize data resources as well.

The deployment diagram indicates a key aspect of the system—the user client software interacts with components on the server side. This connection represents any and all service protocols used to link clients and data servers in the system. These services define interfaces that decouple the clients and servers. Upgrades or modifications to client or server software that do not change the operations and behavior of their service interface will not break the system. This loose coupling is a key design feature necessary to allow the system to evolve as technology and user requirements change.

4.2.2 Node-In-A-Box Software Stack for Nodes and Central Aggregator

Node-In-A-Box (NIAB) is a software stack created on an open-source (OS) platform with OS code and OSGS components, developed as an executable file to install a data management system. Metadata from publications, web services, and other data can to be fed into the main system aggregator, or central node, at http://www.geothermaldata.org easily from a NIAB installation. As system architecture allow the front and back end to be decoupled, use of this software stack is not necessary to be a data provider to NGDS.

- NGDS developed one software stack that provides the functionality for both central node and Node-in-a-Box. The reason for this decision is that so many features of both node types are the same that it is easier to develop one software stack and configure the behavior of the system with a configuration file. In order to do so a new parameter to CKAN's configuration file ("development.ini") that defines the behavior of a node during startup was added. The node can either be configured as Central Node or Node- in-a-Box. As a Central Node the NGDS software provides the harvesting but no uploading capability while as a Node-in-a-Box the NGDS software provides content uploading but no harvesting capability. The NGDS central node (also known as the main aggregator) is

Figure 3Deployment of System components.

at http://www.geothermaldata.org, where discovery of NIAB and other data nodes of differing configurations is accomplished in several ways:

- The NGDS Map Search, http://www.geothermaldata.org/ngds/map, which allows a keyword or drawn bounding box search for GIS data as well as geo-located publications and other resources.
- The NGDS Library Search, http://www.geothermaldata.org/dataset?_tags_limit=0, having keyword as well as faceted search functions. The faceted searches can be performed using authors, contact, broad categories, and by content model (information exchange types).
    - Search hints, using special terms, wildcard searches, and special characters is available at http://ngds.github.io/documents/SearchHints.htm.
- Geothermal Prospector preview, https://maps-stage.nrel.gov/geothermal-prospector/, which is provided in the Dataset Details page for each resource that has a WMS and/or WFS distribution.

As discussed in Section 4.1.6, system design dictates that any software client can access the data in the system given the predictable data types and web service distributions. Some of those client-side access options include:

- The Data Explorer, http://data.geothermaldatasystem.org/
- The USGS National Map Viewer, http://viewer.nationalmap.gov/viewer/
- Google Earth, http://www.google.com/earth/
- Any number of FOSGS (free and open source geospatial software) programs

See http://www.geothermaldata.org/ngds/resources for more information on user access.

4.2.3 Software Operating System Support

Ubuntu 12.04 LTS was chosen as the NGDS reference platform. Due to the nature of NGDS, most users will most likely want to run it on a Linux Operating System. Ubuntu is a well-known and well-documented Linux OS. Also, CKAN is optimized for Ubuntu or other Debian-based Linux distributions.

NGDS has been developed with Ubuntu Version 12.04 LTS because it has a long support cycle. When the next LTS version becomes available the NGDS development environment will be upgraded. Installation files, etc. are written in such a way that they will likely work on any Debian-based Linux version. Therefore, porting to other Linux platforms is a minor effort. NGDS is also frequently tested and installed on the Mac OS X platform but NGDS is not test installed the system on Windows. Most likely it is possible to start the system on Windows but to fully support that would drain too many development resources and it is questionable if this form of testing adds much value.

4.2.4 Software Base Back-end Technologies

NGDS relies heavily on Python and CKAN. This defines the architecture of NGDS to a great extend because CKAN is a framework with well-defined extension points where NGDS functionality has been added. Further, CKAN is used as-is without modifying the CKAN core. Bugs are found in the CKAN core are reported and tracked through the Open Knowledge Foundation's support team and their tools.

Since CKAN is targeted towards Postgres, NGDS's reference database is also Postgres. The PostGIS extension of Postgres is used for geographic features.

In order to serve OGC services NGDS uses Geoserver which runs in its default setting on jetty.

For indexing of metadata and full-text indexing NGDS uses SOLR (on jetty). SOLR is configured according to the CKAN recommendations.

4.2.5 Software Base Front-end Technologies

NGDS uses HTML5 and CSS3 as the base technology for the frontend. HTML5 and CSS3 are now supported by all major browsers (even later versions of Internet Explorer). In addition NGDS uses various JavaScript Libraries. Specifically, NGDS uses JQuery and various libraries based on JQuery. Furthermore, NGDS uses CSSless in order to reduce CSS complexity. At the time of writing of this document these are the state-of-the-art technologies for developing Web applications.

For the production of the HTML pages, NGDS uses the Jinja2 templating system that is built into CKAN. The templating system can be compared to PHP and allows to bring backend-information (made available via Python) into the frontend HTML content.

Jinja2 is very flexible and works well in the CKAN environment. The reason for using Jinja2 is that it is the reference solution for CKAN. Building other templating systems into CKAN would be extra effort and causes problems regarding maintainability.

In many cases it is possible to create the page by either using Jinja2 templating or JavaScript. Jinja2 is preferred rather than JavaScript whenever possible because Python code and Jinja2 code is in general easier to maintain than JavaScript code.

The Leaflet Map Widget is used to present maps (http://leafletjs.com/). This widget is currently popular and has an active developer community. OpenLayers (http://openlayers.org/) was considered but due to higher experience with Leaflet among the NGDS developers, it was decided to go with Leaflet.

Naturally it is not possible to simply replace Leaflet with another map widget solution. However to the extent possible, the Leaflet-specific code is encapsulated in separate classes and modules so that the classes can be replaced in the future in case that the map widget needs to be exchanged.

5. Summary

The central idea of the data access architecture proposed here is that data providers and client applications should be linked through open source interfaces that decouple clients and servers such that they can evolve independently without breaking the system. The hypertext transfer protocol (http) and hypertext markup language (html) are the established protocols and interchange formats in use on the internet, and in the near term these will probably continue to be the mainstay of most interaction in the NGDS.

The OpenGeospatial Consortium Catalog Service for the Web (CSW), currently at version 2.0.2 is the required catalog search and discovery service. The lowest common denominator metadata interchange format using this service to achieve interoperability between metadata provided by various servers is outlined at http://schemas.usgin.org/models/#Metadata and https://github.com/usgin-models/MetadataCompilation, which uses the USGIN profile for ISO metadata.

6. Glossary

Definitions here are meant to clarify the usage of terms in this document.

Artifact: A thing created by humans, usually for some practical purpose. (Source: http://www.merriam-webster.com/dictionary/artifact)

Attribute: A binding between a property, a data type, and a data item; an implementation of a property.

Cardinality: A constraint on the number of instances of assigned property values associated with an individual data item. A cardinality of 1 indicates exactly one value is required; 0..1 indicates an optional single value; 1..n indicates that one or more values is required; 0..n indicates that a value is optional, and multiple values may be specified.

Content model: A model that identifies and defines the data items and the properties (with cardinality) associated with each data item.

Data integration: the process matching field or element names in the schema for various data sets, selecting those that contain the information of interest, and merging content into a single data set with consistent usage of vocabulary and units of measure in a standardized collection of fields or elements.

Data item: An identifiable unit of information. Generally represents some entity in the world.

Data type: A specification of the representation of a single value in an information system, using integer, floating point, string, Boolean.

Entity service: a service that provides a requested resource packaged in some interchange format in response to a request, as opposed to a

Feature type: Type for representing a feature.

Feature: An information resource representing some identifiable thing of interest in the world.

Functional service: a service that takes some input package of information (message) and produces an output response (message) according to some processing logic operating on the input information.

Information resource: A resource that can be transmitted electronically.

Interface: a point of interaction between components, typically defined by a protocol for transmitting messages and a collection of method names and parameter specifications used to invoke operations executed by a component.

Interoperability: "The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units." ISO/IEC 2382-01 (SC36 Secretariat, 2003)

Observation: an information resource representing the event of observing or measuring and recording properties of some feature (Open Geospatial Consortium, Observations and Measurements (O&M), http://www.opengeospatial.org/standards/om). Observations represent the basic data that are the foundation for scientific knowledge.

Operation: an individual process that a software component may execute

Property: A phenomenon that is inherent in the nature of some other phenomenon, and may be used to characterize it by specifying a value.

Protocol: A set of rules which is used by computers to communicate with each other across a network (http://en.wikipedia.org/wiki/Network_protocol).

Representation: A binding between a symbol or collection of symbols (in language, text, graphics, computer bits, etc.) and a human concept or resource.

Resource: An identifiable thing that fulfills a requirement. Usage here is close to definition used in RDF (http://www.w3.org/TR/REC-rdf-syntax), generalized from ISO19115, which defines resource as an 'asset or means that fulfills a requirement' without defining asset or means. "An object or artifact that is described by a record in the information model of a catalogue." (OGC 07-006r1)

Schema: A formally structured representation of a conceptualization. A model presented using some specific notation.

Service: A system that provides one or more functions via a network interface designed for machine interaction; utilization involves some agent making a request and possibly providing some input, at which point the service executes the requested procedure with some predictable result

Specification: A document that describes the technical characteristics of an artifact, possibly including a description of what it should do, or an explicit set of requirements that it must satisfy. (Based on http://en.wikipedia.org/wiki/Specification).

Type: Specification of a collection of attributes and cardinalities for those attributes used to represent a data item.

References

Dublin Core Initiative, 2008-01-14 Dublin core Metadata Element Set, Version 1.1: Dublin Core Metadata Initiative, accessed at http://dublincore.org/documents/dces/.

Franklin, Michael, Halevy, Alon, and Maier, David, 2005, From databases to dataspaces: a new abstraction for information management: ACM SIGMOD Record, V. 34, No. 4, ISSN:0163-5808.

Open Geospatial Consortium Catalog Service for the Web (CSW): http://www.opengeospatial.org/standards/cat

Open Archive Initiative Protocol for Metadata Harvest (OAI-PMH): http://www.openarchives.org/pmh/

OpenSearch protocol: http://www.opensearch.org/Home

ISO 19115:2005, Geographic information - Metadata

ISO 19115/Cor.1:2006, Geographic information – Metadata, Technical Corrigendum

ISO 19119:2005, Geographic information - Services

ISO 19119:2005/Amd 1:2008, Extensions of the service metadata model ISO 19108 designates:

OGC 07-006r1¸ OpenGIS Catalog Services Specification version 2.0.2, Corrigendum 2 release, 2007

OGC 07-045, OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile, Version 1.0.0, 2007

USGIN Team, 2010, Use of ISO 19139 xml schema to describe geoscience information resources v1.1, Current version at http://repository.usgin.org/uri_gin/usgin/dlio/337.

USGIN Team, 2010, Metadata Recommendations for Geoscience Resources, current version available at http://repository.usgin.org/uri_gin/usgin/dlio/335.

How to be an NGDS node

NGDS Specification Version 1.0.
Last edit: 2016-10-04

History:

- 2013-05-17 Accepted by the NGDS GDSDP Workgroup.
- 2013-10-08 Minor revision to manifest file provisions made by SMR based on comments from Matt MacKenzie
- 2013-10-08 Post v1.0 to GitHub

Contents

Introduction

This document proposes requirements that must be met in order to label an internet resource 'NGDS node'.

NGDS Node requirements

An NGDS node is a web-accessible server that hosts at least one of the functional capabilities enumerated below to play a role in making geothermal-relevant data assets accessible.

To be considered a node in the National Geothermal Data System, the following criteria must be met:

1. The capabilities offered MUST play a role in making geothermal-relevant data assets accessible.
2. The capabilities offered MUST be publicly Web-accessible.
3. A node MUST offer at least one of the following capabilities:
    a. Host a Web-accessible repository of geothermal-relevant data assets with metadata conforming to the USGIN ISO profile (USGIN Standards and Protocols Drafting Team, 2010-11) published through an NGDS catalog node.

b. Host a web-accessible folder that 1) contains NGDS-conformant metadata files and 2) is registered for harvesting by an NGDS catalog node.
c. Host NGDS-conformant web services (WMS, WFS, WCS, etc.) that are registered in an NGDS catalog.
d. Host an NGDS catalog node, which is a server operating a CSW 2.0.2 service that offers metadata conforming to the USGIN ISO profile.

NGDS nodes SHOULD self-identify by providing an **NGDS node self-description document**. The NGDS NODE SELF-DESCRIPTION DOCUMENT section (below) outlines proposed content that describes the node and enumerates the NGDS capabilities offered. The template is intended to be a hypermedia document that can be used by web applications to automate connection to node resources. If provided, the self-description document MUST be accessible at http://hostName.xxx/.well-known/host-meta, following the pattern specified by IETF RFC 6415 [http://tools.ietf.org/html/rfc6415].

Details

Hypermedia: Extension of text as an encoding scheme to any media stream that includes information and controls through which the user (or automaton) obtains choices and selects actions. (R. Fielding, 2008, http://roy.gbiv.com/untangled/2008/rest-apis-must-be-hypertext-driven (and discussion))

Metadata conforming to USGIN ISO profile: See "Use of ISO metadata specifications to describe geoscience information resources" (USGIN Standards and Protocols Drafting Team, 2010-11) for detailed guidance on this profile, and "Guidelines for implementation of USGIN metadata content recommendations in ISO 19139 XML metadata" (USGIN Standards and Protocols Drafting Team, 2012) for additional implementation details.

Metadata registered in an NGDS catalog node: an arrangement is in place so that the registering party creates metadata records that are harvested into the target catalog node. Metadata may be created through a forms interface directly to the catalog, harvested from a non-NGDS catalog end point (e.g OAI-PMH, open-search, ATOM etc.), harvested from a Web Accessible directory, or harvested by a file transfer/upload workflow

NGDS catalog node: a server operating a CSW 2.0.2 service that offers metadata conforming to the USGIN ISO profile that describes resources intended as part of the NGDS. Such nodes MUST be registered with the NGDS core aggregating node, either directly or indirectly through another node, such that its metadata may be harvested.

NGDS core aggregating node: The central system administration node, includes a catalog repository that harvests metadata from all NGDS catalog nodes. The NGDS core portal accesses this repository for system-wide searches.

NGDS conformant web service: a web service that implements an NGDS data exchange, defined by a service protocol, interchange format, and exchange-specific profile.

NGDS portal: a web application (NGDS client) that provides a user interface for searching the entire NGDS catalog.

NGDS home page: a web site that is the public face of the NGDS, providing access to the portal, information about the system, tutorials, governance information, and specifications.

NGDS roles:

- **originator**: agent that created the intellectual property (work)
- **contributor**: agent that prepares a resource for publication and makes it available
- **submitter**: agent that generates metadata to register the resource
- **owner**: agent that holds the intellectual property rights for a resource. Copyright holder, legal role.
- **provider**: agent that operates the service endpoint from which a resource is obtained
- **steward**: agent that is responsible for the maintenance/update of a resource

Discussion

A node is a server that provides NGDS resources, not a client that uses the offered resources. A server that hosts applications utilizing NGDS services is a vital part of the system, but is not considered a node. Such applications would be expected to be registered in the system by posting metadata to an NGDS catalog.

Applications enabling search across the entire NGDS catalog are an essential part of the system. At least one such search application will be recognized as the official portal to the system, accessed from the NGDS home page on the web.

Each node is under the stewardship of a data provider. Individual organizations may host more than one node. For example different servers may host OGC web services, a CSW service, and a repository containing document resources, all under the stewardship of a single data provider. These would be considered three distinct nodes. A node may host data from more than one data contributor.

If a server is hosting resources that are described by metadata in an NGDS catalog, and those resources meet the 'geothermal-relevant' criteria, then the operator of that node MAY declare it to be an NGDS node, but doesn't have to.

Putting an NGDS self-description document on the server at the specified location constitutes a declaration that that server considers itself an NGDS node. One of the functions of the self-description document would be that a system registry of known NGDS nodes could have a web crawler application that searches for such self-description documents and starts a node-registration/approval/validation process when it finds new ones. This would enable a 'pull' model registration.

Functional NGDS node categories

Both data resources and metadata can be made accessible either packaged in files or through services that offer capabilities extending HTTP. Thus NGDS nodes can be categorized according to whether they present information in files or through web services. An individual node my offer both kinds of information access.

Files-only node

A publish-only node hosts web accessible files that are either information resources registered in an NGDS catalog, or NGDS-conformant metadata records in a web-accessible folder registered for harvest to an NGDS catalog.

Service node

A service node offers at least one NGDS service that is registered in an NGDS catalog, either through the metadata for datasets exposed by the service, or as a catalog service.

Catalog Node

A service node that offers a CSW service to search the metadata collection hosted by that node

Data node

A service node that offers at least one registered data service (e.g. WFS, WMS, WCS, OpenDAP)

NGDS Node Self-Description Document

Every NGDS node SHOULD offer an NGDS Node self-description document available via http GET at standard path. Recommended practice is to use the IETF "/.well-known/host-meta" path (see http://tools.ietf.org/html/rfc6415).   Because many web-server configurations are set up to hide dot files (files or folders with names beginning with a '.') from directory browsing or even direct access, it is also recommended to a simple JSON-formatted manifest file in the root directory of the server. This file should contain the same information.

This document is a representation of the node resource.   The purpose of this document is to make the node self-describing, both as a Hypermedia document for machine clients and as a web page for people to read. The document would be a publicly accessible HTML document, discoverable using standard web search technology (Google, Yahoo, Bing, etc.), with text description of the node, the node operator, node capabilities, expected lifetime, acknowledgements for support, etc. The template below defines an HTML profile for including information and links that a software agent could use to inspect the capabilities and access desired function. This might be encoded in HTML link elements in the head section, or using 'rel', 'class', and 'type' attributes on HTML anchors or other elements in the document. Some work needs to be done to define the vocabulary of 'rel' or 'class' attributes, or microdata properties

The content for the self-description page would include the following:

1. Node Name: [Name to identify this node]

2. Supporting organization: [name, contact point to report issues]
3. Description: [is there any topical focus, group of contributors, other special characteristics of capabilities offered]
4. Node type: term from controlled vocabulary to indicate type of node. One of {filesOnly, catalog, dataService, catalogAndDataService}
5. Resources:
    a. Data services: List of service types, links to end points, like ArcGIS server REST page?
    b. Metadata publishing profile: how exposed (WAF, CSW, other, including specific profiles used?)
6. How can content of this node?
7. URL for RSS feed for service announcements and information broadcasts.

See the accompanying example document. When the content is established, this could be converted to a template to facilitate creation of new documents for NGDS nodes.

Standards Statements

Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in IETF RFC-2119 [RFC2119].

Status of this Memo

This NGDS Specification licensed under the CC BY license. NGDS Specification Drafts are working documents of the National Geothermal Data System Technology Working Group. Note that other groups may also distribute working documents as NGDS Specification Drafts. Current NGDS Specification Drafts can be accessed in the NGDS/System-design gitHub repository.

NGDS Specification Drafts are draft documents valid until rejected or adopted by the NGDS GDSDPWG, and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use NGDS Specification Drafts as reference material or to cite them other than as "work in progress."

References

USGIN Standards and Protocols Drafting Team, 2010, Metadata Recommendations for Geoscience Resources: U.S. Geoscience Information Network Best Practices Document, Doc ID gin2010-11, v. 1.0.3, available at http://lab.usgin.org/profiles/doc/metadata-content-recommendations (accessed 2010-11-18).

USGIN Standards and Protocols Drafting Team, 2012, Guidelines for implementation of USGIN metadata content recommendations in ISO 19139 XML metadata: U.S. Geoscience Information Network Best Practices Document, available at http://repository.usgin.org/uri_gin/usgin/dlio/499.

USGIN Standards and Protocols Drafting Team, 2010-11, Use of ISO metadata specifications to describe geoscience information resources, Doc ID gin2010-009, available at http://repository.usgin.org/uri_gin/usgin/dlio/337.

Appendix 5

Sustainability Plan

Contents

M. Lee Allison, PI

Arizona Geological Survey Geoinformatics Section
Stephen M. Richard, Chief
Christy Caudill Daugherty, Deputy Chief

Acknowledgements and Disclaimer

The National Geothermal Data System is based on a collection of protocols for metadata and data inter-change. There is no single system data model, rather there is a metadata interchange format and a col-lection of data interchange formats (the NGDS content models, see Data Access Phase 1 deliverable document). Each interchange format defines a data model for an information item of interest, and is de-signed to access particular information tailored to some use case or use cases. Data providers and client applications are linked through open source interfaces that decouple clients and servers such that they can evolve independently without breaking the system. The OpenGeospatial Consortium Catalog Service for the Web (CSW), currently at version 2.0.2 is proposed for catalog search and discovery service. The USGIN profile for ISO metadata defines the metadata interchange format for the NGDS, and is already in use by the NGDS catalog deployed for the AASG Geothermal Data Project. Initial data services are being implemented using WFS 1.1.1 simple feature services, with interchange formats defined for key geo-thermally interesting datasets. Interchange content is encoded as GML Simple Features, an XML application scheme. File-based data are incorporated into the NGDS by creating a metadata record conforming to the metadata specification and inserting it into the NGDS catalog. Files containing the data must be in a web-accessible location specified by a URL in the metadata record.

Background

Since the 2009 American Recovery and Reinvestment Act (ARRA) the U.S. Department of Energy's Geothermal Technologies Office has funded $33.7 million for multiple data digitization and aggregation projects focused on making vast amounts of geothermal-relevant data available to industry for advancing geothermal exploration and development. These projects are collectively part of the National Geothermal Data System (NGDS), a distributed, online networked system for maintaining, sharing, and accessing data in an effort to lower the levelized cost of electricity (LCOE). Most of the projects involved in the NGDS conclude on or before April 30, 2014. The exception is the "State Geological Survey Contributions to the NGDS" project which received a no-cost extension through December 31, 2014 to deploy the Node-in-a-Box (NIAB) developed under the Architecture, Design, and Maintenance award to Boise State University. The investment in building and populating the NGDS has been substantial, both in terms of dollars and time; it is critical that this investment be protected by ensuring sustainability of the data, the software and systems, and the accessibility of the data. Only then, will the benefits be fully realized.

The NGDS vision is that the functions, accessibility, and availability of data will encourage continued participation within the NGDS, from all system users – data consumers, data providers, and application developers. As each data node adds to its data repositories, the system-wide NGDS functions become increasingly valuable to it. Each data provider will have created a value-added service that is transportable and scalable to cover all data in its possession, an example of such is the recent adoption by the State of Arizona to deploy a network built on the NGDS underlying framework for natural resource, environmental, and transportation data. Thus, there are benefits to each participant to continue to add data to the system and maintain it. The long term goal is that the data network reach a 'tipping point' at which it becomes like a data equivalent to the World Wide Web – where everyone will maintain the function because it is expected by its clientele and it fills critical needs. Applying this vision to the NGDS, it also opens the door for additional data providers external to geothermal development, thus increasing the value of NGDS and its underlying data integration platform, USGIN.

This plan is submitted to the NGDS Architecture, Design, and Testing project as a final deliverable for the April 30, 2014 project end-date. It should be noted that the plan is a living document and will be continuously updated throughout the course of 2014 to reflect the most up-to-date developments, information, and resources available to the NGDS. The plan builds off of the May 2013 vision paper "Sustainability Plan for National Geothermal Data System (NGDS) Operations," available via the NGDS GitHub, and describes the four pillars of sustainability: System Operations, Maintaining Content from Providers, Organization and Management, and Business Models.

In May of 2013, it was anticipated that NGDS would not be in full independent sustainable mode upon the completion of the Architecture, Design, and Testing project in April of 2014; this conclusion remains. This is due to a number of factors including the original ambiguity on the ownership and functionality of NGDS prior to the redirection of the Architecture, Design, and Testing project, which completed development of one of the core pieces of the NGDS in April of 2014, an easy mechanism for nodes to provide and share data.

Ownership of the NGDS has been assigned to the USGIN Foundation, Inc. the non-profit organization that serves as the support group to the USGIN platform. USGIN Foundation, Inc. is set up initially under the auspices of the Arizona Geological Survey. Additional details are provided in the text of this document. As anticipated in May 2013, there is a required transition period post April 30, 2014 to continue maturation of the product, including deployment of the NIAB, requiring transition and maintenance support from the primary funding agency, US DOE. Transition and maintenance support is currently guaranteed through April 30, 2015 through a contract from Sandia National Laboratories (option to extend support up to 4 years in one-year increments).

Summary Conclusions

Based on the recommended actions in the May 2013 Sustainability Document, and the current state-of-development for the NGDS, the following can be concluded:

- USGIN Foundation, Inc. will be the long term owner of the USGIN data integration platform and the NGDS software, including the CKAN node package.
- USGIN Foundation, Inc. will include a Geothermal Advisory Committee to ensure that NGDS interests are pursued
- USGIN Foundation, Inc. will market the USGIN system as a solution to the Federal Open Data Access Executive Order and OMB Implementation Memo
- Upon conclusion of the primary data digitization and aggregation projects a snapshot repository of the data available via the NGDS will be backed-up via cloud storage and utilized should one of the contributors nodes fail
- Upon conclusion of the primary projects at least 8 data nodes and 3 regional hubs will be in operation with the hubs providing replication and disaster recovery

- During the transition period, , the NGDS CKAN extension will be cleaned of unused and redundant code to ensure ease of use for external developers in the free and open source CKAN network, enhancing opportunities for reuse (and thus updates to the code)
- During maintenance mode the NGDS CKAN extension will receive added improvements based on user testing, including web site revisions to improve the user experience

The National Geothermal Data System (NGDS) Vision

Goal

The ultimate goal of the National Geothermal Data System (NGDS) is to support the discovery and generation of geothermal sources of energy. The NGDS will provide online access to important geothermal-related data from a network of data providers in order to:

- Increase the efficiency of exploration, development and usage of geothermal energy by providing a basis for financial risk analysis of potential sites
- Assist state and federal agencies in making land and resource management assessments
- Foster the discovery of new geothermal resources by supporting ongoing and future geothermal-related research
- Increase industry and public awareness of geothermal energy development potential

User Communities

The National Geothermal Data System is a network of three linked communities:

4. **Data providers** who will expose information to the system through standardized, internet-accessible interfaces and interchange formats
5. **Software developers** who will build applications that utilize the data in the system, and make it easier for end-users to interact with the system.
6. **End-users** who will utilize the software and information provided by the system in order to understand and develop geothermal resources.

System Architecture

The NGDS includes data covering a wide range of topics, from well logs and drilling data to temperature, geochemical, and geophysical measurements. Standardized data access to important datasets will facilitate utilization of these information resources.

A key component of the system is the catalog service through which data providers register the availability of resources, and through which users discover, evaluate and access resources. A resource is considered part of the system when it can be located by searching the catalog service, which returns a metadata record describing how the resource can be accessed. Data providers maintain nodes in the network, connected through the use of standardized metadata for describing resources, content models for geothermal data, and common web-service protocols for exchanging information. These standards were developed in conjunction with the US Geoscience Information Network (USGIN), thereby providing interoperability with a wider range of geoscientific information.

Data Provider Software Package

A redistributable, open-source software package has been created to give data providers a simple way to register data sources, load data and expose those data as a node in the NGDS. The software supports batch import and upload of shared datasets in supported formats adhering to standard content models. The use of this software is not required in order to participate as a node in the network; data providers may use whatever tools they wish to expose their data, as long as they utilize interchange formats and web-service protocols conforming to NGDS specifications.

End User Software

End-users may interact with the system through a variety of entry points, but the project implemented two primary access points. As much as possible, these will be integrated in order to appear to the end-user as a single web-based experience. The two primary access points are as follows:

NGDS Website

The website is designed to provide information about the NGDS. It serves as an entry point to the system, allowing users to discover data and applications that utilize NGDS resources. The site includes information on the project's progress, NGDS specifications, access to the Map-Centric Search Application described below as well as other software applications utilizing NGDS data, presentations, documentation and tutorials, a catalog of NGDS nodes, and any other results as they become available.

Map-Centric Search Application

A user-friendly, web-based application has been created to support finding, visualizing, mapping, and acquisition of data by end-users. This application allows users to discover and access resources made available across all NGDS nodes, and to search for data across the system based on topic, location, time, provider, or key words. Standardized metadata describing each dataset provide the user with the information necessary to determine the utility of that dataset for their purposes. Geographic datasets visualized through a map interface also allow users to inspect the details of individual data points (e.g. wells, temperature measurements, etc.) from properly formatted datasets. In addition to visualization within the application, the interface provides the information necessary for users to access the data from other, third party software applications.

US Geoscience Information Network

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. NGDS is based on the U.S. Geoscience Information Network (USGIN) which enables users to efficiently find, access, and share geoscience data, reducing the time and effort spent locating and integrating useful information and document new data by providing for information registration (by providers) and discovery (by users) based on standardized catalog services and metadata. The system accommodates resources in various forms, from unstructured text and images to documented, community Web services and interchange formats. To simplify, USGIN is a collection of Web-accessible resources that are registered in online catalogs and conform to data-sharing practices. Developer specifications for USGIN are available via the USGIN GitHub site at https://github.com/usgin.

The core components of the network are information exchange specifications and the catalog function. Information exchange specifications are community agreements on the conventions necessary for the interoperable exchange of some particular information. By using an interchange exchange it allows data in a variety of formats, organization, and structure to be integrated without having to manually transform it. The catalog function comprises a collection of metadata records that describe resources accessible through the network, and a special information exchange that defines metadata content, how the metadata collection is searched, and how metadata are encoded in search responses.

As a system, USGIN is a loosely coupled system of independent data providers, client applications, and infrastructure. The infrastructure includes 1) tools for registration of new resources, searching metadata catalogs, authentication, and resource validation; 2) registries for vocabularies, agents, specifications, and interchange schema; and 3) documentation and educational resources. Because network operation is based on information exchange specifications that are independent of any particular hardware or software implementation, all of the operational components can evolve as technology evolves. Use of standard protocols enables data access using off-the-shelf software, both commercial and open-source.

The distributed nature of the system means that stewardship of resources is determined by the resource owner. Participation in the network requires that a resource provider create metadata that conforms to the profile, and make the metadata and the described resource available. The network is open; anyone can deploy new nodes and components that implement one or more USGIN specifications, without requiring approval. New specifications can be introduced for service protocols, interchange formats, or vocabularies. Similarly, the system back end is decoupled from the front end, meaning anyone conforming to system configuration can build a portal into the system, consuming the data feeds and adding their own applications, visualizations, or other value-added functions. The International Renewable Energy Agency is already taking advantage of this, consuming key data sets for inclusion in the Global Atlas of Renewable Energy, accessible via the IRENA website www.irena.org. Keeping resources under the stewardship of the parties responsible for the information promotes system sustainability because the stewards have a direct connection with the quality of the product rather than submitting the data to a centralized database and related data manager.

NGDS Data Provider Software Package and End User Software

While USGIN permits application of a variety of service protocols, to simplify participation by NGDS data providers a simplified software stack, Node-in-a-Box (NIAB) was developed under the NGDS Architecture, Design, and Testing Project (NGDS Design Project). The NIAB 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
- NGDS-conformant Web APIs (application programming interface) that enable access to NGDS data (WMS, WFS, WCS)
- NGDS-Conformant Web APIs that support the discovery of NGDS resources via catalog service (CSW)
- A web site that supports discovery and understanding of NGDS resources.

While a number of different open-source frameworks were identified for the development of this software stack, including existing systems in use by the NGDS, the NDGS Design Project determined to use the Open Knowledge Foundation's CKAN (http://ckan.org) implementation as it provided the closest match between out-of-the-box functionality and the NGDS NIAB requirements (see https://github.com/ngds/documents).

A major consideration in the evaluation of the software system was the long-term viability (sustainability) of the eventual application. By building on an existing, active and widely used open-source project, including by the US Office of Management and Budget for use on http://data.gov, it ensures that external developers can assist in the ongoing maintenance of the system. Other factors considered in the development framework included; 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.

CKAN is written in Python and makes use of a variety of open source frameworks including 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 and includes file storage, metadata management, and management of structured data. In addition, it offers a plug-in mechanism enabling developers to rapidly extend CKAN's core functionality; most importantly to the NGDS, the ability to support geographic features as well as exposing metadata according to the OGC standard catalog (CSW). CKAN implements crucial housekeeping features such as user management and logging, which can be tedious to implement but crucial for the site's usability.

While there are many attributes to the out-of-the-box CKAN functionality, there are a number of items that required NGDS developer modification including:

1. Modification of the User Interface (UI) to reflect the user experience testing as completed through the NGDS Design Project UI testing
2. Creation of the NGDS/USGIN metadata, including spatial extent
3. Consistency check for well-known structured data files (described in the next section)
4. Providing OGC services for uploaded structure files with geospatial 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 NGDS CKAN code is currently available for deployment at https://github.com/ngds/ckanext-ngds/ as well as registered with the CKAN developer code repository (https://github.com/okfn/ckan). During the sustainability maintenance phase, USGIN developers will interact directly with the Open Knowledge Foundation to provide the following:

1. A more robust installation script
2. A mechanism for script updates to installed instances
3. Clean-up unused/redundant code
4. Ongoing organization and cleanup of the install instructions and repository

Using CKAN as the NIAB solution and primary UI meets many of the needs and software specifications outlined in the NGDS Software Requirements Summary, version 2.7, November 5, 2012 developed by the NGDS Design Project.

NGDS Software Requirements & Achievements

Based on the Requirements Summary, the following have been achieved and are presented as they have implications on the maintenance and sustainability of the network.

2.1 Maintenance:

- All project source code has comments, at least on a per-class level
- The system's architecture and configuration parameters are well documented
- The system's source code is covered by unit tests to at least 50% coverage

2.2 Usability and Accessibility:

- The system has a reasonably simple to use installation tool, with detailed instructions that guide the user through the process of installation and joining the network.
- The system is cloud ready, archived, and available via a Virtual Machine using Amazon Web Services
- The system uses project developed graphical user interfaces for a uniform look-and-feel for web applications
- The project-developed applications provide online help explaining how to perform user-related functions
- Key data import operations are transactional, i.e. the user can abort operations before completion without any negative consequences
- The project has a status indicator for data uploads
- The system underwent review for compliance with section 508 amendment to the Rehabilitation Act of 1973 and ISO/TS 16071 "Ergonomics of human-system interaction – Guidance on accessibility for human-computer interfaces."

Performance and Scalability:

- The system was designed to support a maximum of 1,000 concurrent users
- Each data provider node is capable of maintaining at a list of at least 100 other NGDS nodes for harvest or distributed search
- Each data provider node is capable of supporting up to 50 simultaneous authenticated, logged-in users, at least 50 HTTP requests every minute, and takes no more than 10 seconds to respond to each request
- They system can handle the import of data files up to 2BG in size, import up to 1,000 data files in any one import operation, store up to 100,000 data files in the import director of each data provider, and store up to 500GBs of data files (this is completed by the use of Amazon Web Services and thus is expandable to meet the needs of the system)

Security:

- Valid login authentication is required for all data submitters, stewards, and administrator functions
- Permits only valid users to write data they have permissions to write and delete data they have permissions to delete
- The data provider node maintains the integrity and availability of all data stored in its local data store and repository

Supportability:

- NGDS components are written using standard coding styles for the programming languages used (Python, Java)
- NGDS software is designed utilizing the concept of encapsulation, that is, components are encapsulated by related functionality

System Operations

Code Repository

As part of the systems commitment to open source software development two GitHub repositories for system design, specifications, documents, and code were created and are maintained by the Geoinformatics staff at the Arizona Geological Survey. The repositories are open for participation and available for public viewing.

Main NGDS Repository: https://github.com/ngds
CKAN Development Repository: https://github.com/ngds/ckanext-ngds/
NIAB Installation Repository: https://github.com/ngds/install-and-run

Programming languages used for the NGDS CKAN extension build are Python and JavaScript, with OS (open-source) components Apache SOLR and Tomcat, PostGIS, GeoServer, and GDAL. These were selected due to their support and use in the open source community and have strong communities of practice surrounding their use. Likelihood that the languages and components will be obsolete in the near future is slim.

NIAB Code

This section is specifically on the maintenance of the NIAB code and does not include details on the data maintenance and installed nodes. For information on data maintenance and installed nodes, please view the next section "Maintaining Content."

AZGS is currently in the process of working directly with the Open Knowledge Foundation to increase the viability and usability of the NGDS NIAB as well as to better align the NGDS extensions to the CKAN project code. This work will occur from May 2014 to December of 2014 to help improve the installation of at least six NIAB at State Geological Survey's across the country (additional information provided in the Maintaining Content section). This work will also make additional developers aware of the NGDS extensions, adding developers to the code team, and thus the maintenance of the extension code.

The current version of the code is tagged at version 1.0.1 and available at https://github.com/ngds/ckanext-ngds/. As improvements are made and new code is pushed to the GitHub master branch, newer versions will always be tagged. Current users will be alerted by RSS feed, email, or other notification system when a new version is available, along with a list of upgrades and changes indicated for that version.

Licensing

Per the requirements of the NGDS Design Project, the software produced would have an open source license variant allowing users to copy, distribute, and transmit the software; adapt software for other applications; and make commercial use of the software under the condition that the attribution of the software is included in any copy or derived work. Thus, NGDS selected the GNU Afferno General Public License v.3 and the Creative Commons Attribution 3.0 Unported License. All NGDS software and development materials are licensed in this capacity, including the website and site materials.

- GNU Afferno General Public License, v.3 – specifically for software designed to ensure cooperation with the community, ensuring that users can share and change all versions of a program while maintaining its free use in the community
- Creative Commons Attribution 3.0 Unported License – requires that users provide appropriate credit, provide a link to the license and indicate if changes were made when they share (copy and redistribute the material) or adapt (remix, transform, and build upon the material)

Currently NGDS does not track or monitor usage or popularity of search terms due to the intellectual property associated with exploration activity. The NGDS website at http://geothermaldata.org does monitor page visits, DNS locations, and response time of the site to assist with improvements of the site.

In addition to the licenses for NGDS materials, the website is dedicated to the public domain through the Creative Commons 0 1.0 Universal (CC0 1.0) Public Domain Dedication. Thus the http://geothermaldata.org site includes the following statement:

> *The content of this web site is dedicated to the public domain. All rights to the work are waived worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.*
>
> *The NGDS requests that it be acknowledged as the source in any subsequent use of its information. Some materials on this site have been contributed by private individuals, companies, or organizations and include a copyright notice. It is the user's responsibility to contact copyright owners and obtain the written permission required under U.S. copyright law before using these materials.*
>
> *Links may be made to the NGDS Web site from personal and organization Web pages. NGDS requests that you link to its site rather than downloading portions of the site to another Web server so viewers will see the most up-to-date information.*
>
> *NGDS materials may not be used to state or imply the endorsement of DOE EERE, AZGS or any of their employees of a commercial product, service, or activity, or be used in any other manner that might mislead the public.*

Website and User Portal

The site http://geothermaldata.org is the primary site that provides information about the NGDS to all participants as well as serves as the gateway to the system to discover data and applications that utilize NGDS resources. In addition the site provides informational resources on the project's progress, access to NGDS specifications (on the GitHub), access to a map-centric and library style search interface, access to other software applications utilizing NGDS services (Resources), NGDS tutorials (via YouTube and USGIN site), and a link to the catalog of NGDS nodes.

While the primary site is http://geothermaldata.org, we have also procured the following domain names and provided redirects to the primary site:

- Geothermaldatasystem.org
- Usgeothermaldatasystem.org

In addition, we have the http://stategeothermaldata.org and http://usgin.org domain names (as well as all associated subdomains of the five total URLs). All domain names are registered through the year 2020 to provide consistency in URLs. Currently all domains are hosted via Site5 on a shared server with unlimited disk space, bandwidth, and websites. However, due to recent issues in support with Site5 it was decided that the NGDS and related sites would move to KnownHost which guarantees an independent Virtual Private Server, thus providing additional power and speed to the sites. The system at KnownHost provides 1536MB of guaranteed RAM, 60GB of disk space, and 4000GB of Bandwidth for approximately $40 per month. This should accommodate the NGDS and related sites well into the future. However, should the sites require additional RAM, disk space, or bandwidth, there are five additional options for hosting at varying costs. Service includes a 99.9% SLA, free backups, high performance VPS hosting, and 24/7 support.

Additional advantages to managed hosting includes no hardware requirements, all downtime is covered by the hosting company, and the information/sites are easily transferrable should a new entity require management of the sites. Finally, the new hosting service provides a full dashboard for site statistics and monitoring.

In addition to general website hosting and maintenance, the NGDS requires cloud services for the NIAB GeoServer and multitude of Web Services. Throughout the duration of this project we have used Amazon Web Services (AWS) to meet these needs. AWS costs are established via the traffic and requested redundancy. Currently NGDS services cost approximately $150 per month. While we expect traffic to increase due to the v1 release of NIAB, we anticipate costs of approximately $200 per month for AWS for the near future (12-18 months). During the maintenance phase through April 30, 2015 we will evaluate traffic and adjust this plan accordingly.

Service Notifications

NGDS nodes and hubs are encouraged to use the USGIN RSS feed for service notifications when deploying web services. The tool is available at: https://github.com/usgin/service-notifications. This permits data providers the ability to quickly distribute notifications regarding the services that they host, such as planned downtime, unforeseen server issues, new service availability, etc. Subscription to the service notification is available via the USGIN notifications RSS feed on the GitHub.

Maintenance Staffing

Ongoing maintenance of each of these components will be the responsibility of the Arizona Geological Survey and in time, the USGIN Foundation Inc. It is expected that pure maintenance of such a system will require approximately ¼ developer/sys admin time. This does not include significant updates to the software code or user-experience, nor does it include help-desk or technology transfer assistance.

Maintaining Content: Data, Services, Applications

Regional Server Hubs

Four NGDS Regional Server Hubs (hubs, IL, KY, AZ, and NV) have been thus far tasked with hosting web services for their surrounding data providers (other state surveys or institutions) who lacked the technical capability to host their own data in web service format. Deployment of the Server Hubs was estimated at $100,000 per year. As the hubs move into a maintenance mode, we estimate maintenance at approximately $30,000 to $50,000 per year depending on the institution. This includes limited troubleshooting time, monthly data updates, backup and disaster recovery costs, and institutional facilities and administration costs. Currently 2 of the 3 hubs are using shared ArcGIS servers (1 has a dedicated ArcGIS server) and independent institutional infrastructure, as the concept of USGIN enables participants to use existing infrastructure systems with specified protocols and standards.

NGDS tier 3 data services are discrete data types where aspects are controlled by static schemas. One aspect that is specified in the schema is the layer name of a given data type; this assists with search and access of the data. As the hubs host multiple Tier 3 services from multiple states, they therefore serve multiple services of the same data type. This becomes an issue with the current CKAN NIAB application.

The current construct of the CKAN NIAB application is problematic for hubs or other entities who might serve multiple Tier 3 web services of the same type. This is due to the GeoServer web server platform used in NIAB where only one layer name can exist in any given GeoServer instance. Thus, even though NGDS hubs could use NIAB for service deployment and management, there are undesirable consequences. First, the web services may become too large for the current processing capabilities of most desktop/laptop machines, and second, the data management scheme (each state having its own dedicated service per data type) and proven successful workflow of the hubs would need to change. Over the life of the project, the hubs have been functioning through use of individualized workflows, which usually include ArcGIS Server and PostGIS for web services management coupled with metadata management at the state geothermal data repository (http://repository.stategeothermaldata.org/repository/).

Considering the possible NIAB limitations for those entities serving multiple datasets of the same type, two options are recommended for hubs moving forward; 1) continued use of the state data repository for metadata management and ArcGIS Server for web service deployment; 2) continued use of the current ArcGIS Server workflow for web service deployment coupled with NIAB metadata management.

Scenario 1: Hubs continue to use ArcGIS Server; Hubs continue to use the UI at repository.stategeothermaldata.org

Currently, NGDS hubs use ArcGIS software to deploy web services and manage the associated data formats (geodatabases, feature classes, data tables). When project deliverables (datasets) are deployed as web services, a metadata entry is made at repository.stategeothermaldata.org for that dataset with distributions that include the links to the live web services (WMS, WFS, ESRI rest page).

The current workflow of AASG deliverables (data) processing includes the creation of metadata for the services, where a 'New Resource' is created at repository.stategeothermaldata.org. Moving forward with this scenario, hubs will be responsible for creating metadata at repository.stategoethermaldata.org, which will require minimal additional training for hubs as they are already familiar with the interface. The biggest hurdle will be to assist the hubs in properly integrating the metadata entry point into their current workflow.

Scenario 2: Hubs continue to use ArcGIS Server; Hubs use NIAB for creation and management of metadata for services, but not service deployment

As creating Tier 3 web services using NIAB becomes problematic for hubs, this scenario proposes that services are deployed as usual with ArcGIS Server. Instead of entering the metadata for the services at repository.stategeothermaldata.org, the hubs would use NIAB simply for the creation and management of that metadata.

When a “Link to a data service” is added as a resource in NIAB (creating a metadata record), it is possible to add multiple service distributions. These distributions are the WMS/WFS links. Once a metadata record is created for a service in NIAB, locate the Edit button on the page of the resource. Here, go through the “Link to a data service” workflow once more to add an additional link. This will be added to the same metadata record. The hurdle here is much the same as Scenario 1; hubs will decide which methods of metadata management best fit their workflow.

Until the NIAB can handle multiple layer names or GeoServer instances, a potential development target over the next year, we recommend that the hubs maintain Scenario 1 as installing the NIAB simply for metadata management would cause unnecessary costs for the regional hubs (for additional cloud services and server maintenance on services that they will not be implementing, i.e. GeoServer).

Regional Server Hardware

While the USGIN system is designed to accommodate a variety of data sharing “nodes” and platforms, due to the importance of the multi-state hosting of the Regional Server Hubs, we are recommending a more standardized platform for the hubs while moving forward.

In order to increase the Regional Network Hubs reliability, maximize uptime, and ongoing sustainability we recommend that the Hubs virtual machine host servers be dedicated services rather than using a shared server solution (as is currently the case). We estimate that a Dell tower server in the $2,000 range would be powerful enough to support this type of solution. We also recommend that each of the Hubs use a USB hard drive ($150) attached to the server as a backup drive. The free Windows Backup software included with Windows Server 2012 has the ability to perform scheduled backups to a USB drive that includes a complete image of the entire Hyper-V host server including any virtual machines on the server.

We also recommend setting up Hyper-V Replication (next section) between the hubs Hyper-V host servers for redundancy. If a hub server or virtual machine goes offline, Hyper-V Replication would provide a copy of the hub that could be quickly booted up on the backup Hub server until the main Hub server problem is resolved.

Replication and Disaster Recovery

AZGS is assisting with the replication and disaster recovery for the other hubs. Delays in virtualizing, due to institutional requirements, have postponed full implementation, although as of March of 2014 each hub was running a virtual machine. Hyper-V is a virtualization solution from Microsoft designed to create and operate virtual machines. Hyper-V replication is a feature of Hyper-V that provides the ability to make a copy of a virtual machine at a different geographic location for disaster recovery purposes. In 2013, AZGS installed Hyper-V 2012 on an Arizona Geological Survey server to support the hubs.

In September of 2013, the NGDS virtual server at the Nevada Bureau of Mines and Geology (NBMG) failed catastrophically with a loss of 200GB of data. Fortunately, Hyper-V replication was engaged and the server mirrored, insuring that all NBMG data survived. Using the Hyper-V mirror to restore 100% of the data avoided a catastrophic loss of data that would have taken hundreds of hours to reconstitute; in this case it took approximately 3 days. AZGS has published a guide for installing and creating a Hyper-V replication environment. It is available via the AZGS repository at http://repository.azgs.az.gov/uri_gin/azgs/dlio/1567.

Nodes on the Network

An NGDS node is a web-accessible server that hosts at least one of the functional capabilities enumerated below to play a role in making geothermal-relevant data assets accessible.

To be considered a node in the NGDS, the following criteria must be met:

4. The capabilities offered must play a role in making geothermal-relevant data assets accessible.
5. The capabilities offered must be publicly Web-accessible.
6. A node must offer at least one of the following capabilities:
    a. Host a Web-accessible repository of geothermal-relevant data assets with metadata conforming to the USGIN ISO profile (USGIN Standards and Protocols Drafting Team, 2010-11) published through an NGDS catalog node.
    b. Host a web-accessible folder that 1) contains NGDS-conformant metadata files and 2) is registered for harvesting by an NGDS catalog node.
    c. Host NGDS-conformant web services (WMS, WFS, WCS, etc.) that are registered in an NGDS catalog.
    d. Host an NGDS catalog node, which is a server operating a CSW 2.0.2 service that offers metadata conforming to the USGIN ISO profile.

NGDS nodes should self-identify by providing an **NGDS node self-description document**. Additional information on Node Specifications are available at: https://github.com/ngds/install-and-run/tree/master/Becoming%20an%20NGDS%20Node

There are currently two types of NGDS nodes; the NIAB installs and organizations that use their in house systems to meet the requirements above. Currently, the following nodes are hosting their own data via NGDS Tier 3 web services:

- Southern Methodist University's Geothermal Laboratory (independent project with independent sustainability plan, SMU has guaranteed to house the web services and data for at least 10 years; how the SMU node will respond to issues with metadata and general server maintenance are unknown)
- Illinois Geological Survey (also a hub)
- Kansas Geological Survey
- Kentucky Geological Survey (also a hub)
- Minnesota Geological Survey
- New Hampshire Geological Survey
- Nevada Bureau of Mines and Geology (also a hub)
- Oregon Department of Geology and Mineral Industries
- Washington Department of Natural Resources, Geology & Earth Sciences
- Wisconsin Geological Survey
- Arizona Geological Survey (also a hub)
- Minnesota Geological Survey

As part of the NGDS Design project additional installations of the NIAB were to be deployed, including the metadata aggregator node. Currently the metadata aggregator node is housed at the Arizona Geological Survey (AZGS). Two instances of the NIAB are successfully installed at the Arkansas Geological Survey and the New Mexico Bureau of Geology and Mineral Resources. An additional node at the University of Utah's Earth and Geosciences Institute was to be installed by the time of project completion however, ongoing server issues at EGI prevented the completion of this node by April 30. We will continue to attempt NIAB installation at EGI, if at all possible.

During the May to December 2014 time period, AZGS will work to install a minimum of six NIAB installations across the country at State Geological Surveys. During that time we will also work to maintain the NIAB installs. The following Surveys have been identified as targets:

First Tier Targets:

- Alaska Division of Geological & Geophysical Surveys
- University of Hawaii, Manoa, Geological Department
- Idaho Geological Survey (Confirmed)
- Oregon Department of Geology and Mineral Industries (Confirmed)
- Utah Geological Survey
- Virginia Division of Geology and Mineral Resources (Confirmed)
- Wyoming Geological Survey

Second Tier Targets:

- Geological Survey of Alabama
- Indiana Geological Survey
- Missouri Geological Survey

For the one time setup and maintenance of the NIAB installations, we estimate approximately $12,500 to $25,000 a year in virtual server costs, developer/sys admin time, and facilities & administration costs. As with all costs, this will vary dependent on salary and F&A rates at the institutions; however, institutions should estimate approximately an 80%/20% split between personnel and equipment. Participation in the setup and ongoing testing (May to December 2014) may run as high as $50,000 to $75,000 a year as this requires significant developer/sys admin time.

Possible Extensions & Improvements on the NIAB

The software specifications and requirements document outline a series of requirements that were not met for the Version 1 release. For example, the bulk upload of data sets was a target that was not met. In addition, testing of the v1 release over the next 6-12 months will generate a series of additional user-inspired requirements that will require implementation. One such example is for a metadata repository extension of the NIAB, that is, a NIAB that does not require installation of the GeoServer platform and can be used by entities or organizations that simply want a mechanism for contributing metadata to the network while hosting their own services. As these requirements are developed, issues will be logged and prioritized on the NGDS GitHub. The AZGS and USGIN Foundation Inc. team will accommodate as many of these priorities as possible.

Content Model (Information Exchange) Maintenance & Sustainability

Content models, or information exchanges, in use for the NGDS were established using a community developed mechanism and approved by the Department of Energy's Geothermal Data System Development and Population Working Group (GDSDPWG), which was comprised of the technical leads from each project and a series of technical monitors. When available, existing exchanges were used. Information in deprecated exchanges were generally synthesized with another exchange and thus, are still available. All available exchanges are listed at http://schemas.usgin.org/home/.

Should the likely event occur that additional exchanges are required to expand the usage and applicability of the system, the following methods are proposed. Full details are available at https://github.com/usgin-models.

The decision to define a new information exchange should be based on the likelihood that others will want to publish similar datasets in the future. Members of the USGIN community propose specifications for data sharing exchanges. Exchange documents are developed and reviewed using a publicly accessible repository on GitHub (https://github.com/usgin-models). Each exchange has a separate repository associated with the USGIN-models pseudo organization. A proposed model must have an identified steward, and a working group of at least three participants with relevant domain knowledge and understanding of the interchange technology. There is no formal process for defining workgroup membership; normally the challenge is finding a sufficient number of qualified individuals to provide meaningful reviews and comment.

The exchange steward is responsible for assembling the workgroup and assuring sufficient expertise in the group to generate a sound content model and implementation. The exchange steward requests creation of a new model repository at the USGIN-models GitHub from the organization members (currently primarily AZGS Geoinformatics staff), and identifies workgroup members who will have commit privileges on the repository. Any community member can create a repository branch to propose changes using standard GitHub procedures, and request consideration for merging back into the developing model.

After review and approval by the workgroup, a call goes out to a USGIN technical review e-mail list or by RSS feed (http://notifications.usgin.org/) for comments from the community. An open review period of 4 weeks is normal, after which any comments from the community must be resolved to the satisfaction of the commenter. When issues are resolved to the satisfaction of the stakeholders (workgroup and engaged community), the exchange specification is adopted.

When a specification is adopted, all associated documents are copied to a 'tag' branch in the GitHub repository and are not changed after they are 'tagged'. The Specification documents are also copied to the exchange repository at http://schemas.usgin.org, which is a web site set up to provide public access to exchange specifications and any related xml schema documents or other artifacts required for the deployment and operation of the information exchange.

This mechanism ensures that the development of specifications continues to be community led as well as providing a source-trail for the development of such exchanges.

Maintaining the Existing Data

Experience has dictated that server outages can and will occur. In addition, the potential loss of data from a distributed network, while not as great as "closing the doors" on a traditional database, may occur. In order to preserve the vast amount of data digitization and preservation, the Arizona Geological Survey plans to conduct a one-time data harvest to capture the data available within the NGDS and place it as a back-up and disaster recovery mechanism on both cloud storage and external hard drive. This process will be complete by May 30, 2014. The external hard-drive will be submitted to DOE as a deliverable under the State Geological Survey Contributions project (end date 12/2014).

Maintaining the Client Developed Applications
NGDS and USGIN support client side developed applications similar to applications for the iPhone or Android phones. As such, externally developed applications are subject to the maintenance schedules of the developers and not the NGDS or USGIN.

Organization & Management

Non-Profit Organization

Per the May 2013 sustainability paper, we suggested that a stand along organization offered the greatest independence and likely adherence to the organizations vision, goals, and missions and thus the longevity of a system such as the NGDS. In the fall of 2013, the Arizona Geological Survey, with the guidance of the Arizona Attorney General appointed counsel of Rusing, Lopez, and Lizzardi, PLLC, have successfully incorporated the USGIN Foundation, Inc. with the State of Arizona's Corporation Commission as well as filed initial requests with the Internal Revenue Service. Not only is this step in alignment with the recommendations of the May 2013 paper, but also aligns with the Hutchison and Richard 2011 "Recommendations for the Future of the U.S. Geoscience Information Network," joint report to the Association of American State Geologists and U.S. Geological Survey.

Thus far, we have confirmed the following for the USGIN Foundation Inc.

- Articles of Incorporation with the Arizona Corporation Commission, 12/05/13, file number 1890284-5 (full details are available at: http://starpas.azcc.gov/scripts/cgiip.exe/WService=wsbroker1/names-detail.p?name-id=18902845&type=CORPORATION)
    - Articles of Incorporation
    - Publication of the Articles of Incorporation
- Completed SS4 and received EIN from the IRS: 46-4569299
- Corporate by-laws signed
- Board members being identified – solicitations to the AASG
    - Advisory board members being identified – potential overlap with GRC/GEA

We are still awaiting IRS designation as a 501c3 scientific organization. In the interim, we have been working with our marketing consultants to establish proposed service rates, similar to the Socrata structure or Open Knowledge Foundation service package (additional information on the service packages is described below in the Technology Transfer section).

At this time, the USGIN Foundation Inc. is co-located with the Arizona Geological Survey at 416 W. Congress Street, Suite 100, Tucson, AZ 85719. The State of Arizona has precedence for such activity as the Arizona Game & Fish Department and the Arizona Department of Tourism have related non-profit organizations. While the USGIN Foundation Inc. is co-located in the physical property of the AZGS, the eventual goal is to move the Foundation off state property to ensure that the Foundation is not subject to political activities.

The Foundation provides the following expertise:
- Earth Science Data
- Project Management Excellence & Streamlining of Workflows
- Managing Geospatial Data
- Knowledge of Unstructured AND Structured data formats
- Interoperable Systems
- Data Provider Community Building
- Visions for Long-Term Sustainability

Duties of the USGIN Foundation Inc. will include:
- Management of the Foundation
  - Strategic Planning
  - Fiduciary
- Client Systems Development & Maintenance
  - Programming/Development
  - Data Curation
  - Metadata Maintenance and Curation
  - Systems Architecture
  - Systems Administration
- Business Development, including Outreach & Training
  - Contract Development and Negotiation
  - Government and Foundation Grant Development and Negotiation

Technology Transfer (Education, Outreach, Training)

Foundation Website

As part of our improved outreach and future business development we have created a more user friendly website to introduce potential clients to the system. The revised page is at the existing http://usgin.org page with the technical specifications and tutorials moving to a subdomain at http://tech.usgin.org. All product descriptions, once finalized, will be available and accessible via the USGIN homepage.

In addition to the traditional website at http://usgin.org, specifications are available via http://github.org/usgin, and scenario and user video tutorials at www.youtube.com/geothermaldata. The videos available on the YouTube site include 13 data access scenarios guiding users through the search interface as well as a variety of software clients to show the diversity of the system.

Products & Services

In cooperation with our marketing and product development firm, Blue Canoe Marketing, we have four deliverables, in varying stages of completion, outlining next steps for business development, outreach, and training. These include: 1) the USGIN Suite of Solutions for Open Data Access, 2) a Compliance Guide for Federal Open Data Compliance, 3) the NIAB Marketing Plan, and 4) a 12 month marketing and communications plan for business development.

USGIN Suite of Solutions

There three types of services that the Foundation will offer: Marketing/Pre-Consultation, Consulting Services, and Support Services. Brief information on each of the types of services is offered below. Note that the services are designed to build upon each other.

Marketing/Pre-Consultation

- Workshops
- Training & Education

Consulting Services

- Assessments – provides a plan and recommendations for Open Data Compliance
- Open Data Strategy and Action Plan – provides a scope of work, specifications, and a proposal to build a basic data sharing system
- Interoperable Data Strategy and Action Plan - provides a scope of work, specifications, and a proposal to build an advanced interoperable data sharing system
- Deploying Your Own Node – provides individualized data node with documentation (installation guide, technical support, help desk requirements) and independent data control for the host
- Hosting – provides a migration plan for sustainability of the independent node

Support Services

- Open Data Development – provides hands-on modification of data, catalog creation, and customized workflows (based on the results of the Assessment)
- Interoperable Data Network Development - provides hands-on modification of data, catalog creation, customized workflows (based on the results of the Assessment), community building efforts (data sharing affiliates), and information exchange specifications
- Annual Subscription & Support – provides hosting services, data maintenance, replication services

Fees associated with each service will depend on the scope and size of the proposed project. These services, while intended to be stand-along services offered to potential clients, can also be revised to respond to funding opportunities in government and foundations. For example, the State of Arizona's Natural Resources Review Council recently approved a plan to develop a Natural Resources Decision Support System (NRDSS) to link data, documents, and GIS layers among nine state natural resources, environmental, and transportation agencies, as well as with corresponding agencies, using a USGIN style system. In response, we will be able to provide a best estimate for total cost of such a system using the clearly delineated services defined above. In this instance, we would include services for assessment of available data resources, and then provide development of the interoperable data network, concluding with a set of proposed annual subscription fees for maintenance and support.

Compliance Guide for Federal Open Data Compliance

To aid and assist with marketing of a USGIN style network, and to communicate the benefits of a USGIN style network to non-data or technical stakeholders, we have developed an Open Data Compliance Guide in response to the Executive Order 13642 and OMB Memorandum M-13-13. The Guide provides descriptions on the open source tools and resources available for potential client applications that go beyond the OMB requirements to create an interoperable, federated data network, similar to the NGDS. Details described within the guide discuss the benefits of USGIN and

NGDS, information on how to create and maintain an Enterprise Data Inventory, how to achieve interoperability with that inventory, and how to create and maintain a public data listing (or USGIN style catalog). In addition, the Guide describes mechanisms for engaging with stakeholders and how to document data that is currently under moratorium. Once completed, this guide will also provide a brief overview of how USGIN can help provide these services.

NIAB Marketing Plan

As described in the Data Maintenance section of this plan, a series of State Geological Survey's will be selected to deploy the NIAB software stack and migrate their services from the Regional Server Hubs to their own node. In order to assist them in this effort, we will provide an easy to follow web page with concept graphics, a series of overview slides, and then work with the Survey's to conduct continual user-feedback which we can incorporate both into the software and the long-term marketing of the NIAB software stack.

12 Month Marketing & Communications Plan

Target markets for the NGDS and USGIN Foundation Inc. are other geoscience data providers in the Federal, State, and Private Sectors. We plan to maintain and build our connections within the Geothermal Energy Sector and strategically expand to other Natural Resource sectors, then the Natural Sciences, and broader Environmental communities. To do this, we are currently crafting a 12 month marketing strategy for NGDS and USGIN that will include the following components:

- Prioritized customer segments and targets
- Development of core messaging and unique selling positions for the top tier segments
- Development of an integrated marketing and communication strategy aimed at the target segments that addresses: email marketing, national and international open data/digital data events and conferences; trade association participation and events; public relations and press releases; social media strategy; digital and traditional advertising; and lead generation materials
- A set of recommended marketing activities by month
- Estimated budget to execute the marketing plan

We have already begun implementing the revised marketing strategy with a test-exhibit at the American Association of Petroleum Geologists Annual Meeting & Expo with a revised exhibit pop-up and literature.

Business Models and Client Development

The primary services offered by USGIN Foundation Inc. in support of new interoperable data systems are primarily described in the "USGIN Suite of Solutions" section. However, the section does not address maintenance of existing systems, beyond a general overhead expenditure of the Foundation or through sustained annual subscriptions. As such, we have proposed the following mechanisms for sustained funding:

- Sponsorship – from interested companies such as Esri, Microsoft Research (which offered Azure cloud computing storage), Google
- Donations – from individuals similar to Wikimedia Foundation or Foundation support such as through the Gordon & Betty Moore Foundation or Sloan Foundation
- Advertisement – primarily web based
- Premium User Subscriptions – through the development of improved applications interfaces

These funding mechanisms will be tested through user-feedback during the maintenance phase of the NGDS, and refined based on the success of the 12 month marketing plan.

Conclusions

The National Geothermal Data System and USGIN Foundation Inc. are still largely in development mode; however, with the release of the Node-in-a-Box v1 and NGDS user interface v1 in April, as well as the closure of the primary data collection efforts, we can now focus on the next steps to ensure NGDS and USGIN are sustained and maintained well into the future. These steps include significant user experience testing, marketing of the existing system, and promotion of new interoperable data networks.