Developing a Virtual Data Subsurface Framework for the U.S.- Combining Products of Fossil Energy R&D with Advanced Data Computing

Rose, K.[1], Justman, D. [4], Bauer, J. [1], Baker, D.V. "Vic" [2], Barkhurst, A.[2], Sabbatino, M. [4], Romeo, L. [4], DiGiulio, J. [4], Creason, G. [4], Jones, T.J. [2], and Rowan, C. [3]

[1]US Department of Energy, National Energy Technology Laboratory, Albany, OR; [2]Mid-Atlantic Technology Research & Innovation Center (MATRIC), Morgantown WV; [3]Optimal Solutions and Technologies, Inc., Morgantown, WV, [4]AECOM, Albany, OR

NATIONAL ENERGY TECHNOLOGY LABORATORY

Research & Innovation Center

Abstract: The data revolution has resulted in a proliferation of resources that span beyond commercial and social networking domains. Research, scientific, and engineering data resources, including subsurface characterization, modeling, and analytical datasets, are increasingly available through online portals, warehouses, and systems. Data for subsurface systems is still challenging to access, discontinuous, and varies in resolution. However, with the proliferation of online data there are significant opportunities to advance access and knowledge of subsurface systems. DOE's Energy Data eXchange (EDX) is an online data computing platform designed to improve access to fossil energy R&D products, support multi-organizational R&D through an increasing suite of virtual sharing and analytical capabilities, through public and private online workspaces.

The EDX team has been developing a virtual subsurface digital data framework to support FE R&D for subsurface energy research. The system curates both structured and unstructured data about subsurface systems. These resources span petrophysical, geologic, engineering, geophysical, interpretations, models, and analyses associated with carbon storage, water, oil, gas, geothermal, induced seismicity and other subsurface systems to support the development of a virtual subsurface data framework. The EDX team is also currently developing custom machine learning algorithms and capabilities to enhance user experience, make access and connection to relevant, open-source, subsurface data resources more efficient for research teams to use, analyze and draw insights. EDX content and capabilities are continuously evolving. The development of a virtual subsurface data framework to support DOE FE data science and computing is the next step in the public and private resources in EDX seek to make subsurface energy research more efficient, reduce redundancy, and drive innovation.

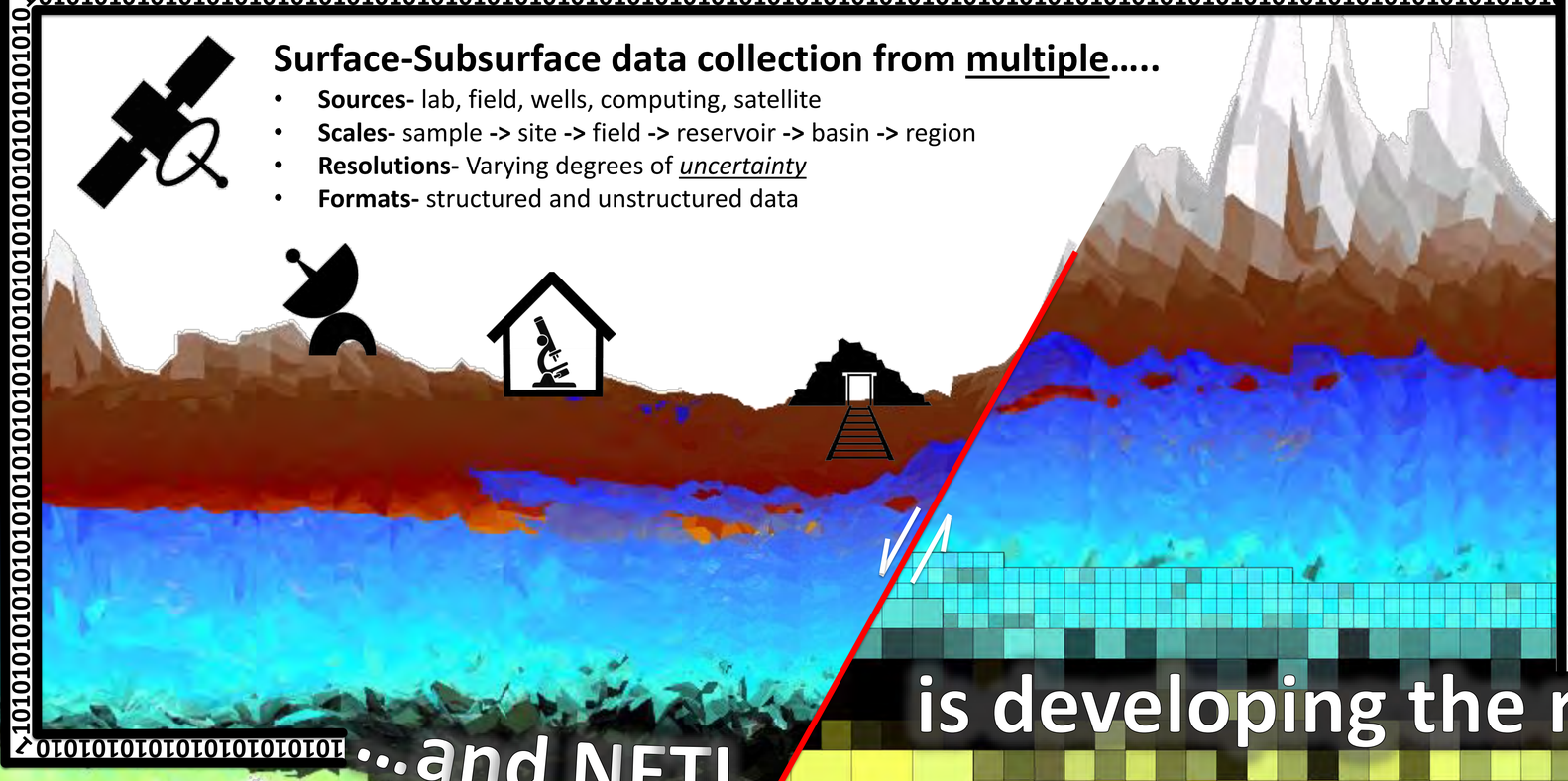
EDX™ Energy Data eXchange

Need - Employ Data for New Insights

- **Volume of data is growing:** Scientific data is projected to exceed more than 40,000 exabytes by 2020.
- **Finding older R&D data is hard:** As published research ages, access to the underlying datasets decreases.
- **20%** of world's data are stored online while **80%** are being privately held

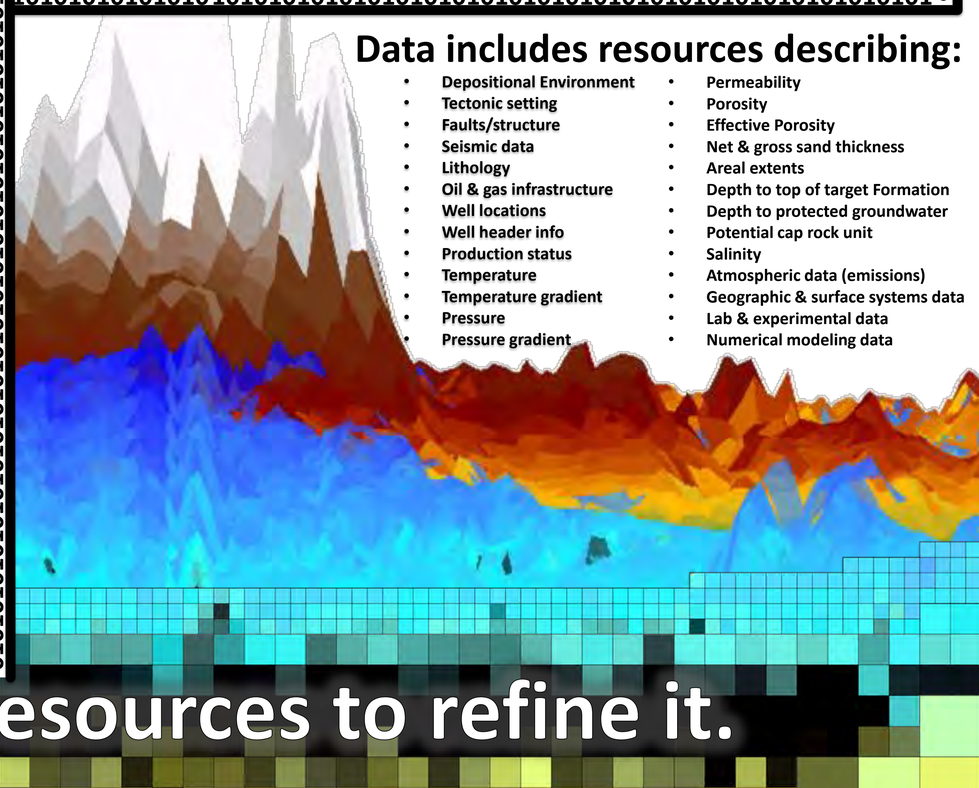
Surface-Subsurface data collection from multiple…..

- **Sources-** lab, field, wells, computing, satellite
- **Scales-** sample -> site -> field -> reservoir -> basin -> region
- **Resolutions-** Varying degrees of *uncertainty*
- **Formats-** structured and unstructured data



Data includes resources describing:

- Depositional Environment
- Tectonic setting
- Faults/structure
- Seismic data
- Lithology
- Oil & gas infrastructure
- Well locations
- Well header info
- Production status
- Temperature
- Temperature gradient
- Pressure
- Pressure gradient
- Permeability
- Porosity
- Effective Porosity
- Net & gross sand thickness
- Areal extents
- Depth to top of target Formation
- Depth to protected groundwater
- Potential cap rock unit
- Salinity
- Atmospheric data (emissions)
- Geographic & surface systems data
- Lab & experimental data
- Numerical modeling data



Developing a Virtual Subsurface Data Framework – For FE R&D

Data Science Hierarchy of Needs

Scientific data poses unique challenges for organizations that need to ***store, share, and curate data***. The **Data Science Hierarchy of Needs** outlines the fundamental concepts of data management that NETL is applying to its ever increasing data stores. This includes developing the tools, data, and procedures needed to move data from disparate data sets and schemas into a unified **Living Database**. This Living Database includes a continuous cycle of data collection and analysis to satisfy the needs of data science.

Inform
Decision making, artificial intelligence, & machine learning

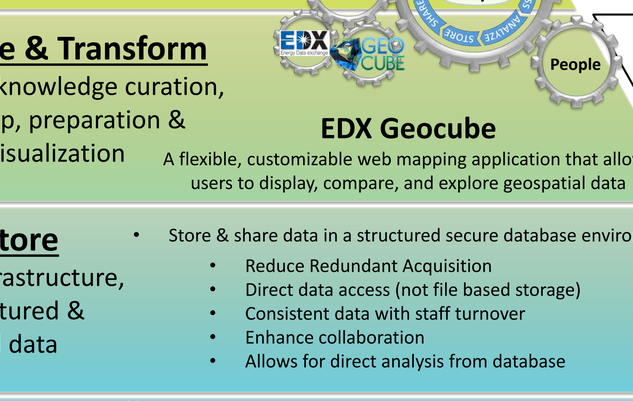
Analyze & Optimize
Testing, experimentation, & validation

Integrate & Label
Analytics, metrics, segments, aggregates, features, & training data

Explore & Transform
Data and knowledge curation, cleanup, preparation & visualization

Move & Store
Data workflow, infrastructure, pipelines, structured & unstructured data

Discover & Collect
Instrumentation, lab reports, sensors, external data, user generated data

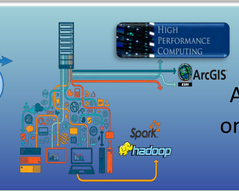


EDX Geocube

A flexible, customizable web mapping application that allows users to display, compare, and explore geospatial data

- Store & share data in a structured secure database environment
  - Reduce Redundant Acquisition
  - Direct data access (not file based storage)
  - Consistent data with staff turnover
  - Enhance collaboration
  - Allows for direct analysis from database

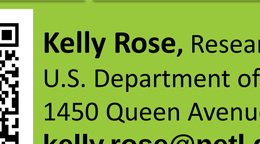
Machine Learning Web Search
Expert-driven Web Search

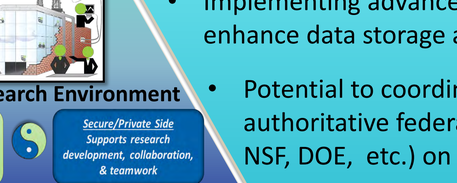


EDX Smart Search

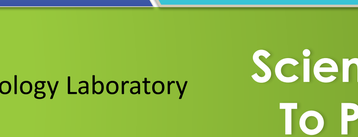
A machine learning, big data tool for rapid, online, .zip, & FTP spatial & non-spatial data mining with Hadoop + Bing + ESRI

Billions of Records
Employing "smart" search tools to include open resources
Machine Learning Web Search
Millions of Records
Oil & Gas, Geothermal Data & Resources





Carbon Storage Data & Resources
Millions of Records



Millions of Records

Ultimate Goal… To maximize time and resources for users to spend on data analytics & decision making

Framework leverages NETL tools and capabilities within EDX to streamline and automate data *curation*, *storage*, and *processing*

BLOSOM · BSEE · CO_2 STORE · MMVOPS · eia · EDX · FRACGEN NFFLOW · GEO CUBE · geoWELL · GLOBAL CO2 STORAGE PORTAL · GRIIDC · MFiX · NATCARB viewer · NETLEC · papaya · ParaView · Freemarkly Geoscience · UCR · VARIABLE GRID METHOD · Well Log Viewer

EDX - A Virtual Library & Laboratory for Energy Science

- Virtualizing team analytics
- Continuing innovations to connect researchers to online Earth-Energy system resources
  - Increasing number of tools and apps for use in team workspaces

A Secure Research Environment

Public Side — Enables knowledge transfer, data preservation, reuse & discovery

Secure/Private Side — Supports research development, collaboration, & teamwork

Next Steps

- Next 18 months focused curation of DOE FE data to seed virtual subsurface data framework further, including resources from RCSP's, Carbon Safe, oil/gas field projects, etc
- Implementing NETL "Smart Search" tool via EDX to facilitate adaptive, real-time integration of open, online data from externally hosted sources
- Implementing advanced data computing solutions to enhance data storage and usability
- Potential to coordinate or actively integrate other authoritative federal subsurface systems (e.g. USGS, NSF, DOE, etc.) on private side of system

Acknowledgment: This technical effort was preformed in support of the National Energy Technology Laboratory's ongoing research under the RES contract DE-FE0004000.

U.S. DEPARTMENT OF ENERGY

Kelly Rose, Research Geologist
U.S. Department of Energy, National Energy Technology Laboratory
1450 Queen Avenue SW Albany, OR 97321
kelly.rose@netl.doe.gov http://orcid.org/0000-0001-6130-4727

Science & Engineering To Power Our Future