



Going Beyond the Publication NETL's Energy Data eXchange

a coordination and collaboration platform to support science driven R&D

Kelly Rose

Geology & Geospatial Team Lead

National Energy Technology Laboratory

CAES Big Data Workshop Tuesday, May 19th, 2015

11)Xportfoli





NETL Research and Development



- Of DOE's 17 national labs only 1 dedicated to fossil energy R&D
 - Growing R&D in EERE and BES arenas
- One lab, three R&D locations, one management structure
 - Government owned and operated (GOGO)
- Both program management/execution and fossil energy R&D
 - Fundamental science through technology demonstrations





Albany, Oregon



Morgantown, West Virginia



Pittsburgh, Pennsylvania



NETL ORD Research



Engineering Solutions using an Integrated Approach

Validated **simulation** tools to predict behavior of engineered natural systems: reservoir to receptor

> **Field-scale data** to calibrate and validate predictions and to establish baselines

System & Process **Optimization Field** Data Collection Monitoring & Data Mining **Geomaterials** Science at **Conditions**

Scientific understanding of the behavior/properties of real **materials** at conditions







multi-organizational

multi-disciplinary



energy

computing/information sciences





Office of Research and Development

Increasing collection and creation of data, tools, and models related to improve R&D efficiency





Numerous methods to access, interact with, and publish data, tools and models

Several limitations to current methods





Image from : http://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416







EDX – Why cont.

Office of Research and Development











NETL

A National Look at Carbon Sequestration

Regional Carbon Sequestration Partnerships



EDX – Why...

Office of Research and Development

Research teams often encounter problems:

- Finding & accessing pertinent, relevant datasets
- Accessing previously developed data
- Sharing data resources amongst the team
- Conducting research across multi-organizational teams

In 2011, **EDX** was developed for NETL/DOE R&D as an *innovative* solution to these challenges by offering:

- Better coordination of data among NETL research teams and with outside collaborators
- Advanced tools for fossil energy research
- Enduring and reliable access to data

EDX is <u>evolves</u> in response to the needs of its users and NETL's knowledge transfer goals.



Image from : http://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416









Built by researchers for research

In 2011, the **Energy Data eXchange (EDX)**, was developed for NETL/DOE R&D as an *innovative* solution to these challenges by offering:

- A secure, online *coordination and collaboration platform* supports energy research, knowledge transfer and data needs
- Enduring and reliable *access* to historic and current R&D *data, data driven products, and tools*
- Both *public* and *secure, private* functionalities

Public Access Enable knowledge transfer, data reuse & discovery <u>Secure/Private Access</u> Support research development, collaboration, & online analytics





EDX aligns to Executive and DOE orders





EDX aligns to Executive and DOE orders

What is the Technology Roadmap?

The DOE Technology Roadmap will outline the future state of IT that will advance DOE's mission over the next five years It includes:

- Consensus future states for technology domains
- Focus areas for each technology domain,
- Recommended steps within each focus area
- Highlights and success stories of early adopters and technology pioneers across the complex

DOE

Technology

Roadmap

Technology

Advisory

Board

Partner

Engagement

Data Strategy



Built by researchers, For research....

ENERGY | the ENERGY lab National Energy Technology Laboratory

their success inside and outside DOE

3. Foster collaboration among diverse stakeholders across the complex



Office of the Chief Information Officer

Officer

Core Values y mission needs s the pace of innovation

DOE advances through collaboration 3.

The Technology Roadmap will outline the future state of IT that will advance DOE's mission over the next five years

The Technology Advisory Board (TAB) identifies innovative IT solutions that can be deployed across the DOE

Identifying and engaging strategic vendors and industry partners to identify optimal solutions to fulfill DOE needs

Define Enterprise Data Strategy to enhance information sharing and safeguarding

Empowerin Next 0 D 6 Ð noissin



Key Philosophies



CKAN

PUBLISH

INTERACT



EDX aims to improve efficiency and reduce redundancy by:

- Building EDX on the open source platform, CKAN
 - Highly customized
 - Enhanced search functionality of existing data repositories built on the CKAN platform
- **Discoverability**, **connecting users to existing online tools and** data repositories built on different platforms (e.g., BOEM, BSEE, NOAA, NETL, NASA, USGS, etc.)
- Accessibility & Reduce Redundancy, using existing web services to connect our users to additional data, while avoid conflicting versions of the data
- Leverage private-secure capabilities
- Ensuring long term access to federally funded products



Governance, Planning, Coordination













EDX constantly *evolves* in response to the needs of its *users ongoing R&D* and *DOE's goals*

Goal, to serve as an R&D coordination-collaboration platform for NETL affiliated R&D

- EDX past, version 1 released end of FY11
 - Discoverability resources & tools, public facing
 - **Preservation of data resources** resulting from NETL affiliated R&D
 - Connecting EDX users to publically available, authoritative data resources
- EDX present, version 2 released in FY13
 - Incorporated private sharing capabilities for multiorganizational R&D teams





EDX Present





EDX Goals:

- Improve<u>accessibility</u>to data resources
- Improve<u>discoverability</u> resources&tools
 - <u>Preservation</u> of data resources resulting from NETL affiliated R&D
 - <u>Connecting</u> EDX users to publically available, authoritative data resources
- Offer <u>private sharing</u> capabilities for multiorganizational R&D teams
 - Facilitating <u>efficient</u>, ongoing R&D





EDX – Workflow

ffice of Research nd Development

EDX serves as a **liaison** between datasets and their utility to future research

- Contribute Public Data
- Contribute Private
 Data until it is
 ready to be
 pushed to Public
- Access outside resources







Many Ways to Use...





LOBERARY
 LABOR AND A CONTRACT AND A CONTRA







Public Sharing & Accessing



Aligns to Executive and DOE orders



Search or sort for data based off *keyword, file type, data source,* and *spatial location*

Canada	👬 EDX 🍈 EDX (private) 👽 NGDS	2387 🔮 Data.gov 1393:	0 🔚 Other			
the state	Search		Q			
United States			Advanced Search			
Mexico	All datasets	Order by:	Relevance •			
and the second	Not finding what you're looking for? Try one of the additi	onal tabs above.				
ap data © OpenStreetMap						
ntributors						
es by mapbox	1:250,000 Geology of Tennessee	e of the printed 1,250,000 and	logic many from the			
Keywords Clear All	Tennessee Department of Environment and Conservation. Division of					
/ata (396)	HTML					
download (348)						
ais (311)	1971–1982 Alaska Gas Pipeline Wars					
Pennsylvania (258)	The 40-year-long epic quest to build an Alaska natu	ral gas pipeline started with a b	attle royal in the mid-			
teology (241)	1970s. The pipeline project would be one of the larg	est privately				
1505 (166)						
13g5 (100)						
ASDA (103)	1982-2001 The Yukon Pacific Era	ific Corn revived interact in an	Alarka gar pipeline in			
pil (147)	1983. Yukon Pacific was born amid doubts among so	ome Alaskans that	Alaska gas pipeline in			
jas (131)	HTML					
vater (114)						
Show More Keywords	2000 Resource Assessment of Selected Coa	Beds and Zones in the No	orthern and			
	A report detailing the surveying and resulting predic	tions for coal beds in regions o	f the Appalachian Basin.			
Eormatr of All	Data downloads include Coal Bed Thickness (Isopach	s), Mined Area,				
Clear All						
HTML (803)	HTML					









Contributing Resources





ontribut n <i>EDX a</i> p <u>rivate</u> d	e datasets a <i>ccount</i> to th collaborative	nd links with e <u>public</u> or a e workspace
Registe	er for an Account	
Username:	username	
Full Name:	Joe Bloggs	
Email:	joe@example.com	
i -	Please check this box if you do not have an NE address.	ETL or contractor NETL email
Password:	•••••	
Confirm:		
NOTICE TO USI Users who provide th access to EDX once t check your email inb creation form.	ERS: heir "@netl.doe.gov" or "@contr.netl.doe.gov" email hey have validated their email address. To validate ox for a validation message from EDX once you've	address will have instant your email address, please submitted your account
Users who do not ha Point of Contact (PO(confirmed your affili your account.	ve an NETL or NETL contractor email address are re C)'s name and email address to gain access to the ation with your listed NETL POC, you will receive an	quired to provide an NETL system. Once we have email validation to activate
		Create Account

https://edx.netl.doe.gov





CR

or im to this



Click on a dataset to learn more – can *preview* datasets and *access* links or *download* data

Share data through *social media*, *groups*, and *workspaces*

Save datasets to your *library* or *follow* datasets for updates

asets		Add to Library Nominate to Grow, O Foll
set has not yet d, why not add one te Dataset	Submarine C	Cables
	Dataset extent	
k		
eas, news stories have been related set yet, why not w?	Vap dat & OpenStreetilize contributors	
Add Item	Tiles by MapBox	
	Occurrence of submarine cables in an	nd around U.S. navigable waters.
	No description for this resource	Preview © Downlo 1
	Marine -> Infrastructure ->	
	License Restrictions	
	Categories	
	Marine Man Made Features, Marine, C Locations	ables
	Mississippi, Texas, Florida, Alabama,	Louisiana, United States, North America
	Mississippi, Texas, Florida, Alabama, Additional Info	Louisiana, United States, North America
	Mississippi, Texas, Florida, Alabama, Additional Info Field	Louisiana, United States, North America
	Mississippi, Texas, Florida, Alabama, Additional Info Field State	Louisiana, United States, North America Value active
	Mississippi, Texas, Florida, Alabama, Additional Info Field State Ecological Region	Louisiana, United States, North America Value active Marine
	Mississippi, Texas, Florida, Alabama, Additional Info Field State Ecological Region Is NETL associated	Louisiana, United States, North America Value active
	Mississippi, Texas, Florida, Alabama, Additional Info Field State Ecological Region Is NETL associated Organization	Louisiana, United States, North America Value active Marine No National Oceanic and Atmospheric Administration
	Mississippi, Texas, Florida, Alabama, Additional Info Field State Ecological Region Is NETL associated Organization Organization Acronym	Louisiana, United States, North America Value active Marine No National Oceanic and Atmospheric Administration NOA







Sharing & Accessing Resources



Keep track of public or private resources you have *contributed, followed,* or added to your *library*













Private Collaboration

News Archive Refer a Friend Contact

Powe<u>r Our Futur</u>

Foster & nurture ongoing R&D with outside partners through secure collaboration

Tools & Apps

earch EDX now



https://edx.netl.doe.gov





If EDX does not have access to the data you are looking for, we work to point you in the right direction...

Groups EDXtools EDXportfolios

🚠 EDX 🍰 EDX (private) 🔮 NGDS 22287 🔮 Data.gov 129350 🗮 Other

powered by

National Estuarine Research Reserve System, background, search icon, background ... arrow, space

U.S. Total Crude Oil Proved Reserves, Reserves Changes, and ...

Proved Reserves as of 12/31. 19,121, 20,882, 23,287, 28,544, 30,529, 33,371, 1899-2013. Cha



Integration with in house and external tools, models and data directly within EDX





News Archive Refer a Friend Contact

Q 0 12-

Q

About FAQ

Sort by: Relevance +

Custom Search

NETL's Energy Data eXchange

Contribute

reserves

My EDX

OWebsites to Search

BOEM BOEM

BSEE BSEE eia EIA

R NETL NETL

🗑 👧 NASA

✓ ■USGS USGS

My Workspaces

About 38,800 results (0.52 seconds)

nerrs.noaa.gov/ReservesMap.aspx

Reserves During Year . South Slough, OR

nerrs.noaa.gov - NERRS Reserves Map

backgroud. Home > NERRS Reserves. NERRS Reserves

www.eia.gov/dnav/pet/pet_crd_pres_dcu_nus_a.htm

















EDX – Next Steps

Office of Research and Development





Continue to expand multi-organizational collaboration functionality

2015+ Emphasis on online analytics capabilities



- 2011 R&D, planning
- 2012 Version 1 release, data discoverability/accessibility
- 2013 Version 2 release, secure sharing for R&D

2014 – Growing and evolving, more advanced features and functionality, content (data, tools)

2015 and beyond – Version 3, emphasis on "**big data**", online **analytical capabilities, partnerships** with others





Energy R&D and Big Data





Collaborative Workspaces

My Favorite Collaborative Workspaces

Search...



This Collaborative Workspace contains administrative documents that support...



Geothermal - 4D **Monitoring Project**

Datasets



Collaborative Workspac



Q

LOBERGY
 Ithe ENERGY lab
 National Energy Technology Laboratory

https://edx.netl.doe.gov















EDX – Community

Office of Research and Development



EDX Version 3 Architecture

Office of Research and Development

SYSTEM ARCHITECTURE:

3 major components that comprise the enhancement of EDX architecture for version 3:

- 1. Common communication and service invocation components
- 2. Hadoop Ecosystem components
- 3. Collaboration components
- 4. Modeling and Simulation components

- **ESRI Libraries:** Clipping GIS data for target area(s) of interest from national datasets
- HDFS: Hadoop filestore that is replicated (similar to RAID) across the Hadoop cluster nodes. HDFS is self-healing
- **HBase:** Non-relational database that is distributed across Hadoop nodes to enable parallel data query by Hive. Enables users to create tables from data stored in HDFS.
- <u>Hive:</u> Uses SQL-like query language that generates MapReduce code based jobs to analyze data stored in HBase. Jobs run in batch and are not interactive
- Map/Reduce: Programming paradigm on how to approach/solve data problems. It was created by Google as an approach to index and get meaning from all of the data on the internet. Consists of two components: Map() and Reduce():
 - Map() function (typically written in Java) that executes on the data on each node to operate on the data where there data resides in the cluster. Generates output processed by the Reduce() function.
 - Reduce() function aggregates the output of the Map() function to output a combined list.
- **Spark:** In-memory data analysis that is a faster alternative to Hive for Hadoop clusters with high amounts of available RAM

DOE Crosscuts, Subter Example

Office of Research and Development

EDX Version 3 – Big Data, Scalability & Community

- Big Data related challenges will benefit from advanced computer science based solutions to the data gathering, fusion, management and utilization
- Combined with appropriate hardware, has the potential to offer EDX users functionality to support management and R&D analytical needs
- Addresses need for <u>scalable</u> architecture & storage going forward
- Leverage EDX "community"
 capabilities for DOE R&D needs

Evaluating Induced Seismicity with Geoscience Computing & Big Data –

A multi-variate examination of the cause(s) of increasing induced seismicity events

April 28, 2015

Kelly Rose, Principal Investigator National Energy Technology Laboratory Geology & Geospatial Team Lead 1450 Queen Ave. SW, Albany, OR 97321 541-967-5883

Scope of Work – Project Objectives

- **Develop tools and approaches** to manage multiple heterogeneous datasets
- **Develop a probabilistic approach** that can be utilized to assess potential for induced seismicity impacts through big data analyses
- Develop stochastic approaches to **reduce uncertainty** and constrain subsurface trends
- Improve joint analysis of multiple datasets, focus on advancing "Big Data" mining and integration techniques to improve knowledge and reduce uncertainty about subsurface systems through advanced geoscience computing

Seedling Phase Analysis: Spatio-Temporal Earthquake Distributions Vs. Fault and Surface Water Distributions

Scientific/Technical Approach – Big Data

Big Data Research journal

When utilized for subsurface analysis and exploration, spatio-temporal statistics can help characterize the uncertainty and support analysis of one or more data sets...

- Traditional statistical methods focus on drawing conclusions about a population from a limited set of data points
- Opportunity to use advanced statistical principles to downselect and extract information from large heterogeneous datasets at multiple scales – Big data for geosciences
- **Spatio-temporal statistical** models to incorporate both continuous and categorical measurements
 - Generate predictions across regions from both structured and unstructured data
 - Use information from this time or location, to say something about what will happen in the future, or over there!
- Uncertainty characterization, quantification, & communication
- Develop means of providing efficient data reduction and assimilation of large volumes of near-real-time monitoring data for rapidly updating predictions of subsurface behavior
- Develop hierarchical and dynamic inversion approaches
 - Consider the multi-scale nature of the subsurface & information about antecedent conditions and uncertainty quantification

... to effectively guide research, management and policy decisions and drive advance computation analyses to reduce subsurface risk

EDX as a Dashboard for R&D Resources

Office of Research and Development

N≝TL National Energy Tech	nology Laborator	v				News Archive	e ReferaFri	iend Contact
	The Ene	ergy Data eXchange		Search EDX now		Q Log in Join		
Home Search	My EDX	Contribute M	Ay Workspaces	Groups	EDXtools E	DXportfolios	About	FAQ
• Websites to Search		📥 EDX 🥥	NGDS 89	Data.gov 8	19 🗮 Other			
ВОЕМ ВОЕМ					_			
BSEE BSEE		oklahoma in	duced					Q
eia eia		About 2,800 res	sults (0.75 seconds	;)		Sort	by: Relevan	ce *
		Workshop o	n Hazard from I	powered nduced Seis	^{by}		Custom Se	earch
NETL NETL		Nov 14, 2014 hazard from ind	quake.usgs.gov/re Co-hosted by the (uced seismicity for	gional/ceus/wor Oklahoma Geo the United Sta	ksnop/ logical Survey and t tes.	he USGS mod	lels to descrit	be the
🛛 👧 NASA		USGS Relea	ase: 2011 Oklal newsroom/article.a	noma Induce sp?ID=3819	ed Earthquake N	<u>lay Have</u>		
USGS USGS		Mar 6, 2014 I that a human-ind	n a new study invol luced magnitude 5	ving researcher 0 earthquake r	rs at the U.S. Geolo near	gical Survey, scie	entists obsen	ved
		Induced Ea earthquake.usg	r <u>thquakes</u> s.gov/research/ inc	luced/				
		Jan 1 2015 V	as the M5 3 Trinid	ad CO Earthou	iake Natural or Indu	cod? in sever	al locations	

https://edx.netl.doe.gov

Multi-Organizational Team Supports EDX Philosophy

Office of Research and Development

This particular project illustrates the EDX "community" goal of making resources **discoverable and useable efficiently to support DOE R&D needs** to mutual benefit

<u> Task 1</u>

EDX – Feedback & Collaboration

"A model for web security."

"EDX is like FaceBook for nerds!"

"I've wanted a tool like this at NETL for years."

"EDX is a one-stop shop for energy related data and information."

"I was pleasantly surprised at how inviting and user friendly geoWELL is."

"I had a chance to look through your "Energy Data Exchange" site, and it looks fantastic!"

Collaborations and Alignment With:

- AARA Carbon Data (over 10,000 files)
- NRAP
- Futuregen
- DOE Subter Crosscut
- USGS NGGDPP, National Geological and Geophysical Data Preservation Program
- NSF Earthcube
- Data.gov
- NGDS
- Multiple universities, state & federal systems

Office of Research and Development

EDX aligns to Executive and DOE orders

Public Access Enable knowledge transfer, data reuse & discovery

<u>Secure/Private_Access</u> Support research development, collaboration, & online analytics

Built by researchers for research

Contact Information:

EDX Support - EDXsupport@netl.doe.gov

Kelly Rose - Kelly.rose@netl.doe.gov

