

Guidelines for Provision and Interchange of Geothermal Data Assets

Supporting Advancement of Geothermal by
Populating the National Geothermal Data System

US Department of Energy Geothermal Technologies Office

July 3, 2014

Version 2.0

The US Department of Energy Geothermal Technologies Office (GTO) funded the development of the National Geothermal Data System (NGDS), a distributed information system providing access to integrated data in support of, and generated in, all phases of geothermal development. NGDS is built in an open paradigm (the U.S. Geoscience Information Network, USGIN, which is operated by USGIN Foundation, Inc.) and employs state-of-the-art informatics approaches and capabilities to advance the state of geothermal knowledge in the US. This document presents guidelines related to provision and interchange of data assets in the context of the National Geothermal Data System. It identifies general specifications for NGDS catalog metadata and data content; and it provides specific instructions to GTO-funded projects for preparation and submission of data assets to the DOE Geothermal Data Repository (DOE-GDR), which is an NGDS node.

Guidelines for Provision and Interchange of Geothermal Data Assets

Contents

1	Guidelines for Provision and Interchange of Geothermal Data Assets	1
1.1	Introduction	1
1.2	Metadata Provision and Interchange	1
1.3	Data Provision and Interchange.....	1
1.4	Data Plans for DOE-Funded Projects – What to Include.....	3
1.5	Submitting Data Assets from DOE-Funded Projects.....	3
1.6	“Nodes” on the Federated National Geothermal Data System (NGDS)	5
1.6.1	NGDS Catalog	5
1.6.2	NGDS Data Access Web Services	5
1.6.3	NGDS Data Assets	5
2	Appendix A: Data Categories and Attributes	7
2.1	Engineering Data.....	7
2.2	Well Data.....	7
2.3	Business Data	8
2.4	Geological Data	10
2.5	Geochemical.....	11
2.6	Geothermal Data.....	11
2.7	Infrastructure	12
2.8	Production/Disposal/Injection Data	12
2.9	Geophysical & Synthetic Data.....	13
2.10	Direct Use.....	13
3	Appendix B: Metadata Categories and Attributes	14
4	Appendix C: Technical Format Types	15
5	Appendix D: Standards for Metadata and Data Representation.....	16

1 Guidelines for Provision and Interchange of Geothermal Data Assets

1.1 Introduction

The Geothermal Technologies Office funded the development of the National Geothermal Data System (NGDS), a distributed information system providing access to integrated data in support of, and generated in, all phases of geothermal development. NGDS increases the likelihood of success in geothermal development and reduces risk by providing a comprehensive technical and economic knowledge base. NGDS is an environment through which all types of geothermal-relevant data and publications can be disseminated, enhancing understanding about both successes and failures in geothermal research and development and providing indicators concerning the appropriate use and quality of data. NGDS is built using the U.S. Geoscience Information Network (USGIN, operated by USGIN Foundation, Inc.) an open paradigm— supporting broad access by users of all types to data and tools relevant for their work while protecting information when required— and employs state-of-the-art informatics approaches and capabilities to advance the state of geothermal knowledge in the US.

This document presents guidelines related to provision and interchange of data assets in the context of the National Geothermal Data System.

1.2 Metadata Provision and Interchange

Specification of metadata for any data asset provides the information necessary to list that asset in a searchable catalog and enable discovery through various search mechanisms. Metadata specifications abound, many of which draw from the very basic Dublin Core metadata specification¹, extending it in a variety of ways to support specialized subject matter. For Geothermal and other Geosciences data assets see the USGIN [Metadata Recommendations for Geoscience Resources](#)² for a description of Dublin Core extensions supporting this subject area. Specific metadata format requirements and templates are determined by the provider of the repository to which data assets are submitted.

1.3 Data Provision and Interchange

A number of content models for commonly occurring datasets are available at <http://schemas.usgin.org/models/>. Content models define data interchange specifications for commonly occurring datasets, particularly for implementing data access Web services and other exchange mechanisms. Content models enable consistent exchange of data among servers and applications, thereby enhancing interoperability among computing systems. Content models available to date include: **

- Abandoned Mines
- Active Fault/Quaternary Fault
- Aqueous Chemistry
- Borehole Lithology Intercepts

¹ <http://dublincore.org/specifications/>

² <http://usgin.org>

Guidelines for Provision and Interchange of Geothermal Data Assets

- Borehole Lithology Interval Feature
- Borehole Temperature Observation
- Contour Lines
- Direct Use Feature
- Drill Stem Test Observations
- Contour (Isoline) Feature
- Fault Feature/Shear Displacement Structure
- Fluid Flux Injection and Disposal
- Geologic Contact Feature
- Geologic Reservoir
- Geologic Units
- Geothermal Area
- Geothermal Fluid Production
- Geothermal Metadata Compilation
- Geothermal Power Plant Facility
- Gravity Stations
- Heat Flow
- Heat Pump Facility
- Hydraulic Properties
- Physical Sample
- Powell and Cumming Geothermometry
- Power Plant Production
- Radiogenic Heat Production
- Rock Chemistry
- Seismic Event Hypocenter
- Thermal Conductivity Observation
- Thermal/Hot Spring Feature
- Volcanic Vents
- Well Fluid Production
- Well Header Observation
- Well Log Observation
- Well Tests

Additional content models under consideration:

- Financial
- Daily Drilling Report
- Geophysical Survey Results
- Hydraulic Property Observation
- Subsurface Alteration
- Surface Alteration

Guidelines for Provision and Interchange of Geothermal Data Assets

- Well Completion Information
- Well Production Hardware

Note: for any given data asset repository, if there is not an existing accepted format defined for a particular type of data, the content models above may also be used as templates to format data sets for submission to such a repository.

** All content models are available at <http://schemas.usgin.org/models/>. They are under change management and the content model directory will provide access to current and previous releases. Also, new content models are being added to the collection. For a current view of these and any new content models, or to suggest a new model, please see: <https://github.com/usgin-models> as well as documentation at https://github.com/usgin/usginspecs/wiki/_pages.

1.4 Data Plans for DOE-Funded Projects – What to Include

For each task of the project, the following questions should be answered:

- What externally sourced data are input for this task?
 - Describe the data to be acquired (especially structure and meaning).
 - Why are they needed for this task?
 - From where are they sourced?
 - What is known about the origination of the data?
 - Is there a cost for access to the data (licensing)?
- What data will be generated as a result of the task (raw or derived)?
 - Describe the data to be produced (especially structure and meaning).
 - What is the expected value of the produced data?
 - Who is likely to use/benefit from the data?
 - How are the data likely to be used?
 - Provide an example of how the data will likely be used and by whom.
 - Is produced data raw data or derived data?
 - If derived, by what method/algorithm and what version of that method/algorithm? Proprietary algorithms should be identified but their details need not be revealed.
- What categories of geothermal-relevant data will be produced (raw or derived)?
 - See recommended data categories – Appendix A
- In what formats will datasets be produced?
 - See recommended technical formats – Appendix C
- What standards apply to production of datasets?
 - See applicable standards – Appendix D
 - See also expected metadata attributes – Appendix B
- To what NGDS node(s) will each project dataset be published?
 - For datasets from DOE-funded work the default repository is the DOE Geothermal Data Repository (DOE-GDR); alternatives must be negotiated with DOE project officers.
 - Data providers may also chose to become a node on the NGDS by downloading the free software stack “Node-in-a-Box” at <http://geothermaldata.org>.

1.5 Submitting Data Assets from DOE-Funded Projects

Guidelines for Provision and Interchange of Geothermal Data Assets

Why must data assets be submitted?

"Our **success** should be measured not when a project is completed or an experiment concluded, but when **scientific and technical information is disseminated**. Beyond broad availability of technical reports, e-prints and multimedia, and publication in peer-reviewed journals, **open access to experimental data and analysis codes** is increasingly important in policy-relevant research areas. The Department will establish guidelines for use with both grants and contracts to **ensure appropriate access to, and retention of, scientific data and analysis methods**. In more applied areas, **knowledge of what did not work** can be of equal value with **positive results**, for that can prevent the misapplication of significant private resources. The Department will therefore encourage the documentation and archiving of negative results from all its performers **using the most advanced informatics tools**."

DOE Strategic Plan, May 2011

Assure Excellence in R&D Management, pp. 43-44

Who are the players and their roles in the data submission process?

- Submitter – preparing and submitting data assets
- DOE – Oversight, technical monitoring, funding for NGDS and DOE-GDR development
- NREL – hosting and supporting the OpenEI-based DOE-GDR application, assisting with data curation

What data assets must be submitted?

- Data, produced by a DOE-funded project, that:
 - has relevance to any aspect of geothermal development
 - could potentially be useful to others outside of the project
- Data, preferably in a “consumable” and non-proprietary format
 - Data structured to meet one of the NGDS content models are the preferred submission type
 - Internally structured (CSV, Excel, XML, LAS, LIS, ...) commonly used formats are preferred over unstructured (detailed below)
 - Unstructured (PDF, Word, text file...) commonly used formats are acceptable
 - Proprietary raw data formats are acceptable, only when conversion to non-proprietary formats is impossible or if such conversion incurs a loss in data fidelity
 - If a particular kind of data asset is required by some government regulation, data asset formats required by regulation will be considered satisfactory for submission to DOE-GDR
- Accompanying metadata that helps with data discovery by others – tied to each data asset submitted. Metadata includes items such as:
 - Author(s) – for due credit
 - Title
 - Description
 - Publication Date

Guidelines for Provision and Interchange of Geothermal Data Assets

- Geographic coordinates – for mapping purposes
- Timeframe – begin and end date/time for activities
- Keywords – for searching
- Categorization – for proper filing

Where are data assets submitted?

- Data assets from DOE-funded projects are submitted to the DOE-GDR at: <http://gdr.openei.org>
- Data assets are preserved intact, exactly as submitted, and may be retrieved at any time by the submitter

When do data assets need to be submitted?

- Within 90 days of completion of a data set
- At termination of the project for any reason

Note: Data assets are made publicly available immediately³ upon submission, unless the submitter negotiates a specific “moratorium” period. If such a moratorium is specified a data asset will be secured in a protected data store until the moratorium expires, at which point the asset will become publicly available.

How do data assets get submitted?

- See the instructions at: <http://energy.gov/eere/geothermal/data-provision-instructions-all-doe-geothermal-technologies-office-funds-recipients>

1.6 “Nodes” on the Federated National Geothermal Data System (NGDS)

There are several nodes in the federated NGDS network. Each node on NGDS hosts one or more of the following capabilities:

1.6.1 NGDS Catalog

Like a library catalog an NGDS catalog lists the data assets available in repositories covered by the catalog, captures metadata describing those data assets, provides methods for discovering relevant data assets, and indicates the means for acquiring a data asset once discovered.

1.6.2 NGDS Data Access Web Services

These are functional Web-based programming interfaces that enhance interoperability among the nodes that comprise NGDS. Instead of directly accessing data repository content, end user applications access NGDS content through these data access Web services.

1.6.3 NGDS Data Assets

NGDS data assets may be either unstructured or structured. Handling is somewhat different for these, as described below.

³ a submitter may request – without negotiation– that public availability be postponed for a year

Guidelines for Provision and Interchange of Geothermal Data Assets

1.6.3.1 Unstructured (and Semi-Structured) Data Assets

Unstructured NGDS data assets are those data assets whose content is not easily parsed into separate and distinct data items. Common unstructured technical formats include: Adobe PDF, Microsoft Word, Microsoft PowerPoint, various image formats, various audio formats, etc. Each unstructured data asset must be registered in an NGDS catalog using sufficient metadata and must be directly internet accessible - in other words, a unique HTTP URL can be used to access a data asset using a browser.

1.6.3.2 Structured Data Assets

Structured NGDS data assets are those data assets whose content is easily parsed into separate and distinct data items. Common structured technical formats include: XML, JSON, LAS, Microsoft Excel, Microsoft Access, various database formats, etc.; some of these are file-based and some exist as database content. Each structured data asset must be registered in an NGDS catalog using sufficient metadata, and must be either directly internet accessible (required if the asset is file-based), accessible through appropriate data access Web services, or both. Structured data sets in proprietary and/or binary data formats will be treated in the same manner as unstructured file-based data assets.

2 Appendix A: Data Categories and Attributes⁴

The following data types are common data types that may be created during geothermal exploration or surface/subsurface exploration and research. Data collected during the following research activities are suggested for submission to the NGDS. To aide in data submission, **highlighted** data types can be submitted in NGDS content models.

2.1 Engineering Data

- Engineering Information
- Technology
- Engineering Modeling
 - Reservoir Modeling
 - Heat Loss
 - Tracers
- Engineering Designs

2.2 Well Data

- Well Type (s) – Could be more than one, potentially
 - Water
 - Oil
 - Disposal
 - Gas
 - Injection
 - Geothermal
 - Slim Hole
 - Monitoring
 - Test
- Well Status
 - Producing
 - Plugged/Shut In
 - Plug Date
 - Plug Depth
 - Casing Left
 - Nearest Water Zone
 - Mud Filled
 - Number of Plugs
 - Size of Plugs
 - Temporarily Abandoned
- Depth
 - Production
 - Interval 1
 - Interval 2
 - Interval 3
 - Injection/Disposal
 - Elevation
 - Sea Level (±)
 - DF (Drill Floor)

⁴ Siemens Corporate Research, “GTDA DATA REQUIREMENTS SPECIFICATION, Version 2.1,” 07/8/2010, Appendix B, pp. 25-31. (includes minor terminology corrections)

Guidelines for Provision and Interchange of Geothermal Data Assets

- KB (Kelly Bushing)
 - Vertical Depth
 - Well Bore Length
 - Bottom Hole Location (lat/long - for deviated boreholes)
- Casing
 - Size
 - Weight
 - Depth
 - Amount of Cement
 - Type
 - Date
- Tubing
 - Size
 - Depth
 - Date Installed
- Liners
 - Size
 - Depth
 - Type
 - Date Installed
- Drilling Segments
 - Size
 - Bit Size
 - Direction(s) (could be multiple directions at different depths)
 - Mud Weight
 - Density
 - Viscosity
 - Date Commenced
- Identifiers
 - API (or other identifier if API does not apply)
 - Sidetrack Identification
 - Well Number
 - Name
 - Field
 - Drilling Contractor
- Logs Run (there are many possible well log types; the following are examples)
 - Nuclear Magnetic
 - Temperature Log
 - Resistivity
 - Compensated Dual Resistivity
 - Micro Resistivity
 - Induction
 - Gamma Ray
 - Neutron
 - Electromagnetic
 - Dual Spacing Neutron
 - Porosity Log
 - Spontaneous Potential

2.3 Business Data

- Financial Information (content model in development)
 - Construction costs
 - Pre-Feasibility

Guidelines for Provision and Interchange of Geothermal Data Assets

- Feasibility Studies and/or Pilots
- Land Costs
 - Drilling Site
 - Power Plant Site
- Drilling Costs
 - Exploratory Drilling
 - Production Drilling
- Equipment Costs (Capital)
 - Drilling Site
 - Power Plant Site
- Electric Transmission Costs
- Employment – Construction Phase
- Legal and/or Regulatory Construction Costs
- Ongoing Operational Costs
 - Equipment Costs (Maintenance, Repair, etc.)
 - Drilling Site
 - Power Plant Site
 - Employment – Operational Phase
 - Legal and/or Regulatory Ongoing Costs
- Consumer Cost / Pricing Data
- Insurance Costs
 - Liability
 - Financial/Loan insurance
- Financing Arrangements
 - Government Cost Sharing (Non Tax)
 - Built to Own, Operate, Transfer, etc.
 - Loans
 - Bonds
 - Joint Ventures
- Regulatory
 - Environmental
 - Geological (mudslides, earthquakes, etc.)
 - Emissions and/or Pollution (Air and/or Water)
 - Impact Analysis
 - Mitigation
 - Compliance
 - Protected Species or Areas (Plant, animal, marine, wetlands, population centers, etc.)
 - Impact Analysis
 - Mitigation
 - Compliance
 - Permits
 - Drilling
 - Transmission Facilities
 - Power Plant
- Well Owner Information
 - Rights
 - Surface
 - Mineral
 - Water
 - Geothermal
 - Oil/Gas Field

Guidelines for Provision and Interchange of Geothermal Data Assets

- Name
- Operator
- Lease
 - Owner
 - Terms
- Taxes
 - Energy Taxes
 - Property Taxes
 - Tax Incentives
 - State
 - Federal
 - Local
- RFP/PPA/Demand

2.4 Geological Data

- Stratigraphy
 - Depth
 - Lithology
 - Induration
 - Mineralogy
 - Framework Mineralogy
 - Cement Mineralogy
 - Grain Size
 - Formation Name
- Hydrology
 - Saturation
 - Depth to Water Table
- Well Cores and/or Cuttings
 - Where stored
 - Images
 - Depth Ranges
 - Diameter
 - Full Diameter
 - Side wall
- Pressure
 - Bottom Hole
 - Wellhead
 - Shut In
 - Injection/Disposal
- Porosity
 - Depth
 - Source
 - Type
- Permeability
 - Depth
 - Method Used
 - Direction
 - Horizontal
 - Vertical
 - Unspecified
- Density
- Structure
 - Reservoir

Guidelines for Provision and Interchange of Geothermal Data Assets

- Enthalpy
- Pore Volume
- Size
- Thickness
- Faults

2.5 Geochemical

- Chemistry
 - Na
 - K
 - PH Level
 - Hardness
 - Alkalinity
 - Ca
 - Mg
 - Other Elements
- Gas Saturation
- Total Dissolved Solids
 - TDS Measured
 - TDS Calculated
- Medium Sampled
- Collection Method
- Type of material
 - rock
 - fluid
- Conductivity
- Turbidity
- Specific Gravity
- Isotope

2.6 Geothermal Data

- Heat Flow
 - Corrections Applied
 - Climate
 - Terrain
 - None (Uncorrected)
 - Calculation Type
 - Averaged
 - Interval
 - Quality Rating
- Heat Generation/Production
 - Element Content
 - Thorium (Th)
 - Uranium (235U, 238U)
 - Potassium (40K)
 - Calculated Value
- Thermal Conductivity
 - Measurement Type
 - QTM
 - Needle Probe
 - Divided Bar
 - Estimated

Guidelines for Provision and Interchange of Geothermal Data Assets

- Sample Depth Interval
- Number of Samples used to Calculate
- Temperature at Depth
 - Measurement Type
 - Continuous
 - Equilibrium
 - Non-Equilibrium
 - Discrete
 - BHT
 - Corrected/Uncorrected
 - Date/Time

2.7 Infrastructure

- Transmission
 - Type
 - Capacity
 - Owner(s) and/or Operator(s)
- Land Classification
 - Private
 - Federal
 - State
 - Adjacent
- Roads
- Power Plants – get more specific to geothermal – see examples below
 - Types
 - Nuclear
 - Geothermal
 - flash steam
 - binary
 - hybrid
 - other
 - Wind
 - Coal
 - Natural Gas
 - Biomass
 - Hydro
 - Others
 - Capacity
 - Owner(s) and/or Operator(s)
- Pipelines
 - Type
 - Distance

2.8 Production/Disposal/Injection Data

- Oil
- Gas
- Water/Brine
- Radioactive Isotopes
- Condensate
- Other
- CO₂
- Geothermal

Guidelines for Provision and Interchange of Geothermal Data Assets

- Flow Rate

2.9 Geophysical & Synthetic Data

- Seismic Reflection Survey
 - 2D vs. 3D
- Remote Sensing
- Gravity
- Magnetics
- Geodetic
- Stress
 - Method Used
 - Magnitudes
 - Orientation
 - Frequency
 - Spacing
 - Stress Axis
- Seismicity
 - Source
 - Natural
 - Induced
 - 'Propping'
 - Production
 - Injection
 - Magnitude

2.10 Direct Use

- Applications
- Equipment/Materials

3 Appendix B: Metadata Categories and Attributes⁵

- Source
 - Author
 - Affiliation
- Quality
 - Subjective score
 - Number of flagged values
 - Percentage of missing values
- Time/Date
 - Time of measurement
 - Time of last change
 - Time of submission
- Geographic/Location Data
 - Datum (SRS)
 - Lat/Long (potentially a range of coordinates to express an area vs. a specific point)
 - Surface
 - Lat/Long Coordinates at Surface
 - Elevation of Surface Coordinates
 - At Depth
 - Lat/Long Coordinates at Depth
 - Elevation of Coordinates at Depth (Negative #)
 - Elevation of Seafloor at Depth (Negative #)
 - Legal or Map Descriptors
 - District
 - County
 - State
 - Section
 - Survey
 - Block
 - Abstract
 - Township/Range
 - Plot
 - Tract
 - Name
 - Current
 - Prior
 - Region/Area
 - UTM, etc.

⁵ Siemens Corporate Research, "GTDA DATA REQUIREMENTS SPECIFICATION, Version 2.1," 07/8/2010, Appendix C, pp. 31-32.

4 Appendix C: Technical Format Types

- Web Sites (.html, .aspx, ...)
- Files
 - Unstructured data files
 - Documents (.doc, .pdf, .ppt, ...)
 - Images (.tif, .jpg, .png, ...)
 - Audio (.mp3, .wav, .wma, ...)
 - Video (.mp4, .mov, .wmv ...)
 - Structured data set files
 - Generic structure data sets (.xls, .xml, .json, ...)
 - GIS (maps, shape files, features, ...)
 - Non-GIS specialized data sets (.las, lis, ...)
- Databases
 - Traditional Relational (Access, SQL, Oracle, ...)
 - non-Traditional (HBase, Virtuoso, MongoDB, ...)
- Software Applications (Data Access, Analysis, Modeling, Data Submittal, etc.)

5 Appendix D: Standards for Metadata and Data Representation

- ISO
 - ISO 19115/19119/19139 (metadata)
- Dublin Core
 - Metadata Specification
 - DC-RDF
- Federal Geographic Data Committee (FGDC)
- Open Geospatial Consortium (OGC) – Web API Protocols
 - Catalog Service for the Web (CSW)
 - Web Map Service (WMS)
 - Web Feature Service (WFS)
 - Web Coverage Service (WCS)
- HDF Group
 - HDF5
- NetCDF