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

*Catalyzing Collaboration & Accelerating Innovation*
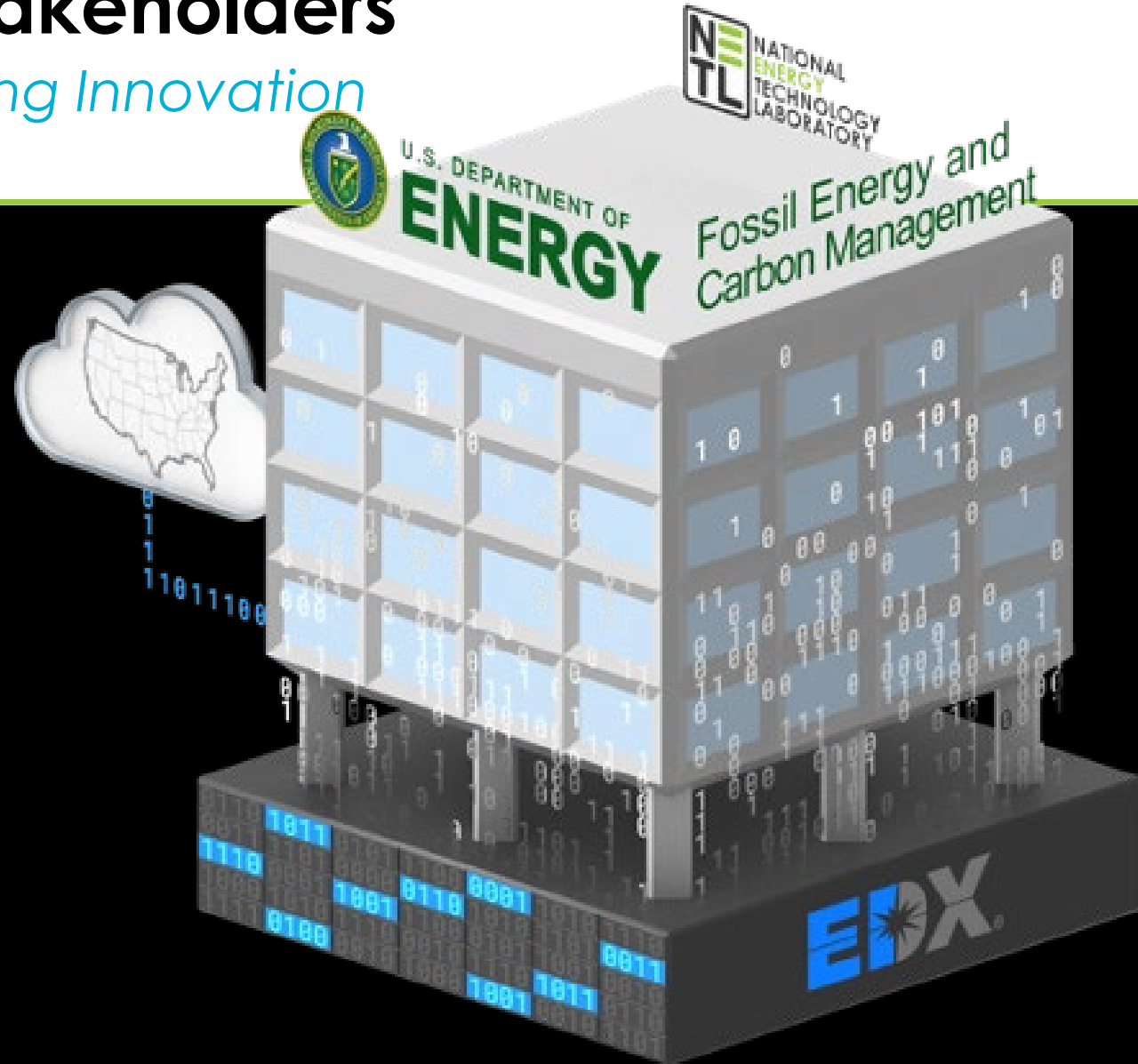
## EDX®
Energy Data eXchange

### Chad Rowan

National Energy Technology Laboratory (NETL)

Computational Science & Engineering

Advanced Computing & Artificial Intelligence

# Disclaimer
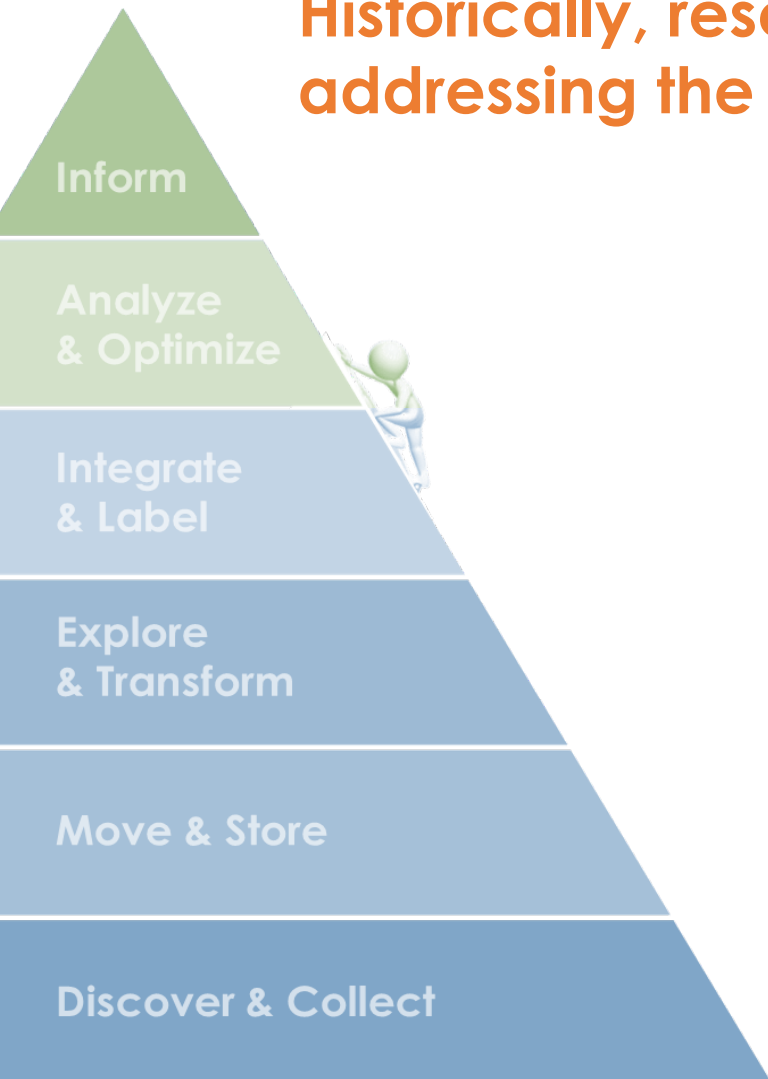
# Did you know?

**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?

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



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

**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*

**EDX** Energy Data eXchange ®

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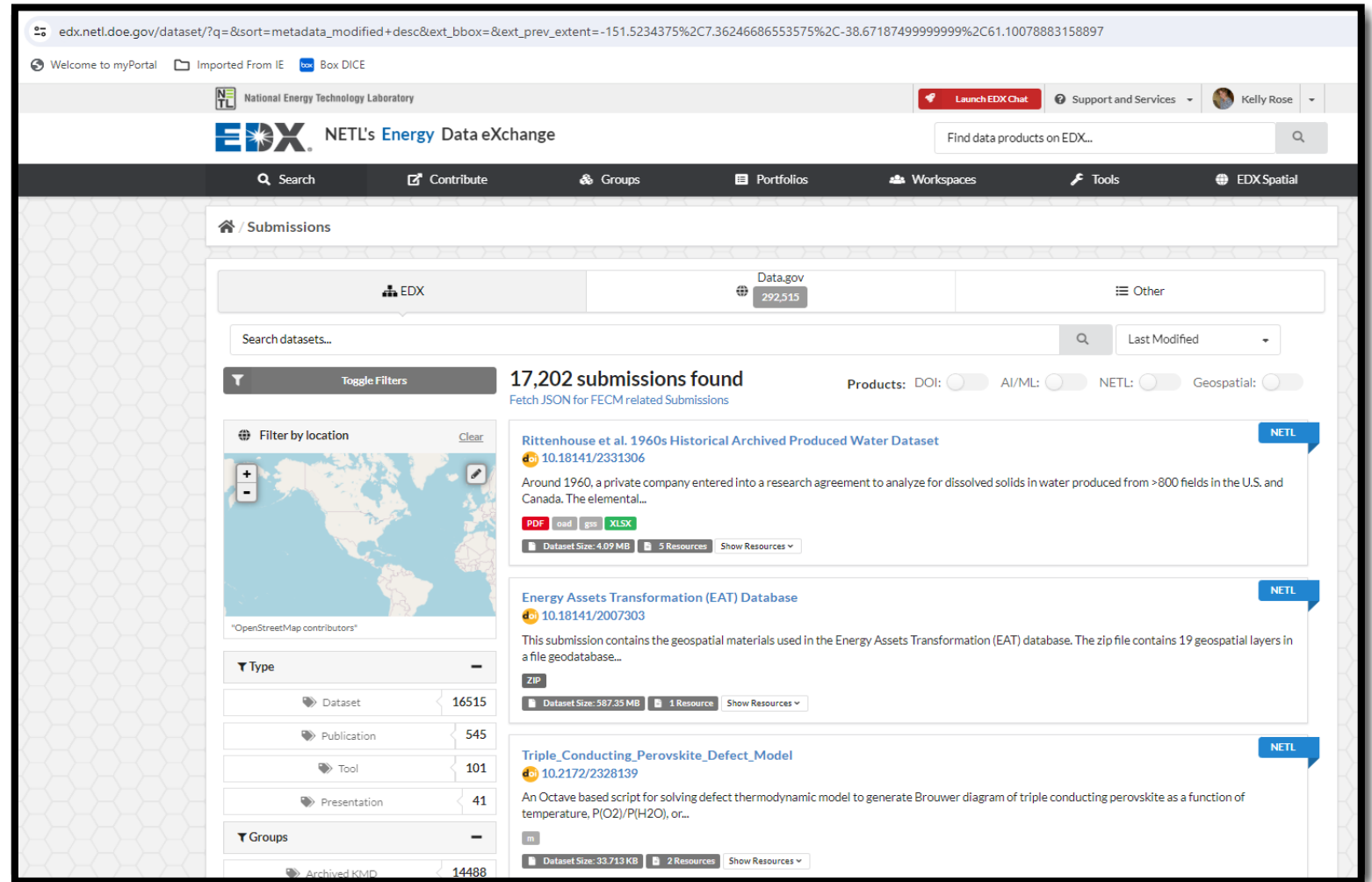


v.1 2012          12+ years          EDX++ v.4 2024

# How is EDX currently supporting its community?

| Preservation | Collaboration | Discoverability | Compliance | Management | Computing |
| --- | --- | --- | --- | --- | --- |
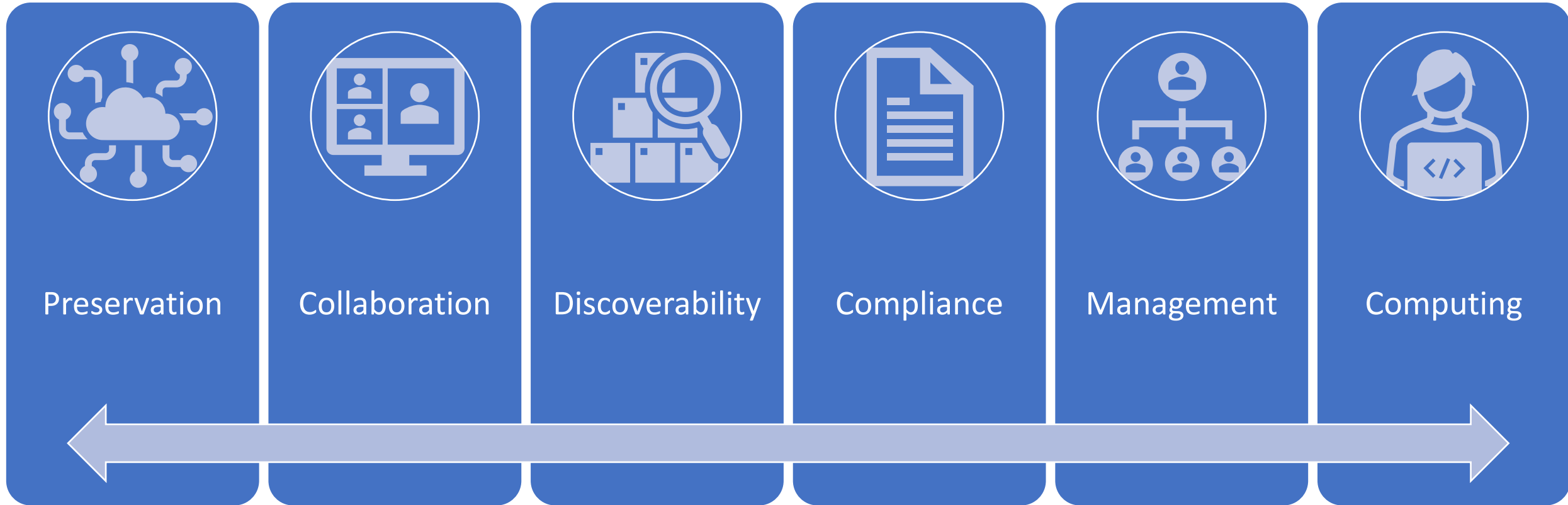
**NETL/FECM has invested in EDX to serve the FECM community as a virtual data library and laboratory**

# EDX Public Product Inventory



Wordcloud representing weight of tags/keywords in public submissions.

| Total Submissions | Total Resources | Approximate File Size (TB) |
|---|---|---|
| 17,268 | 33,533 | 324 |

**Number of Submissions by Type**

| Type | Count |
|---|---|
| Datasets | 16,557 |
| Research Documents | 550 |
| Tools | 120 |
| Presentations | 41 |
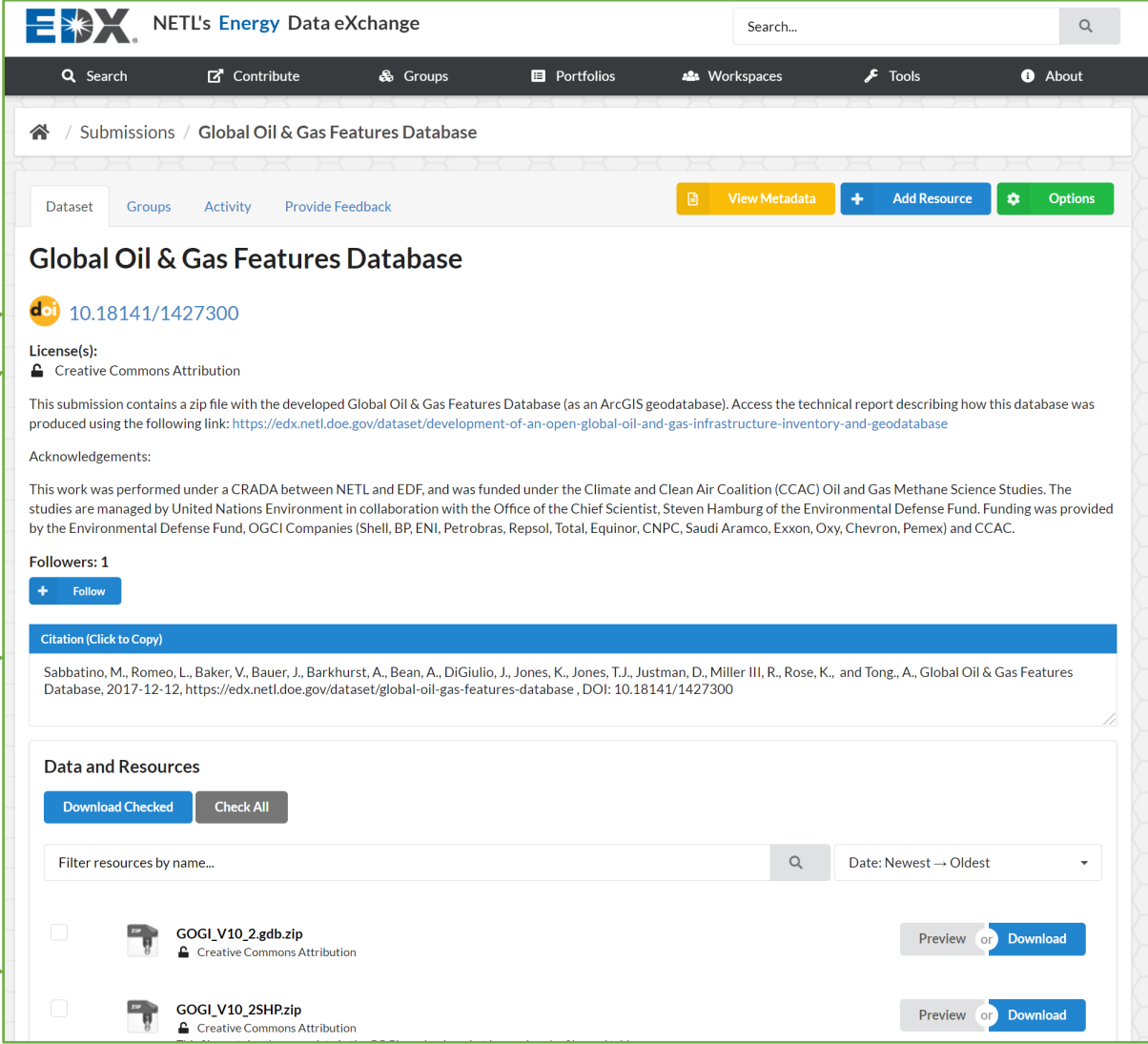
# Advantages of publishing data products



OSTI DOI Number

Data License

Data Citation

Data Access

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

*can help!*

# Who is part of the EDX community?

✓ NETL/FECM employees

✓ External collaborators

✓ General public
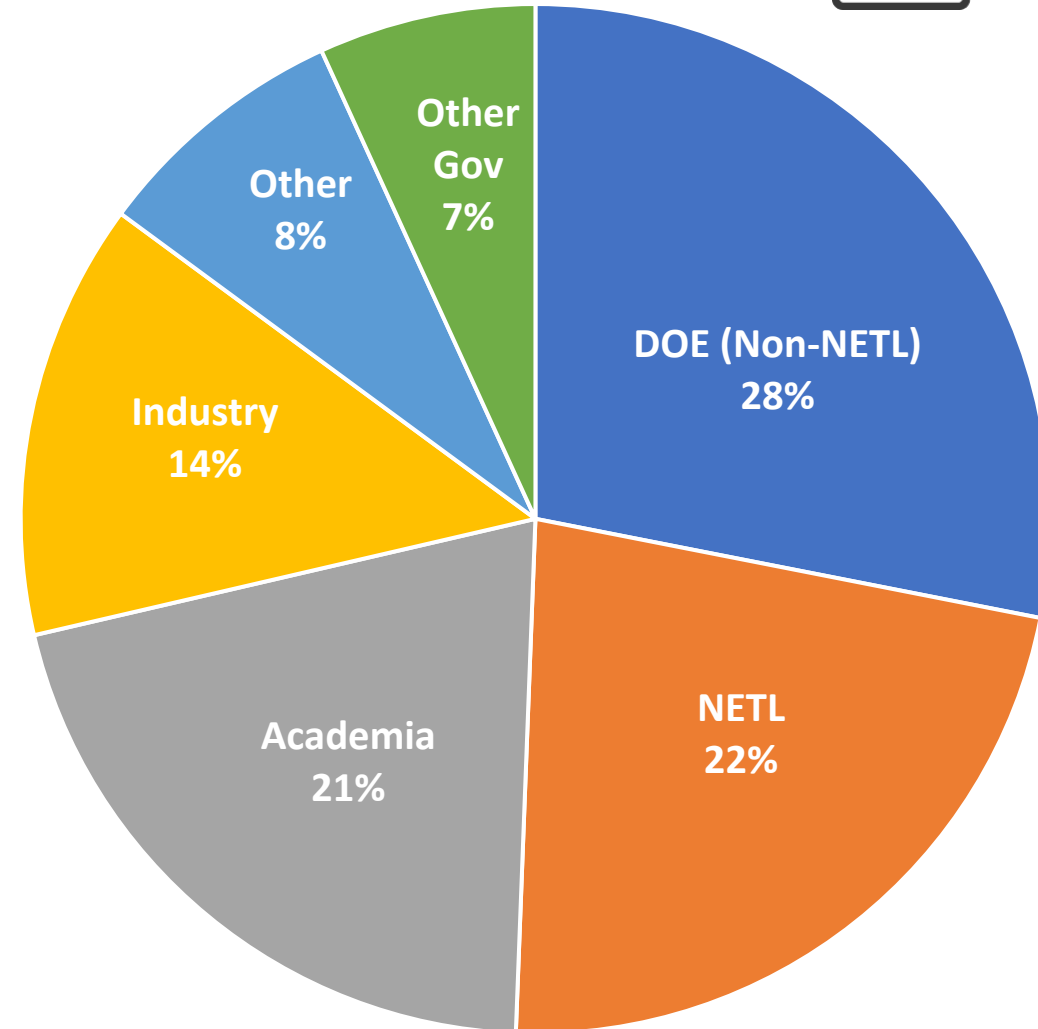
**2600+ Active Registrants**



Pie chart:
- DOE (Non-NETL) 28%
- NETL 22%
- Academia 21%
- Industry 14%
- Other 8%
- Other Gov 7%

# 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
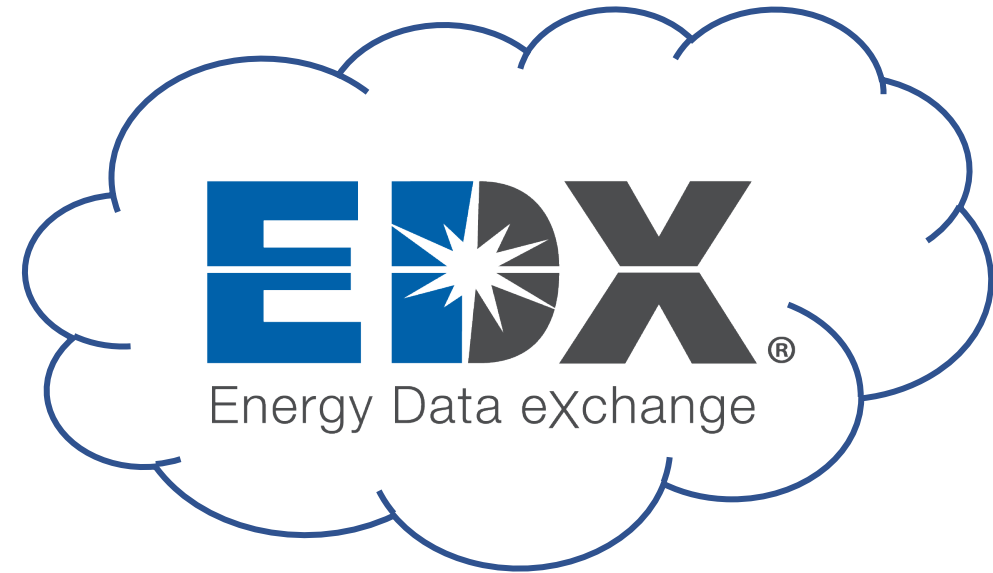
# Why migrate to the Cloud?



## 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 <mark>use since in 2011</mark>**

- 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
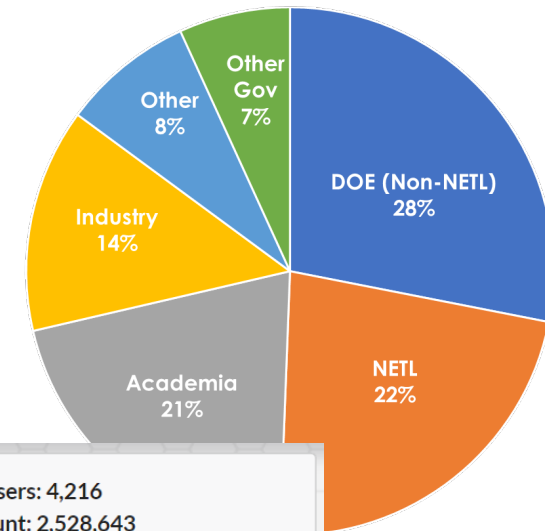
- Public side hosts and serves as an **internationally registered and recognized** data repository.
  - Curating individual data products coupled with Datacite.org citation standard, as well as connections to multi-cloud virtualized interoperability features (e.g. APIs, analytics, Cloud services, etc)

- **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 <u>all</u> 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**

Pie chart:
- DOE (Non-NETL) 28%
- NETL 22%
- Academia 21%
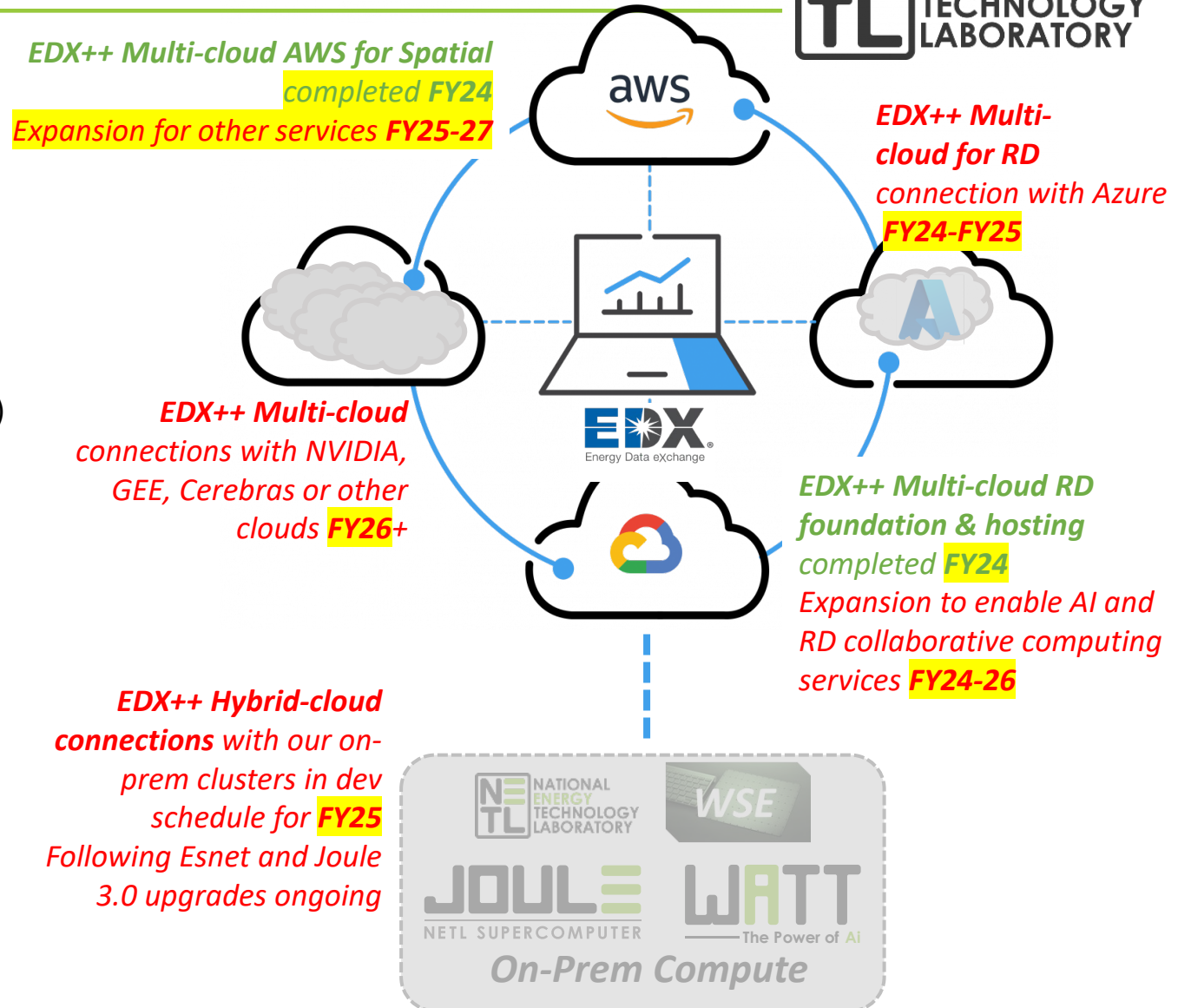- Industry 14%
- Other 8%
- Other Gov 7%

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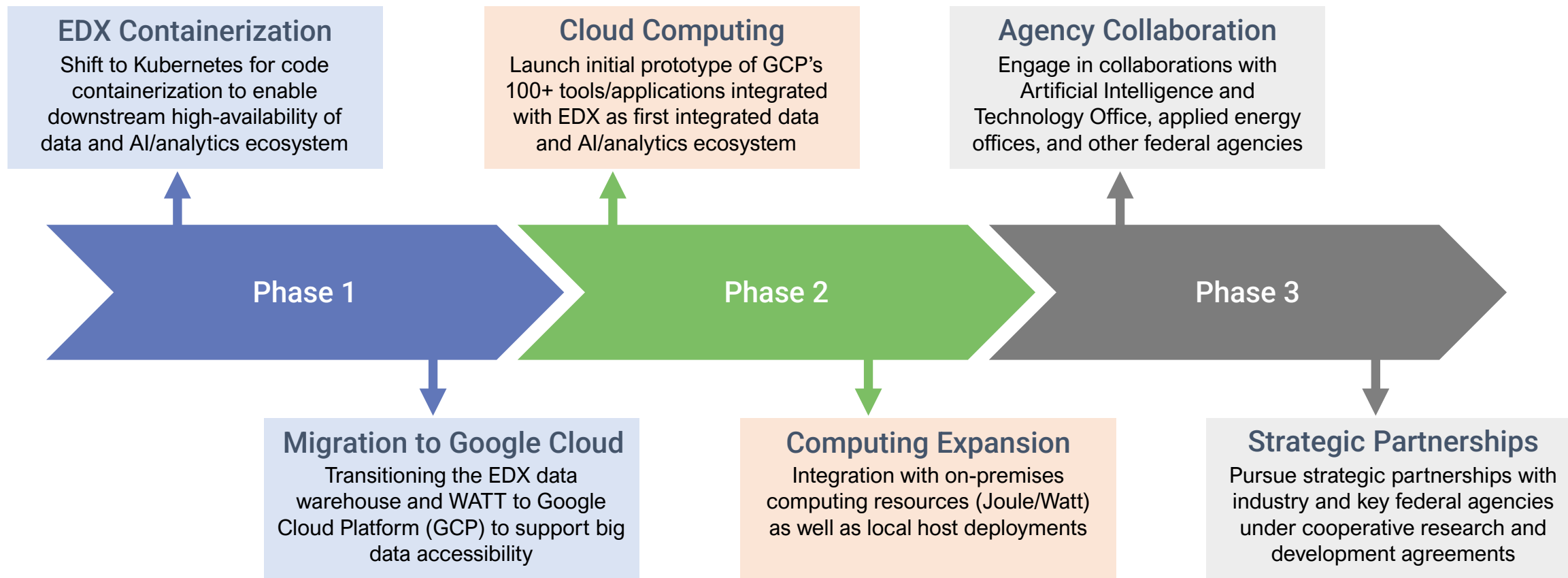Resource Downloads: 3,129,206

# 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)



*EDX++ Multi-cloud AWS for Spatial completed FY24*
*Expansion for other services FY25-27*

*EDX++ Multi-cloud for RD connection with Azure FY24-FY25*

*EDX++ Multi-cloud connections with NVIDIA, GEE, Cerebras or other clouds FY26+*

*EDX++ Multi-cloud RD foundation & hosting completed FY24*
*Expansion to enable AI and RD collaborative computing services FY24-26*

*EDX++ Hybrid-cloud connections with our on-prem clusters in dev schedule for FY25 Following Esnet and Joule 3.0 upgrades ongoing*

*On-Prem Compute*

# Building upon a foundation of success

**EDX Containerization**
Shift to Kubernetes for code containerization to enable downstream high-availability of data and AI/analytics ecosystem

**Cloud Computing**
Launch initial prototype of GCP's 100+ tools/applications integrated with EDX as first integrated data and AI/analytics ecosystem

**Agency Collaboration**
Engage in collaborations with Artificial Intelligence and Technology Office, applied energy offices, and other federal agencies

**Phase 1**

**Phase 2**

**Phase 3**

**Migration to Google Cloud**
Transitioning the EDX data warehouse and WATT to Google Cloud Platform (GCP) to support big data accessibility

**Computing Expansion**
Integration with on-premises computing resources (Joule/Watt) as well as local host deployments

**Strategic Partnerships**
Pursue strategic partnerships with industry and key federal agencies under cooperative research and development agreements
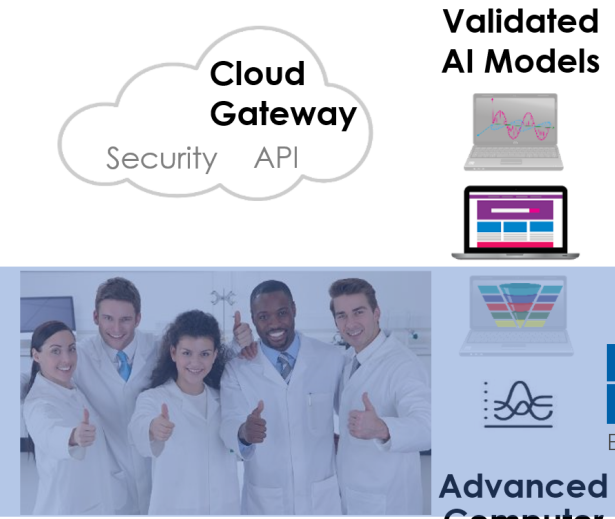
# Why EDX++ Matters Today and Tomorrow

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



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

Hidden Data & Resources

EDX® Energy Data eXchange

Private Data

Shared Program /Agency Data

Legacy Data

Cloud Gateway
Security    API

Validated AI Models

Advanced Computer Resources

EDX++® Energy Data eXchange

Faster
Cheaper
Improved
Outcomes

**Siloed AI Environment**
*Needless Duplication & Frustration*
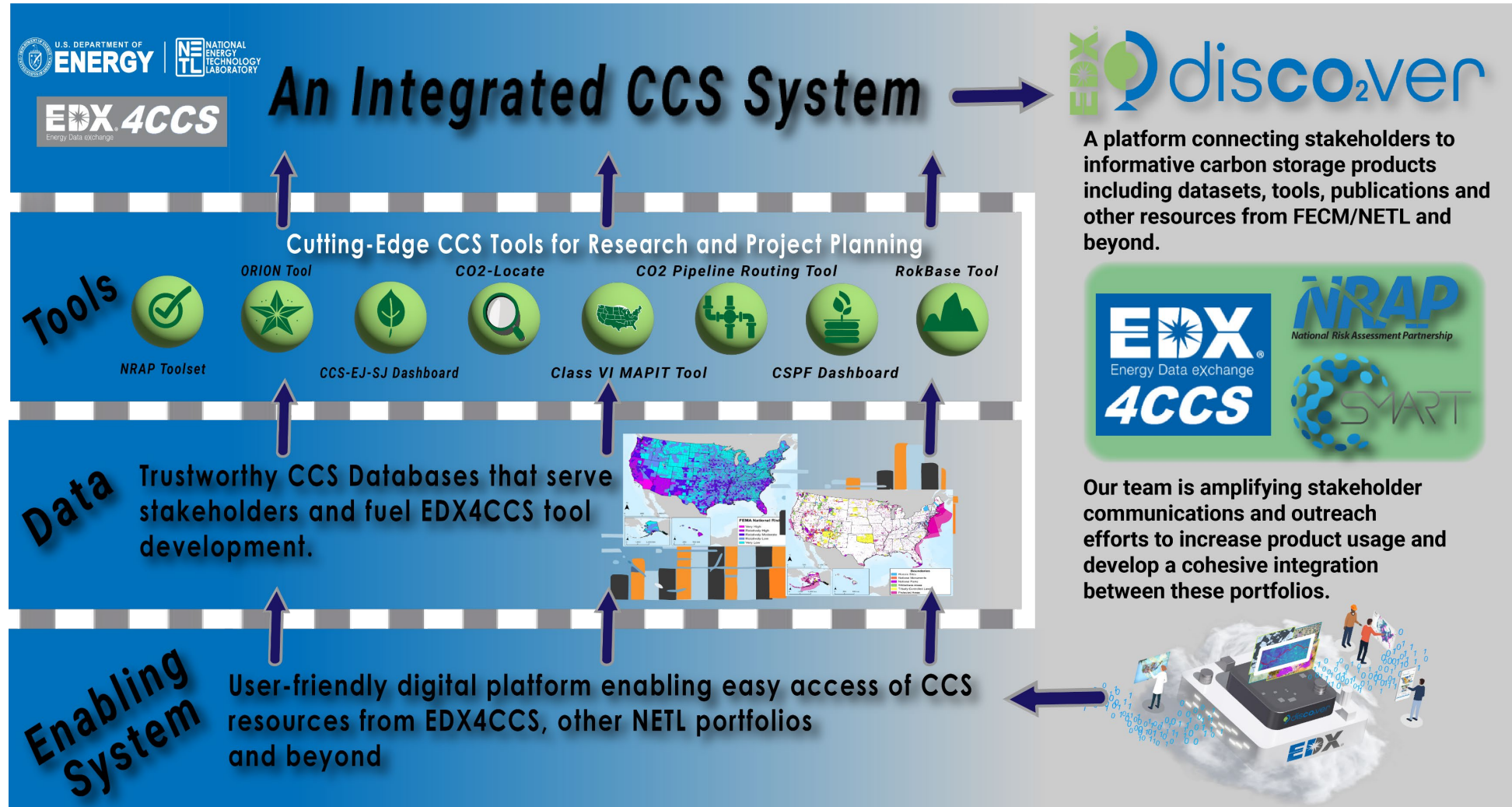
**Integrated AI Environment**
*Interoperable & Seamless Access*

# Accelerating Research with Cloud Capabilities

## Connecting the CCS Community to Authoritative Data Products



An Integrated CCS System

EDX disco₂ver

A platform connecting stakeholders to informative carbon storage products including datasets, tools, publications and other resources from FECM/NETL and beyond.

**Tools**

Cutting-Edge CCS Tools for Research and Project Planning

ORION Tool    CO2-Locate    CO2 Pipeline Routing Tool    RokBase Tool

NRAP Toolset    CCS-EJ-SJ Dashboard    Class VI MAPIT Tool    CSPF Dashboard

**Data**

Trustworthy CCS Databases that serve stakeholders and fuel EDX4CCS tool development.

Our team is amplifying stakeholder communications and outreach efforts to increase product usage and develop a cohesive integration between these portfolios.

**Enabling System**

User-friendly digital platform enabling easy access of CCS resources from EDX4CCS, other NETL portfolios and beyond

# Actionable Data Science for CTS – *Highlighted* products



## RokBase

**Virtual data portal and tool**

**Explore & query available core and rock property data**

## CO2-Locate

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



**Well Density** where **Total Vertical Depth** is 2,500 – 5,000 ft
*Analysis performed at 20km² grid scale*

## CTS MapIT

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

## CTS Pipeline Planning

**Database & Tool** for the spatial routing of pipelines and transport of $CO_2$

## CTS EJ-SJ

**Database & Web Tool** for social & environmental justice CTS decision making
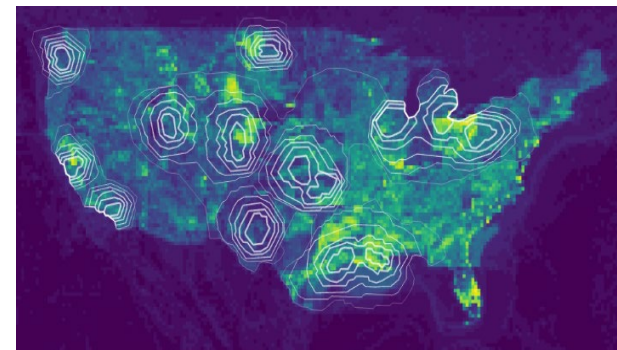
## CS Technical Viability Assess.

**National geodatabase & interactive tool** of open source geologic, geophysical, structural, hydrologic, energy extraction, transportation infrastructure, & environmental data
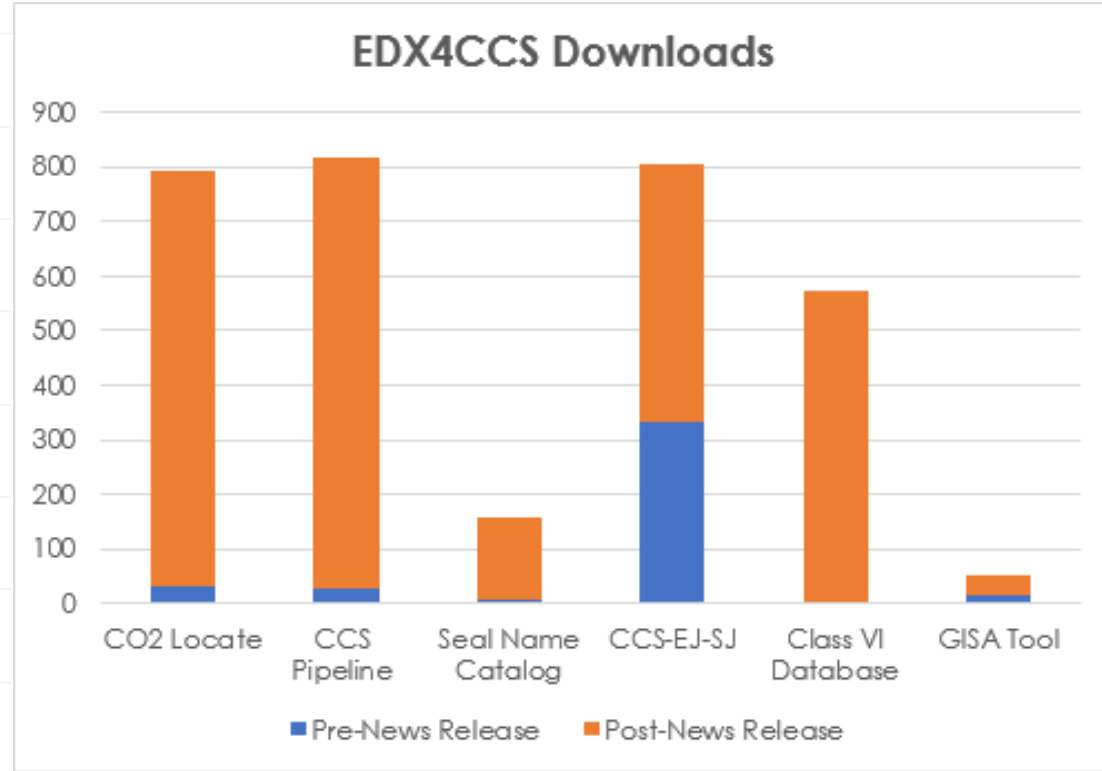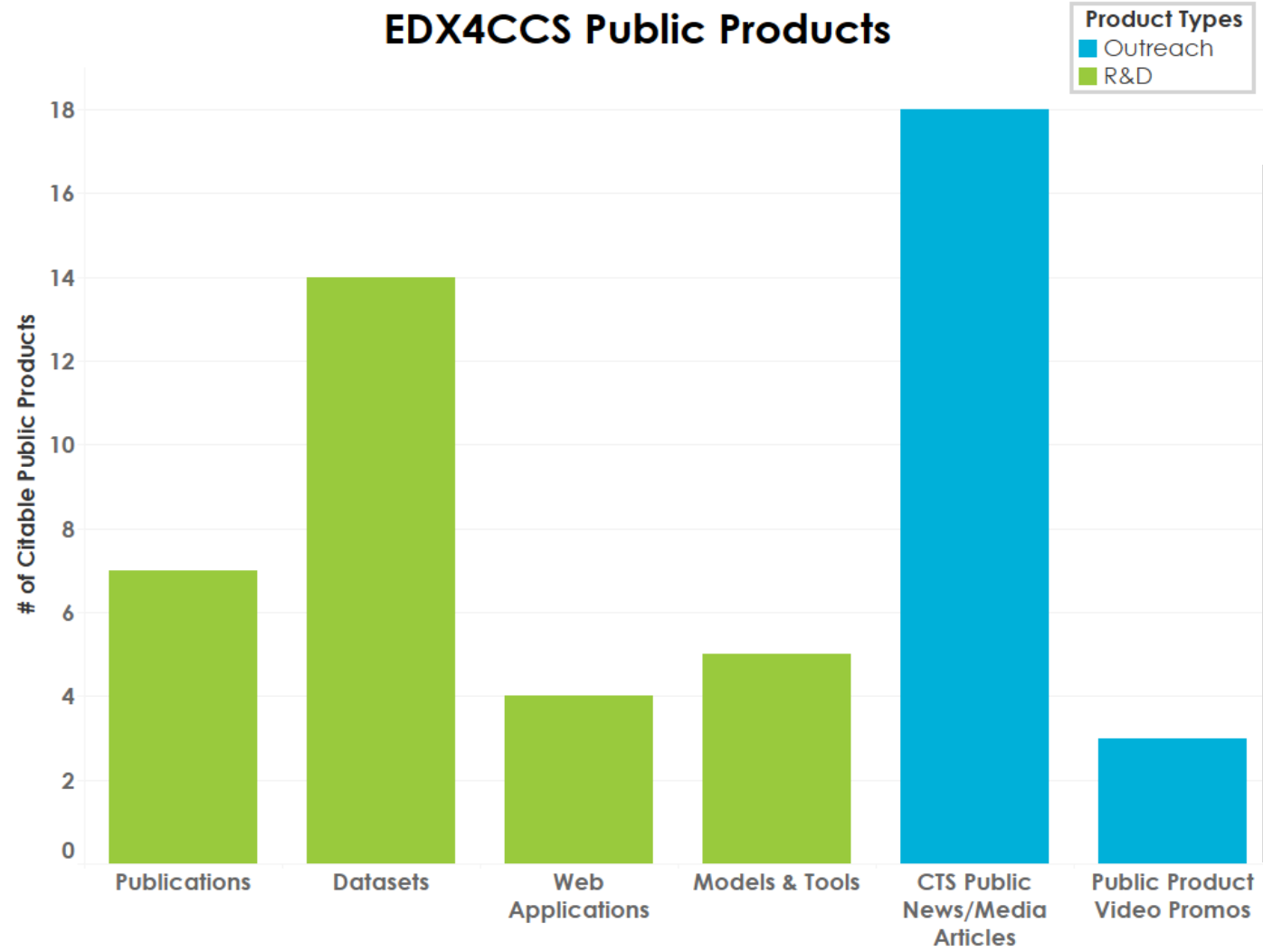
# Actionable Data Science for CTS

EDX4CCS Public Products

Product Types
- Outreach
- R&D



EDX4CCS Downloads

Pre-News Release | Post-News Release

# 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](mailto:EDXSupport@netl.doe.gov) to begin your journey into the cloud.**