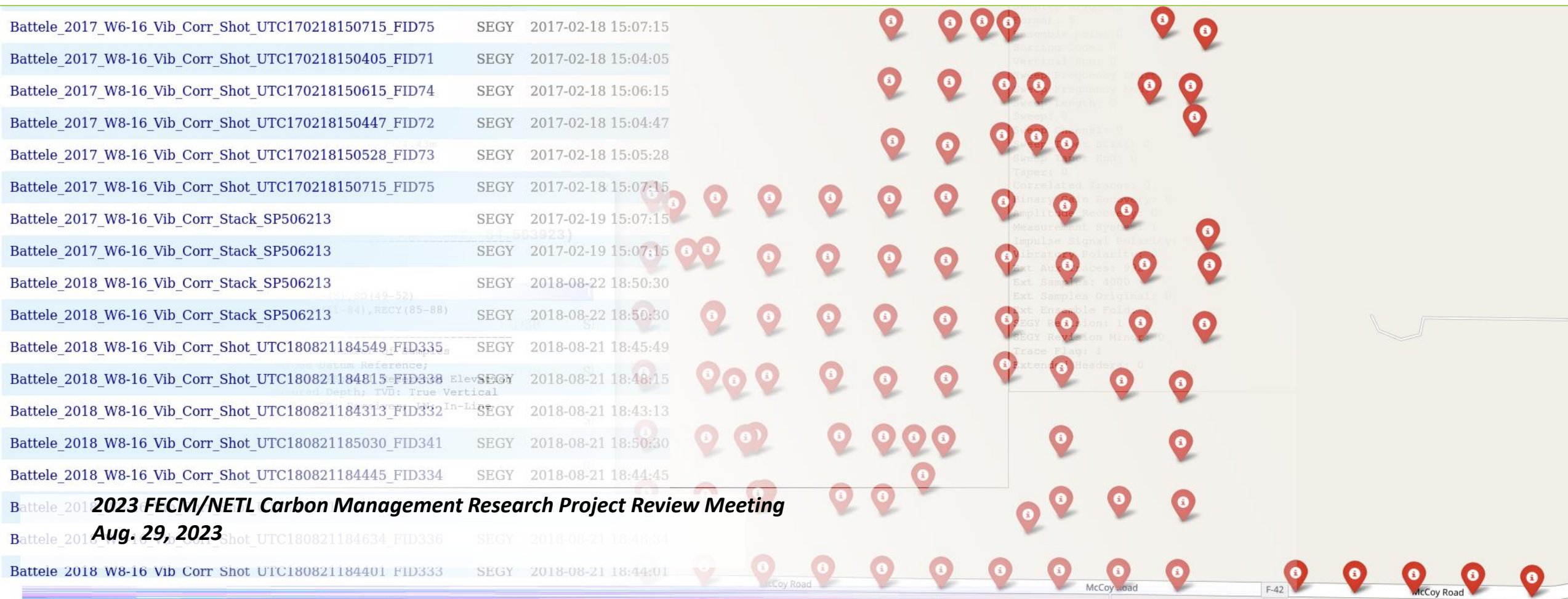# EDX4CCS 3D Data Preview Tool

Enabling Discovery of Large Seismic Datasets

Dakota Zaengle
NETL Support Contractor

2023 FECM/NETL Carbon Management Research Project Review Meeting
Aug. 29, 2023

Authors and Contact Information

Patrick Wingo[1,2]; Dakota Zaengle[1,2]; Maneesh Sharma[3,4]; Michael Gao[1,2]; Stephen Leveckis[1,2]; Julia Mulhern[1,2]; Jennifer Bauer[1]; Kelly Rose[1]

[1]National Energy Technology Laboratory, 1450 Queen Avenue SW, Albany, OR 97321, USA

[2]NETL Support Contractor, 1450 Queen Avenue SW, Albany, OR 97321, USA

[3]National Energy Technology Laboratory, 3610 Collins Ferry Road, Morgantown, WV 26505, USA

[4]NETL Support Contractor, 3610 Collins Ferry Road, Morgantown, WV 26505, USA

Disclaimer

What is being addressed?

So much data…

Several public seismic datasets exist, but they are hard to access
- ~ 32 million files (~215.5 TB) of data across five datasets
- Size constraints make it difficult to download first to explore
- Data are *heterogeneous;* different types of data; different file types
- Each dataset is constructed differently; challenge to navigate

Project Name	Total Disk Usage (TBs)	Resources
Illinois Basin Decatur Project (IBDP)	69.1	13,959,746
Stress in Deep Subsurface	2	2,335,875
MRCSP	8.1	596,778
HFTS-1	30.5	7,685,332
MSEEL	105.8	7,338,464
TOTALS	215.5	31,916,195

How can we make the search for desired data easier?
- Provide a *web service* to select and preview data assets
- Provide means to find data **by location**
- Provide lightweight preview for each selected asset, with preview style relevant to the type of data
- Provide means to download data

High Level Objectives

Provide a Tool to Locate, Preview, and Download Data

Develop a tool that will allow users to preview and select data from large CS datasets

Optimizing the search process by:
- Reducing the time it takes to locate the desired resources
- Avoiding unnecessary downloads when working to locate specific assets
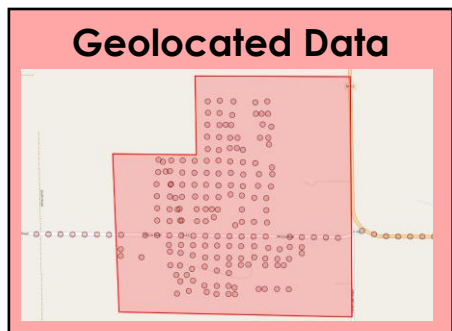- Improving overall access to DOE-funded datasets
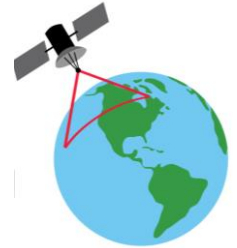
This tool is NOT:
- A comprehensive data analysis tool
- Applicable to arbitrary seismic datasets

Use **Open Source** technologies:
- Python (https://www.python.org/)
- Folium (https://python-visualization.github.io/folium/)
- NumPy (https://numpy.org/)
- Pandas (https://pandas.pydata.org/)
- GDAL (https://gdal.org/)
- Leaflet (https://leafletjs.com/)
- Flask (https://flask.palletsprojects.com/)
- Plotly (https://plotly.com/)
- ObsPy (https://github.com/obspy/obspy/wiki/)

General Workflow

Geolocated Data

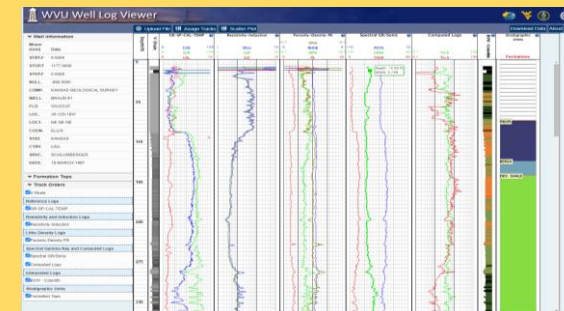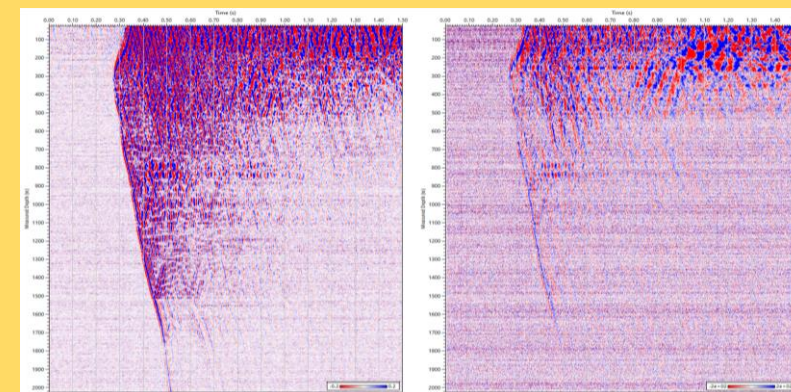*User selects data, routed to preview*

2D Data
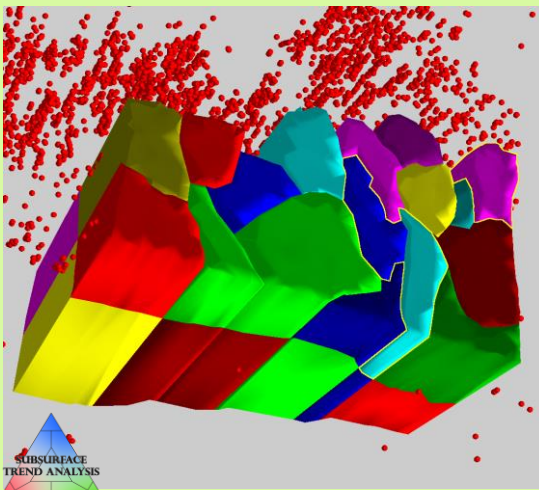
- Well Logs
- Seismic Profiles
- Well figures/ diagrams…
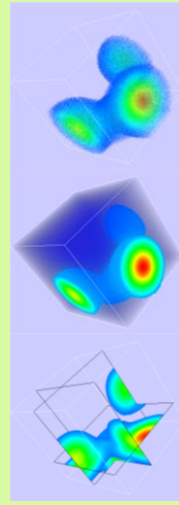
(*http://wvulogviewer.com/*)

3D Data

Mesh
- Boundaries
- Formation footprints

Volumetric (voxel)
- Structure

Other
- Point
- Line-string (well path)…

User downloads desired data
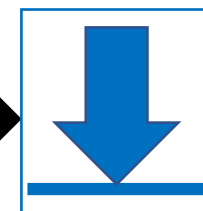
Component Flow Diagram
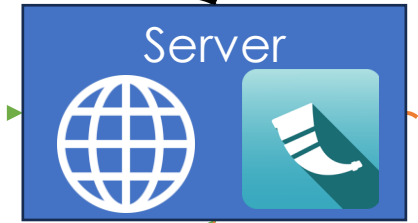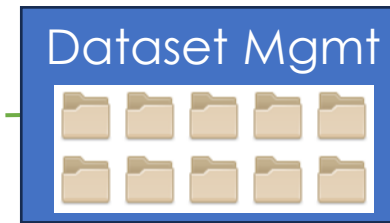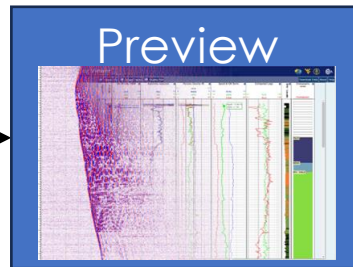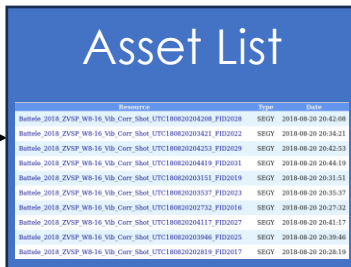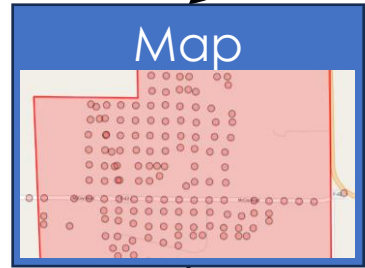
General Project Schedule

Theme: Incremental Progress

- **Completed Goals:**
  - Move datasets to Watt to be staged for development efforts
  - Provide alpha proof-of-concept
    - Single map layer for one of the datasets, compiling and anchoring specific file types
    - ~573K files were geo-anchored (~96% of resources)
    - Beginnings of map viewer

- **Upcoming Goals:**
  - Integrate WVU's well log viewer into tool (10/23)
  - Release beta version on DisCO$_2$ver platform, if available (1/24)
  - Release v1.0 of tool (1/25)

- **Final Outcome:**
  A tool allowing stakeholders to explore select datasets first with a map, preview assets with an appropriate visualization, and download as desired

Conceptual Work Categories

- The necessary work to accomplish this task can be broken down into the following categories:

1. Traverse Datasets	2. Geo-anchor Assets
3. Summarize Assets	4. Provide Previews

1. Traverse Datasets

Challenge:

- Each dataset is structured differently; no "one size fits all"
  - There is some overlap with assets, but that's not guaranteed
  - Even if file types are common, they may contain entirely different types of data

Solution:

- Each dataset needs a custom collection of logic to be evaluated and serve up available data

- While the logic for serving up the data may be different, the results can be served up in a common way

Target Datasets

- DE-FE0031686
- IBDP
- MRCSP
- MSEEL

Project Name	Total Disk Usage (TBs)	Resources
Illinois Basin Decatur Project (IBDP)	69.1	13,959,746
Stress in Deep Subsurface	2	2,335,875
MRCSP	8.1	596,778
HFTS-1	30.5	7,685,332
MSEEL	105.8	7,338,464
TOTALS	215.5	31,916,195

2. Geo-Anchoring Assets

Challenge:

- Not all data assets are provided with explicit geospatial coordinates
- Even if provided, coordinate system projection may not be obvious
- Four types of coordinate assignments:
  - Explicit in file spec
  - Explicit in loose text
  - Implicit based on reference to another resource
  - No direct reference
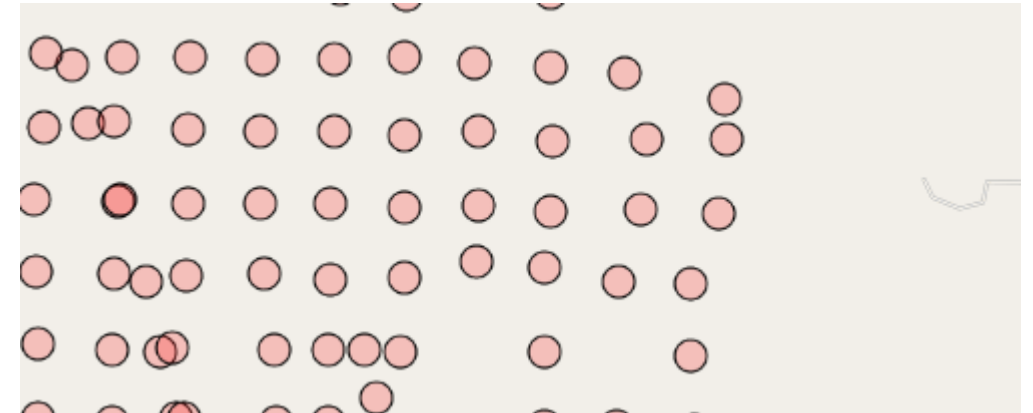
Solution:

- Associated metadata will often have some identifier that can be mapped to a location using another collection of assets
- Final reports and other descriptive documents can contain information on how to interpret geospatial information

3. Summarize Assets

Challenge:

- Specify files/assets for a given geographic location

- Bridge between point selection and asset preview

Solution:

- Presently, only listing files (naïve approach)

- Better approaches may exist for specific collections of data, such as:
  - Datasets where each file is a timeslice in a larger collection
  - Supplementary files that inform a bigger picture

Resource	Type	
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820204208_FID2028	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820203421_FID2022	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820204253_FID2029	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820204419_FID2031	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820203151_FID2019	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820203537_FID2023	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820202732_FID2016	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820204117_FID2027	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820203946_FID2025	SEGY	2018
Battele_2018_ZVSP_W8-16_Vib_Corr_Shot_UTC180820202819_FID2017	SEGY	2018

4. Provide Previews

Challenge:
- Many different asset/filetypes
- Different types of conceptual data

Solution:
- Provide a variety of previewers
  - Server queries data manager for appropriate viewer
  - Data is served to user with specified previewer
  - If no previewer is supplied, then use a placeholder

- Implemented previews:
  - ***.las files**
  - ***.sgy files**
  - Text files
  - Table files
  - Image files

Example Placeholder Output

Data View

- Data Layer: mrcsp
- Point Index: 85
- Record Id: 7

Download File
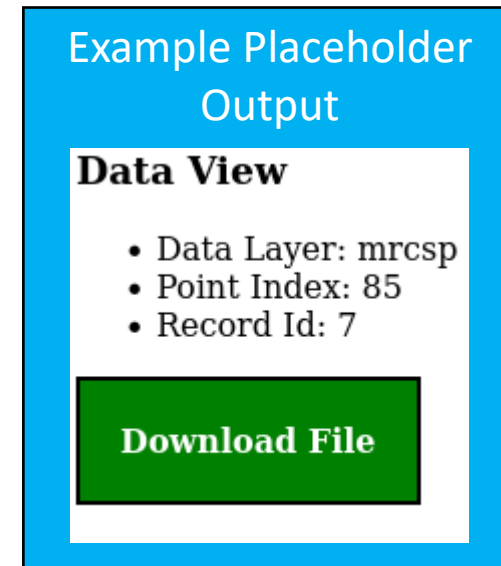
Example Data Visualization

Well Log (*.las) files.

- Well log files (*.las) are common

- Text-based format

- West Virginia University's *Well Log Viewer* is an existing web-based tool for viewing well logs
  - This tool has been incorporated into the previewing tool with only minor edits

Source: http://wvulogviewer.com/
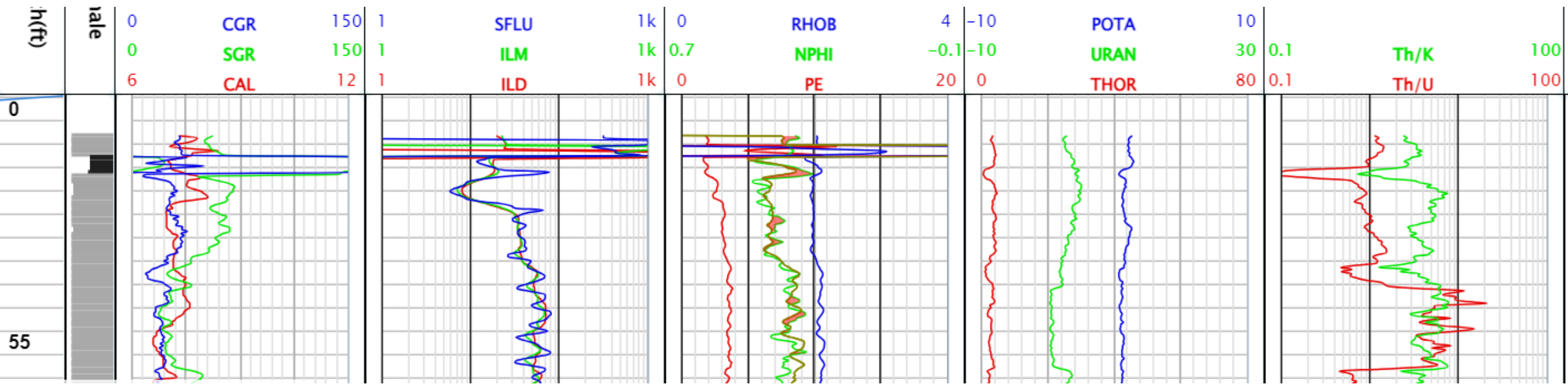
Example Data Visualization

Seismic (*.sgy) data

- SEGY files (*.sgy) are very common in seismic data, and vary wildly
  - Text headers do not have a standard structure, but can contain important metadata
  - Binary headers do have a standard structure, but are not always followed by vendors

- For visualizing SEGY files, a combination of text display and charts are used
  - Text header and (standard) binary field values are displayed
  - 2D or 3D charts are supplied, with the type of charts determined by heuristics, or data provided by dataset logic
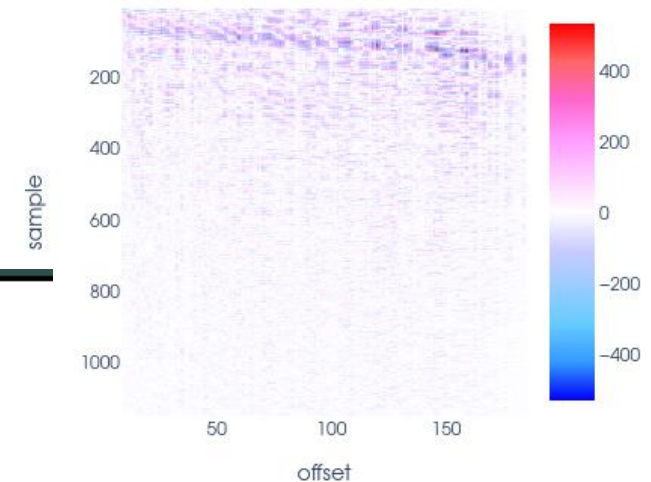
Text Header

```
C01 ----------------------------- JOB.INFO ------------------------------------
C02 CLIENT: Battelle, FIELD: Chester16Unit, WELL: 8-16
C03 CONTRACTOR: Silixa, FIELD ENG1: Nanut G., FIELD ENG2: Jeroen B.
C04 PD: MSL, RDR: KB, SDR: MSL, RRE: 407.21m, SRE: 402.64m; SD: 0.00m
C05 UNITS: Meter
```

Default Binary Header

```
Job ID: 0
Line Number: 0
Reel Number: 0
Traces: 954
Aux Traces: 0
Interval: 1000
```

Future Challenges and Developments

NETL NATIONAL ENERGY TECHNOLOGY LABORATORY

Challenges:

- Tool cannot become available until source datasets can be hosted on the EDX DisCO$_2$ver platform

- Each included dataset requires custom logic and code for parsing
  - The following datasets have *some* logic implemented:
    - Illinois Basin – Decatur Project (*IBDP*)
    - Midwest Regional Carbon Sequestration Partnership (*MRCSP*)
  - Logic for the following datasets is pending:
    - Marcellus Shale Energy and Environment Laboratory (*MSEEL*)
    - Stress in Deep Surface

- "Asset" classification is file-centric
  - Conceptual presentation will likely need to reflect how stakeholders will utilize data

Development Aims:

- More format types will need to be included
  - *.kml, *.xml, and others are candidates for implementation
  - Additional preview types
- Publish API for including new datasets
  - Custom logic will be required, but will be known
- Publish API for including new previewers
- Better downloading approach
  - Checkboxes?
  - Shopping cart?
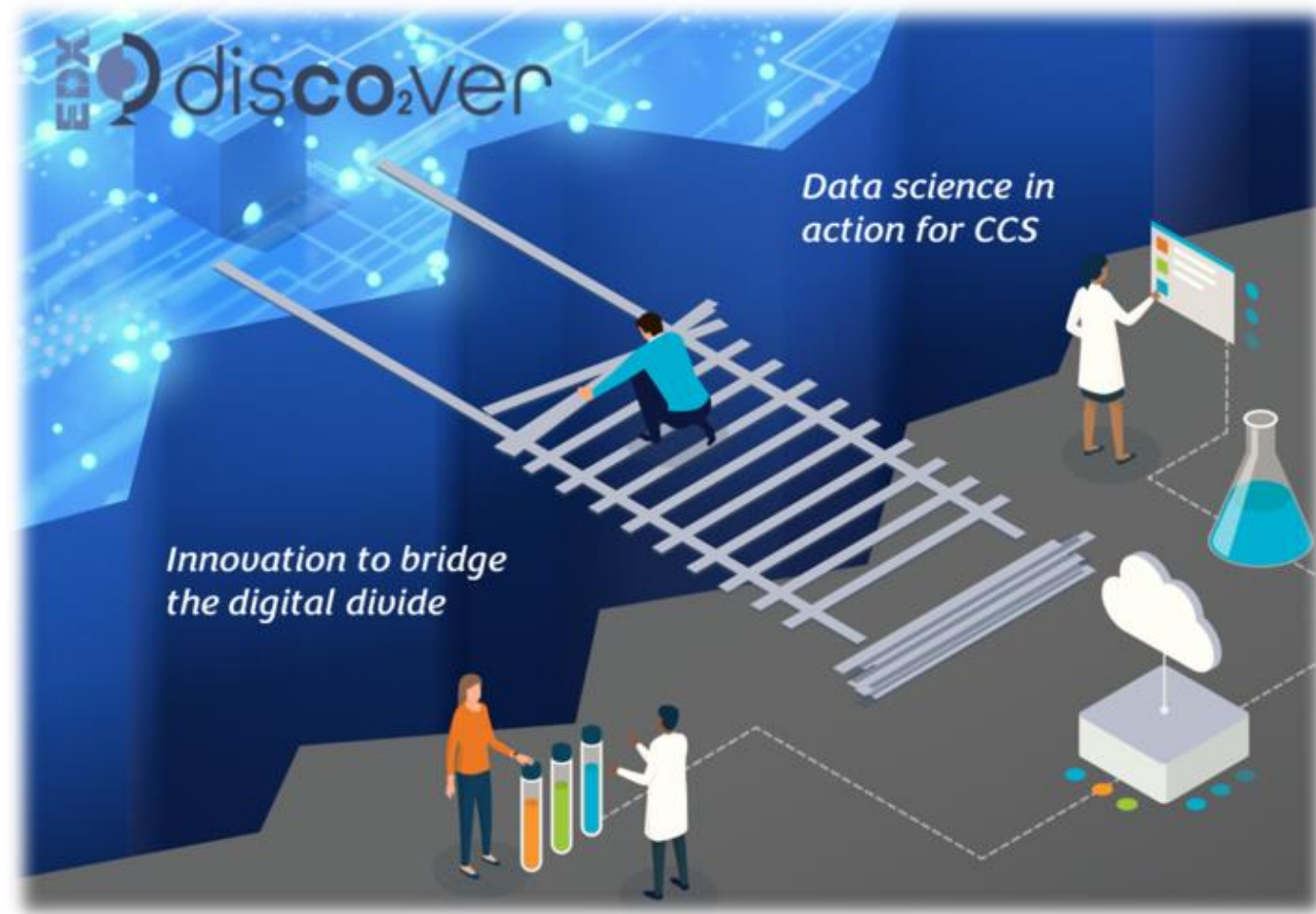
Tuesday Evening - Live Tool Demos!

When: 5:45 - 7:45 p.m.

Where: The Ballroom Foyer and East/West Atriums

What:

- **Environmental Justice and Social Justice** for CS Systems
- The **international offshore CS** and web-database and tool
- RokBase, Virtualizing **CS Rock Property Data** platform
- **Class VI Data Support Tool** for regulatory requirements
- **CO2 Pipeline Routing** Smart Tool
- Co2Locate - **Class II Well Reuse and Regional Evaluation Tool**
- **Carbon Storage Planning Framework Dashboard**
- **3D Data Viewer** and Preview Capability
- AIIM Model, **Assessing Infrastructure Reuse Potential** for CS
- **EDX disCO2ver,** a one-stop tool for CO_2 digital resources

➕ In demo "theater room" **EDX** Energy Data eXchange **support team** will offer **in person demos & Q&A**

EDX disCO2ver — Data science in action for CCS — Innovation to bridge the digital divide

Acknowledgments

NETL
RESOURCES

VISIT US AT: www.NETL.DOE.gov

🐦 **@NETL_DOE**

📷 **@NETL_DOE**

f **@NationalEnergyTechnologyLaboratory**

U.S. DEPARTMENT OF
ENERGY