The DisCO2ver Platform, a Virtual Carbon Storage Data Laboratory, and Infrastructure

Unlocking data-driven capabilities for the entire CCS community

Kelly Rose[1]
Paige Morkner[1,2]
Chad Rowan[1,2]

Carbon Management Project Review Meeting
Aug. 17, 2022

[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*

U.S. DEPARTMENT OF ENERGY

EDX Energy Data eXchange

NETL NATIONAL ENERGY TECHNOLOGY LABORATORY

Disclaimer

This project was funded by the United States Department of Energy, National Energy Technology Laboratory, in part, through a site support contract. Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

A history of DOE Program Product Curation

DOE Carbon Storage

NATIONAL ENERGY TECHNOLOGY LABORATORY

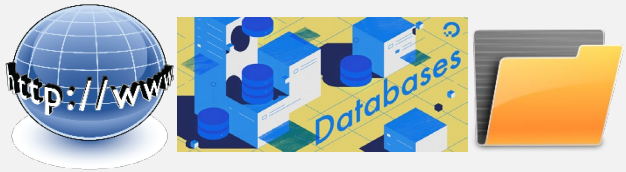
2015 **2016** **2017** **2018** **2019** **2020** **2021** **2022**

- 2015 DOE Program managers **initiated requirement** to contribute & curate CS data products from CS program in EDX
- Included requirements in FOAs
- Funded NETL RIC EDX team to provide customized support to the CS-affiliated performers, as well as the development of data science-informed optimizations in EDX for the CS Program

NETL SmartSearch automates data discovery using AI/ML by …

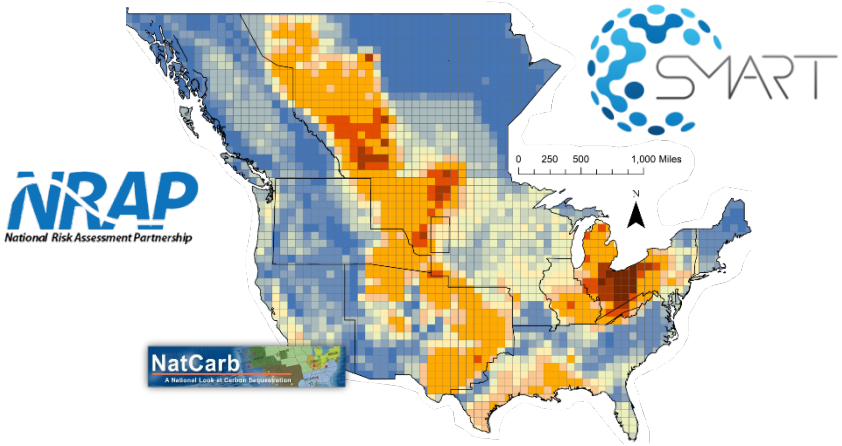
1) **Analyzing** content you like

2) **Finding** new, targeted content



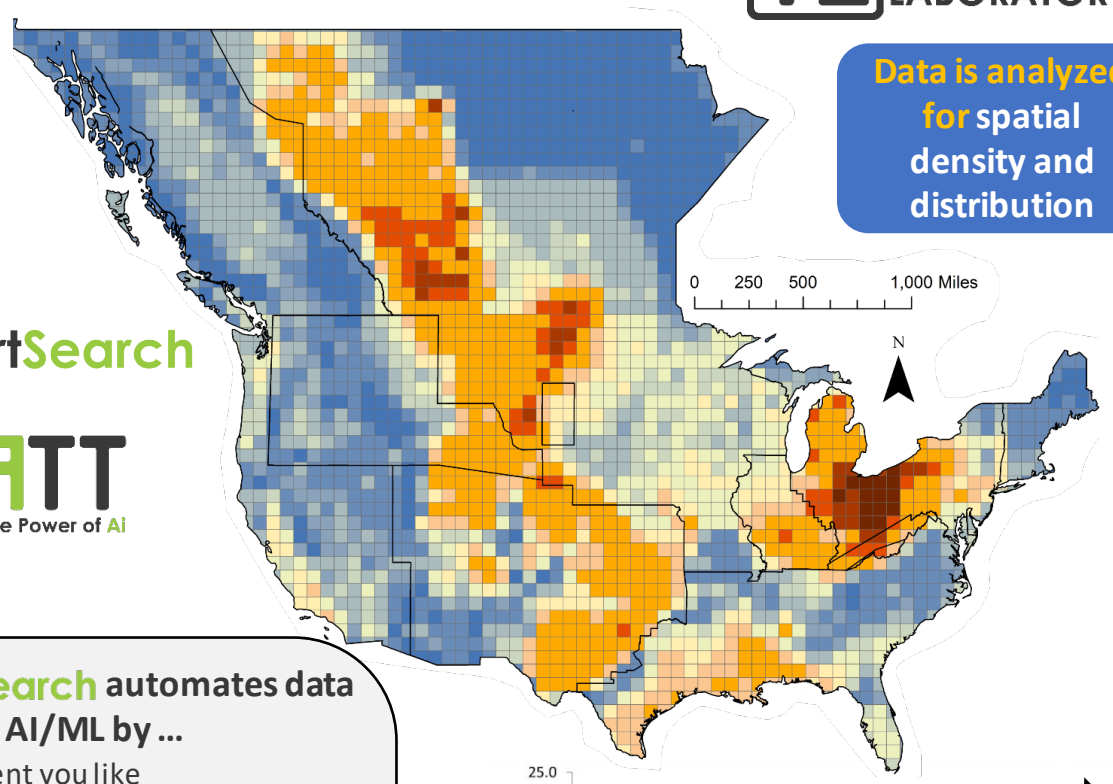
- Expanded to CarbonSafe and additional extramural programs
- EDX used by each CS Program extramural team (**51 individual Private Workspaces**)
- NETL geo-data science and EDX teams launched the development of AI and data-science enhanced efforts to acquire, tag, organize, refine, and improve virtualization of CS-program products as they were accumulating

- 2022 ASFECM endorses FOA/FWP EDX requirement
- 2020 CS Program had a **foundation** of data resources from both DOE-funded projects and extramural AI-discovered assets
- NETL deployed **Open-carbon storage data collection** via EDX as a refined, integrated database (included Program & AI-discovered assets), with **hundreds of thousands of attributes, from thousands of data resources**
- In use accelerating **DOE's SMART-CS initiative, regulatory, and commercial (SBIR)** efforts too



Work to Date Recap – Creating an FECM Carbon Storage Data Ecosystem

- **PB of carbon storage data preserved** using the **public and private** sides of **EDX**
  - **TB are public**
  - **971 RCSP resources are public (data, models,+)**
  - **Over 600,000 spatial data attributes**
  - **Includes 120+ TB of seismic data** from MGSC and MRCSP accessible via EDX-Watt as well
- **AI/ML/NLP enhanced tools to drive data discovery and knowledge extraction**
  - Beta version of **SmartSearch©** completed
- **EDX GeoCube v4.0, Geospatial data, and CS web mapping tool**
  - **User interface built** for **CO2-SCREEN** and the **Offshore CO2 Saline Calculator** tools
  - **Links** to access **15** different **CCS tools**



Data is analyzed for spatial density and distribution

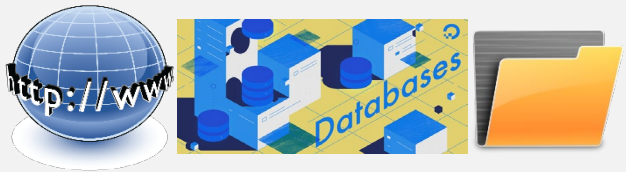
NETL SmartSearch

WATT The Power of AI

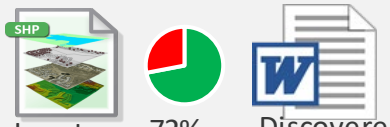
NETL SmartSearch automates data discovery using AI/ML by …

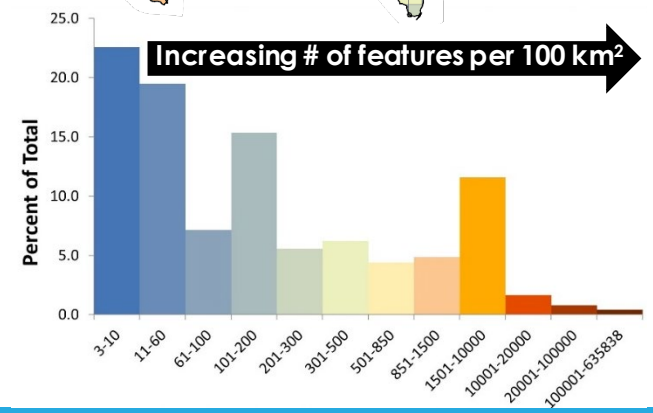
1) **Analyzing** content you like

2) **Finding** new, targeted content



3) Understand **how relevant** the discovered data are to your targeted content



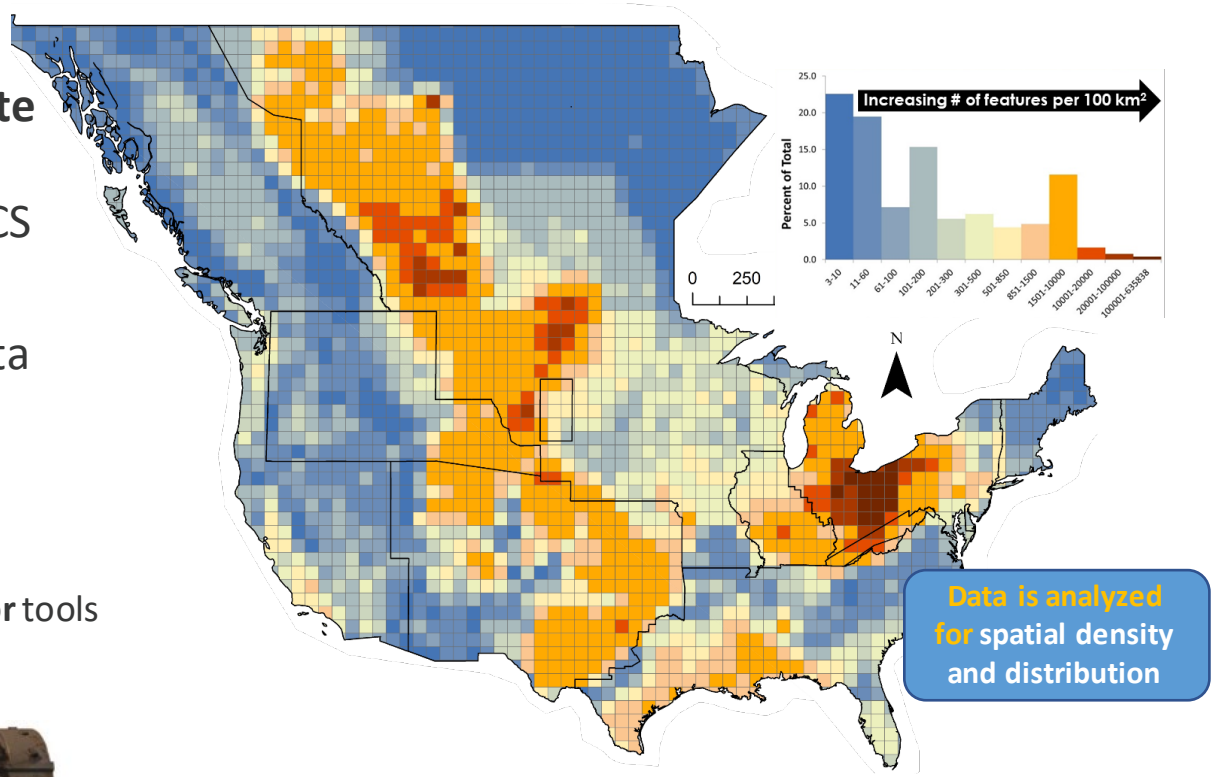


Morkner, P., et al. **2021,** Distilling Data to Drive Carbon Storage Insights, journal: *Computers & Geoscience*

FECM has Invested in Creating a Carbon Storage Digital Resource-base

Need to advance for democratized Use

- **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
- **Enabling geospatial data and CS web mapping for select CCS resources to date**
  - **User interface built** for **CO2-SCREEN** and the **Offshore CO2 Saline Calculator** tools
  - **Links** to access **15** different **CCS tools**



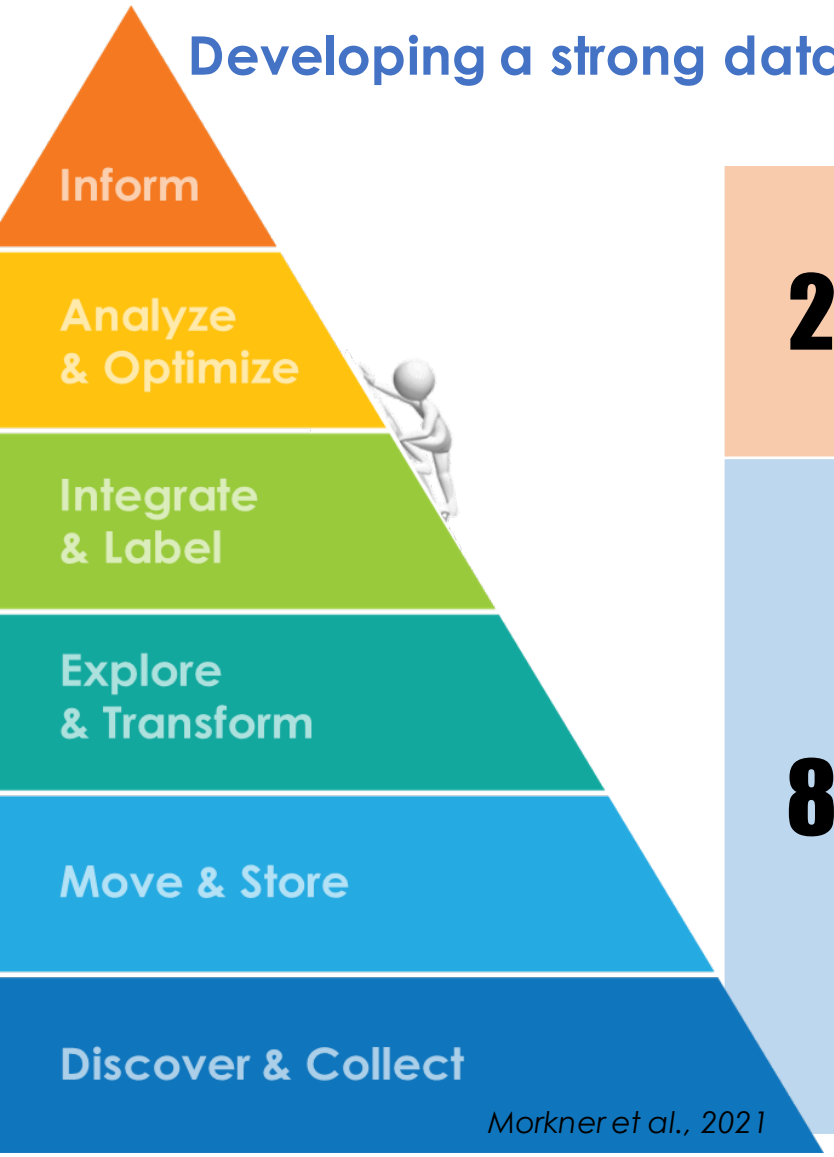
Current limitation of these data, tools, and capabilities…
…they ***still largely require the end-user to have access to the right expertise and computational resources to put them to use….***

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

Data are the *Energy* for Analysis & Inquiry

NATIONAL ENERGY TECHNOLOGY LABORATORY

Developing a strong data foundation is key to any program/project's success



Morkner et al., 2021

20%

80%

→

EDX has been used by other DOE Programs to reduce this data access barrier/overhead

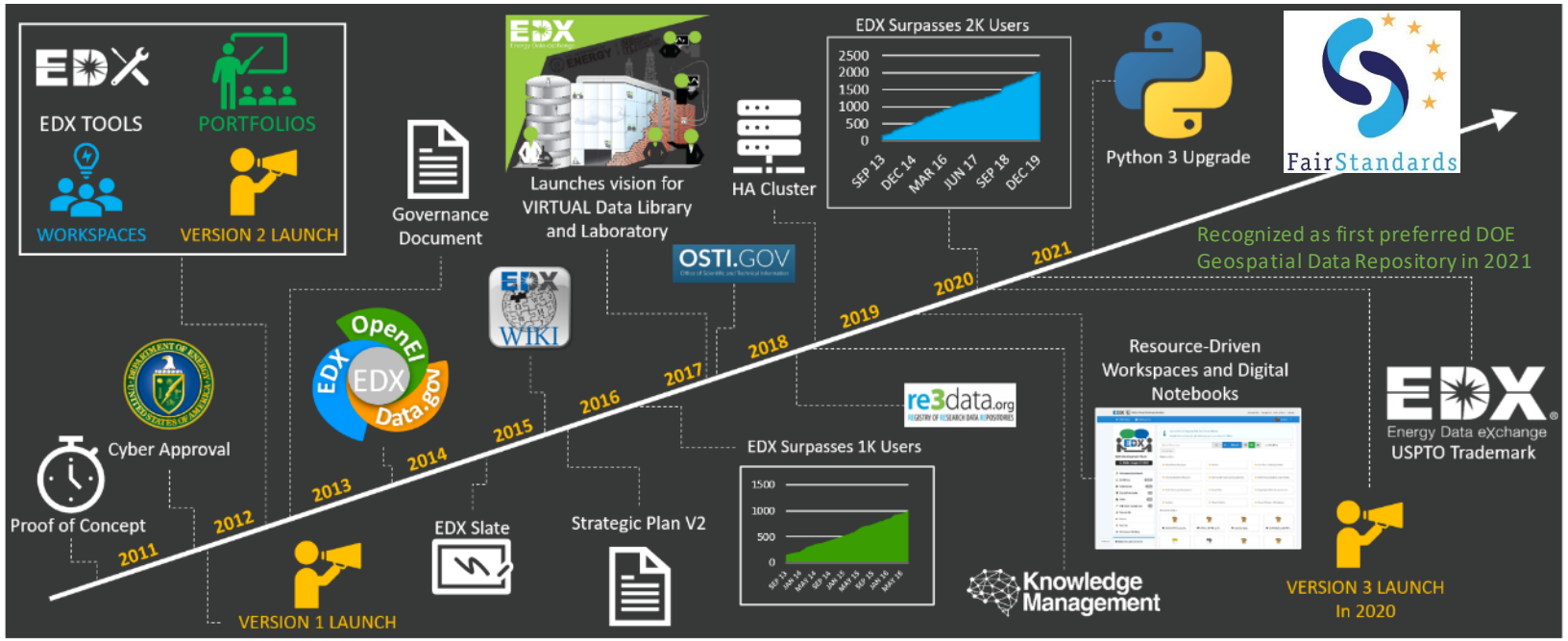
Presently data-driven teams spend ~80% of their time addressing the bottom components of the "data pyramid"

Crowdflower 2016

Initiated in 2011, the Energy Data Exchange (EDX®) supports the entire life-cycle of data

- **Public dissemination** of data, publications, presentations, tools, and software
- **Private** and multi-organizational collaboration
- **Secure and accessible** platform to internal and external users
- DOE and Federal **compliance**
- **Scalable architecture**
- **Agile development process** to meet the needs of users
- **Enterprise capabilities for** data curation, software/GIT, and cloud services in an integrated platform (EDX++)

Over a Decade of Experience Building a Virtual Laboratory for R&D



EDX, FECM's investment in their data foundation

NATIONAL ENERGY TECHNOLOGY LABORATORY

Registered Users: 3,260
Resource Count: 979,810
Resource Downloads: 2,223,142

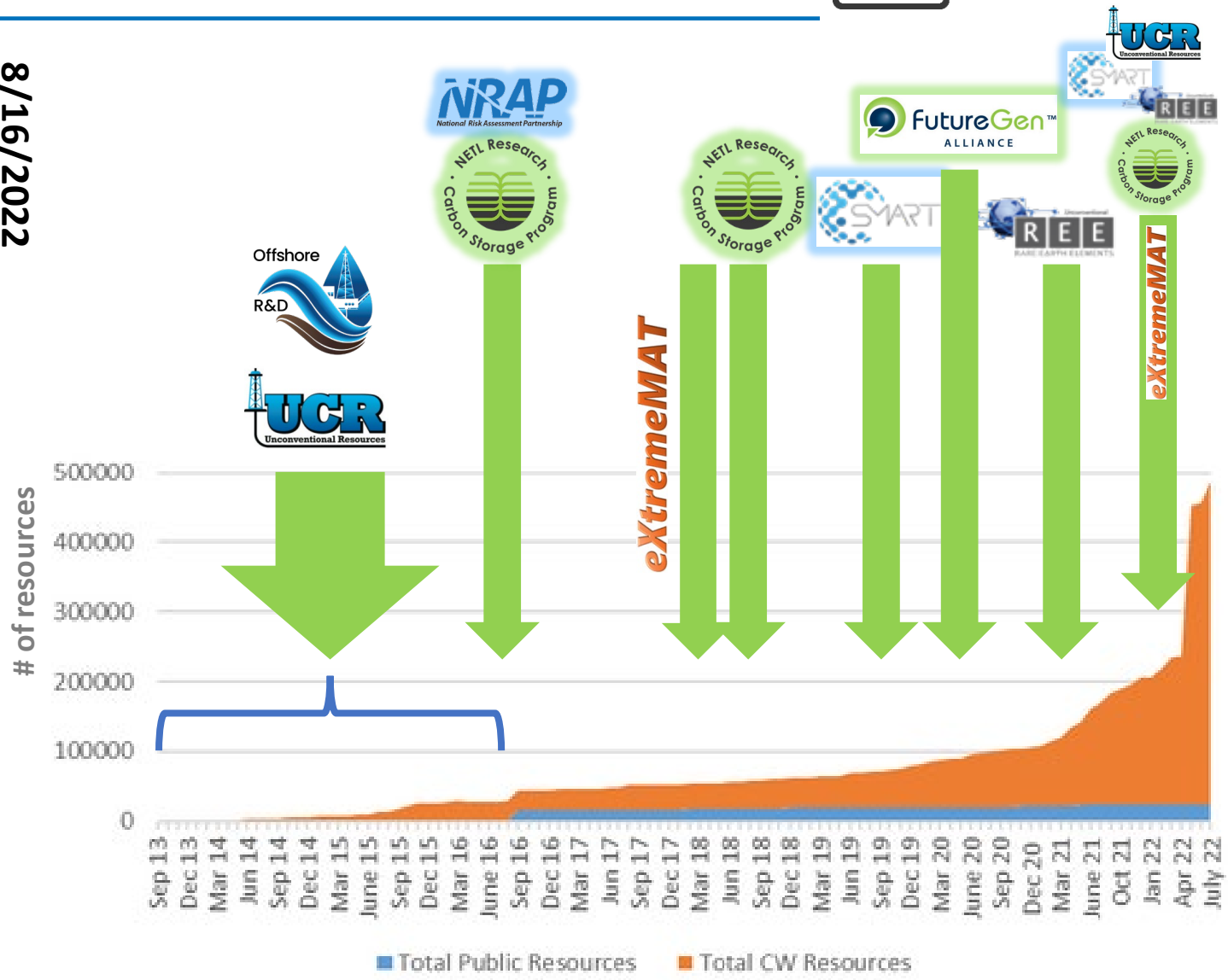
8/16/2022

Access to Published Data, Tools, Publications, Presentations, and resources such as:

- Over 23,000 published data resources (datasets, tools, apps)
- Almost 25 million federated data resources
- 42 Research Tools integrated with EDX for access & virtual use
- 48 EDX Groups where R&D communities share data & information
- 10 Research (FWP) Portfolio Sites technical R&D collections/sites
- Key public data from EDX via free commercial cloud services

Secure, Private Collaboration

- Over 3000 Active Research Projects via EDX Collaborative Workspaces
- Over 209,000+ secure, private data files

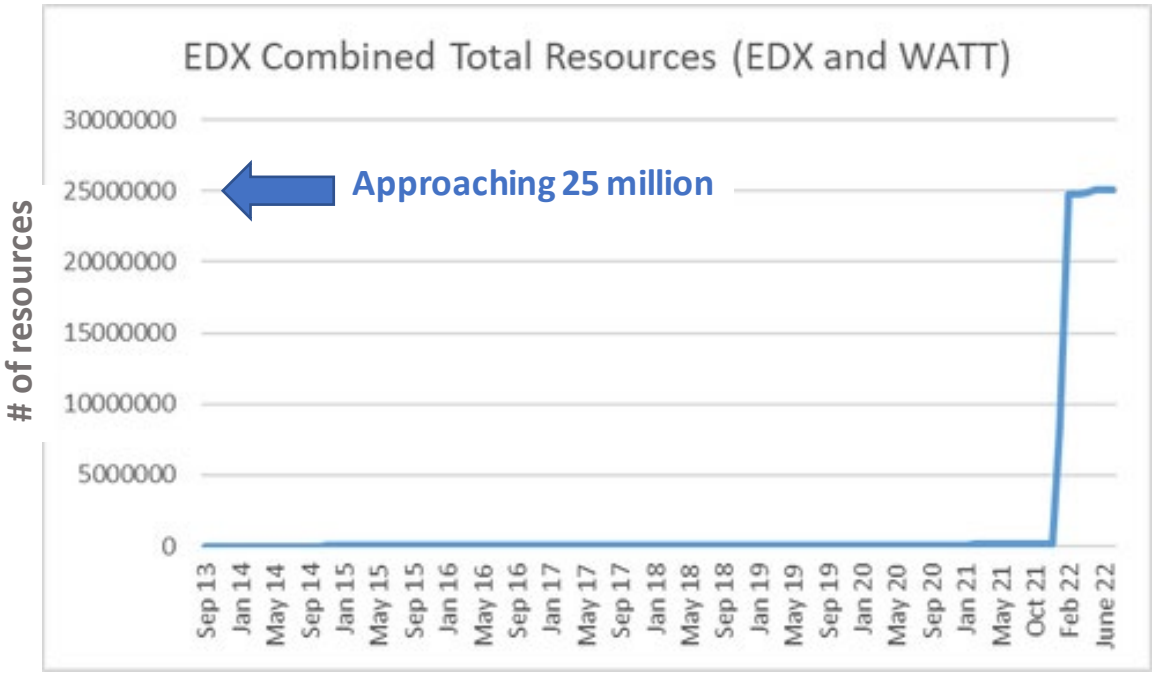


EDX, FECM's investment in their data foundation

NATIONAL ENERGY TECHNOLOGY LABORATORY

Registered Users: 3,260
Resource Count: 979,810
Resource Downloads: 2,223,142

8/16/2022



Access to Published Data, Tools, Publications, Presentations, and resources such as:

- Over 23,000 published data resources (datasets, tools, apps)
- Almost 25 million federated data resources
- 42 Research Tools integrated with EDX for access & virtual use
- 48 EDX Groups where R&D communities share data & information
- 10 Research (FWP) Portfolio Sites technical R&D collections/sites
- Key public data from EDX via free commercial cloud services

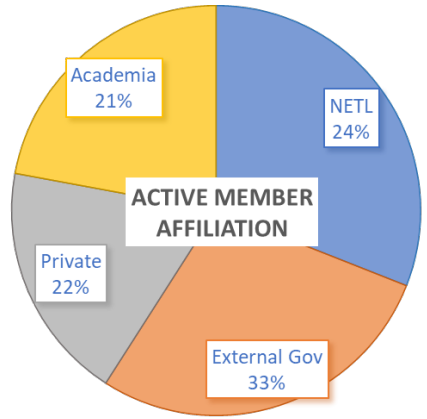
Secure, Private Collaboration

- Over 3000 Active Research Projects via EDX Collaborative Workspaces
- Over 209,000+ secure, private data files

aws Open Data on AWS
https://opendata.aws

Google Cloud Public Datasets

Azure Open Datasets
Curated open data made easily accessible on Azure



Democratizing Access to Digital Resources for the CCS Community

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

POLITICS

Major internet providers to slash cost of broadband service for low-income Americans

Michael Collins
USA TODAY

Published 5:02 a.m. ET May 9, 2022 | Updated 9:10 a.m. ET May 9, 2022

Here's why access to the internet is not created equal, how COVID-19 made it worse

As schools debate about returning to online learning, the lack of internet access for many Americans is a big sticking point.

Just the FAQs, USA TODAY

- 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 have 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

Basic Science R&D

Increases the knowledge base of a field of research while applied science uses that knowledge to solve specific problems.

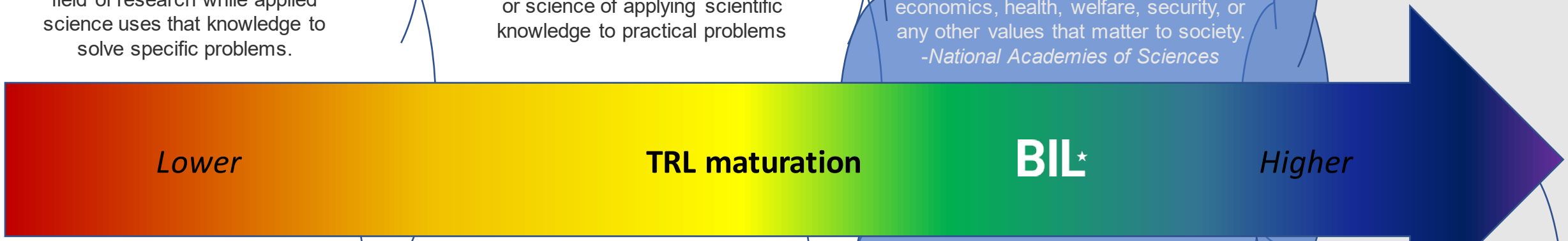
Applied Science R&D

The discipline dealing with the art or science of applying scientific knowledge to practical problems

Actionable Science R&D

Defined as science delivered to a decision-maker that is timely…and meaningful—in terms of safety, economics, health, welfare, security, or any other values that matter to society.
-National Academies of Sciences

Commercial & Regulatory Applications



NETL Research Carbon Storage Program

NRAP National Risk Assessment Partnership

SMART

EDX Energy Data exchange

4CCS

Enable and Accelerate CCS Deployment

NATIONAL ENERGY TECHNOLOGY LABORATORY

Real-time Visualization, Forecasting, and Virtual Learning for Decision Makers

- Science-based ML-based tools (real-time) for decision-makers (site managers, regulators, etc.,) to accelerate field-scale deployment.
- Tools that can transform the interactions in the subsurface and significantly improve efficiency, effectiveness, and safety of carbon storage operations.

Site-specific risk-based decision support tools (open-source) for Stakeholders

- Physics-based tools for risk management (long-term, liability, and financial)—addressing leakage assessments, induced seismicity risk management, and strategic monitoring for uncertainty reduction.
- Addressing stakeholder needs for permitting and project start-up as related to risk assessment and management. DOE NLs are providing TA to EPA—based on expertise gained through NRAP

Putting CCS data resources to work - Stakeholders/Community access to authoritative data, analytical, visualization, and virtual data infrastructure capacity

- CCUS Ecosystem – data curation and collaboration platform, includes both public and private data sharing capabilities, entire life-cycle of data, data discovery, transformation, and integration.
- Multi-cloud deployment to enhance access by the CCUS stakeholder community to virtualized CCUS data and models to accelerate CCUS commercial, regulatory, social, and environmental goals.

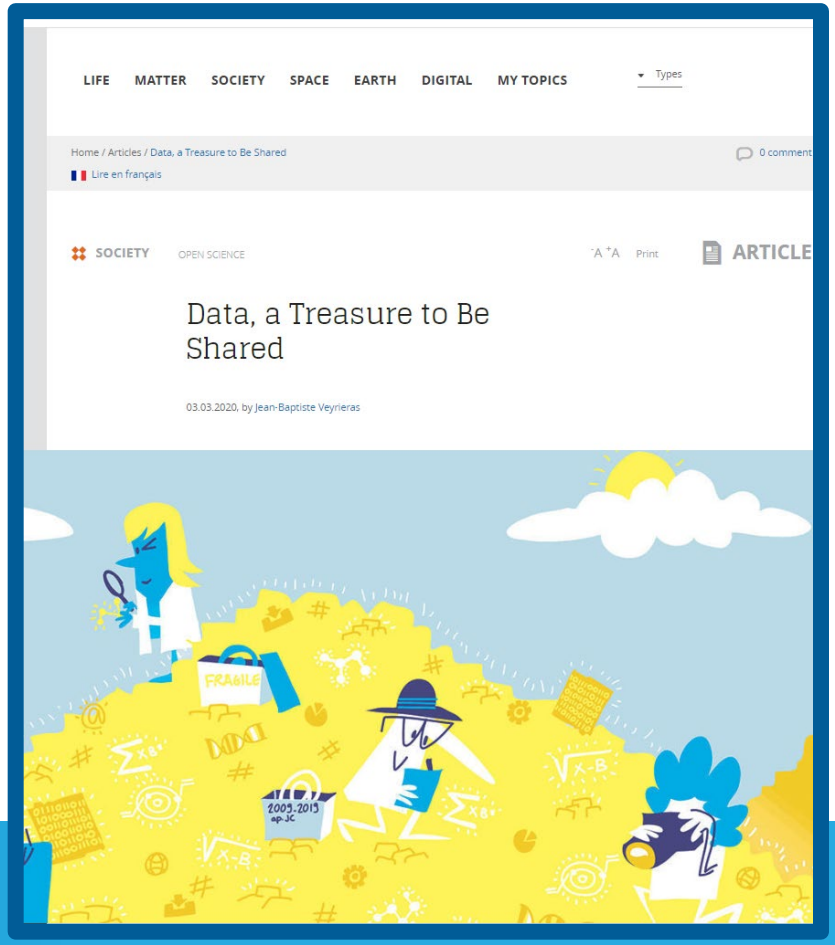
EDX 4CCS
Energy Data eXchange

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 Ecosystem** – data curation and collaboration platform, includes both public and private data sharing capabilities, entire life-cycle of data, data discovery, transformation, and integration.
- **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.

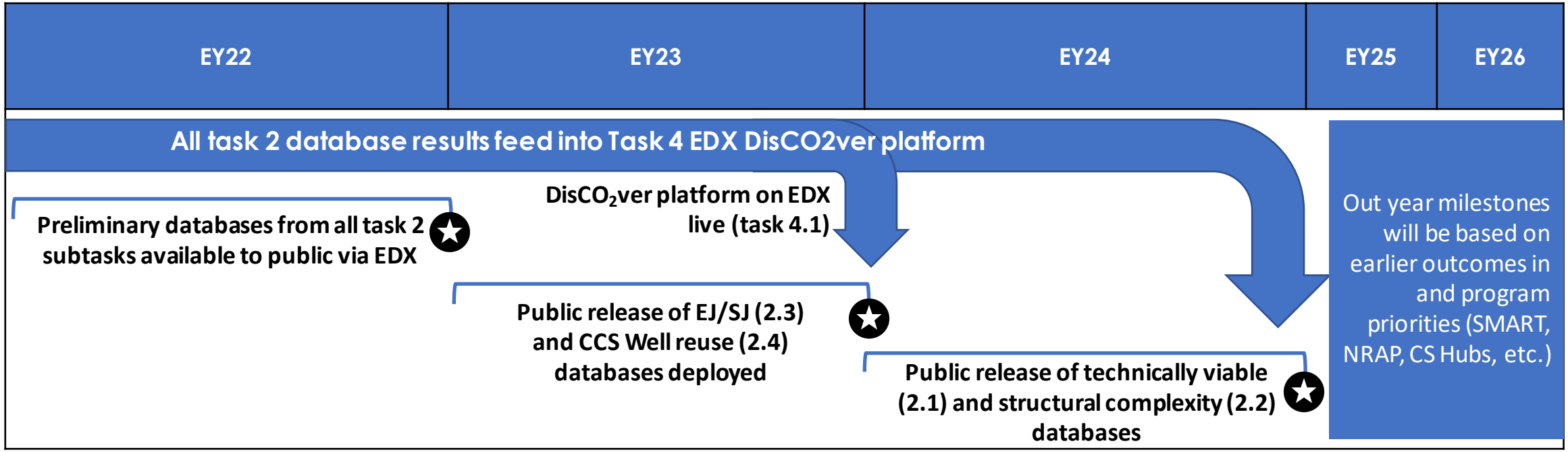
- **CCS Data Resources**
  - Refining, integrating, and generating priority data resources to feed carbon storage commercial, R&D (SMART, NRAP, etc.), and regulatory models
- **CCS Tools**
  - Development and deployment of strategic Carbon Capture and Storage (CCS) data visualization and analytical tools to accelerate CCS commercial and regulatory efforts
- **Core CCS EDX DisCO2ver 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 the deployment of key capabilities and resources for the CS community.
- **Outreach & Capacity Building**
  - Outreach efforts, training, and capacity building within communities across the US.
  - Future CO_2, H_2 and SNG, biomass, and other source points and production/emission rates



2.0 – CCS Data Integration, Generation, and Deployment

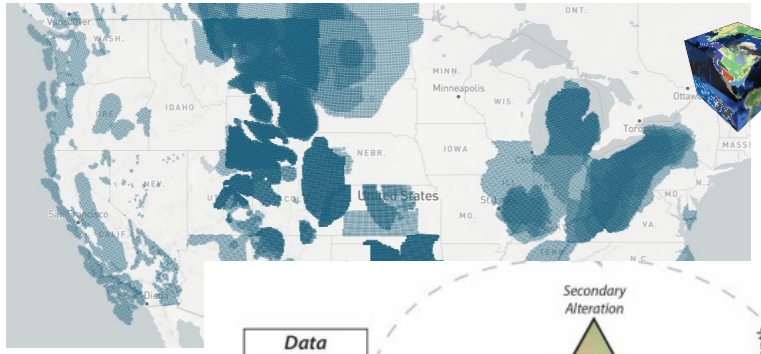
Goal ⇨ DOE's CCS programs have invested in the strategic curation of CCS data products over the last seven years. This suite of subtasks targets the generation of refined datasets and databases to meet key stakeholder data needs.

- Reduces redundant dataset acquisition and integration by each user for priority database needs
- Combining DOE-funded data with other public, authoritative datasets to increase the useability and completeness of resulting integrated databases
- Helps identify remaining data/knowledge gaps

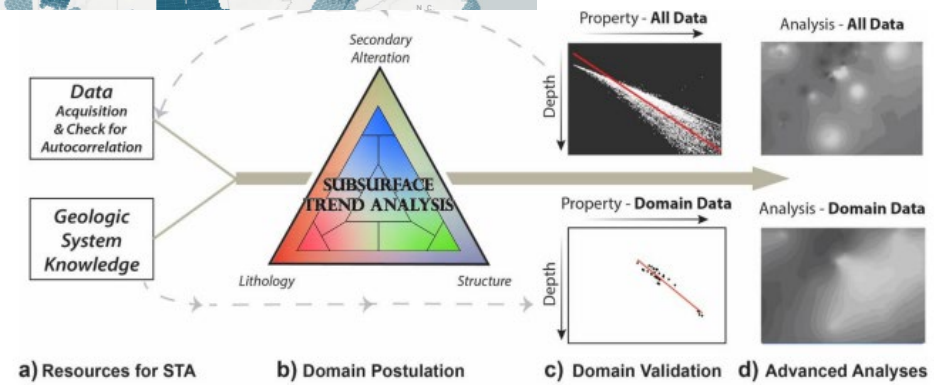


Technically Viable Carbon Storage Resource Geospatial Analysis

Need: Understanding more accurately regions with the highest feasibility for CCS considering storage capacity, environmental & socio-economic factors, and infrastructure



Rose et al., 2021

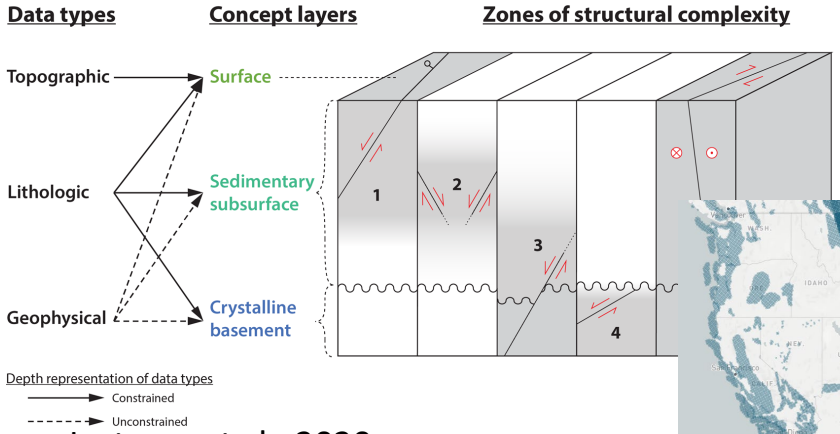


Outcome: A dataset that considers additional factors in storage resource assessments to enable understanding and planning of feasible CCS on a large scale in the USA

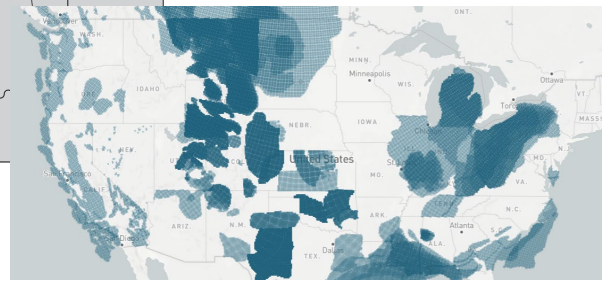
Technically viable carbon storage database published to EDX via DisCO_2ver that will be available for internal and external stakeholder use

National Structural Complexity Database

Need: Datasets to better understand the influence structural complexity will have on long-term carbon storage security



Justman et al., 2020



Outcome: Provides information about areas with limited or poor-quality structural complexity information and assists with the identification of data gaps that may require additional, future fieldwork

Publish structural complexity analyses within a database for selected deep saline basins on the DisCO2ver, EDX

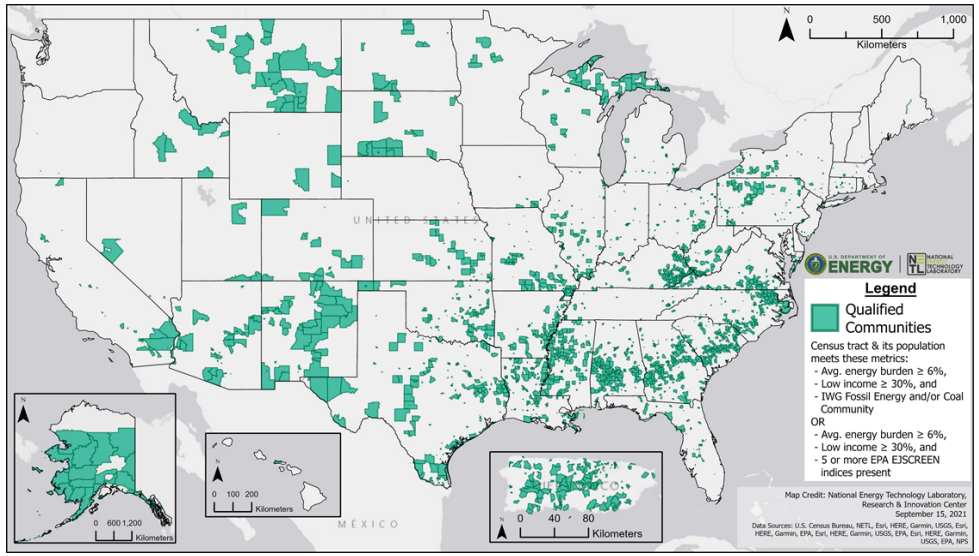
EJSJ Dynamic Dataset for CCS Systems

National Well Database to Support CCS Reuse and Risk Work

NATIONAL ENERGY TECHNOLOGY LABORATORY

Challenge: Lack of information & insight available to better understand the viability of a CCS project's implementation based off existing environmental justice (EJ) and social justice (SJ) challenges, as well as overall community attitudes towards CCS

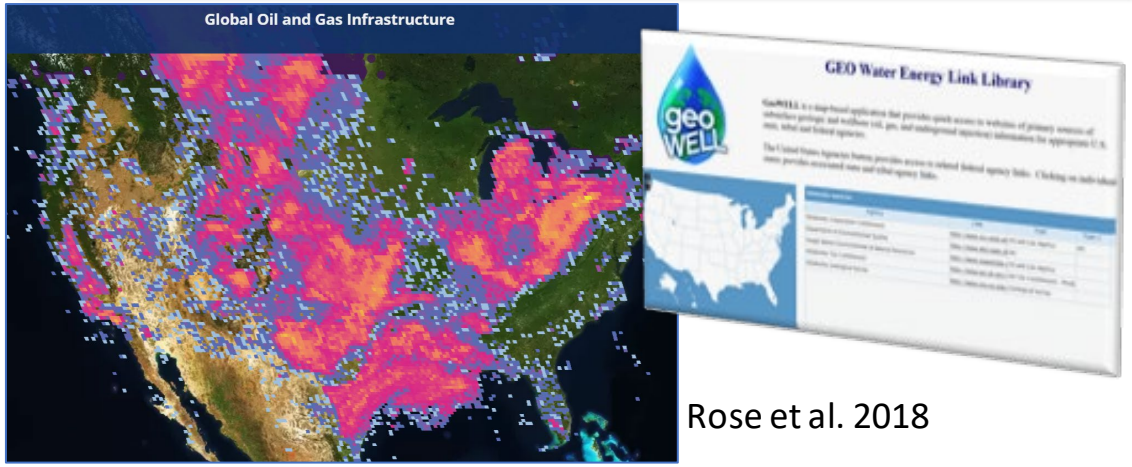
Need: Safe, long-term geologic CS requires comprehensive well information. Currently, well data are disparate, siloed by local entities. Stakeholders* have requested the integration of these resources to better understand national CCS resources and opportunities.



Bauer and Rose 2021

Metrics used in DOE's Communities LEAP Pilot, to which qualify communities have:
High or severe energy burden (≥ 6%), 30% or higher low-income pop., and a historic economic dependence on fossil fuels; or high or severe energy burden (≥ 6%), 30% or higher low-income pop., and a high exposure to environmental hazards

Outcome: Downloadable CCS-EJ-SJ database on EDX DisCO2ver, to support CCS project efficiency and improve CS stakeholders' understanding of EJ & SJ factors and how they may affect commercial, regulatory, and research applications.



Rose et al. 2018

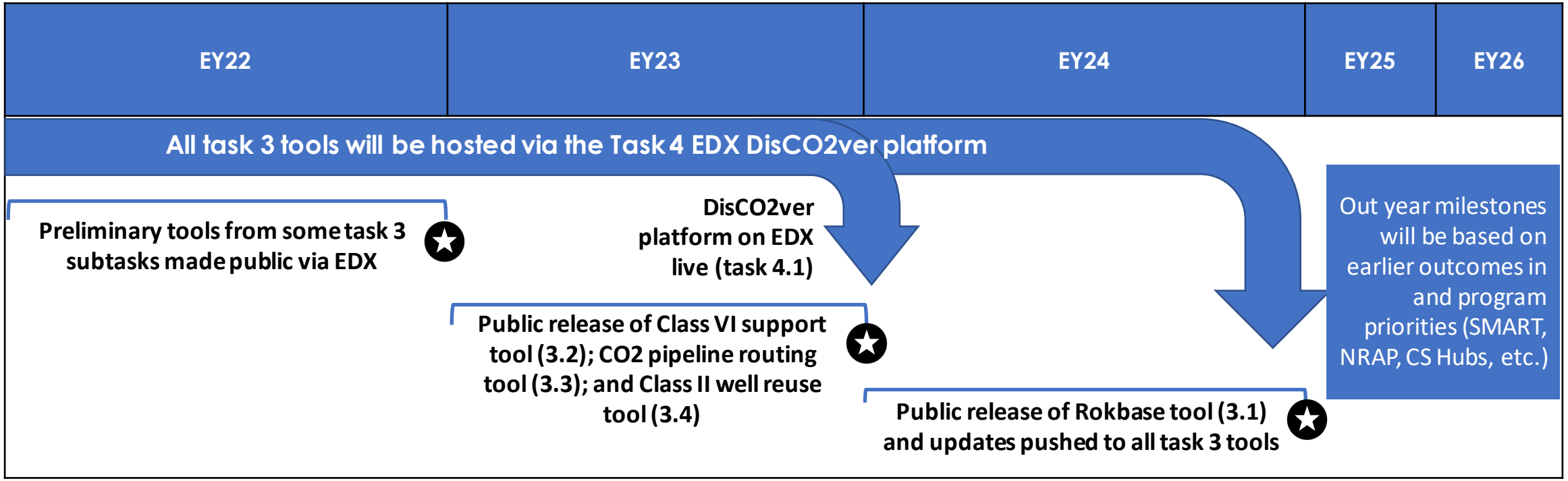
Outcome: The National Well Database will provide an up-to-date, integrated resource supporting CCS R&D across the country. Information gained from the Database will offer critical insights for safe injection site selection, while leveraging existing well infrastructure, limiting costs and the fossil energy footprint.

National CCS Well Database published on EDX®, offering up-to-date well data for CCS R&D

3.0 - Tools - Development of Priority Tools for CCS Data Interaction and Visualization, & Decision Support

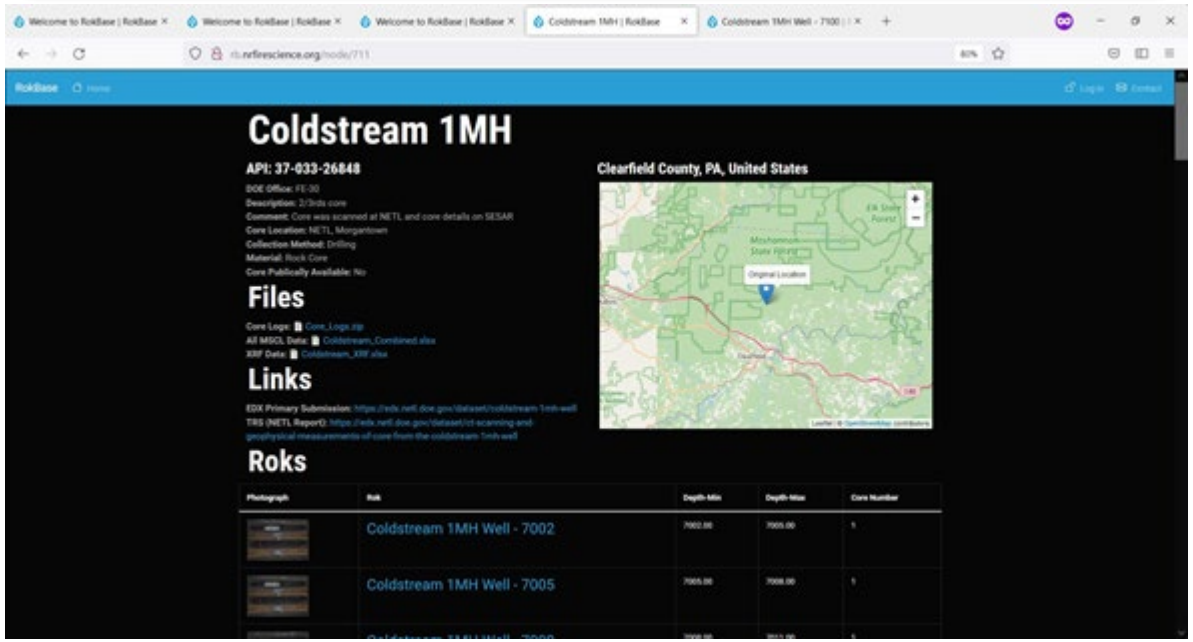
Goal ⇨ Development and deployment of strategic Carbon Capture and Storage (CCS) data visualization and analytical tools to accelerate CCS commercial and regulatory efforts

- Promoting efficient reuse of DOE CCS and open-source datasets in tools/applications for decision support
- Democratizes access to data plus tools for all CCS stakeholders
- Hosted on the Energy Data eXchange (EDX), a publicly available, online platform



RokBase, Virtual Rock Property Tool

Challenge: Ensuring that years of DOE investments into core characterization data collection can be easily found, queried, visualized and reused

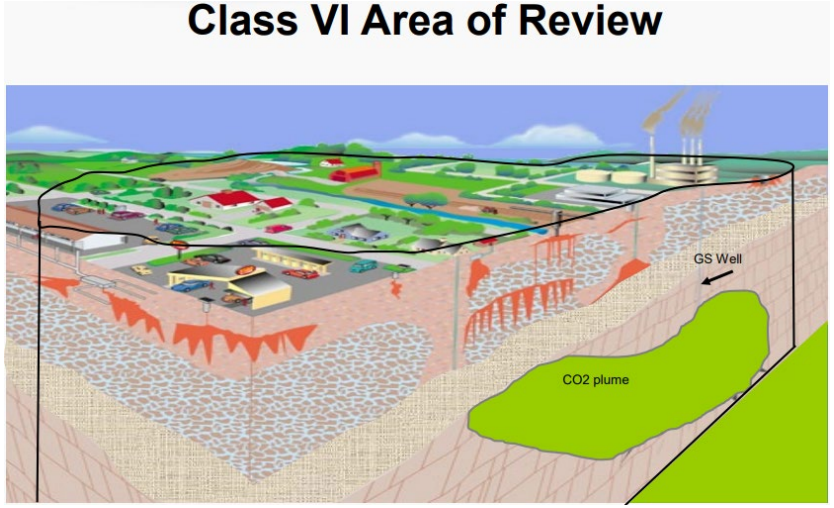


Outcome: A visualization platform that enables users to easily access core characterization data from carbon storage projects.

Tool linked to data on EDX of for exploring and downloading core characterization data, enabling users to reuse detailed subsurface property information from DOE field study areas.

Class VI Data Tool

Challenge: CCS deployment in coming years will require application for Class VI permitting with the EPA to drill injection wells, which requires vast amounts of data



Tool would help user's walkthrough the information required for a Class VI Permit & help generate information or allow manual entry/submission of data to provide a more complete submission packet

Value: **Provide access to necessary publicly-available datasets for the whole of the U.S. in a central tool that can be used when preparing to apply for a Class VI permit**

A data portal that contains spatial data and well logs, enabling users to query, visualize, and download data in formats needed to add data to the EPA GSDT tool during the Class VI application process.

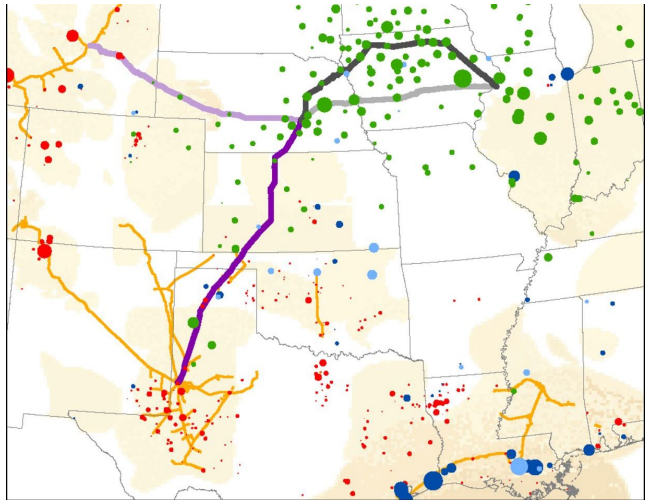
Opportunity to work with Environmental Protection Agency (EPA) and/or U.S. Geological Survey (USGS). **Potential to connect to EPA tool GSDT**

CO_2 Pipeline Routing EJ/SJ Smart Tool

Challenge: Current CCS pipeline models fail to integrate complex economic, social, and environmental variables that can greatly affect the success of developing new pipelines and increase project costs.

Outcome: Interactive tool to assist in planning routes of new CO2 pipelines across the U.S. to accelerate development while considering multiple EJSJ variables.

Interactive tool on EDX DisCO2ver Platform, allowing users to plan routes for CO2 pipelines across the U.S., both onshore and offshore and evaluate routes against EJ & SJ variables that may prohibit successful development.

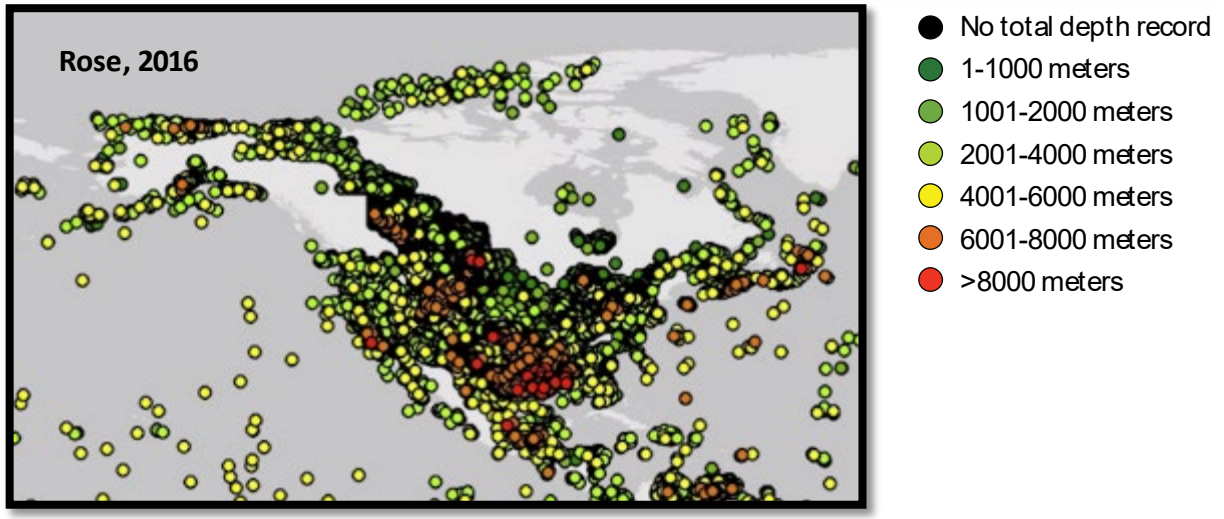


Provide interactive platform to evaluate potential routes and obstacles, evaluate scenarios and solicit stakeholder feedback

Class II Well Reuse and Regional Evaluation Tool

NATIONAL ENERGY TECHNOLOGY LABORATORY

Challenge: Currently unknown if Class II wells are ideal candidates to be repurposed for geologic carbon storage.



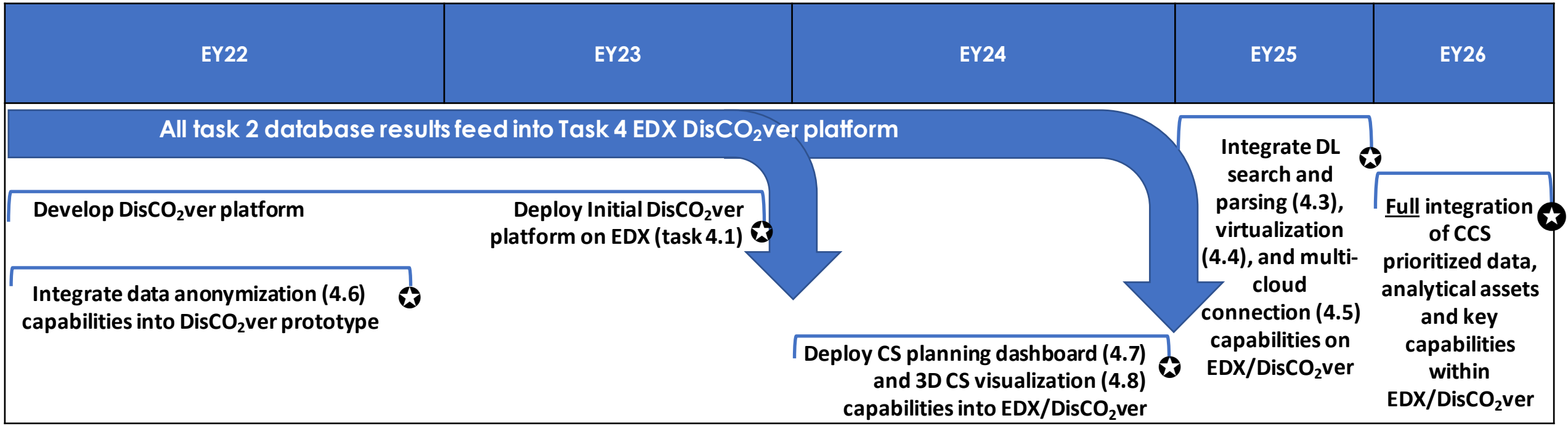
Outcome: Provide a standardized method to review existing Class II wells and characterize their potential for safe carbon storage reuse that will provide critical information to researchers and decision-makers that can help ensure safe and efficient evaluation of Class II wells for CO_2 injection.

Standardized methodology to priority rank Class II wells for reuse potential for carbon storage that can be applied by other CS stakeholders to ensure safe and efficient evaluation of Class II wells for possible CO2 injection.

4.0 - Core Carbon Capture & Storage EDX DisCO2ver Platform

Goal ⇨ Democratizing and streamlining access to curated Carbon Capture & Storage (CCS) data resources through an online/virtual ecosystem

- Connect stakeholders across commercial, regulatory, and research domains to a curated collection of CCS data, models, and capabilities (visualization, analysis, etc.)
- Virtual data library and laboratory to promote efficient reuse and interaction for prioritized capabilities
- Secure, facilitated, architected environment, optimized for CCS data
- Promoting compliance with DOE and Federal requirements and policies

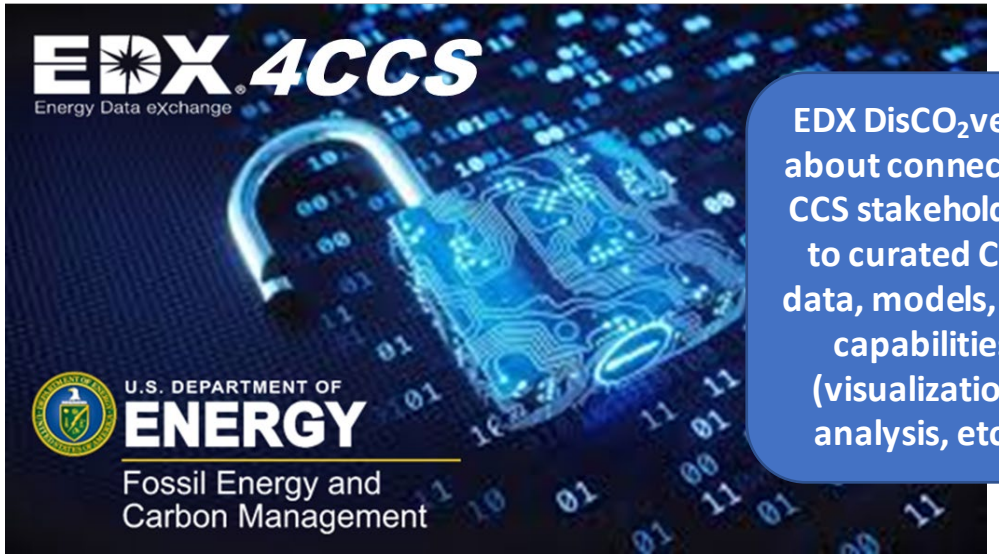


4.0 - Core Carbon Capture & Storage EDX DisCO2ver Platform

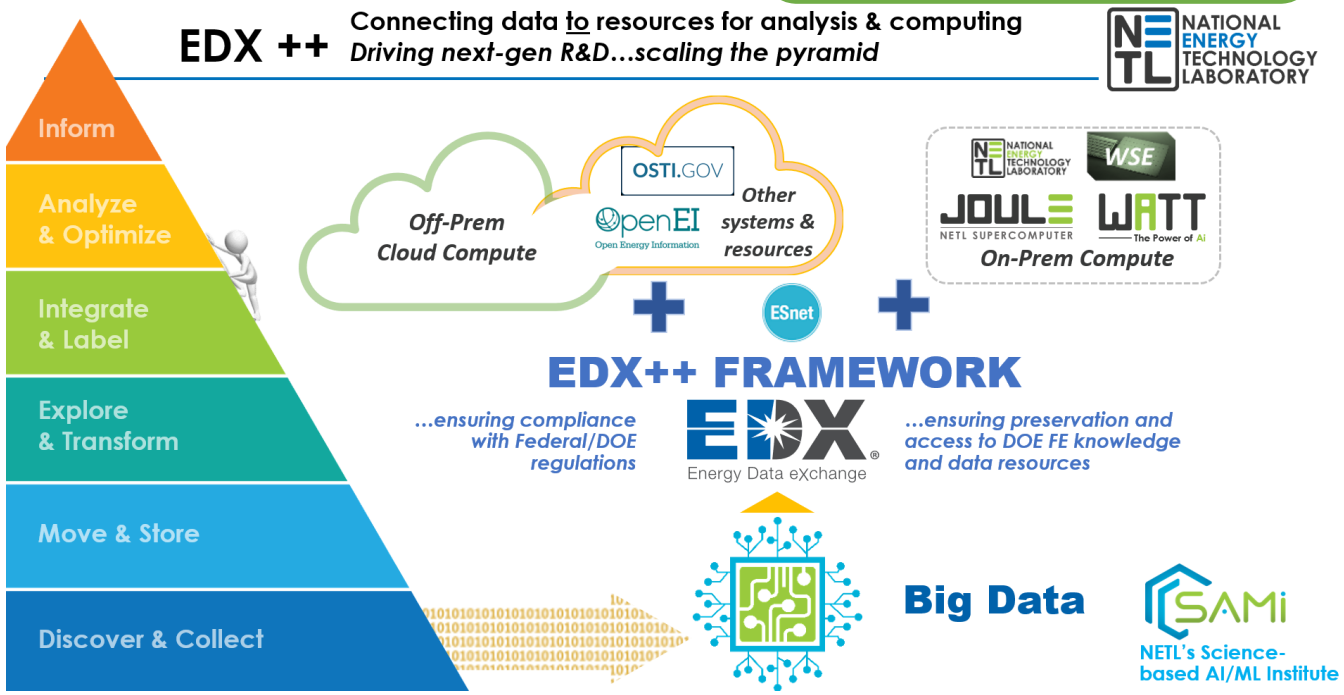
EDX DisCO2ver platform, Broader community virtualized data computing platform—**DisCO$_2$ver**

- 4.1: **DisCO2ver**, a Common Operating Platform for the Next Generation of CO2 Systems (EY23+++)
- 4.2: Integrated Access to CCS-Relevant **Data from Non-DOE Authoritative Data Sources** (EY23+++)
- 4.3: **SmartSearch and SmartParse** Capabilities Integration into the DisCO2ver Platform (EY23)
- 4.4 Integration of **CCS Analytical Models and Tools** into DisCO2ver (EY23+)
- 4.5 EDX **Cloud** Optimization for CCS (EY23)
- 4.6: **Data Anonymization** Tool Integration into DisCO2ver (EY22)
- 4.7: Carbon Storage **Planning Framework** Dashboard (EY23)
- 4.8**: 3D Data Viewer** and Preview Capability (EY24)

EDX++ is a separate effort to upgrade and enhance the EDX core infrastructure, connecting users to multi-cloud and on-premise compute



EDX DisCO$_2$ver is about connecting CCS stakeholders to curated CCS data, models, and capabilities (visualization, analysis, etc.)



https://edx.netl.doe.gov/about

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

these 5 core competencies support

Core Projects/Programs

- **Carbon Storage** (DisCO2ver)
- **Hydrogen** (InsigH2t)
- **Materials** (Metals, Cements, Polymers, Carbon)
- **Minerals Sustainability**
- **Infrastructure**
- **EDX Geospatial**

Ancillary Projects/Programs

- Comp Sci
- Energy Water
- Labs
- Microbiology
- Minerals Sustainability
- Sensors
- Subsurface
- System Analysis
- AND OTHERS…

EDX has fundamental capabilities that can be utilized across many research communities

NETL Research · Carbon Storage Program

SHASTA Subsurface Hydrogen Assessment, Storage, and Technology Acceleration

Natural Gas Hydrates

NRAP National Risk Assessment Partnership

SMART Offshore R&D

GEO CUBE

WATT The Power of AI

SAMI SCIENCE-BASED AI/ML INSTITUTE

REECM

UCR Unconventional Resources

NETL SmartSearch

What's next: EDX DisCO$_2$ver

NETL Research · Carbon Storage Program

NETL SmartSearch

NETL SmartParse

NatCarb
A National Look at Carbon Sequestration

CarbonSAFE Projects

Carbon Storage Open Database

EDX DisCO$_2$ver

- **Development and deployment** of the "DisCO$_2$ver Platform" **user interface** within the EDX ecosystem
- Offering an **integrated**, dedicated user interface
- **Connecting** CCS-community stakeholders to CCS-priority data, tools, and models for use
- Leveraging the core data management, curation, collaboration, and virtualization capabilities offered through the secure, online EDX++ system.

GEO CUBE

EDX++

SMART

NRAP
National Risk Assessment Partnership

NETL RESOURCES

VISIT US AT: **www.NETL.DOE.gov**

https://edx.netl.doe.gov/about
https://edx.netl.doe.gov/carbonstorage

Thank you!

Kelly Rose, Kelly.Rose@netl.doe.gov
Paige Morkner, Paige.Morkner@netl.doe.gov
Chad Rowan, Chad.rowan@netl.doe.gov