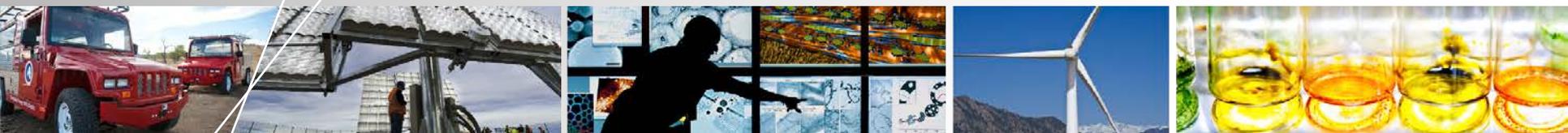




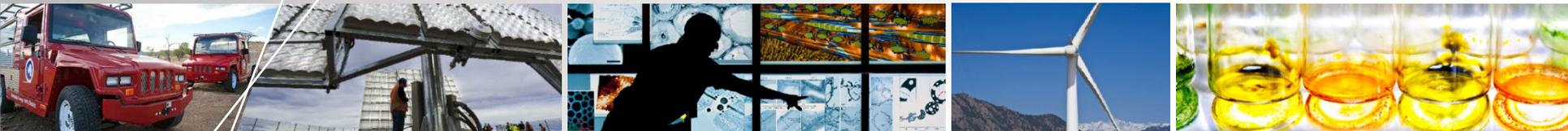
Geothermal Data Repository



Data Management Strategy

Jon Weers

May 11th, 2016



Guiding Principals

- ▶ Open by Default
 - Raw is Preferred
 - Built to Share
 - Keep it Simple
 - Cultivate Quality

Why we do this

Sharing information in accessible and useable ways to:

“reduce the inherent risk in early stages of development and encourage an independent investment market”

- Deloitte LLP, *Geothermal Risk Mitigation Strategies* (2008)

“fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives”

- President Obama's May 9, 2013 Executive Order

It is already home to more than 700 GDR submissions, 2,600 resources, and over 30GB of publicly accessible data from over 60 DOE GTO funds recipients.

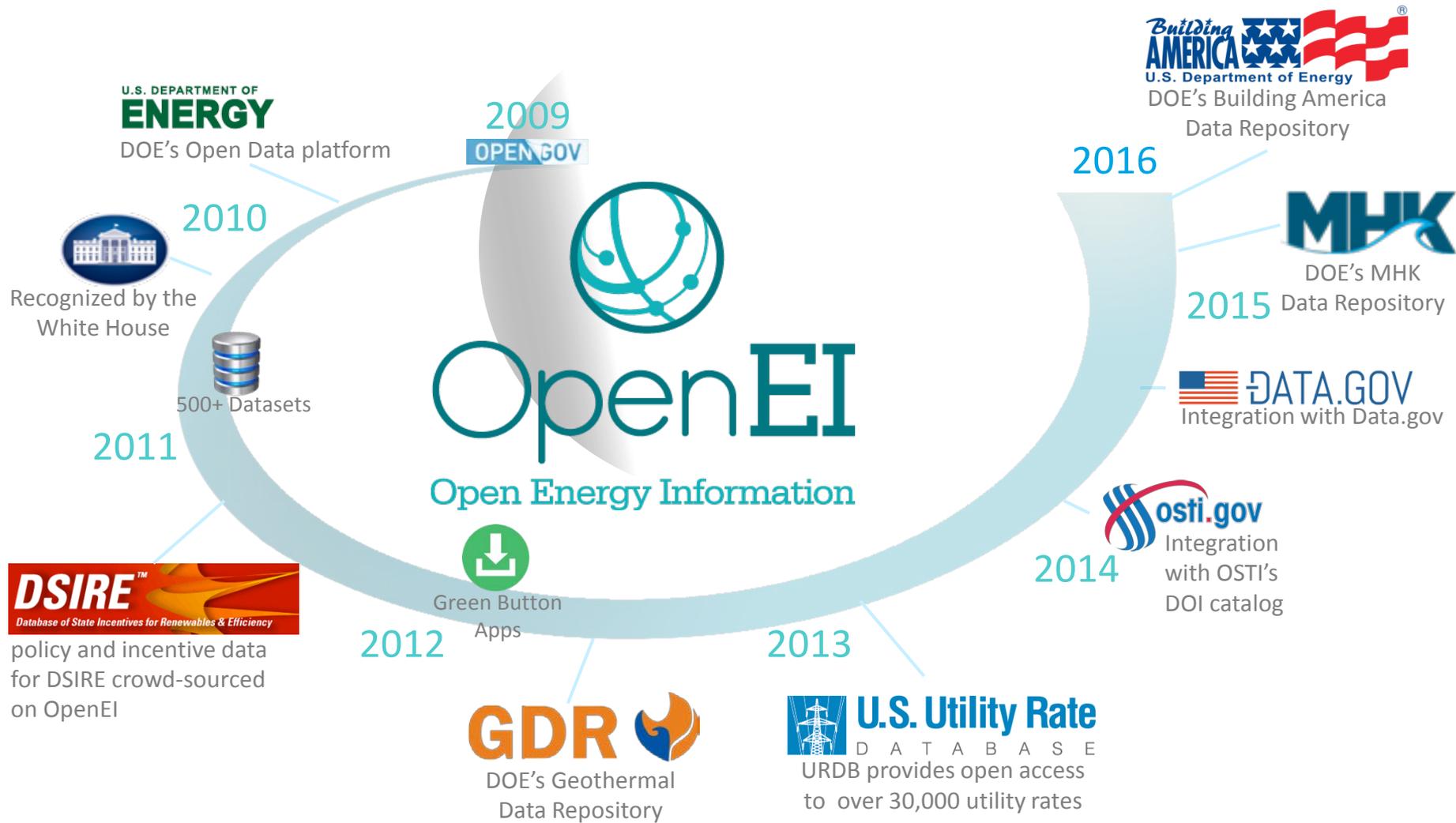


developed on
OpenEI

The DOE-GDR is a node on the NGDS. →

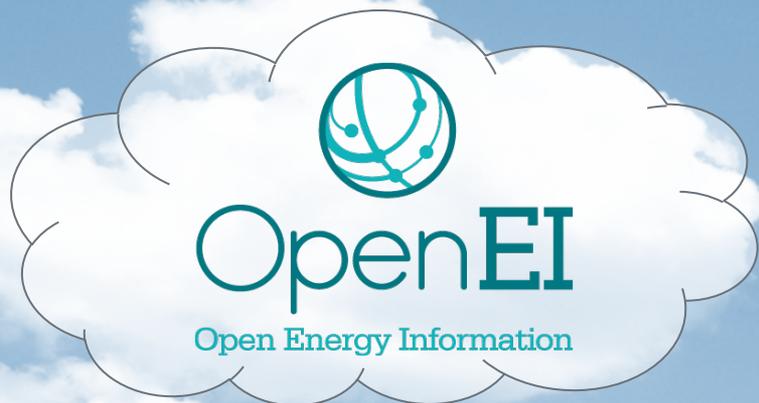


OpenEI: Open Energy Info since 2009



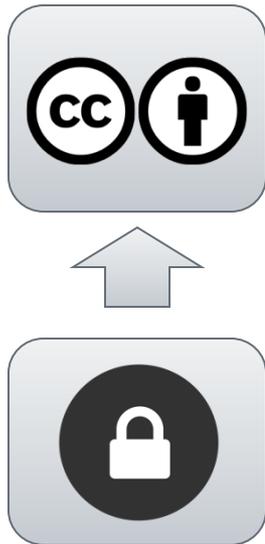
Why OpenEI

- Utilizes a DOE-approved secure Amazon environment
- OpenEI is built on the cloud!



Proper Data Management

- Open by default
- Secure when appropriate
 - All new submissions go into a secure vault until curated
 - Submitters can elect to have their submission stay in the vault for up to 5 years
 - All submission eventually become publicly available



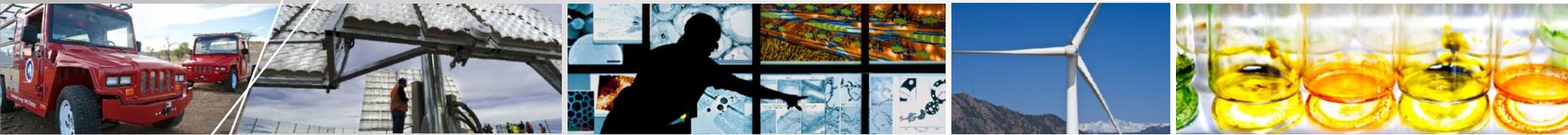
Moratorium

 **This submission is subject to a moratorium.**

Check this box if your submission is considered proprietary for a predetermined amount of time. This should only be checked if a moratorium is explicitly stated in your award, AOP, or contract and you have discussed this with your DOE project officer. You will be required to provide the release date specified in your award, AOP, or contract.

Release Date 

The earliest date on which all resources in this submission will be made publicly available.



Guiding Principals

- Open by Default
- ▶ Raw is Preferred
- Built to Share
- Keep it Simple
- Cultivate Quality

What kinds of data does DOE want?

- Maps, charts, photos, logs, reports, etc.
- “All data generated from projects funded by the Geothermal Technology Office”
- Including raw data



What kinds of data does DOE want?

- Maps, charts, photos, logs, reports, etc.
- “All data generated from projects funded by the Geothermal Technology Office”
- Including raw data



The Value of Raw Data



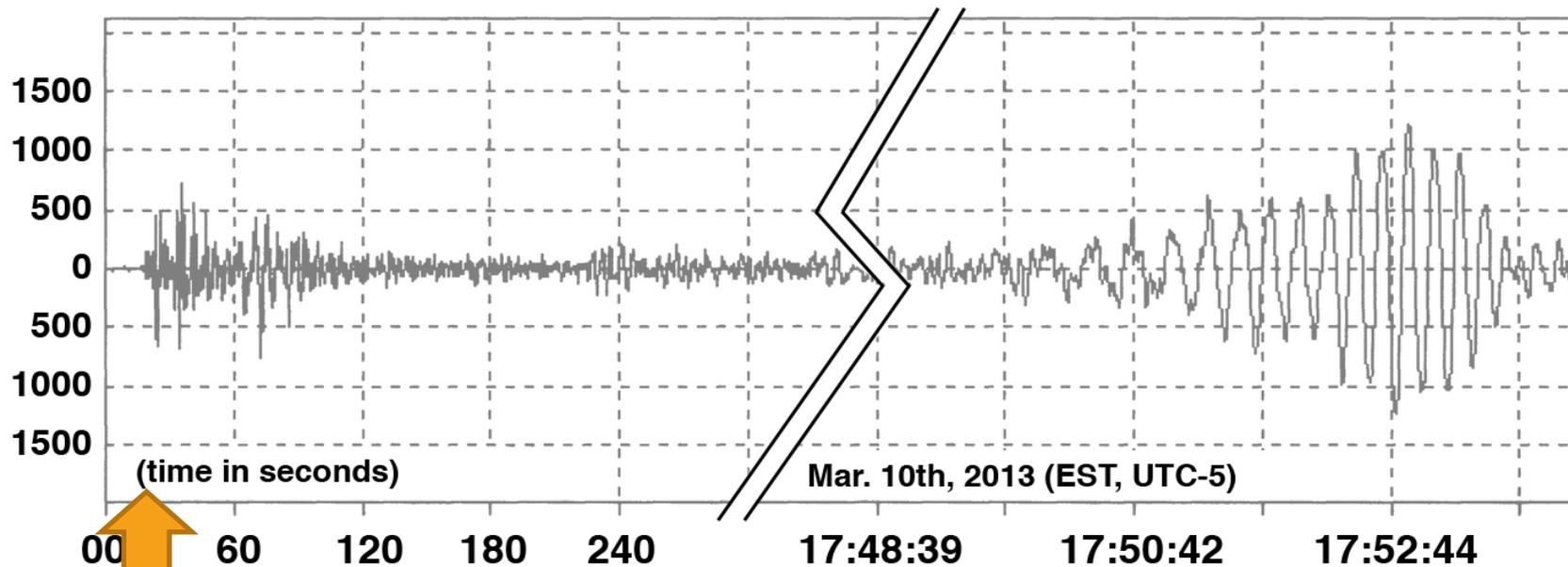
- Raw data can aid the understanding of aggregate findings and summaries
 - *Summary information can be filtered*
- Raw data is *re-useful*
 - *Remains free to be used in any context, including topics unrelated*
- Raw data can be used again and again in unforeseen ways
 - *Fuels innovation*

What kinds of data does DOE want?

There is such a thing as *too* raw.

Data should be finished enough that it is:

- Comprehensible
- Usable
- Useful



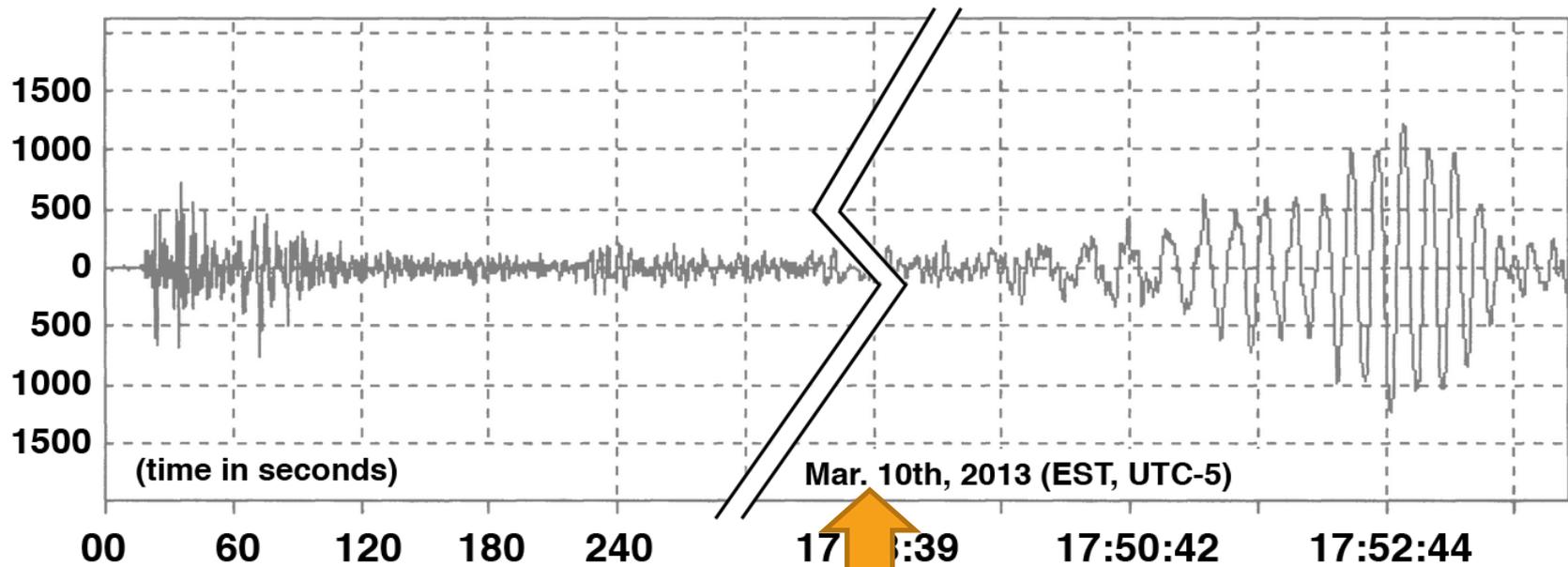
- No context (relative time)
- At least we know the units

What kinds of data does DOE want?

There is such a thing as *too* raw.

Data should be finished enough that it is:

- Comprehensible
- Usable
- Useful

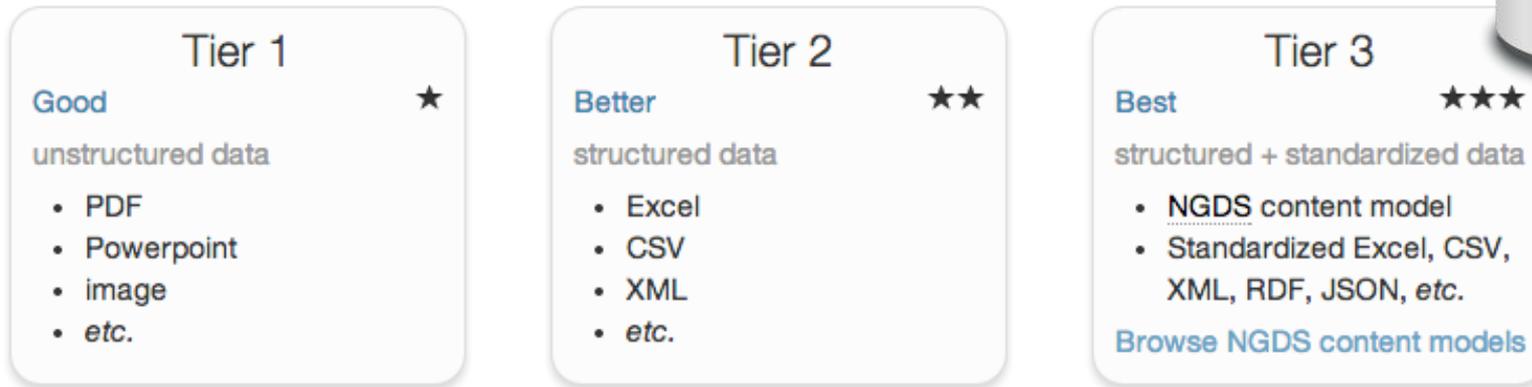


- Precise time (easier to contextualize)
- We know the date, units, and time zone

What kinds of data does DOE want?

The Three Tiers of Data

A simple guide to help submitters choose.



- **Tier One: Unstructured data**
 - PDFs, Powerpoints, scanned images, etc.
- **Tier Two: Structured data**
 - Excel spreadsheets, XML, database dumps, organized text, etc.
- **Tier Three: Structured and standardized data**
 - **Content models**, standardized excel, RDF, LOD, etc.

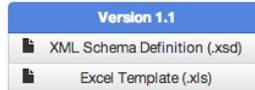
Use of Content (Data) Models

NGDS Content Models

- Standard structure for data collected
 - Organized context for common data
 - Makes the data more useful
 - Allows for like data to be easily combined, analyzed
- Available online at:
 - <http://schemas.usgin.org/models/>
- New Models are actively being developed
 - Based on international standards
 - Community vetted
 - Open for discussion at <https://github.com/usgin-models>



This workbook documents a content model for exchange of information about fluid fluxes into or out of a well. A generic model is intended to allow for oil, gas, or water injection (negative flow). The focus of this observation type is on injection (any fluid flowed into well) or disposal. Because of the wide variety of possible flow measurements (commodities, aggregation type), the model adopts a thin approach, which may result in many records being returned for a single well. Thus most data about the well is linked through the HeaderURI instead of included inline to keep the redundant data volume down.



Version 1.1

> Field List Version 1.1

-- fields in yellow are required --

Layer Name: FluidFluxInjection

Name	Description
OBJECTID (int)	Integer record count, added by ArcGIS servers during service deployment; not required but cannot be removed when deploying from ArcGIS feature class. Not user maintained, so this element appears in XML implementation but not in the Content Model Template.
ObservationURI (string)	Unique identifier for this observation. This is the identifier (Uniform Resource Identifier) that will be used to cross-reference data, samples, and other observations to the observation represented by this record. Best practice is to define an http URI that will dereference to a normative description of the observation. USGIN URI's may be redirected to any URL using the http://resources.usgin.org redirector; contact the USGIN development team at AZGS for instructions.
RecordName (string)	Common or human-readable name for the record. Recommend using only web-safe characters (a-z A-Z 0-9 _-.) in the name. Be consistent in the naming convention used within any given dataset.
Notes (string)	Any additional information to be provided, including description and other data not captured by the template, details about collection method, contact information for

Location as a required field

Location of data collection

- **Greatly increases the value of data**
 - Can associate data with other proximate data
 - Users of data are searching for it by location
 - Empowers map-based searches on NGDS
 - Location-aware devices auto-append your searches
- **Location now required for GDR**
 - GDR now requires submitters to define a location for each resource
 - Location can be defined as a point or an area
 - Area definition can be used to associate data with the US, or entire world

GDR Location Selector

Location Area Point

Map Satellite

Standish Honey Lake

Google

Map data ©2014 Google Imagery ©2014 TerraMetrics Terms of Use Report a map error

Use the map above to define an area or select one from the dropdown below.
You can also [define a point](#).

Geothermal Area Amedee Geothermal Area

[See Geothermal Areas on OpenEI for more information.](#)

Cancel Save changes

* "DOE Geothermal Data Repository" OpenEI: Open Energy Information. NREL. 2-15-2014 Web.

↓

NGDS SEARCH

MED? SEARCH

What kinds of data does DOE *not* want?

- **Personally Identifiable Information (PII)**

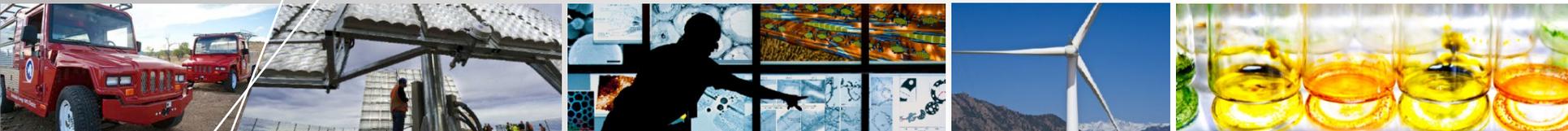
- Social security numbers
- Bank account numbers
- Home phone numbers and personal addresses of individuals not involved directly in the authoring of the data
- *“Any piece of information or combination of pieces that could be used to compromise the identity of an individual”*



- **Data not suitable for eventual public release**

- Precautionary Measures have been taken for protected data





Guiding Principals

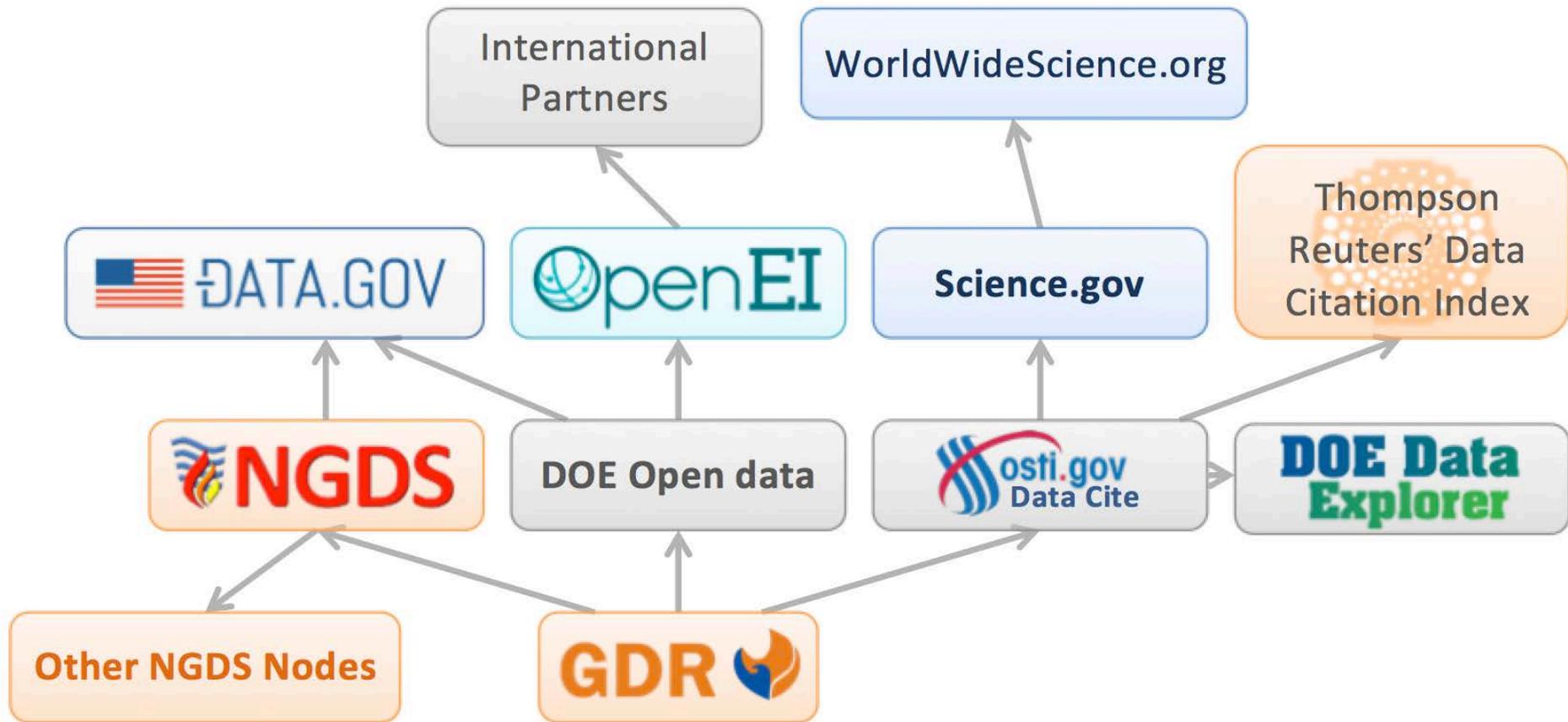
- Open by Default
- Raw is Preferred
- ▶ Built to Share
- Keep it Simple
- Cultivate Quality



Your item is located on isle 36,273.
Have a nice day 😊

- Recognize that there are other players in the data and knowledge brokering game
- Working with them, not against them
- Recalculate the way success is measured
- Be supportive, not duplicative

The far reach of GDR data



- Data added to the **GDR** will be automatically propagated to all the sites above, and more!
- Submitting is communicating to the scientific community

Linking instead of uploading

- Uploading isn't the only way to submit data
- **Some data should not be uploaded**
 - copyrighted materials, data already hosted online, data too large for internet download
- Use links to create a complete submission narrative

Link to an Existing Resource X

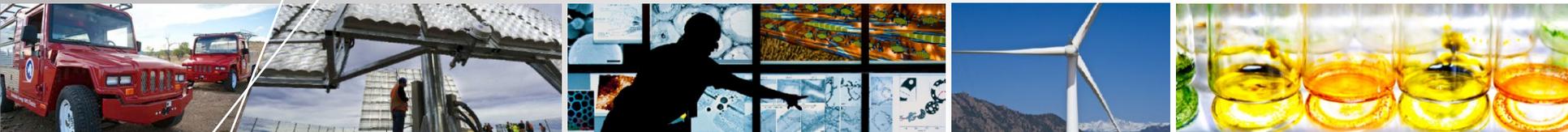
Link URI

Ideally a unique and permanent web address for this specific resource

Display Name

The title, filename, or short name used when displaying this link

- "DOE Geothermal Data Repository" OpenEI: Open Energy Information. NREL. 2016-02-12. Web.



Guiding Principals

- Open by Default
- Raw is Preferred
- Built to Share
- ▶ **Keep it Simple**
- Cultivate Quality

Smart Forms and Intelligent Agents

- Intelligent agents behind the scenes
- Autocompleting based on assumptions
 - User can then change if needed

Sectors

 Bioenergy	 Buildings Efficiency	 Chemical & Materials Science	 Computational Science	 Concentrating Solar Power	 Electricity Integration	 Energy Analysis
 Energy Systems Integration	 Geothermal Energy	 Hydrogen and Fuel Cells	 Photovoltaics	 Transportation	 Water Power	 Wind Energy

Select all NREL technology sectors that apply to this set of data

Keywords

NREL x energy x data x keywords

Supply several keywords that describe the information in this submission.

For example: Idaho Springs pv photovoltaic solar energy

Disciplines

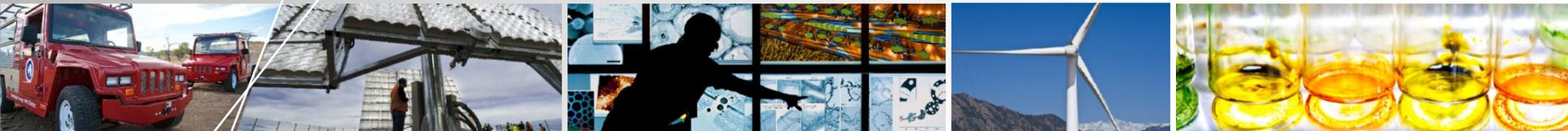
<input type="checkbox"/> Instrumentation	<input type="checkbox"/> Nuclear Science
<input checked="" type="checkbox"/> Knowledge Management	<input type="checkbox"/> Policy/Admin/Law
<input type="checkbox"/> Marine Hydrokinetic	<input type="checkbox"/> Storage
<input type="checkbox"/> Nanoscience/Nanotechnology	<input type="checkbox"/> Superconductivity/Superfluidity
<input type="checkbox"/> National Defense	<input type="checkbox"/> Synthetic Fuels

Select all fields of study that support the technology of the data

Program Codes for DOE

019 00 89 089
019 05 89 089
019 10 89 089
019 20 89 089
019 50 89 089
019 60 89 089





Guiding Principals

- Open by Default
- Raw is Preferred
- Built to Share
- Keep it Simple
- ▶ Cultivate Quality

Creating a culture of good metadata

- Rigorous curation
 - Team of curators
 - Access from anywhere
- Departmental support
 - Leverage department policies and procedures to enforce quality data



Home

Data ▾

Help ▾

About

Search

Submit Data

Curate Submissions

43

Awaiting curation

27

In curation

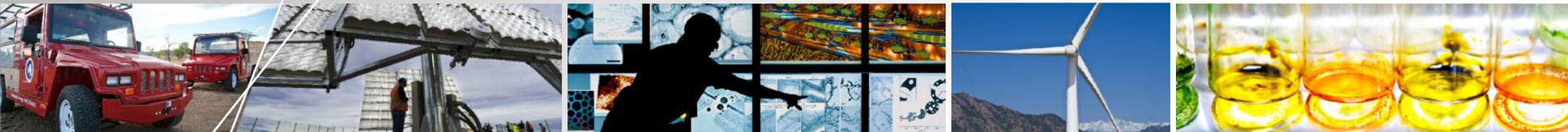
90

Need a DOI

Submission Name	Resources	Project Number	Organization	id	DOI	Submitted	Status
 Strength recovery of cement composites in steam a...	1	FY16 AOP 3.2.1.6	Brookhaven National Laboratory	804	ineligible	04/28/2016	Awaiting curation
 Fluid rare earth element anlyses from geothermal ...	1	EE0006748	University of California Davis	803	Add DOI	04/27/2016	Awaiting release



Thank You!



This work was funded by the US Department of Energy

Questions?

Jon.Weers@nrel.gov

303-275-3701