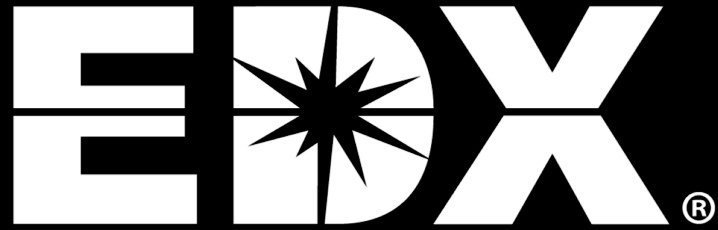


EDX++ Multi Cloud, Unlocking Next Generation R&D Data Management for CTS Stakeholders

Catalyzing Collaboration & Accelerating Innovation



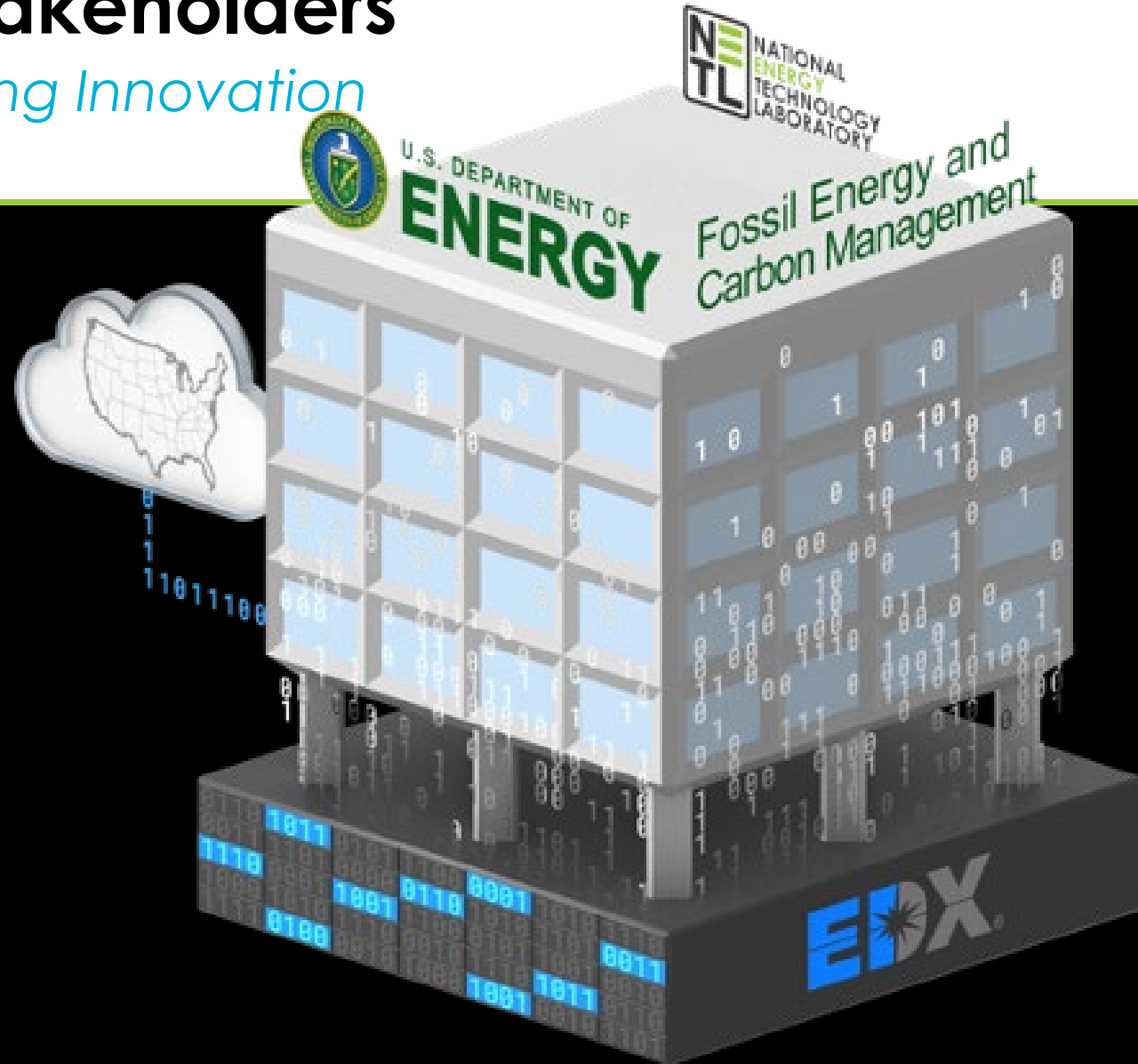
Energy Data eXchange

Chad Rowan

National Energy Technology Laboratory (NETL)

Computational Science & Engineering

Advanced Computing & Artificial
Intelligence

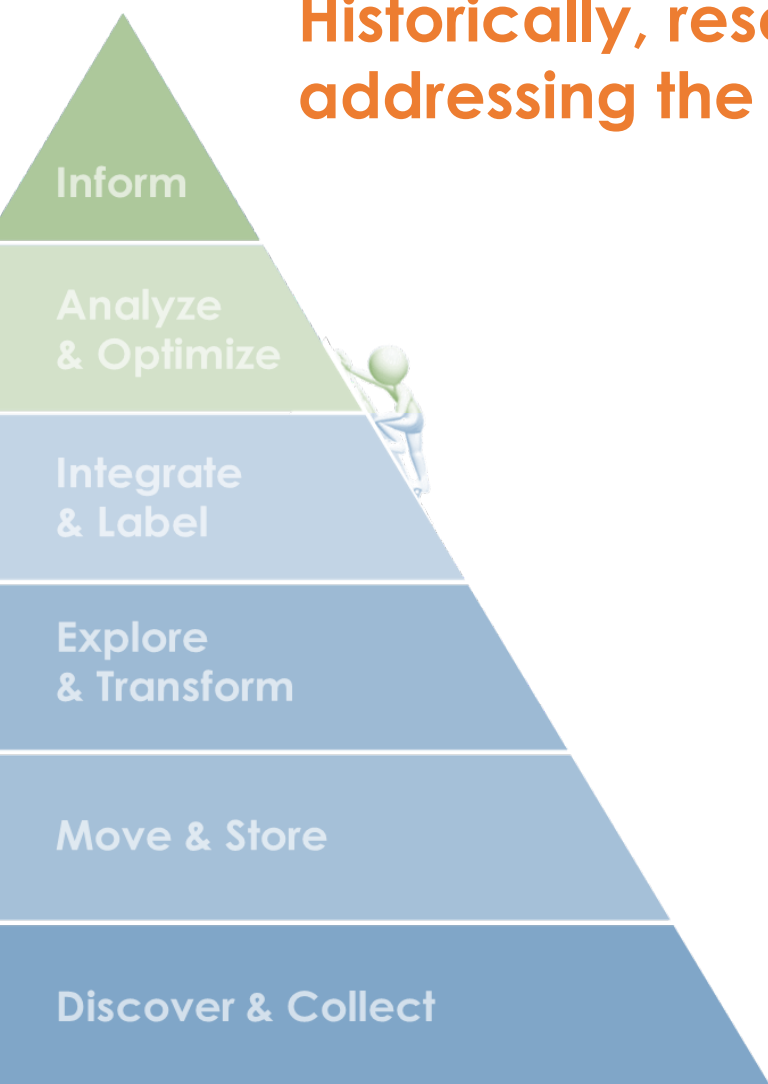


Disclaimer

This work 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 authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Did you know?

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



20%

EDX[®] has been used by DOE Programs to reduce data access barrier/overhead

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?

FECM's foundational R&D digital infrastructure & AI test bed

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



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



v.1 2012

12+ years

EDX++ v.4 2024

How has NETL addressed R&D data needs?

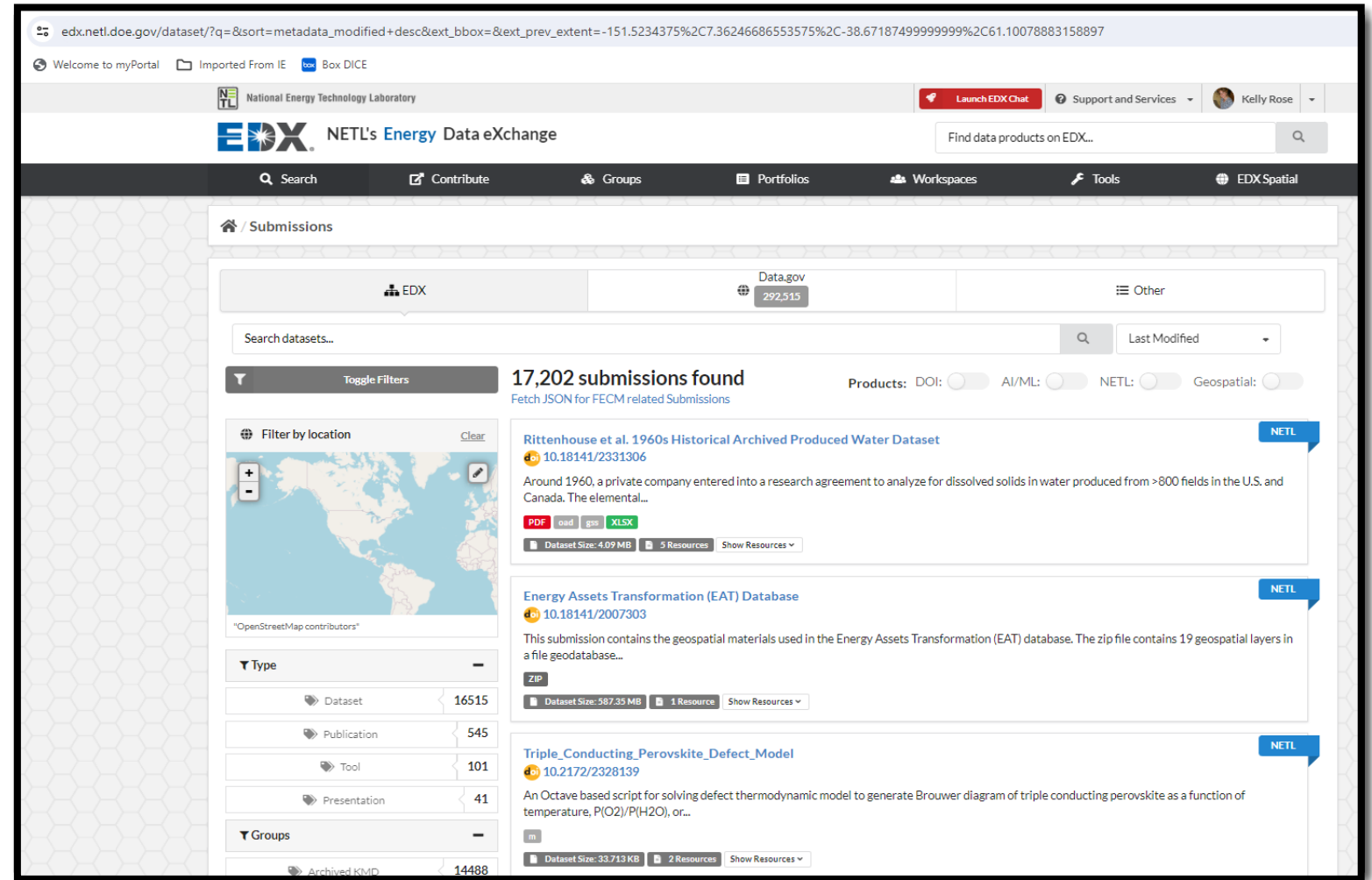


FECM's foundational R&D digital infrastructure & AI test bed

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



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

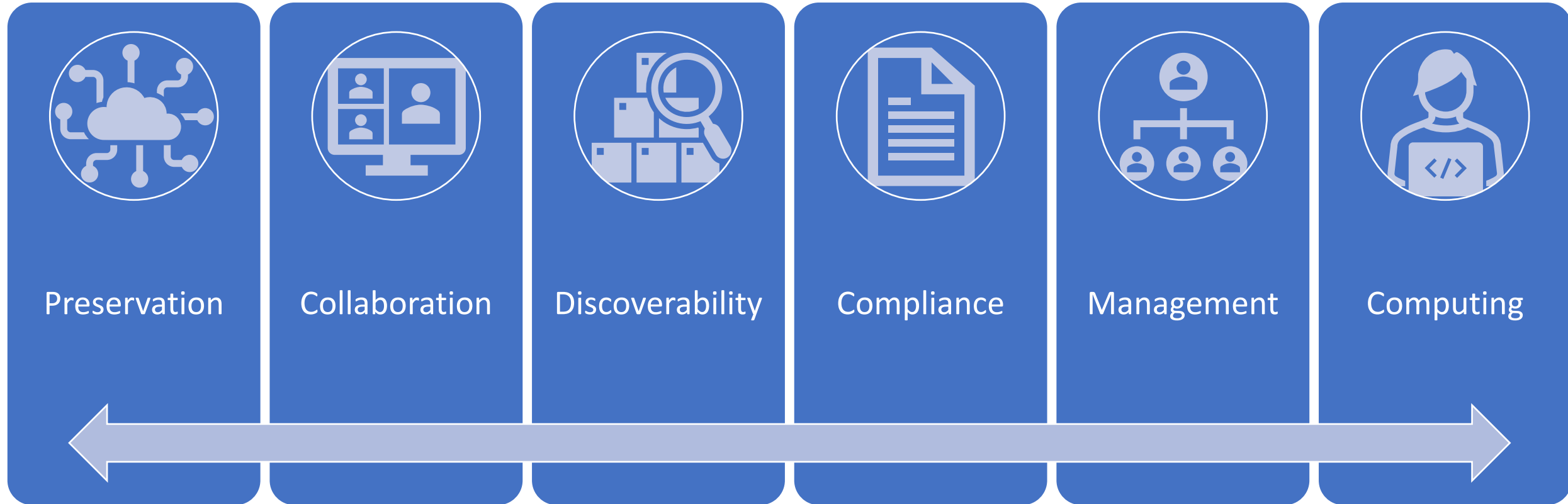


v.1 2012

12+ years

EDX++ v.4 2024

How is EDX currently supporting its community?



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

EDX Public Product Inventory

Total Submissions

17,268

Total Resources

33,533

Approximate File Size (TB)

324

Datasets

16,557

Research Documents

550

Tools

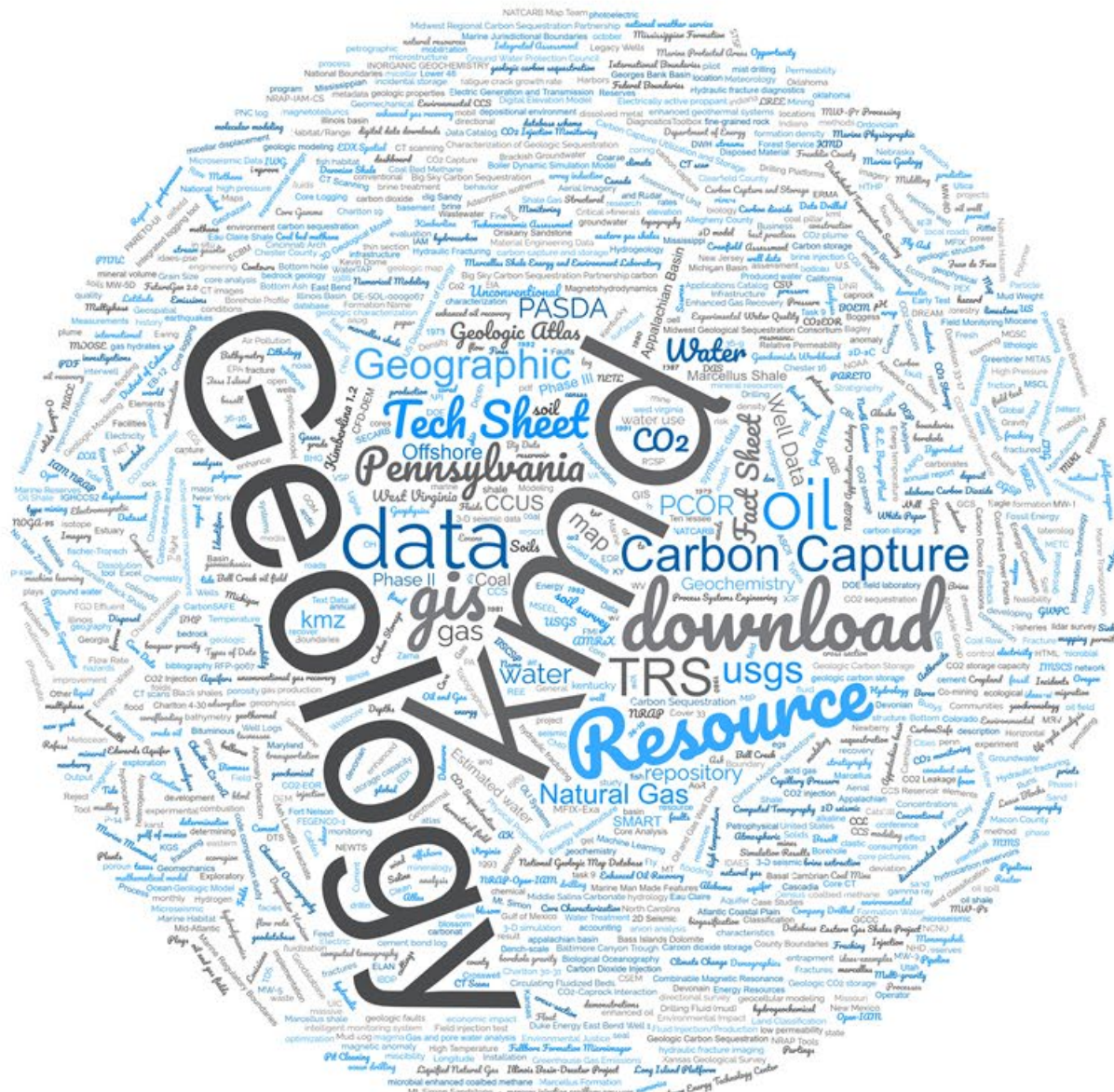
120

Presentations

41



Number of Submissions by Type



Wordcloud representing weight of tags/keywords in public submissions.

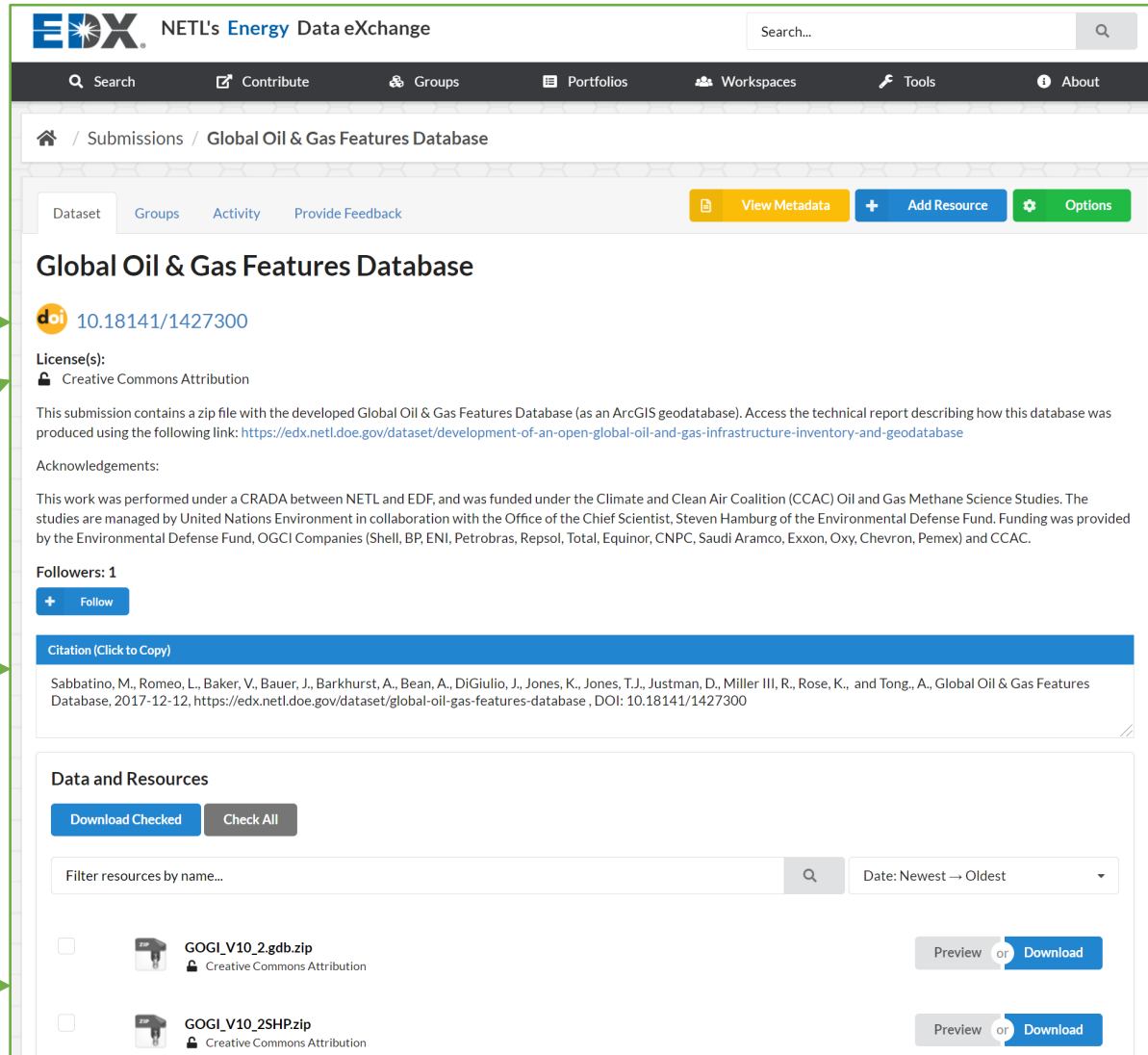
Advantages of publishing data products

OSTI DOI Number

Data License

Data Citation

Data Access

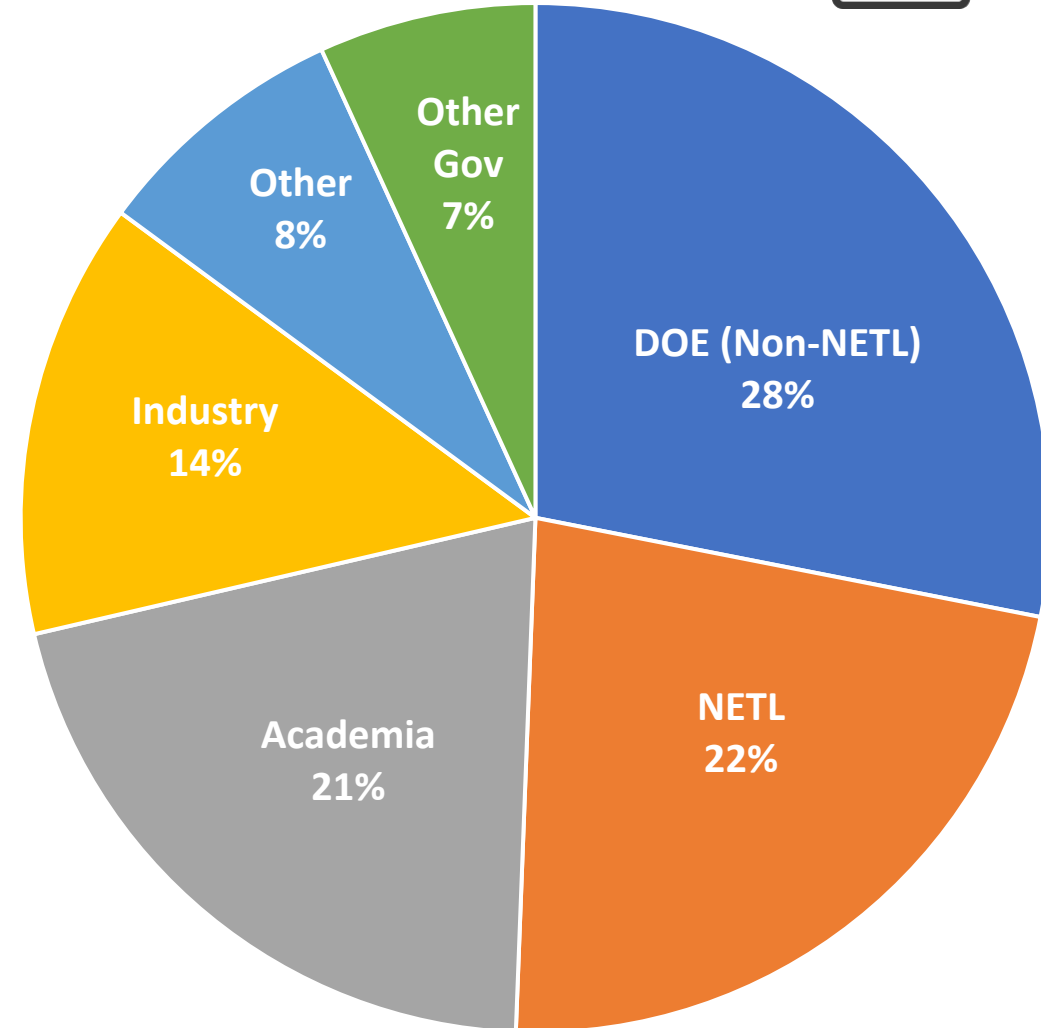


Many journals require models, tools and data be publicly available prior to journal publication.

Who is part of the EDX community?

- ✓ NETL/FECM employees
- ✓ External collaborators
- ✓ General public

2600+ Active Registrants



2-3 years ago, EDX & FECM Stakeholders were at a crossroads

- Data, users, tools, etc. were growing at a rapid pace
- Users needed quick, efficient access to data and information for mission critical programs and projects
- Users needed secure & responsible access to AI/ML tools, compute, hosted applications, etc.
- **Outdated** and isolated on-prem architecture while still relevant is:
 - Not built for multi-scale compute
 - Presently NETL (and most of DOE) supports desktop (small scale) users or HPC, leaving mid-scale multi-scale data computing users without adequate resources
 - Are not built for large data transfer
 - Is built for internal users, not multi-entity R&D data science teams
 - Is also built to spec for key HPC workflows

RIC (EDX) collaborated with FECM (10), NETL IT/Cyber, and DOE OCIO to enable **EDX 4.0 multi-cloud launch, March 2024**



Why migrate to the Cloud?

EDX growth is aligned with Cloud

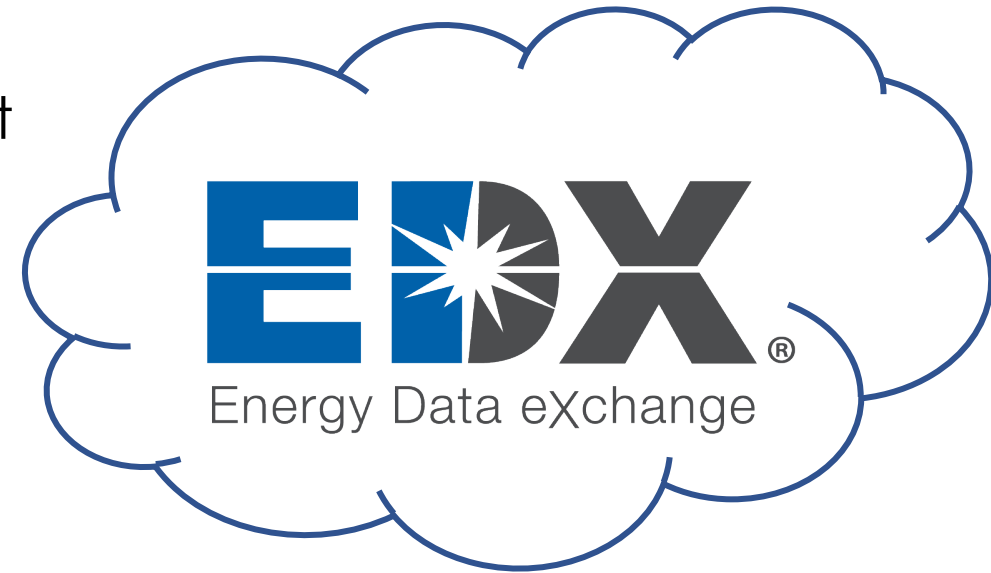
- Rapidly deploy new services
- Data and Services Redundancy
 - Disaster Recovery and High Availability
 - Enhanced Uptime
- Scalability for services
- Storage
- Cloud Provider Managed Infrastructure
 - Guaranteed infrastructure uptime Service Level Agreements



Why migrate to the Cloud?

Benefits for EDX Dev/Ops

- **Security**
 - Artifact Registry Vulnerability Scanner
 - Security Command Center and Cloud Armor
 - Data Security – encryption in transit and at rest
 - Secrets via Secret Manager
- **CI/CD:** Automated Builds and Deployments tied to commits!
- **Unlimited storage** via Buckets
- **Stability:** reliable infrastructure and networking
- **Cloud Network:** Fast, reliable, zonal / regional / global
- **Deep Bench:** broad tech community
- **Infrastructure:** Managed databases, CPU / RAM / GPU options, etc.

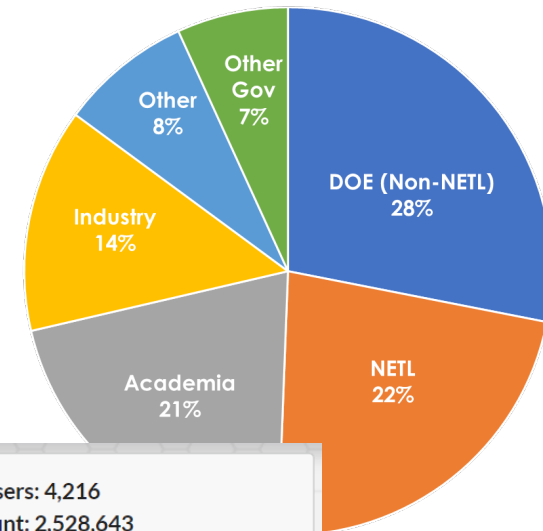


EDX Features and Key Elements Recap



EDX is a DOE, GOGO data curation and collaboration platform, initiated and in use since in 2011

- Built to provide **compliance** with DOE and Federal requirements/orders etc while also addressing **multi-entity research team** needs to ensure more efficient and secure research execution
- Utilizes a data lake approach, but also offers more structured data assets as well as advanced AI/deep learning capabilities for data discovery, transformation, and integration (SmartSearch©)
- EDX **curates** data products of DOE R&D (“data products” = datasets, software, models, tools, links, technical reports, etc) but also uses **services** (APIs, federation etc) to **connect** DOE-affiliated researchers with authoritative, relevant assets from external sources
- Private side uses role-based security to allow for **secure, private, metadata-informed data sharing** by multi-entity teams. Promotes collaboration to mature DOE-affiliated data R&D efforts
- Public side hosts and serves as an **internationally registered and recognized** data repository.
 - Curating individual data products coupled with Datacite.org citation standard, as well as connections to multi-cloud virtualized interoperability features (e.g. APIs, analytics, Cloud services, etc)
- **Amplifies** visibility for DOE data products, including ties to Data.gov, Geoplatform, OSTI, Re3Data, Google search etc
 - Federates and aligns with key DOE, Federal and International platforms
- **Multi-cloud** architecture and approach
 - Continuously evolving with input from the technical/R&D community, **built by researchers** for research and technical users
- **Flexibility** to support and adapt to a variety of uses and Program needs
- History of development through core Operations & Maintenance as well as via individual programs and project teams
- **Vetted** for use by all DOE national labs + HQ and other facilities
 - Endorsed by DOE OCIO, DOE Chief Counsel, and others
- Registered trademark awarded in 2020, recipient of the DOE Secretary of Energy Honor’s team award in 2021
- Over **3 million resources downloaded**; 2.5 million resources hosted



As of 7/25/2024

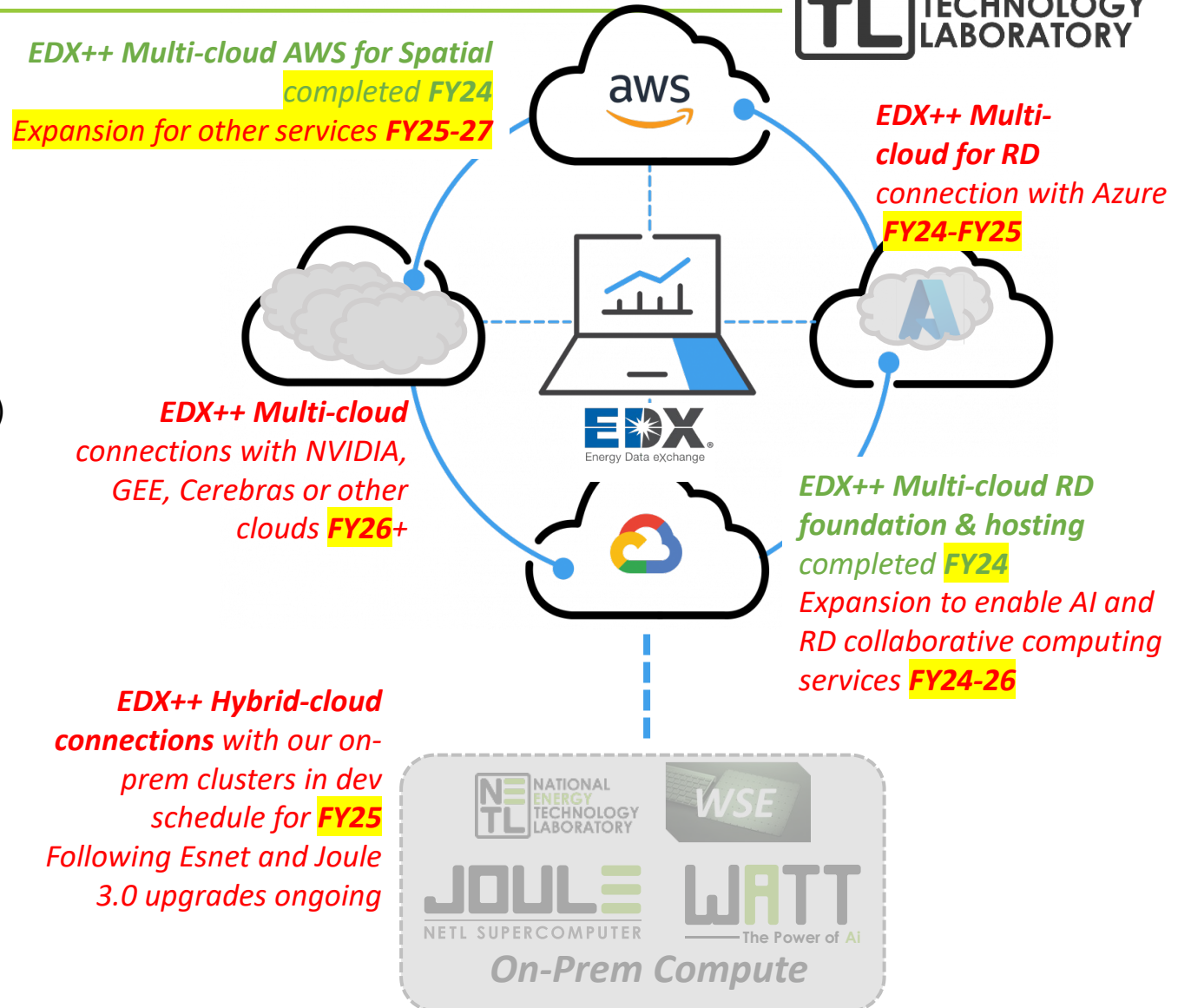


Registered Users: 4,216
Resource Count: 2,528,643
Resource Downloads: 3,129,206

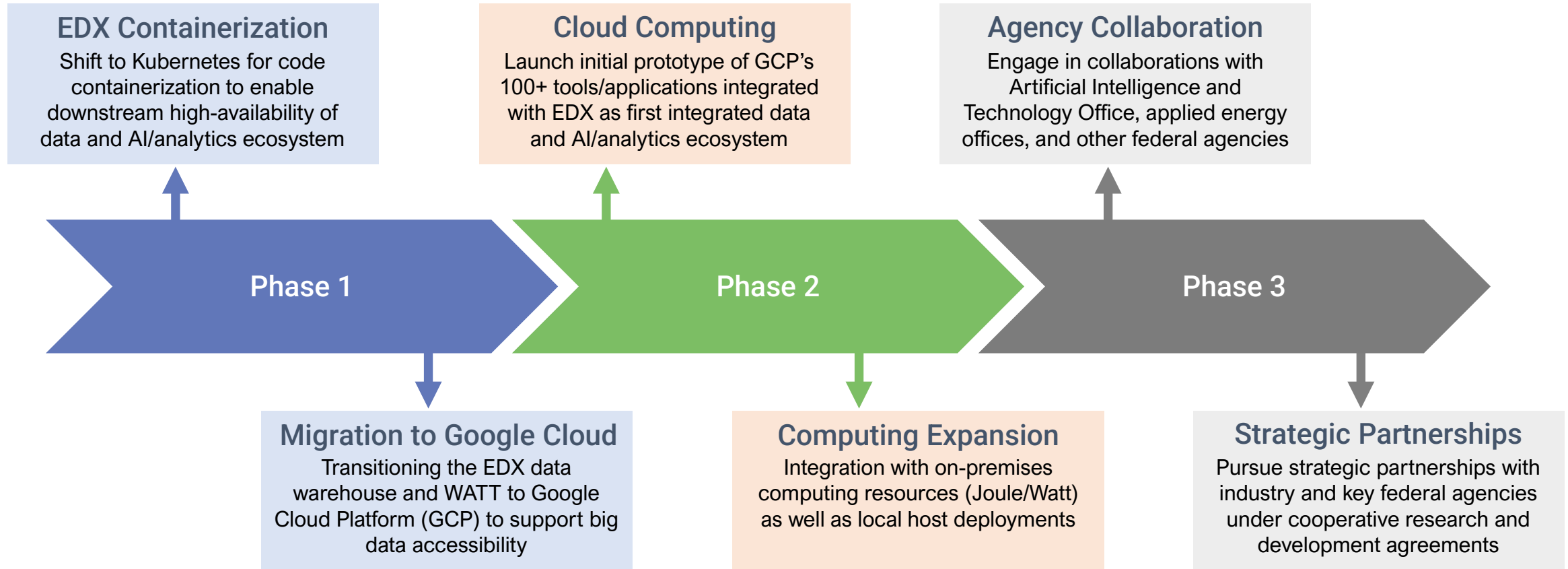
A solid, secure foundation, & next steps

EDX++ 4.0 went live March 18, 2024

- **An enterprise approach to scalable architecture, storage, and compute**
 - Enabling FECM/NETL and stakeholders to collaborate & innovate
 - Handling CUI through Public governance & capabilities
- **A virtual, digital solution for FECM RD that is:**
 - Federal cyber & multi-cloud strategy compliant
 - Supports & enables compliance with multiple executive orders for data, AI, and public digital infrastructure (see [EDX Ref Shelf](#))
- **Fully implemented TIC 3.0 (TIC = trusted internet connection)**
 - DOE's 1st TIC 3.0 compliant RD system
 - Upgrade from 1G shared network connection to 10G
 - Scalable to faster network connection speeds
- **Unlocked foundational multi-cloud capabilities**
 - Cloud compute via GCP, and AWS and Esri for initial spatial use cases
 - Cloud hosting of EDX via GCP
 - Including hosting via Google's Open Data Program of very large (TB and PB) NETL public products previously stranded
 - Cloud connectivity to other CSPs (i.e EDX Spatial on AWS)
 - Multi-cloud strategy enables option to transfer EDX++ data to/from other Cloud Service Providers (CSPs)



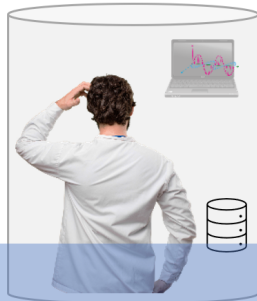
Building upon a foundation of success



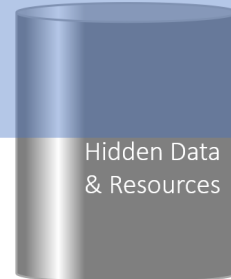
Why EDX++ Matters Today and Tomorrow

Connecting data to resources for analysis and computing for advanced collaboration & accelerated, scalable, secure innovation

EDX
Energy Data eXchange



Limited Access
Higher Costs
Redundant Efforts
Time-consuming Steps
Suboptimal Results



Siloed AI Environment
*Needless Duplication &
Frustration*

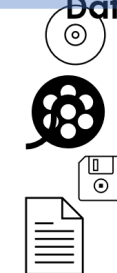
Private Data



Shared
Program
/Agency Data



Legacy
Data



Validated
AI Models



Advanced
Computer
Resources



Integrated AI Environment
*Interoperable & Seamless
Access*

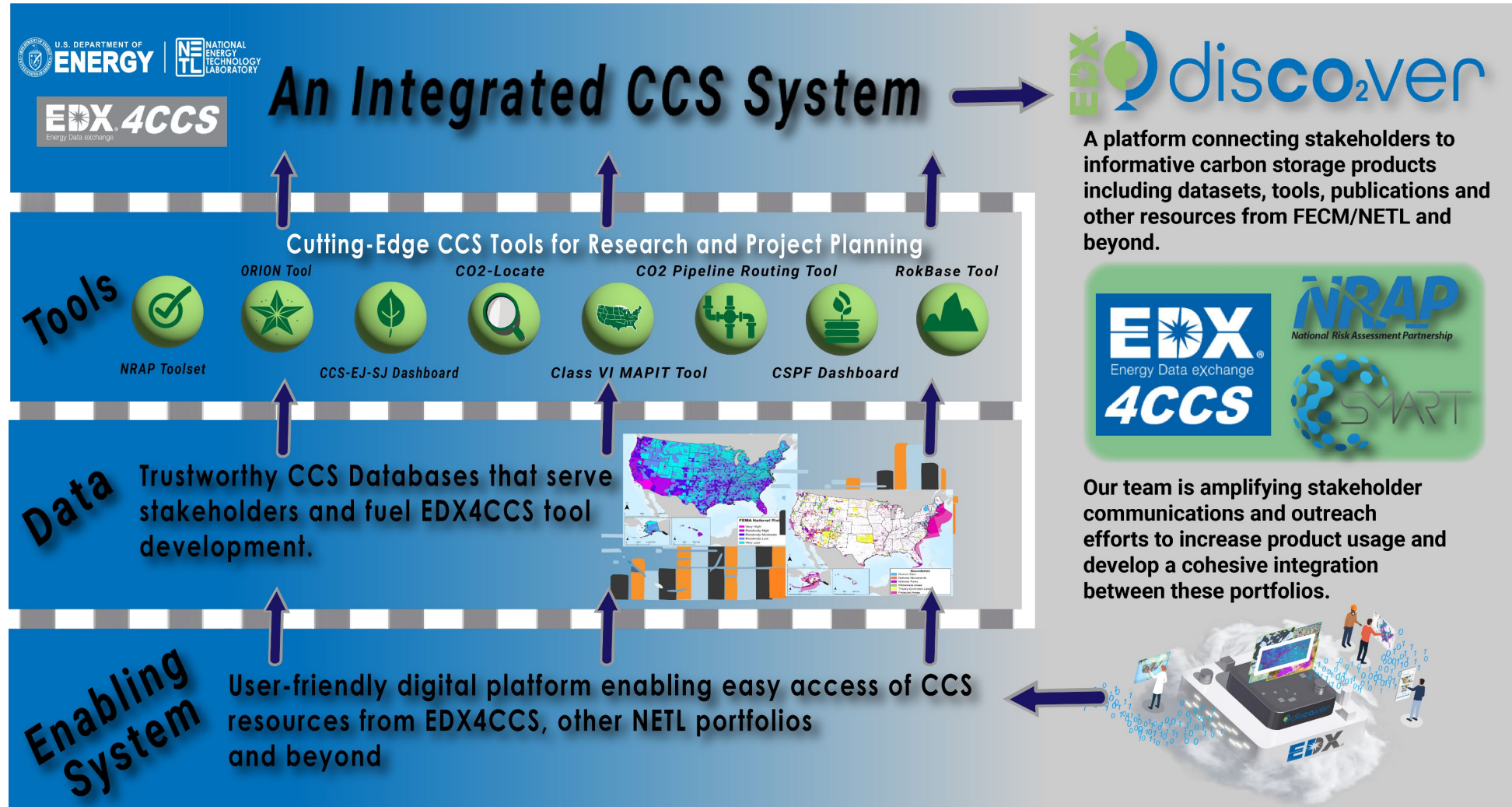


Faster
Cheaper
Improved
Outcomes

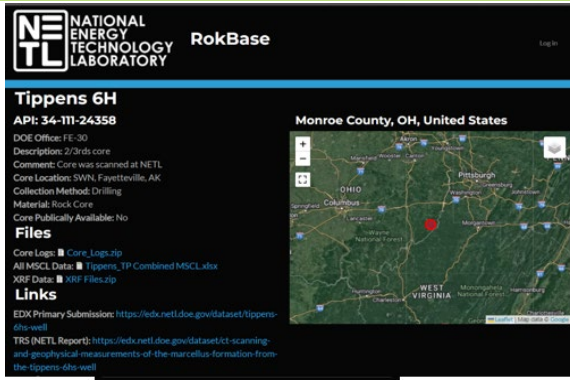
EDX++
Energy Data eXchange

Accelerating Research with Cloud Capabilities

Connecting the CCS Community to Authoritative Data Products



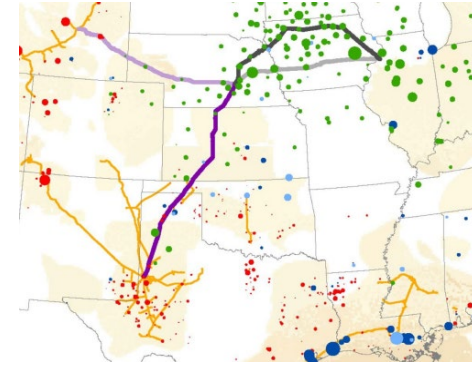
Actionable Data Science for CTS – Highlighted products



RokBase

Virtual data portal and tool

Explore & query available core and rock property data

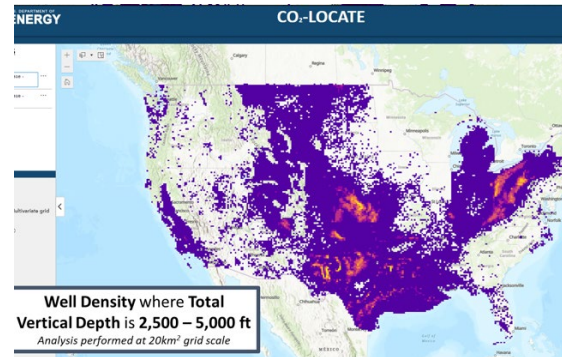


CTS Pipeline Planning

Database & Tool for the spatial routing of pipelines and transport of CO₂

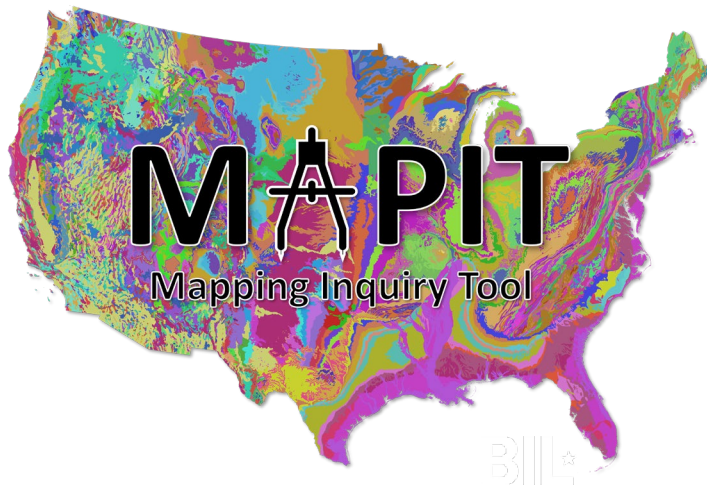
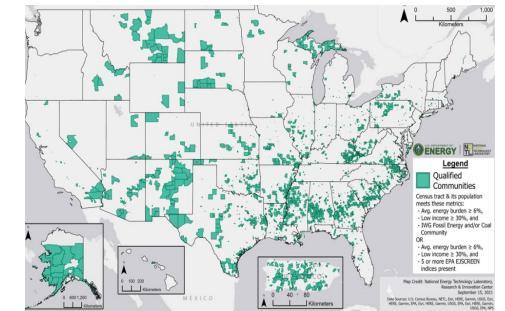
CO2-Locate

Tool & Database of National Well Information, for Reuse and Planning support



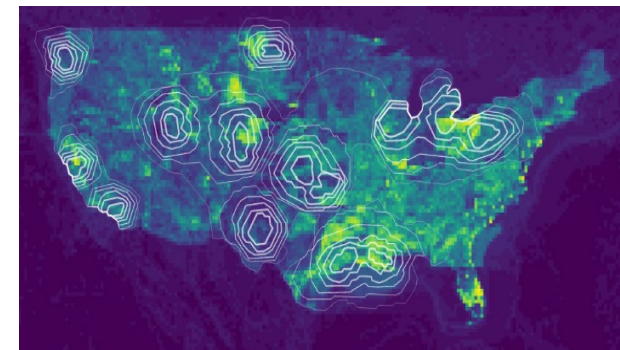
CTS EJ-SJ

Database & Web Tool for social & environmental justice CTS decision making



CTS MapIT

Geodatabase and planning support tool for Underground Injection Control (UIC) Class VI permits



CS Technical Viability Assess.

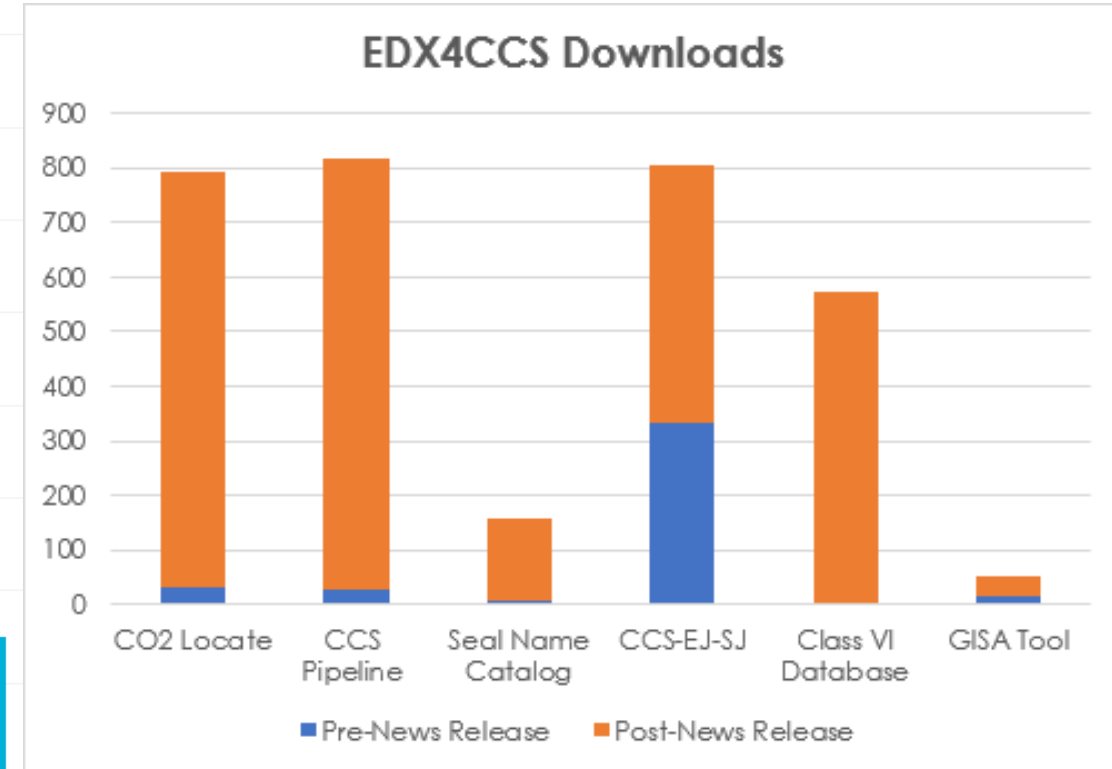
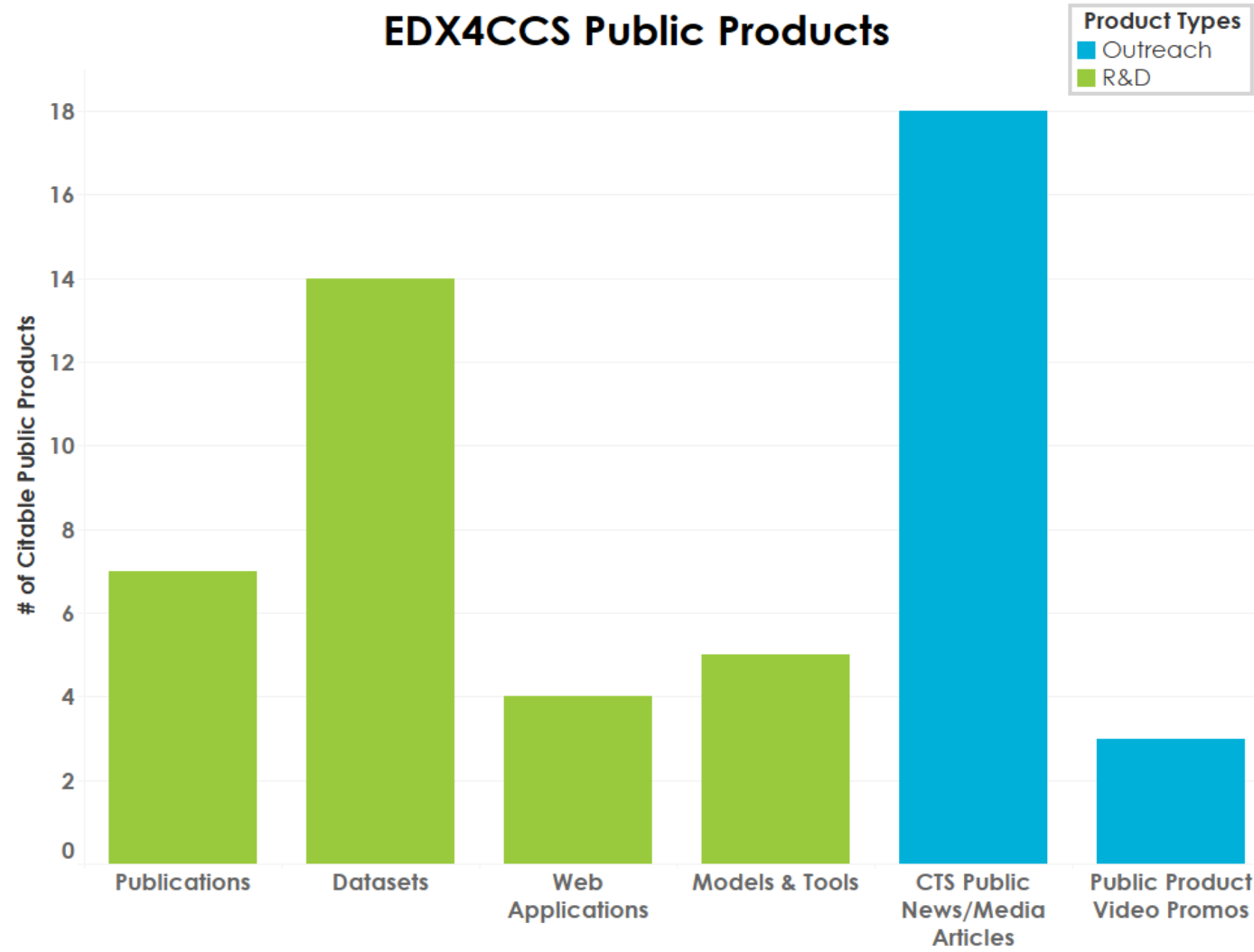
National geodatabase & interactive tool of open source geologic, geophysical, structural, hydrologic, energy extraction, transportation infrastructure, & environmental data

Actionable Data Science for CTS

2 years
30 CTS digital
products via EDX



EDX4CCS Public Products



The Future is Now

FECM is embracing research challenges with state-of-the-art solutions

- ✓ Evolving into a hybrid, multi-cloud solution
- ✓ Accelerating AI/ML
- ✓ Tackling data compute in the cloud and on-prem
- ✓ Improving transfer speed, security, and pipe



Questions?

Please contact EDXSupport@netl.doe.gov to begin your journey into the cloud.