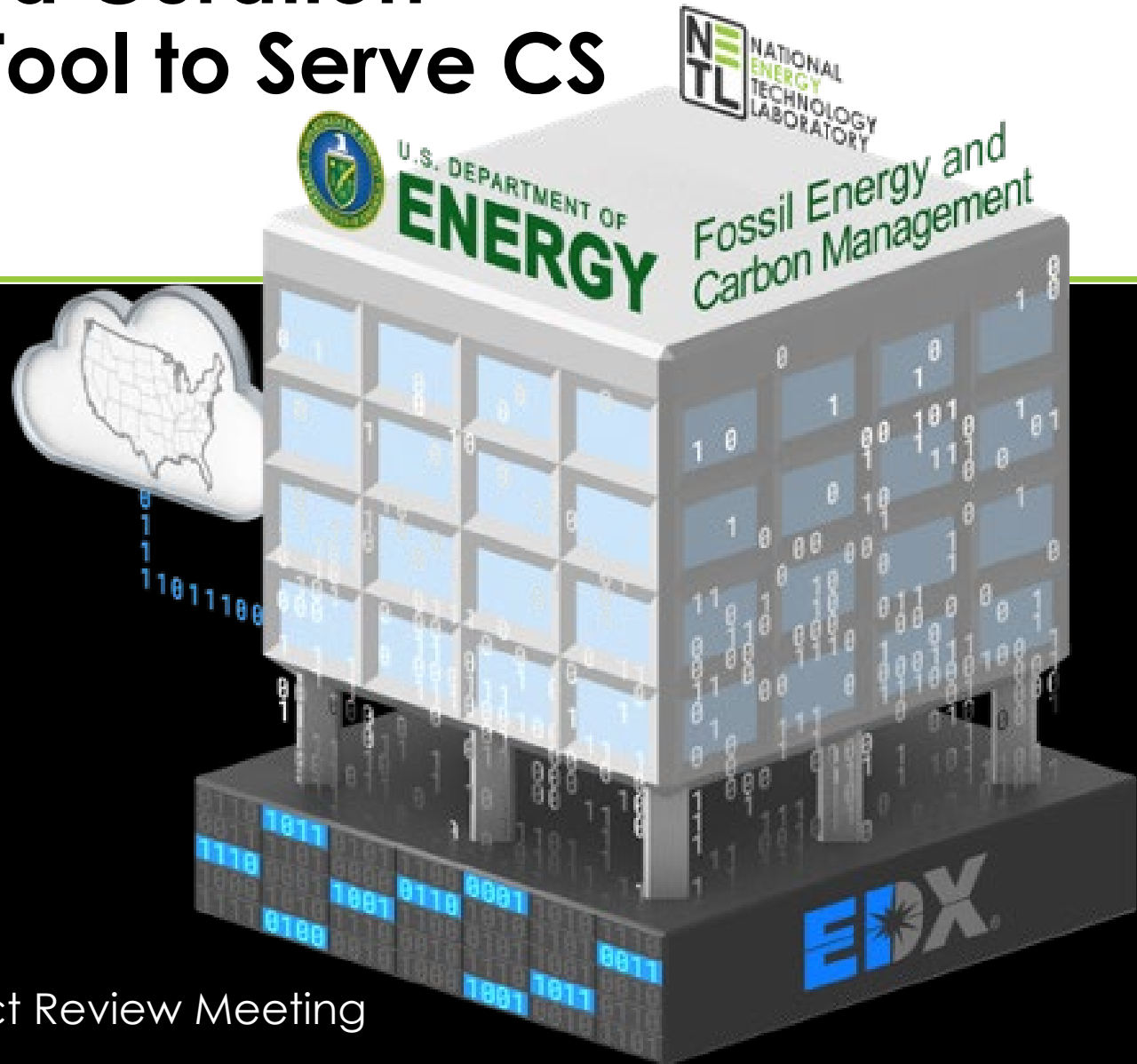


Understanding Federal Data Curation Requirements and EDX++ Tool to Serve CS Data Curation Needs



Chad Rowan, RIC, NETL MGN
Jessica Sinclair, Leidos, NETL MGN
Kelly Rose, RIC, NETL ALB

FECM/NETL Carbon Management Research Project Review Meeting
Aug. 28, 2023



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

Authors and Contact Information



Chad Rowan¹ ; Jessica Sinclair^{2,3}; Kelly Rose⁴

¹National Energy Technology Laboratory, 3610 Collins Ferry Road, Morgantown, WV 26505, USA

²NETL Support Contractor, 626 Cochran Mill Road, Pittsburgh, PA 15236

³National Energy Technology Laboratory, 626 Cochran Mill Road, Pittsburgh, PA 15236

⁴National Energy Technology Laboratory, 1450 Queen Avenue SW, Albany, OR 97321, USA

What is the DOE Public Access Plan (June 2023)?

Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

- Provide free, immediate access to peer-reviewed, scholarly publications;
- Provide immediate access to scientific data displayed in or underlying publications and increased access to other data;
- Use persistent identifiers (PIDs) for research outputs, researchers, organizations, and awards.

Replaces the DOE Public Access Plan of 2014



- **When will requirements of the 2023 Plan go into effect?**
 - new policy and implementation guidance will be issued by December 31, 2024
 - full implementation required by December 31, 2025
- **What data needs to be made publicly available?**
 - unclassified and otherwise unrestricted digital scientific data arising from research and development (R&D) activities funded by DOE
- **Will there be additional funding provided to pay for public access to publications and data?**
 - Reasonable publication fees and data management expenses are allowable costs of an award or contract and can be included in proposed budget requests



Additional Federal Data Curation Requirements

What are they? How do I stay compliant?

- [DOE Public Access Plan](#) (updated June 2023)
- [Digital Accountability and Transparency Act \(DATA\) of 2014](#)
- [Geospatial Data Act of 2018](#)
- [Foundations for Evidence-Based Policymaking Act of 2018](#)
- [Maintaining American Leadership in Artificial Intelligence \(EO 13859\)](#) (February 2019)
- [Memorandum From ASFECM Brad Crabtree \(Aug 2022\)](#)
- And others...



New guidance for FECM funded research



IRM 31452

Department of Energy

Washington, DC 20585

August 22, 2022

MEMORANDUM FOR FOSSIL ENERGY AND CARBON MANAGEMENT DEPUTY ASSISTANT SECRETARIES AND DIRECTOR, NATIONAL ENERGY TECHNOLOGY LABORATORY

FROM: BRAD CRABTREE *Brad Crabtree*
ASSISTANT SECRETARY
FOSSIL ENERGY AND CARBON MANAGEMENT

SUBJECT: Require all new and existing Funding Opportunity Announcements, Field Work Proposals, and other procurement vehicles to include a statement that calls for all resulting data products to be submitted to Fossil Energy and Carbon Management and National Energy Technology Laboratory

Organizations and governments are moving rapidly to harness data-driven artificial intelligence and machine learning. Fossil Energy and Carbon Management (FECM) is already using these tools to significantly minimize the environmental impacts of fossil fuels and smooth the transition toward net-zero carbon emissions; however, FECM research teams must often spend inordinate time and effort to create, locate, or procure the datasets they need. Requiring FECM research partners to submit to FECM and the National Energy Technology Laboratory (NETL) the data products from their funded research would help future researchers rapidly identify and access needed datasets, reducing costs, and accelerating the development of critical technologies.

Currently, research partners in many of our programs are not required to provide FECM and NETL with the data and information products that are generated specifically by or for their Department of Energy (DOE) funded research. This situation forces subsequent projects that might otherwise benefit greatly from that existing data to conduct lengthy searches of other sources or generate it anew.

Effective immediately, all Funding Opportunity Announcements (FOAs), Field Work Proposals (FWPs), and other procurement vehicles will include a data requirement statement unless the responsible Deputy Assistant Secretary (DAS) provides justification for not doing so. The data requirement statement can be tailored to individual program needs, but it should be based on the attached statement reviewed by FECM Headquarters offices and NETL. This requirement pertains to the following:

1. All *new* FOAs, FWPs, and other procurement vehicles.
2. All FOAs, FWPs, and other procurement vehicles that are currently *posted but not yet awarded*.

In addition, at the discretion of the responsible DAS, programs will assess the cost and feasibility of negotiating the addition of this data requirement to previously awarded projects.

An official copy of the memo
can be found in the
SAMI workspace on EDX

<http://edx.netl.doe.gov/resource/39862dd8-4444-4784-8fcf-cdef81b22282/download>

What guidance is provided in the memo?

All data products generated as a result of DOE-funded R&D are to be provided to FECM and NETL

What does the guidance impact?

- All **new** FOAs, FWPs, and other procurement vehicles

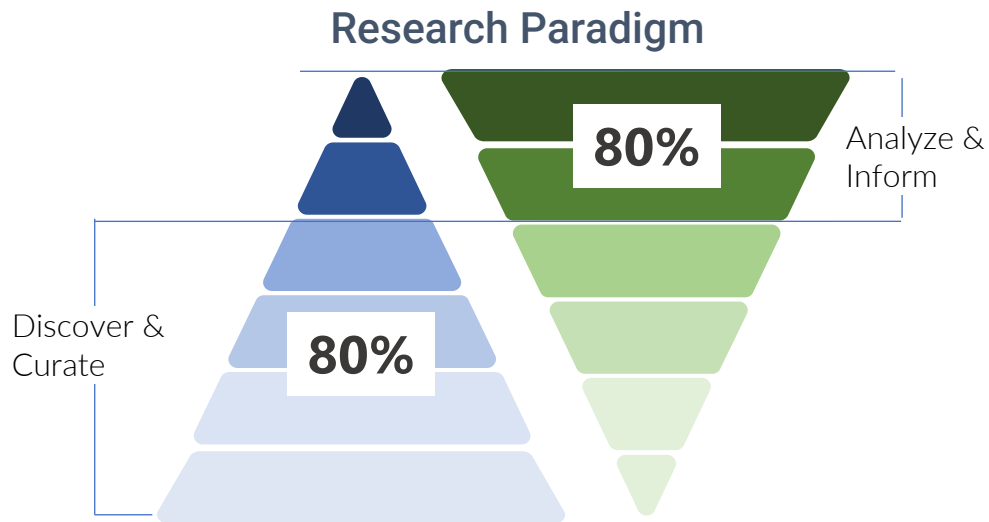
Who does the guidance impact?

- NETL/FECM funded researchers
- Project performers
- Approving managers
- Future research teams

How is EDX addressing these data curation requirements?

Inverting the Data Pyramid

Studies have shown that researchers spend **80% of their time** getting data to a point where they can do something with it



EDX helps teams with compliancy, often without them even knowing it



- Serves as a **data warehouse** for **published data submissions**, allowing users to upload/download data
- Provides a **private sharing environment** facilitates **collaboration** and **curation**
- Provides automated data **searching at scale using AI** via **SmartSearch**
- Offers a **built-in publishing workflow** to allow public access to appropriate data
- Empowers researchers with a **Cyber/IA and cloud-compliant data and analytics platform**
- *In Development*—Enable **access to data, analytics tools, and on-prem/cloud-based computing resources** in a single location

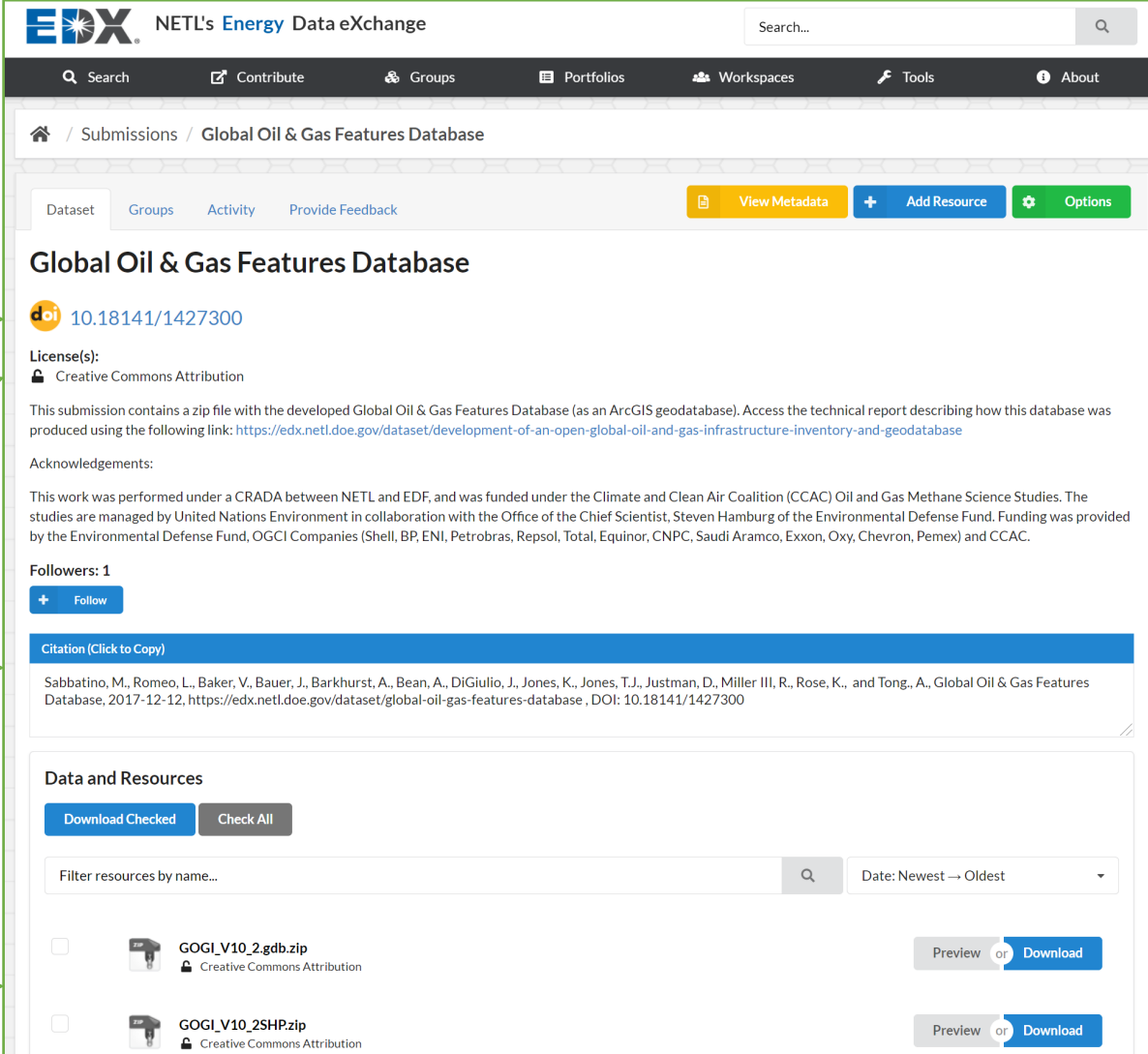
Advantages of publishing data products

OSTI DOI Number →

Data License →

Data Citation →

Data Access →



EDX. NETL's Energy Data eXchange

Search... [Search Icon]

Search Contribute Groups Portfolios Workspaces Tools About

Submissions / Global Oil & Gas Features Database

Dataset Groups Activity Provide Feedback View Metadata Add Resource Options

Global Oil & Gas Features Database

doi 10.18141/1427300

License(s):
Creative Commons Attribution

This submission contains a zip file with the developed Global Oil & Gas Features Database (as an ArcGIS geodatabase). Access the technical report describing how this database was produced using the following link: <https://edx.netl.doe.gov/dataset/development-of-an-open-global-oil-and-gas-infrastructure-inventory-and-geodatabase>

Acknowledgements:

This work was performed under a CRADA between NETL and EDF, and was funded under the Climate and Clean Air Coalition (CCAC) Oil and Gas Methane Science Studies. The studies are managed by United Nations Environment in collaboration with the Office of the Chief Scientist, Steven Hamburg of the Environmental Defense Fund. Funding was provided by the Environmental Defense Fund, OGC Companies (Shell, BP, ENI, Petrobras, Repsol, Total, Equinor, CNPC, Saudi Aramco, Exxon, Oxy, Chevron, Pemex) and CCAC.

Followers: 1
Follow



Citation (Click to Copy)

Sabbatino, M., Romeo, L., Baker, V., Bauer, J., Barkhurst, A., Bean, A., DiGiulio, J., Jones, K., Jones, T.J., Justman, D., Miller III, R., Rose, K., and Tong, A., Global Oil & Gas Features Database, 2017-12-12, <https://edx.netl.doe.gov/dataset/global-oil-gas-features-database>, DOI: 10.18141/1427300

Data and Resources

Download Checked Check All

Filter resources by name... [Search Icon] Date: Newest → Oldest

<input type="checkbox"/>	 GOGI_V10_2.gdb.zip Creative Commons Attribution	Preview or Download
<input type="checkbox"/>	 GOGI_V10_2SHP.zip Creative Commons Attribution	Preview or Download

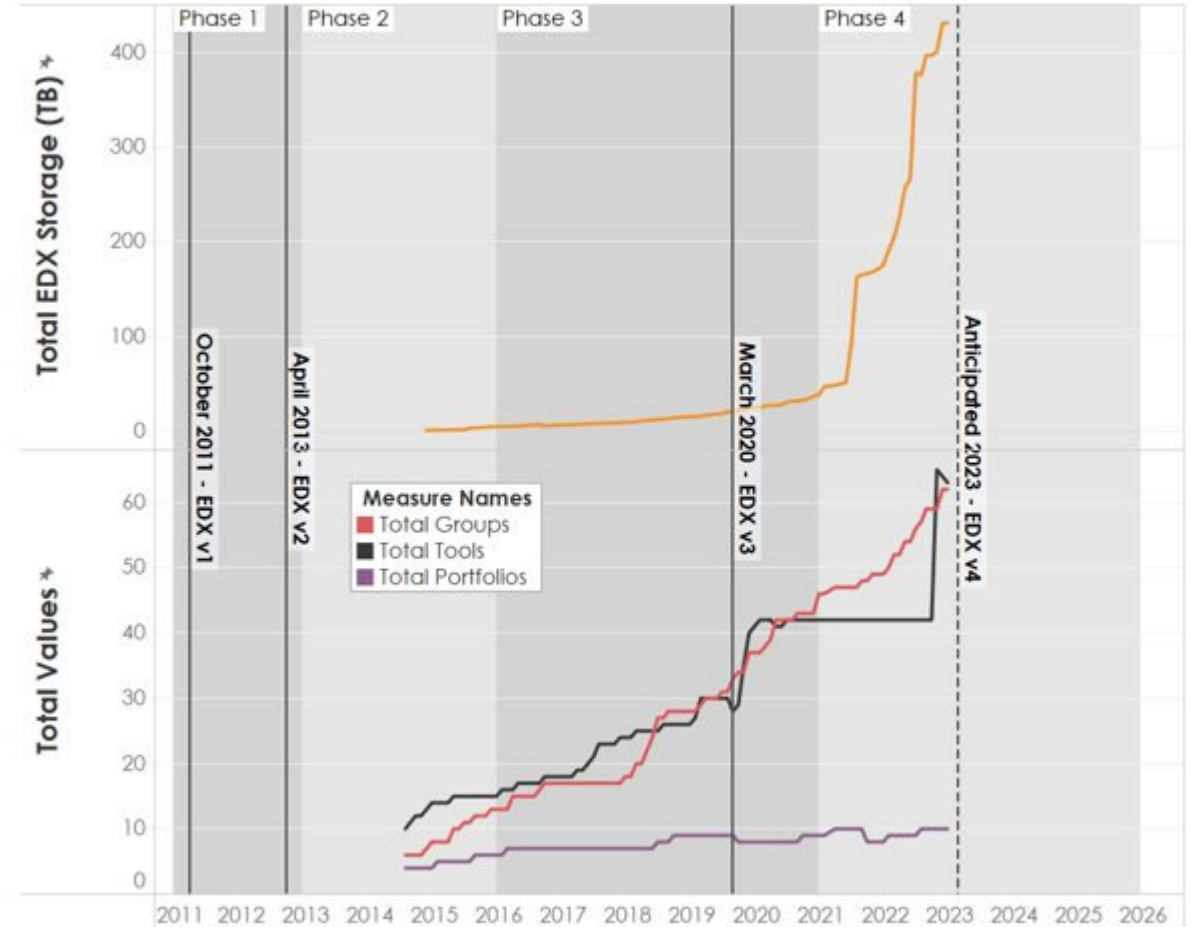
Many journals require models, tools and data to be publicly available prior to journal publication.



can help!

What is curated on EDX?

- **430+** TBs of data
- **60+** published groups
- **60+** published tools
- **10+** research portfolio websites



OVER 34 MILLION DATA RESOURCES!!!

How is EDX supporting its community?

CURRENT STATE

PUBLIC DATA PRESERVATION

- Identified as DOE FECM's data preservation and dissemination platform
- Built-in data compliancy
- Managed submission workflow for review and approval
- Data, presentations, models, tools, etc. are made available
- Submissions are promoted and made available in external data repositories

PRIVATE DATA CURATION

- Uses role-based and tiered access management to maintain secure workspaces
- Facilitates internal collaboration and provides a sharing environment optimized for research scale and processes
- Allows users to upload/download data and resources
- Provides a mechanism to publish data when appropriate

IN DEVELOPMENT

INTEGRATION W/ ANALYTICS & COMPUTE

- Hybrid, multi-cloud data management and compute
- Integration of code collaboration platform will allow for crowdsourcing of AI tools and algorithms by authorized users
- Containerization and virtualization will empower the research community with the ability to bring packaged data, AI models, and tools to both on-prem and cloud-based computing resources

Bottom Line

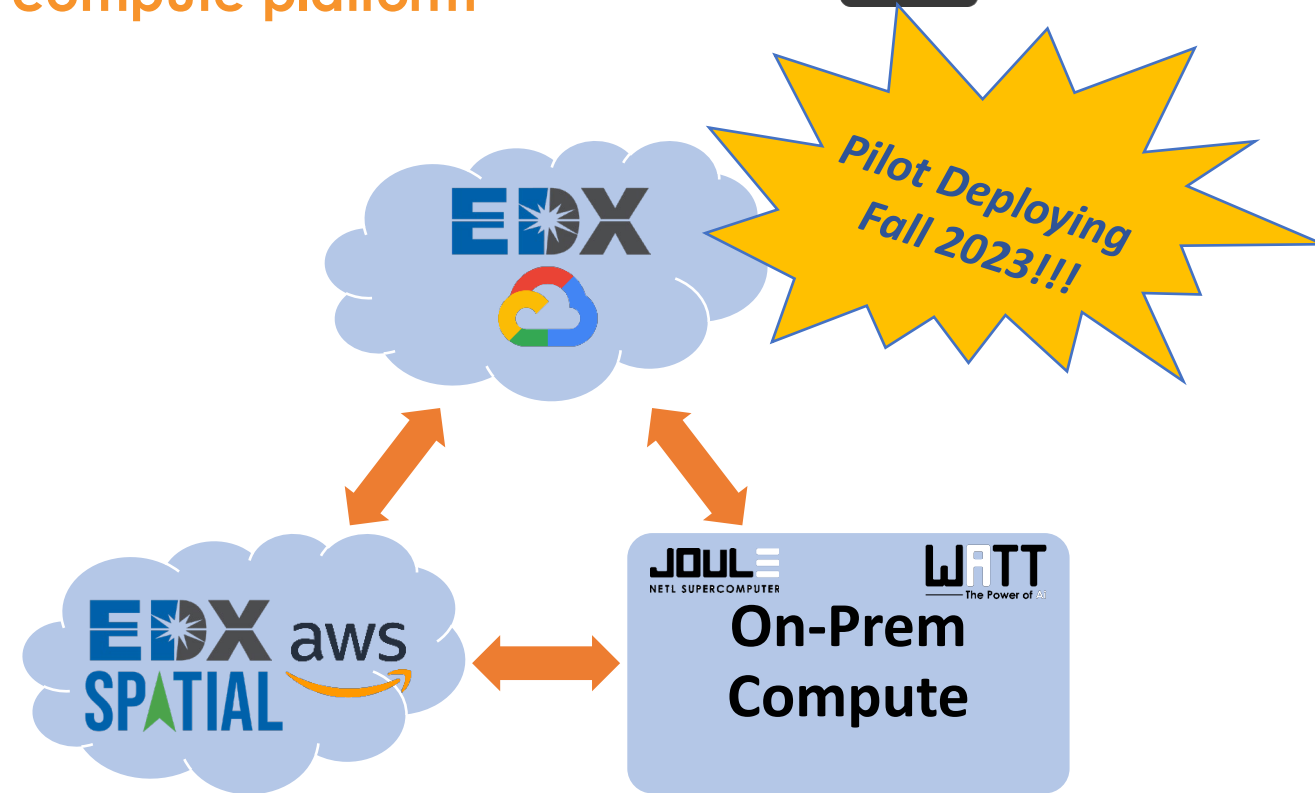
NETL and FECM are building a hybrid, multi-cloud, robust data management and compute platform

What is EDX++?

A hybrid, multi-cloud data management and compute platform

The first cloud instantiation of EDX++ will consist of:

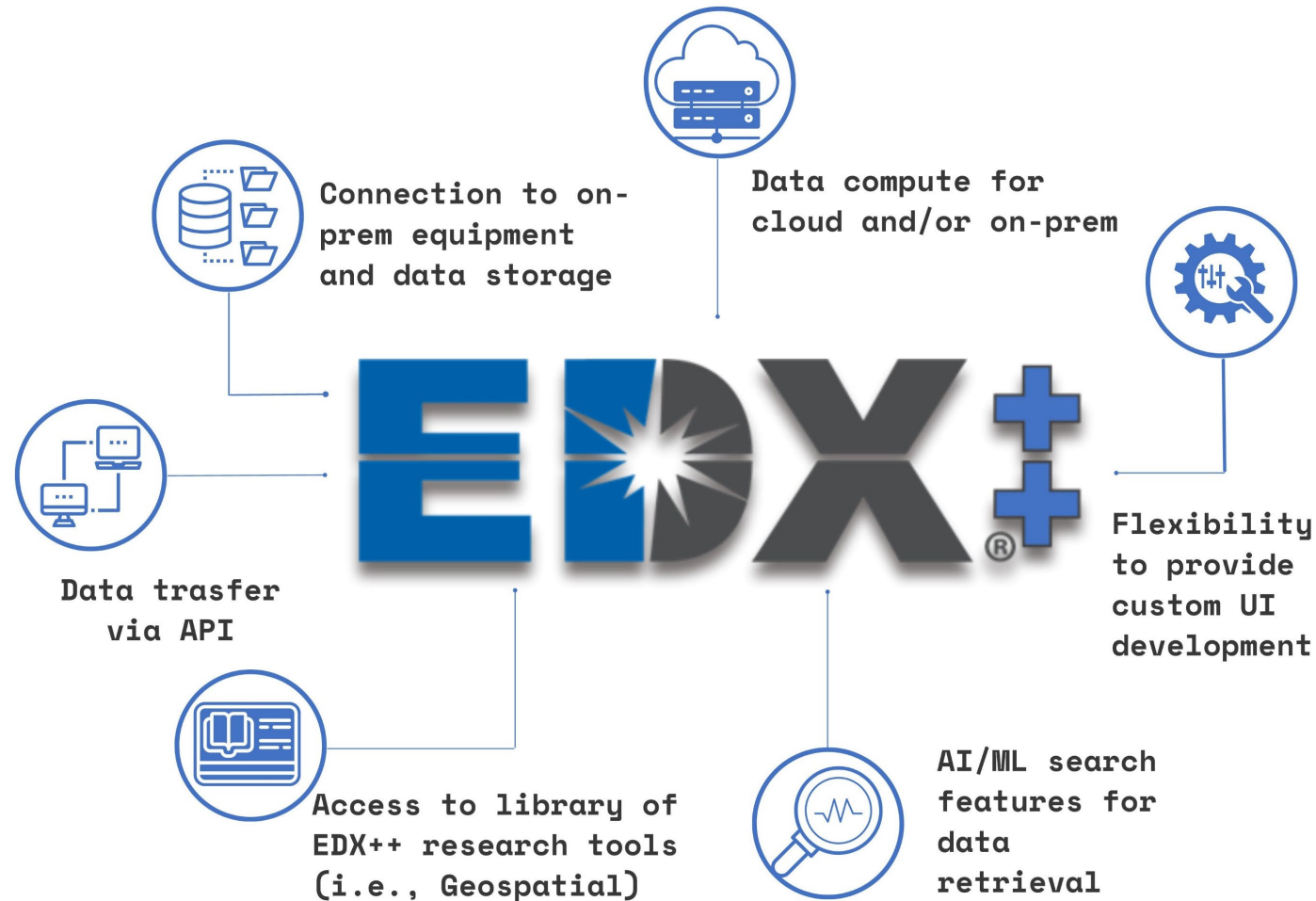
- EDX on Google Cloud Platform (GCP)
- EDX Spatial on Amazon Web Services (AWS)
- EDX Compute on cloud service providers and/or on-prem (Watt & Joule)



Data movement to/from EDX (GCP) to/from EDX Spatial (AWS) or any other cloud service provider using APIs

Research & Development crossover

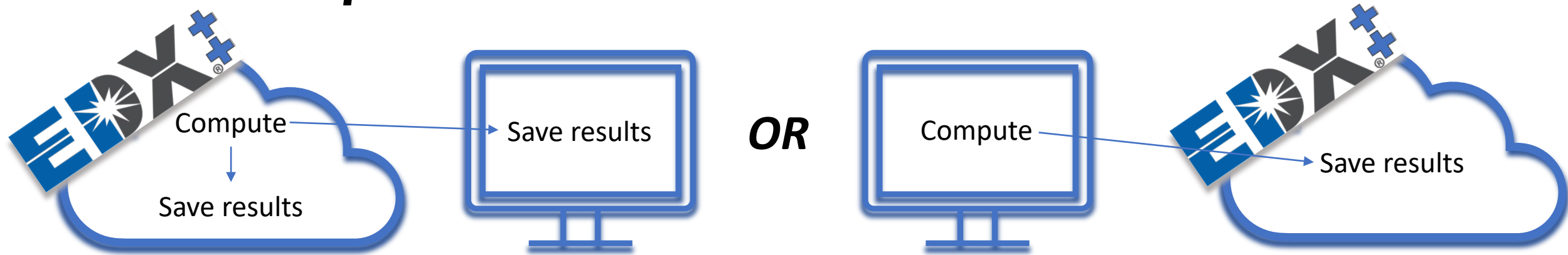
EDX++ provides researchers with features to improve process of their work:



Where is EDX++ going?

Multi-Cloud Infrastructure

Workspaces will be full virtual laboratories



- Git repository listing allows users to post links within their workspaces to externally hosted git repositories.
- Applications will be fully functioning on a multi-cloud infrastructure, including EDX++ on GCP and EDX Spatial on ArcGIS hosted by AWS.

- Users will have options to transfer EDX++ data to other Cloud Service Providers (CSPs) to perform compute outside of the EDX++ infrastructure on GCP.
- Users will have options to transfer external CSP data to the EDX++ infrastructure on GCP.

NETL

RESOURCES

VISIT US AT: www.NETL.DOE.gov

 @NETL_DOE

 @NETL_DOE

 @NationalEnergyTechnologyLaboratory

