EDX Cloud Optimization for Carbon Management

Cloud Computing & AI R&D for Energy & Environmental Innovation
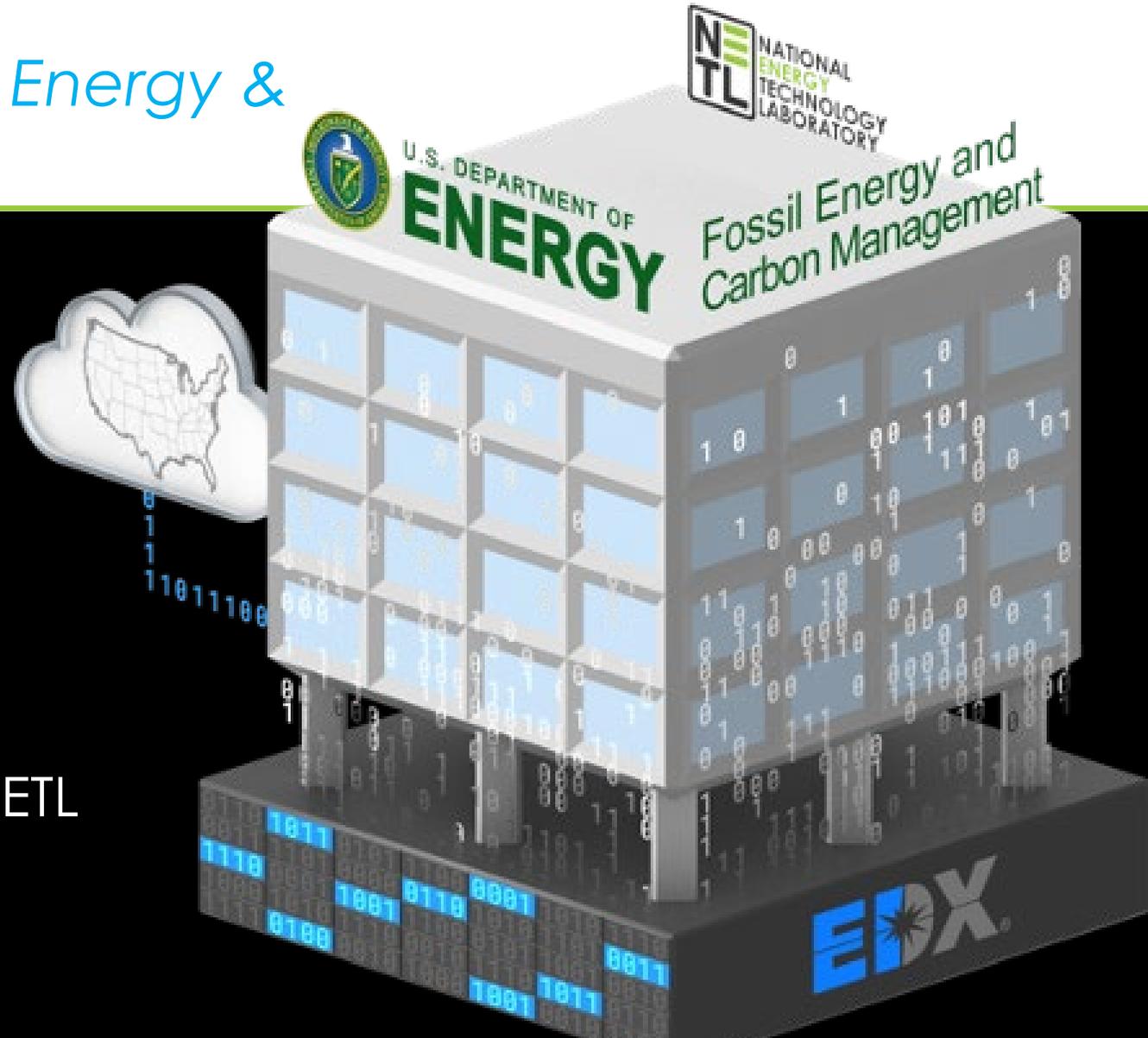
Kelly Rose, PhD

Technical Director, Science-based AI/ML Institute (SAMI), Research Innovation Center, NETL

& Vic Baker

Chief architect & cloud computational engineer, AVN @ NETL

8/2023

NATIONAL ENERGY TECHNOLOGY LABORATORY

U.S. DEPARTMENT OF ENERGY

Fossil Energy and Carbon Management

EDX

DOE's World-Class Experimental & Computational R&D Facilities

- **Operates some of the most capable computing systems and largest collection of advanced experimental facilities**

- **Frontier** is the world's fastest supercomputer & is HPC/exascale system

- Responsible for US nuclear security through deep partnerships across government

- Large producer of classified and unclassified scientific data in the world

- Strong foundation combining physical, biological, environmental, energy, mathematical and computing sciences

- Maintains world-class scientific workforce

- Strong ties with private sector technology and energy organizations and stakeholders
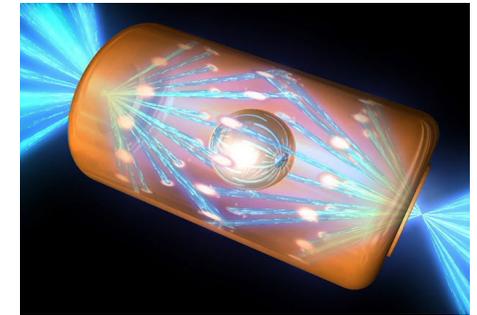
NETL's Advanced Computing R&D

DOE's Only GOGO National Laboratory

JOULE NETL SUPERCOMPUTER — **Center for Computational Science and Engineering**

WATT The Power of AI — **Center for Artificial Intelligence and Machine Learning**

SAMi SCIENCE-BASED AI/ML INSTITUTE — **Science-based Artificial Intelligence/Machine Learning Institute**

3.6 petaflops

13th fastest supercomputer within DOE National Laboratories and **174th fastest** in the world.

High-performance computational power optimized for energy R&D

Links **104 GPUs** with **16 petabytes** of storage.

On-prem AI/ML GPU-optimized R&D cluster

Established in 2020
Catalyzing AI- and ML-driven innovation
Combining established scientific methodologies with **AI/ML & advanced computational approaches**

Need: R&D is increasingly multi-organizational & data-driven, requiring data management & computational resources, in combination with advanced computing to drive innovation
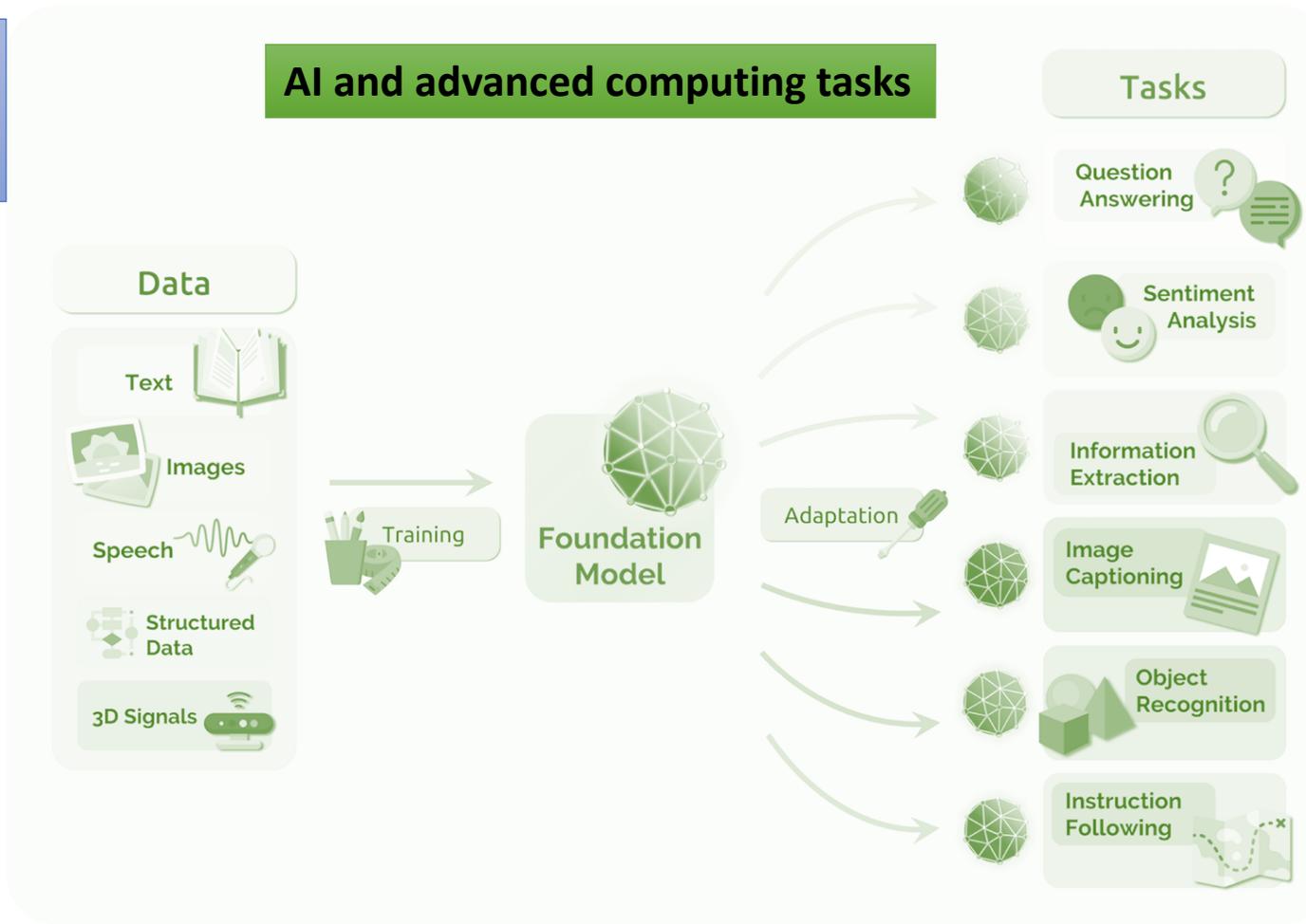
DOE & NETL are early in our Cloud enabled R&D journey

Scientific & Engineering Datasets

Biology
Geoscience
Materials
Chemistry
Particle Physics
Nuclear Physics
Computer Science
Climate
Medicine
Cosmology
Fusion Energy
Accelerators
Reactors
Energy Systems
Manufacturing

AI and advanced computing tasks

Data

Text
Images
Speech
Structured Data
3D Signals

Training

Foundation Model

Adaptation

Tasks

Question Answering
Sentiment Analysis
Information Extraction
Image Captioning
Object Recognition
Instruction Following

DOE Mission Tasks

Scientific Discovery
Digital Twins
Inverse Design
Code Optimization
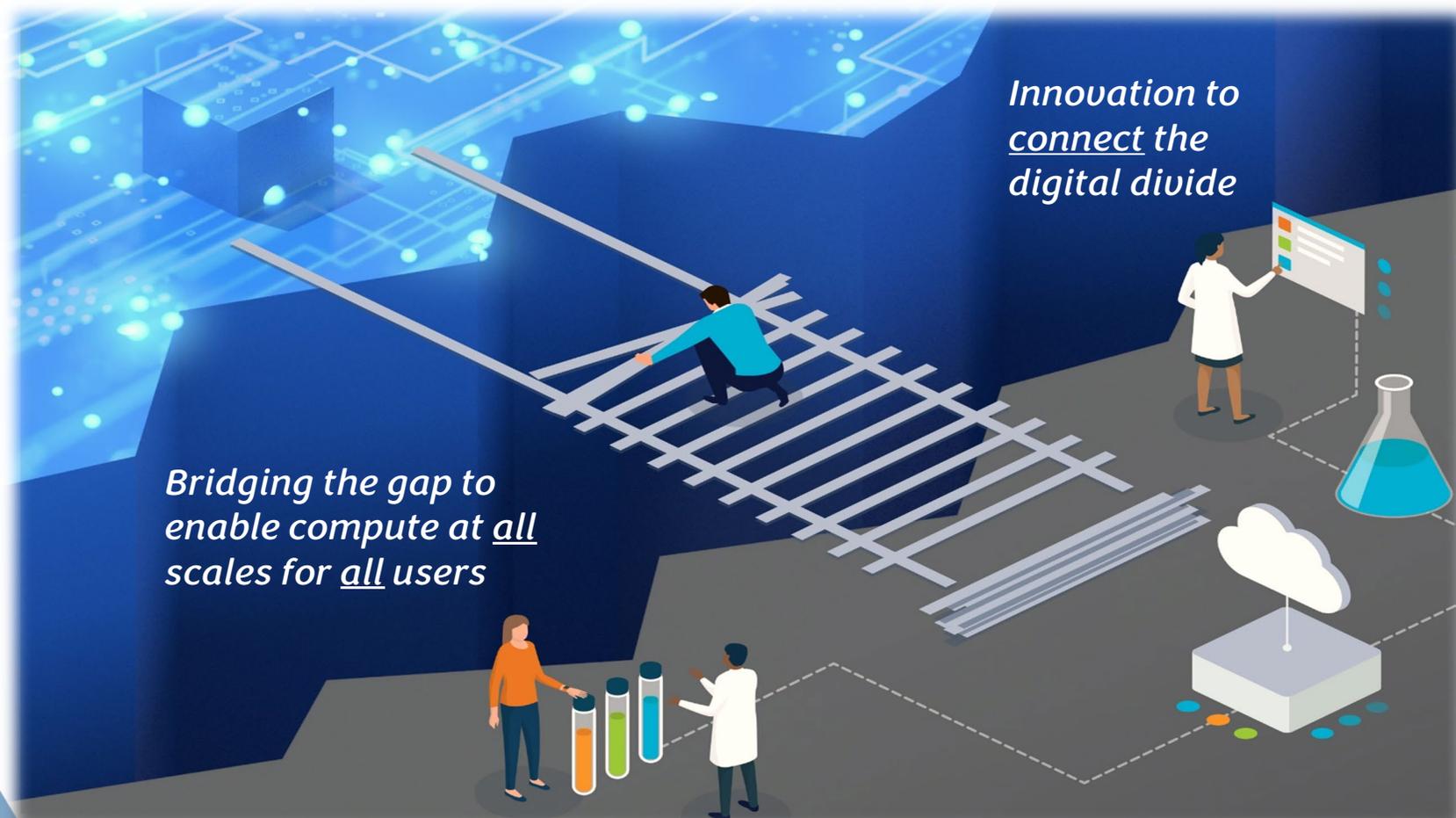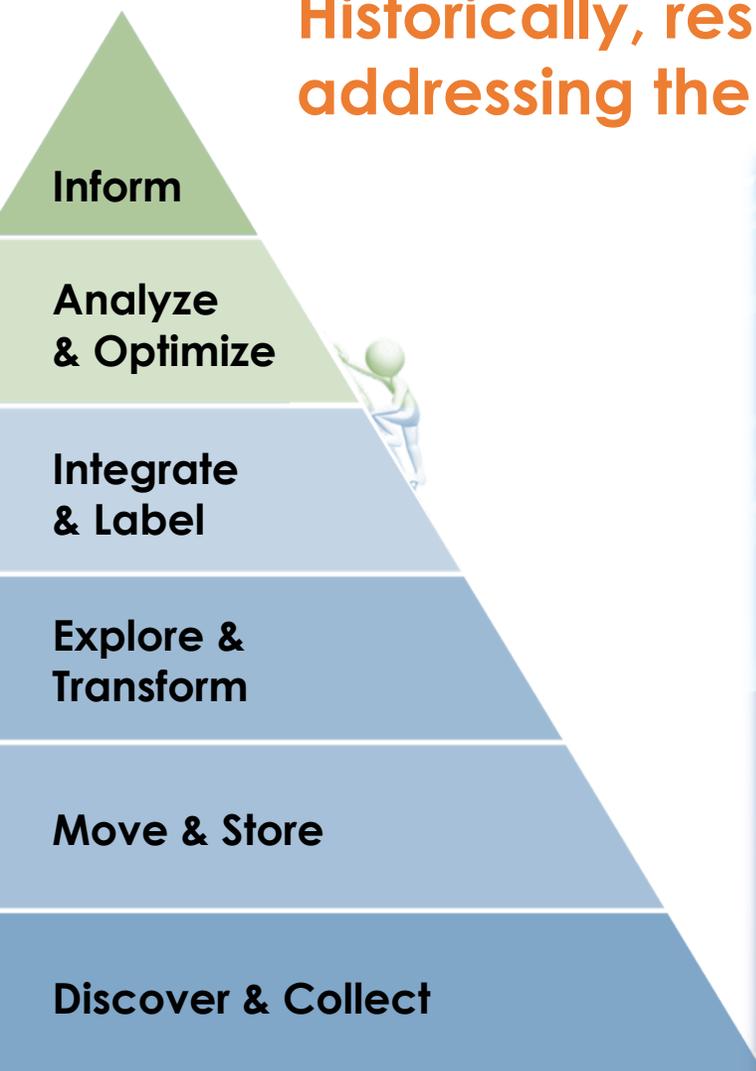Accelerated Simulations
Autonomous Experiments
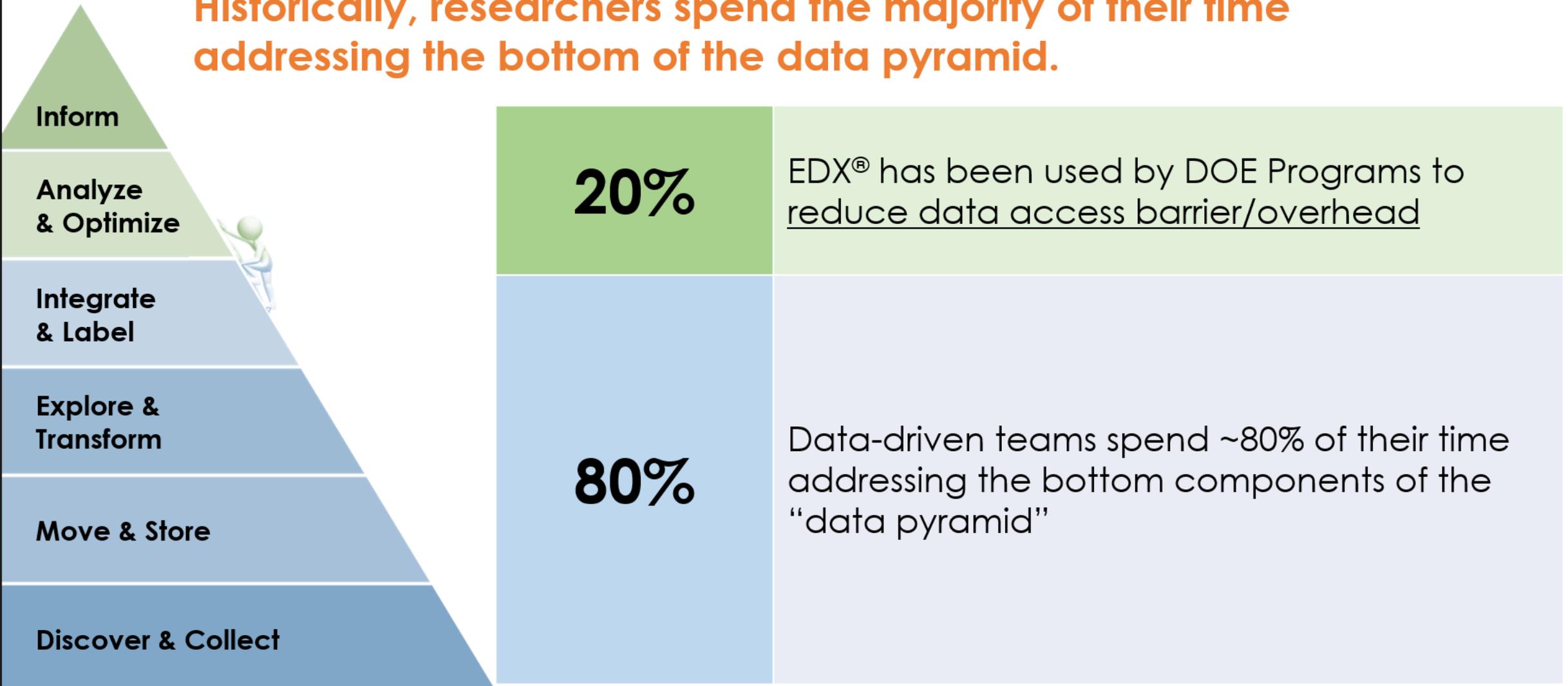Secure Data Infrastructure
Co-Design

Data Resources are the Energy for Innovation

Historically, researchers spend the majority of their time addressing the bottom of the data pyramid.

Inform

Analyze & Optimize

Integrate & Label

Explore & Transform

Move & Store

Discover & Collect

Innovation to <u>connect</u> the digital divide

Bridging the gap to enable compute at <u>all</u> scales for <u>all</u> users

Data Resources are the Energy for Innovation

NETL — NATIONAL ENERGY TECHNOLOGY LABORATORY

Historically, researchers spend the majority of their time addressing the bottom of the data pyramid.

Inform

Analyze & Optimize

Integrate & Label

Explore & Transform

Move & Store

Discover & Collect

20%	EDX® has been used by DOE Programs to <u>reduce data access barrier/overhead</u>
80%	Data-driven teams spend ~80% of their time addressing the bottom components of the "data pyramid"

How has NETL addressed R&D data needs?

2021 *U.S. DOE, Secretary of Energy's Achievement Award winner*

EDX
Energy Data eXchange

a web-hosted, virtual library and laboratory that supports the NETL/FECM community
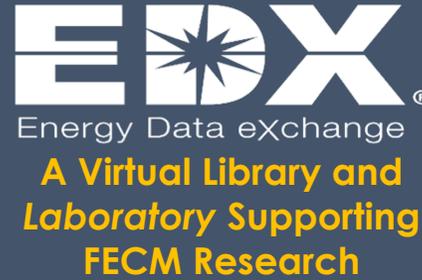
Energy Data eXchange

- NETL/FECM has supported EDX for over 13 years

- Core capabilities & governance available to all stakeholders

- Secure collaboration

- Publishing data products = visibility

EDX®
Energy Data eXchange

A Virtual Library and *Laboratory* Supporting FECM Research

Core Competencies of EDX

AI/ML	Big Data Compute	Data Management

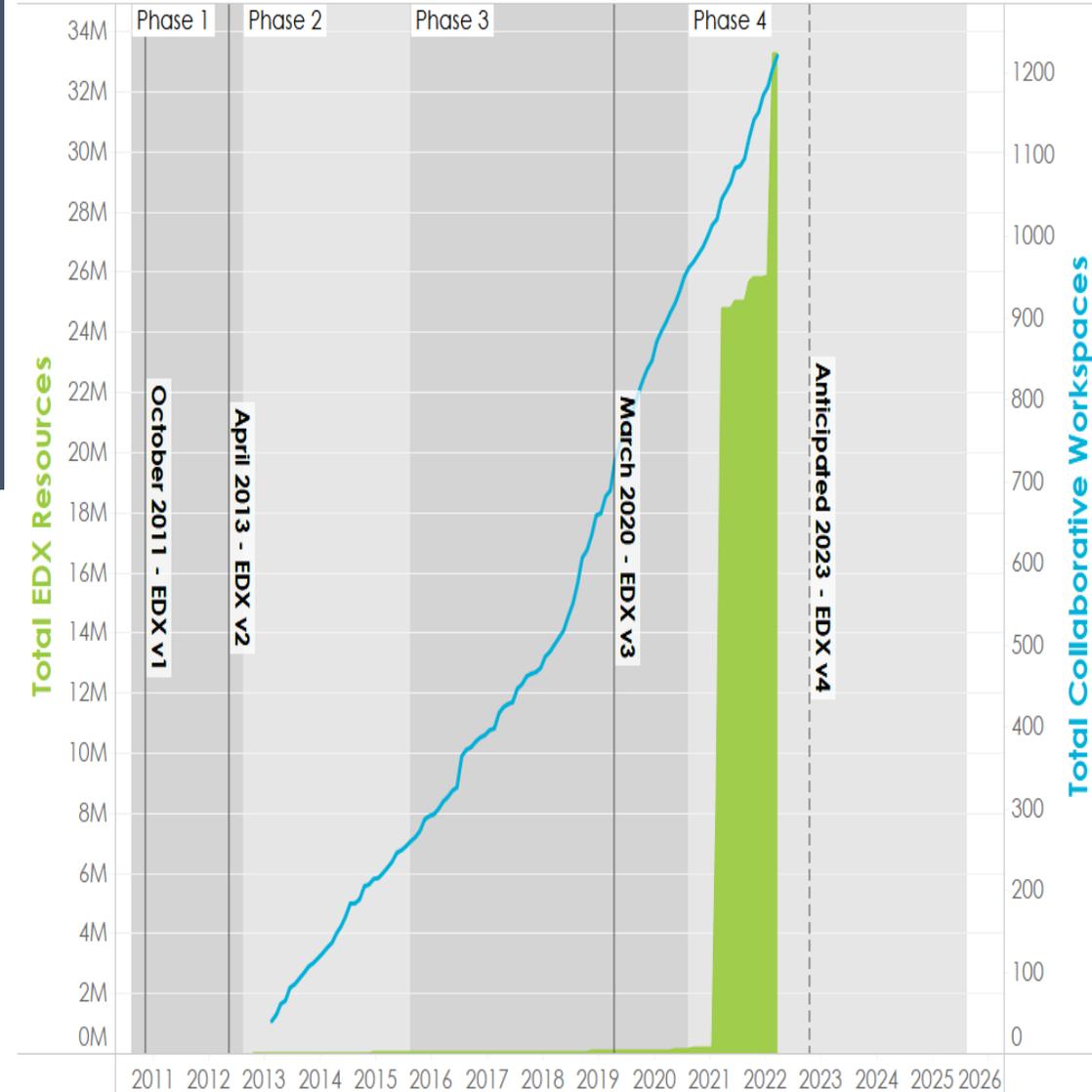| Multi-Organizational Private Collaboration	Public Dissemination of Data Products

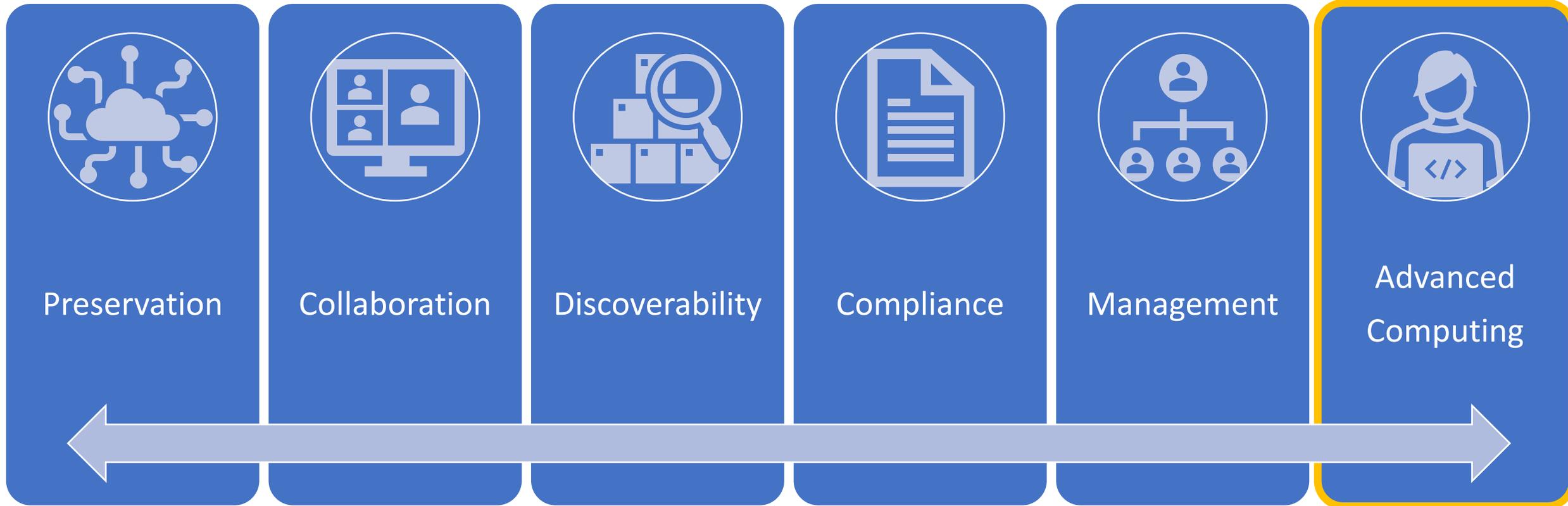**1200+** secure workspaces

33+ million resources

Collaboration among multi-organizational teams

File, link, tool, application or code

How is EDX currently supporting its community?

An "enterprise" system to enable secure, strategic R&D innovation

| Preservation | Collaboration | Discoverability | Compliance | Management | Advanced Computing |

NETL/FECM has invested in EDX to serve the FECM community as a virtual data library and laboratory

We have On-prem Compute so *Why Add the Cloud*?

- EDX & NETL/DOE's data-driven R&D mission and growth is aligned with Cloud <u>plus</u> on-prem compute
  - Enabling **compute where big data reside**
  - Supporting **multi-organizational, secure collaborations**
  - Rapidly deploy new services
  - Data and Services Redundancy
    - Disaster Recovery and High Availability
    - Enhanced Uptime
  - **Scalability** for services
  - **Storage** flexibility
- **Cloud Provider Managed Infrastructure**
  - Guaranteed infrastructure uptime Service Level Agreements
  - Compliance with federal orders/requirements while leveraging adaptability and resources of Cloud to **enable more agile resources for stakeholders**

EDX ®
Energy Data eXchange
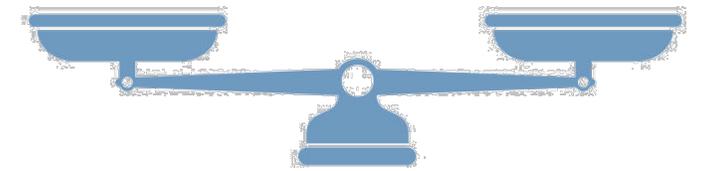
Which Cloud? Why more than one?
EDX++, a multi-cloud solution, connecting data with compute

Freedom for users to use current cloud service providers
- Utilize APIs to compute, transfer data
- Enabling advanced computing resources for AI and data science
- Connecting EDX users with on-premises HPC and advanced compute

Improves flexibility and performance
- Why multi-cloud?
  - Does not limit users to one storage & compute platform
  - Enabling the right compute solution for the need
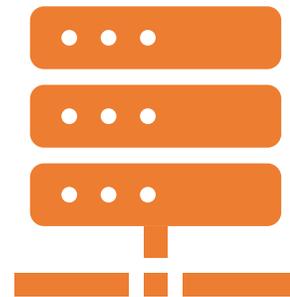- Compute occurs at the data source

Resilience
- Redundancy across multiple regions
- Strategic alignment for data transfer and compute across multiple cloud service providers
- GOGO, NETL partnership with DOE OCIO for enterprise Cloud deployment

Security & Alignment to Federal Reqs
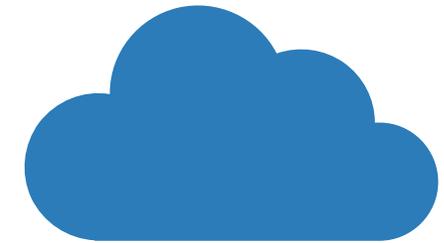
Key Lessons Learned... *insights from earlier adopters*

Why GCP First?

- Our team assisted with DOE HQ OCIO's ATO evaluation of GCP, 2019-2021

- Conducted extensive deep dives into AWS and Azure
  - *"Review & Recommendations of Cloud Platforms for FECM/NETL R&D – 2/9/2022"*

- Provides an ideal environment for cloud native, scalable innovation (R&D benefits)

- Google's Site Reliability Engineering (SRE) approach:
  - Use software as a tool to **manage** systems, **solve** problems, and **automate** operations tasks

- It is Cloud Native -- they create cloud technologies the world uses

- Offers SA, **subscription agreements**, which has cost advantages for gov
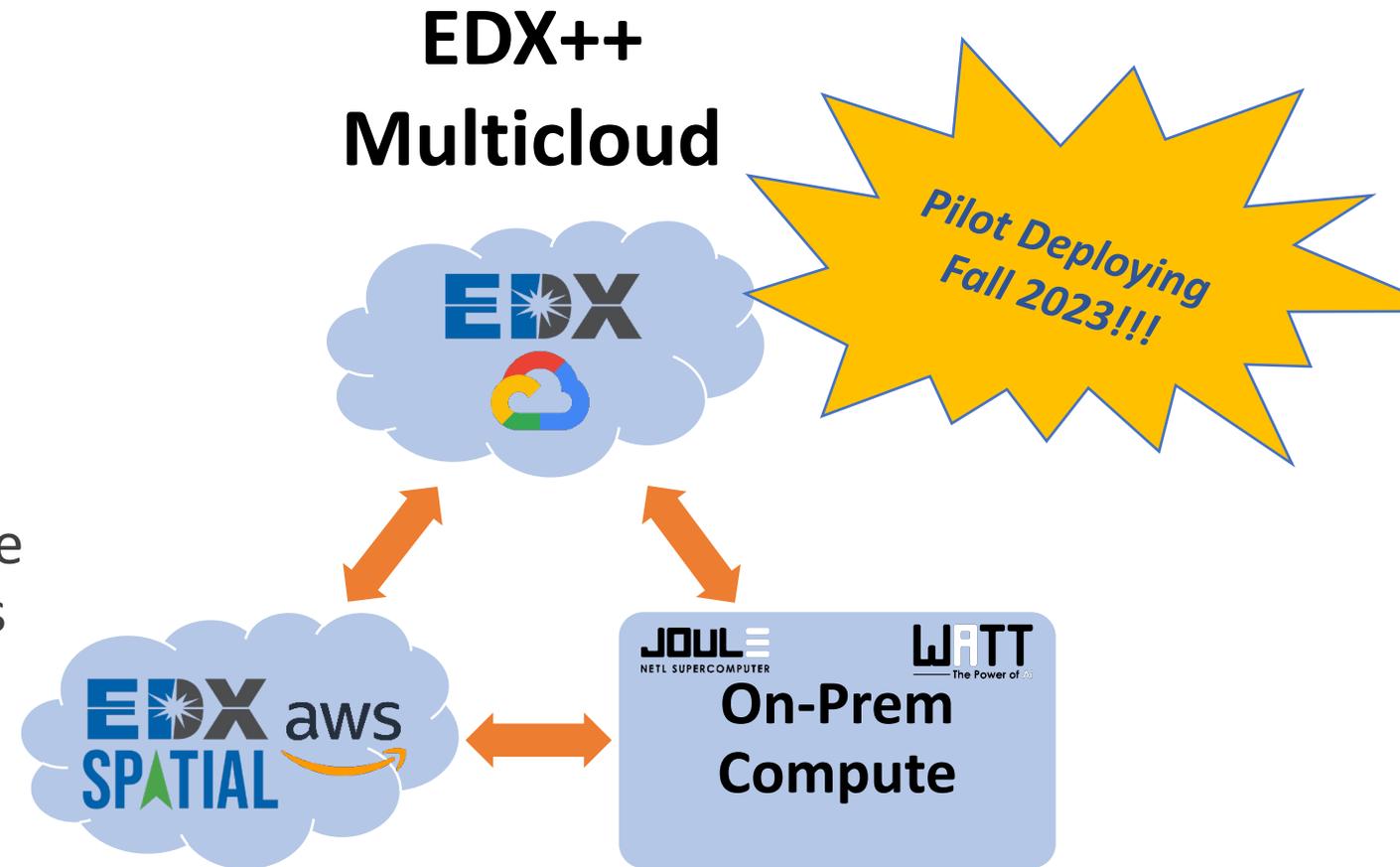
Infrastructure with Cloud

- Latest CPUs, GPUs available as soon as CSP can deploy

- Need a different CPU, more ram?
  - Modifications are immediately available to "right size" your compute
  - **Cut out months of bids, order delays, installation delays, setup delays**

- Turn compute machines on when you need them, off when you don't

- **This is development freedom** - one of the best features of cloud
  - **Spin up what you want**
  - **When you want**
  - **Where you want**
  - **Automatically scale based on load**

9/15/2023   https://edx.netl.doe.gov/sami/

Enabling Capabilities & our R&D Users

- Architecting multi cloud environment to **support research mission**

- Working to a**ccelerate access to commercial and open-source** cloud resources for our community
  - Including training & collaborative opportunities

- **SAMI** is handling administrative and work force engagements to accelerate access responsible to these resources

- **SAMI is sponsoring key trainings** to use AI, data, and cloud compute resources responsibly

EDX++ Multicloud

Pilot Deploying Fall 2023!!!

On-Prem Compute

https://edx.netl.doe.gov

Training, R&D, AI & *Humans in the Loop*

https://azurecharts.com/overview

sami@netl.doe.gov

https://wa.aws.amazon.com/wat.map.en.html

https://cloud.google.com/blog/topics/developers-practitioners/back-popular-demand-google-cloud-products-4-words-or-less-2022-edition

In EDX DOE & federal compliance is built-in

EDX++ is building additional compliance and useability in for our community

EDX helps teams with **compliancy (Cyber & Executive),**

often without users even knowing it

& accelerates collaboration for ongoing R&D

A comprehensive list of orders and initiatives can be found on the EDX Reference Shelf

EDX4CCS & EDX++, FECM Investment in Carbon Storage Digital Resources

Need to advance for <u>democratized</u> use & useability

- **PB of carbon storage data preserved** using the **public and private** sides of **EDX**
- Curating access to **downloadable instances** of NRAP and other CS Program **models & tools**
- **Developed custom, AI/ML/NLP enhanced tools** to drive CCS data discovery and knowledge extraction

  WATT — The Power of AI

  NETL SmartSearch
- **Enabling** geospatial data and CS web mapping for select CCS resources to date

Increasing # of features per 100 km²

Morkner, P., et al. **2021,** Distilling Data to Drive Carbon Storage Insights, *Computers & Geoscience*

Data is analyzed for spatial density and distribution
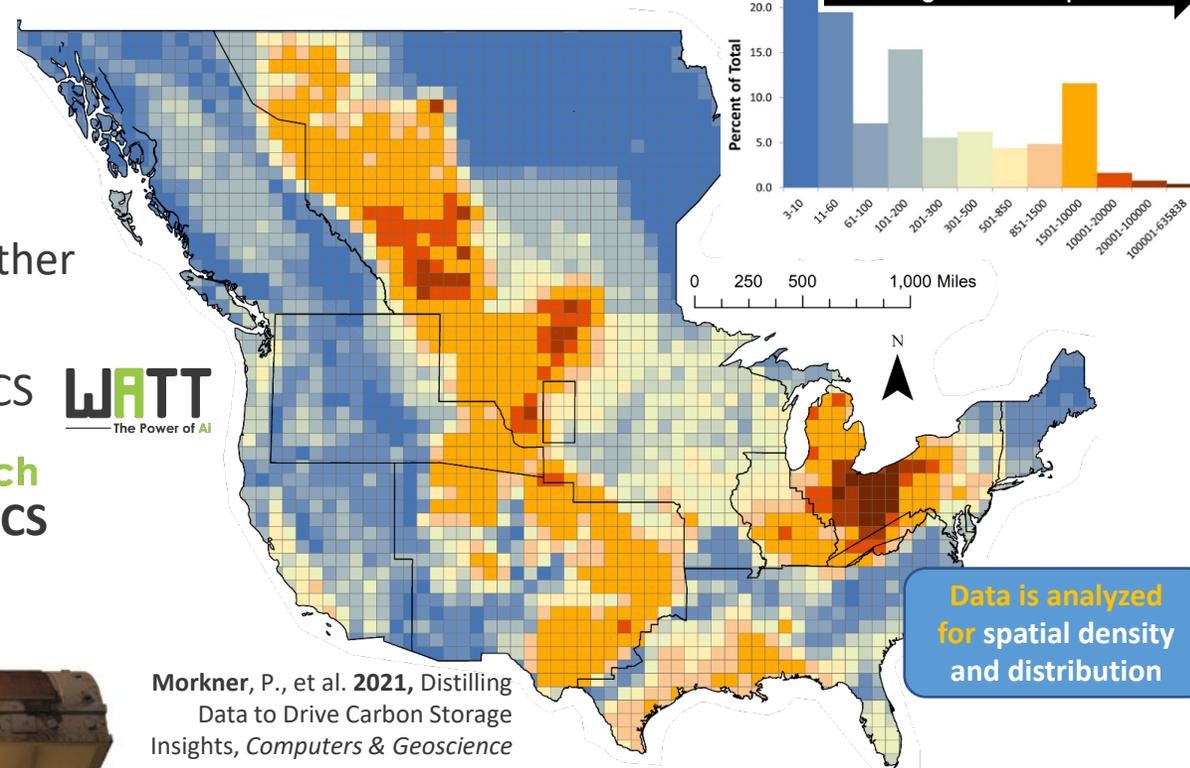
Current limitation of these data resources & capabilities…
… still largely require the end-user to have access to the right expertise and computational resources to put them to use….

Democratizing Access to Digital Resources for the CCS Community

EDX4CCS - Putting the Power of Advanced Data Viz & Analysis in the Hands of Many

- The world is undergoing a digital revolution
- Previously, access to advanced data and models was limited to the domain of scientists/engineers with advanced computing and infrastructure resources on-premises
- But the last 10 years has seen a revolution, **putting the power of data-driven analysis and decision making in the hands of many**

This FWP is about **unlocking that data-driven capability** for the CCS community

- Elevating access to authoritative, robust CCS data, models, and online/virtual computing and visualization
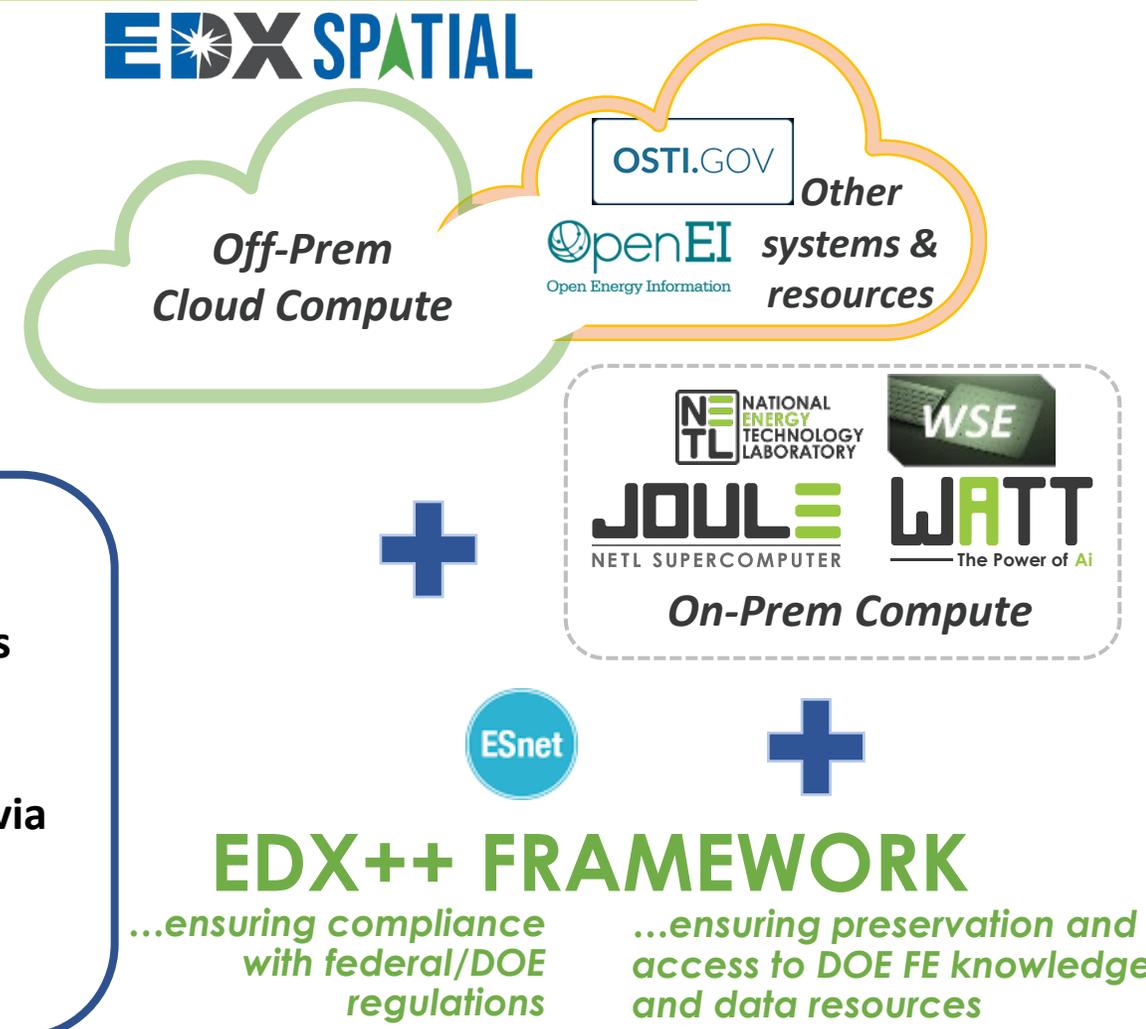- Tailoring digital resources from & for CCS users

Already Benefitting DOE FECM & NETL's Mission

Using AI/ML, millions of data features and attributes have been integrated and preserved across the USA in support of advanced carbon storage projects

This effort has already aided SMART, NRAP and outside entities (e.g., major industry operator) to drive subsurface modeling, machine learning, and insights for a range of end user needs

EDX supports:
- Carbon storage researcher & program data ingestion
- Data mining to aggregate authoritative, open-source resources relevant to CS researchers
- Integration of other FE resources
- Access, visualization, and interaction with CS data collections via NETL EDX mapping platforms Natcarb Viewer and EDX Spatial
- Reuse of data by new FE projects via EDX Collaborative Workspaces and more…

EDX SPATIAL

OSTI.GOV

OpenEI
Open Energy Information

Off-Prem Cloud Compute

Other systems & resources

NATIONAL ENERGY TECHNOLOGY LABORATORY

WSE

JOULE
NETL SUPERCOMPUTER

WATT
The Power of Ai

On-Prem Compute

ESnet

EDX++ FRAMEWORK

…ensuring compliance with federal/DOE regulations

…ensuring preservation and access to DOE FE knowledge and data resources

EDX.4CCS

5 Years, Focus on Delivering…

Putting CCS data resources to work – Democratizing Stakeholders/Community access and use of CCS data, tools, & visualization thru virtual infrastructure capability

- **CCS System** – data curation and virtualization platform, includes both public and private data sharing capabilities, entire life-cycle of data, data discovery, transformation and integration.
- **Using the EDX++ Multi-cloud deployment** to enhance access by the CCS stakeholder community to virtualized CCS data and models to accelerate CCS commercial, regulatory, social and environmental goals.

- **CCS Data Resources**
  - Refining, integrating and generating priority data resources to feed carbon storage commercial, R&D (SMART, NRAP, etc), and regulatory models
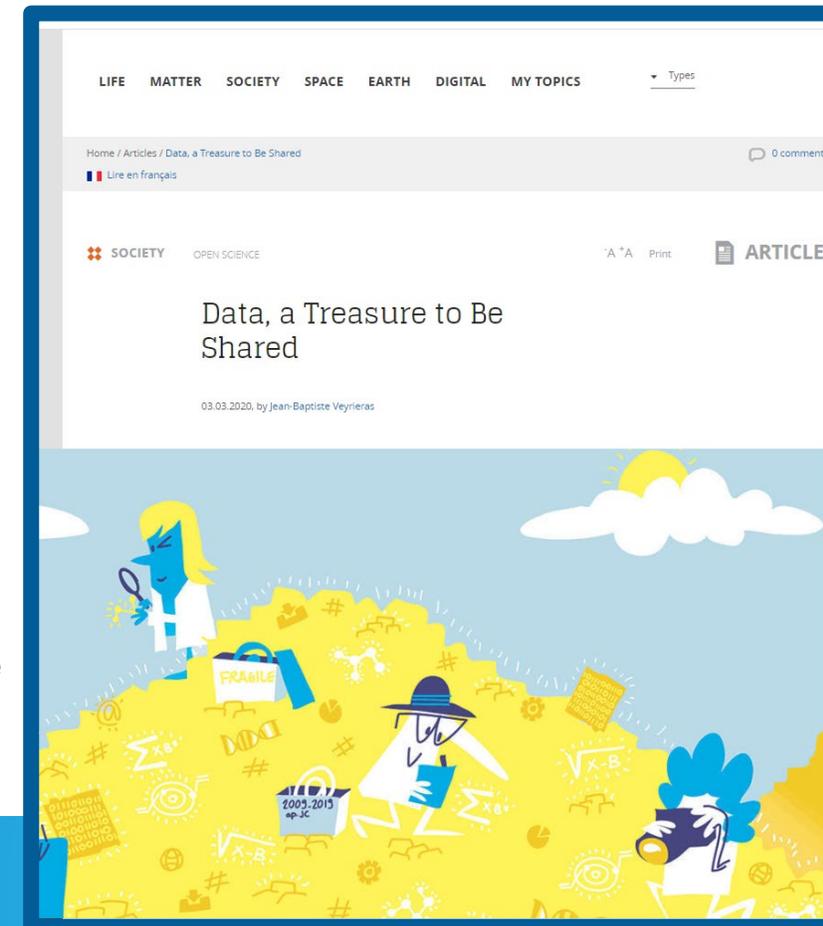- **CCS Tools**
  - Development and deployment of strategic Carbon Capture and Storage (CCS) data visualization and analytical tools to accelerate CCS commercial and regulatory efforts
- **Core CCS EDX disCO$_2$ver platform**
  - Connecting stakeholders across commercial, regulatory, and research domains to a curated collection of CCS data, models, and capabilities (visualization, analysis, etc.)
  - Building off the EDX-FECM Integrated Development Environment core but implementing and expediting deployment of key capabilities and resources for the CS community.
- **Outreach & Capacity Building**
  - Outreach efforts, training, and capacity building within communities across the US.
  - Connecting stakeholders with actionable, digital infrastructure and resources to accelerate the CCS mission

LIFE   MATTER   SOCIETY   SPACE   EARTH   DIGITAL   MY TOPICS   ▼ Types

Home / Articles / Data, a Treasure to Be Shared   💬 0 comment

🇫🇷 Lire en français

⊞ SOCIETY   OPEN SCIENCE   ⁻A ⁺A Print   📄 ARTICLE

Data, a Treasure to Be Shared

03.03.2020, by Jean-Baptiste Veyrieras

EDX4CCS, Actionable Data Science (Public Products)

RokBase

Virtual portal and tool for users to access, explore and query available core and rock property data collected from carbon storage spots across the U.S.

*Web tool available via EDX*

200 Page Views, 5 Downloads
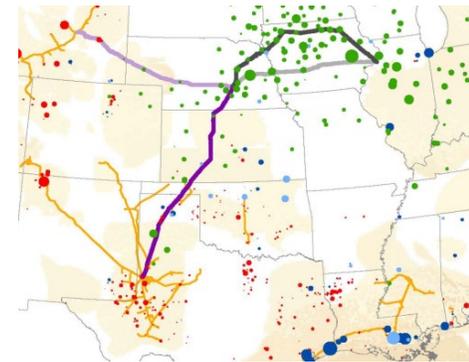
CCS Pipeline Route Planning Database

- Containing >90 gigabytes of data representing critical considerations for the spatial routing of pipelines and transport of CO_2

293 Downloads, Published 3/31/23, NETL News Story

CO$_2$-Locate Database

National Well Database & Class II Well Reuse and Regional Evaluation Tool built to support site selection, risk analysis and other stakeholder needs

263 Downloads, Published 3/31/23, NETL Story

Talk today @ 2:00pm
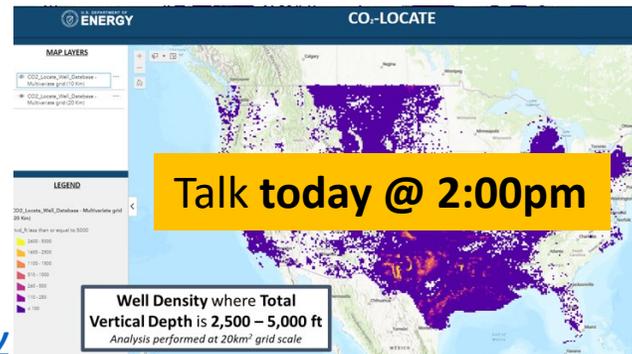
CCS EJ-SJ Database

Social and environmental justice database aligned to carbon storage systems to aid in decision making around key future infrastructure placement & more

302 Web App Views, 62 Downloads, Published 3/31/23, Media Plan in Dev.
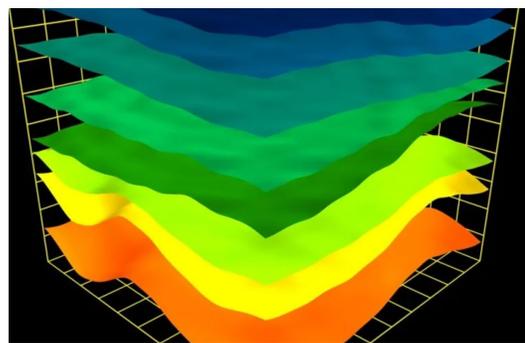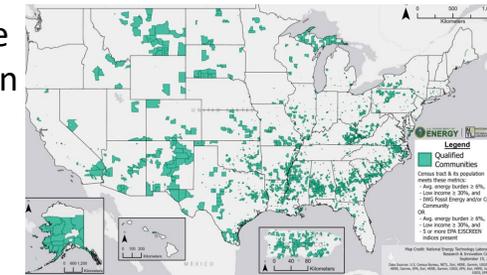
Class VI Data Support Tool

Geodatabase of spatial data layers to support the initial development of EPA Underground Injection Control (UIC) Class VI permits

181 Downloads, Published 3/28/23 NETL News Story
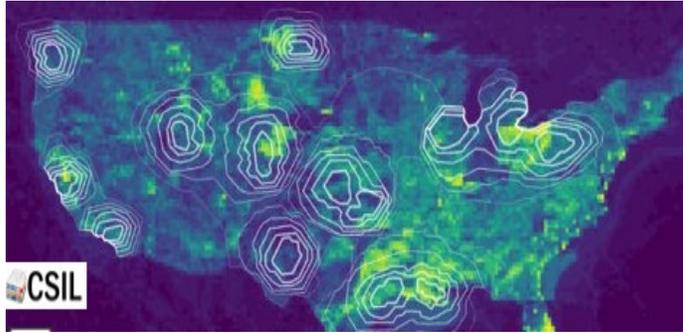
Talk today @ 1:10pm

CS Technical Viability Database V1

Geodatabase of geologic, geophysical, structural, hydrologic, energy extraction, transportation infrastructure, local political boundaries, community and environmental data

Recently Published on 6/13/23, 3 downloads

EDX4CCS, Actionable Data Science

Carbon Storage Viability Matrix V 0.1

- Criteria for systematic evaluation of CS technical viability data
- Includes subsurface, EJ/SJ, and physiographic factors
- Conveys CS viability in a region of interest based on available data and informs storage resource est.

Coming soon - EDX Release in November 2023

Carbon Storage Framework Dashboard
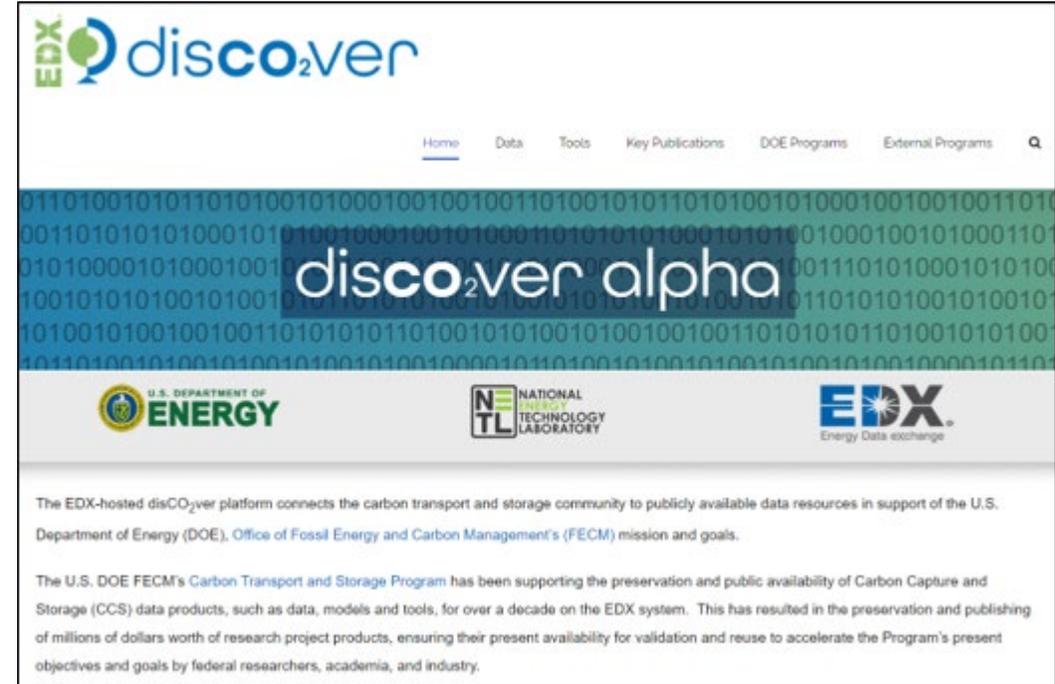

Talk **Tuesday @ 4:10pm**
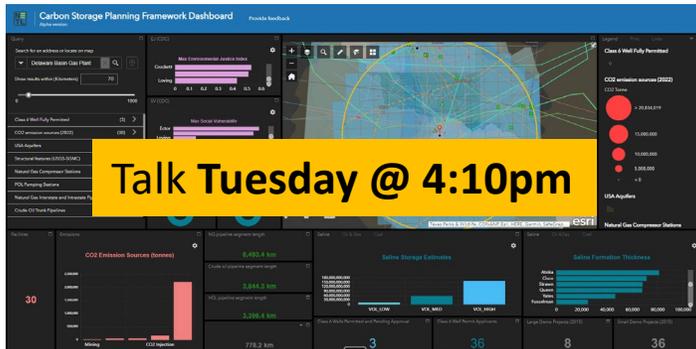
Provides quick and efficient access and insights into

- CS datasets (including geologic, technical, environmental, EJ/SJ)
- To support CCS planning, feasibility, and resource assessment efforts

Planned for EDX Release on summer 2023

EDX disCO$_2$ver - Connecting Users to CCS Digital Resources

The EDX-hosted disCO$_2$ver platform connects the carbon transport and storage community to publicly available data resources in support of the U.S. Department of Energy (DOE), Office of Fossil Energy and Carbon Management's (FECM) mission and goals.

The U.S. DOE FECM's Carbon Transport and Storage Program has been supporting the preservation and public availability of Carbon Capture and Storage (CCS) data products, such as data, models and tools, for over a decade on the EDX system. This has resulted in the preservation and publishing of millions of dollars worth of research project products, ensuring their present availability for validation and reuse to accelerate the Program's present objectives and goals by federal researchers, academia, and industry.

https://edx.netl.doe.gov/disco2ver
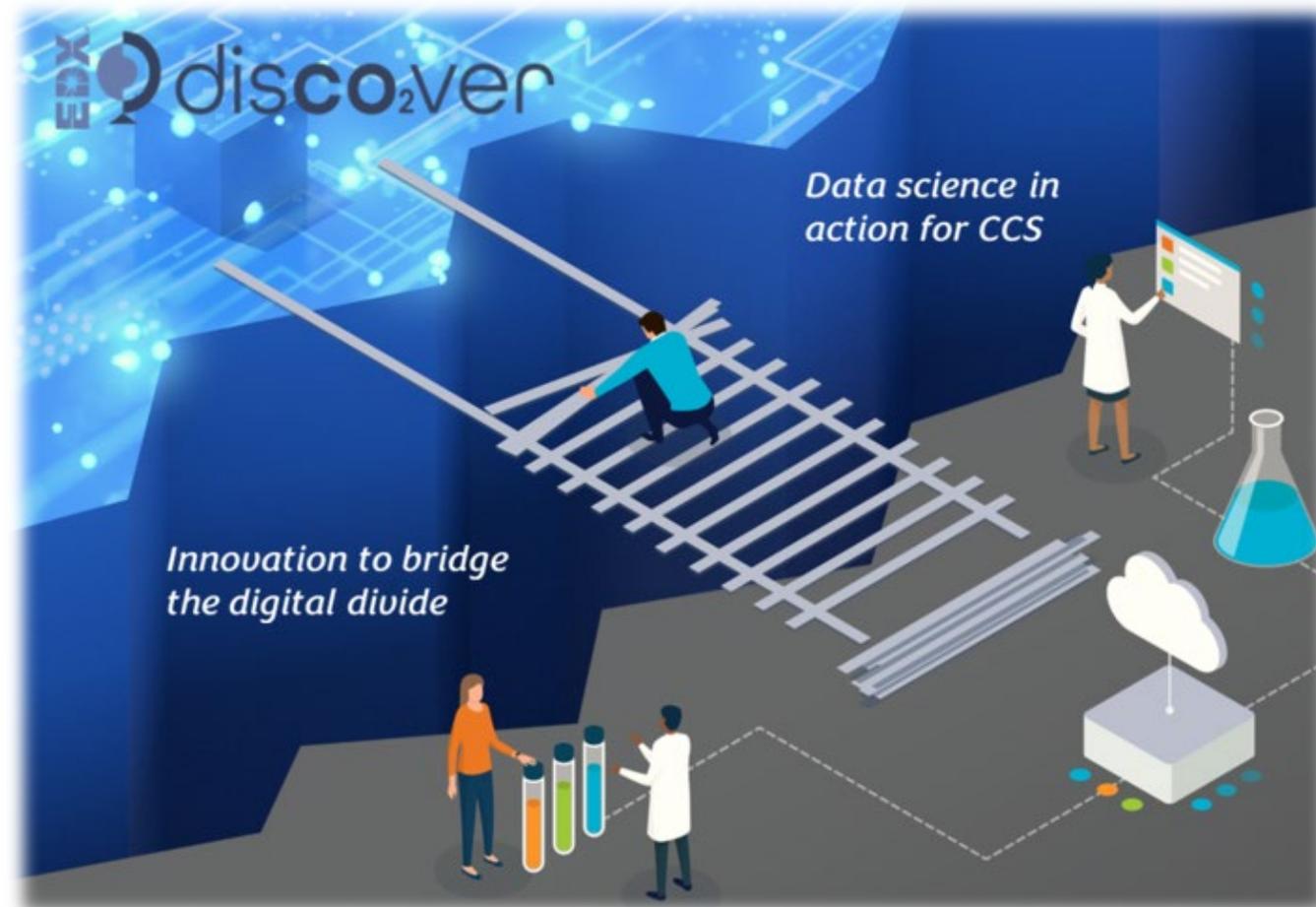
Tuesday Evening - Live Tool Demos!

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

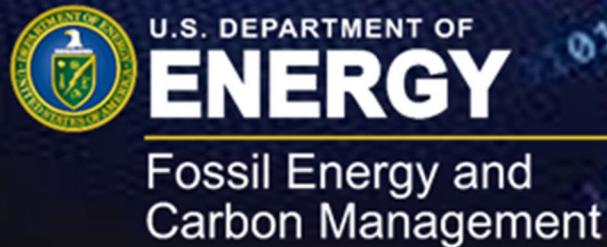**Where:** The Ballroom Foyer and East/West Atriums

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 disCO$_2$ver,** 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 disCO$_2$ver

Data science in action for CCS

Innovation to bridge the digital divide

EDX Energy Data eXchange

EDX Cloud Optimization for Carbon Management

Cloud Computing & AI R&D for Energy & Environmental Innovation

NETL NATIONAL ENERGY TECHNOLOGY LABORATORY

EDX.4CCS Energy Data eXchange

U.S. DEPARTMENT OF **ENERGY**
Fossil Energy and Carbon Management

Thank you!!!

Kelly Rose, kelly.rose@netl.doe.gov

Technical Director, Science-based AI/ML Institute (SAMI), NETL

& Vic Baker

EDX chief architect & cloud computational engineer, AVN @ NETL

Transition takes a team:

NETL **maximus** **AVN**

Office of THE CHIEF INFORMATION OFFICER

U.S. DEPARTMENT OF **ENERGY** Fossil Energy and Carbon Management