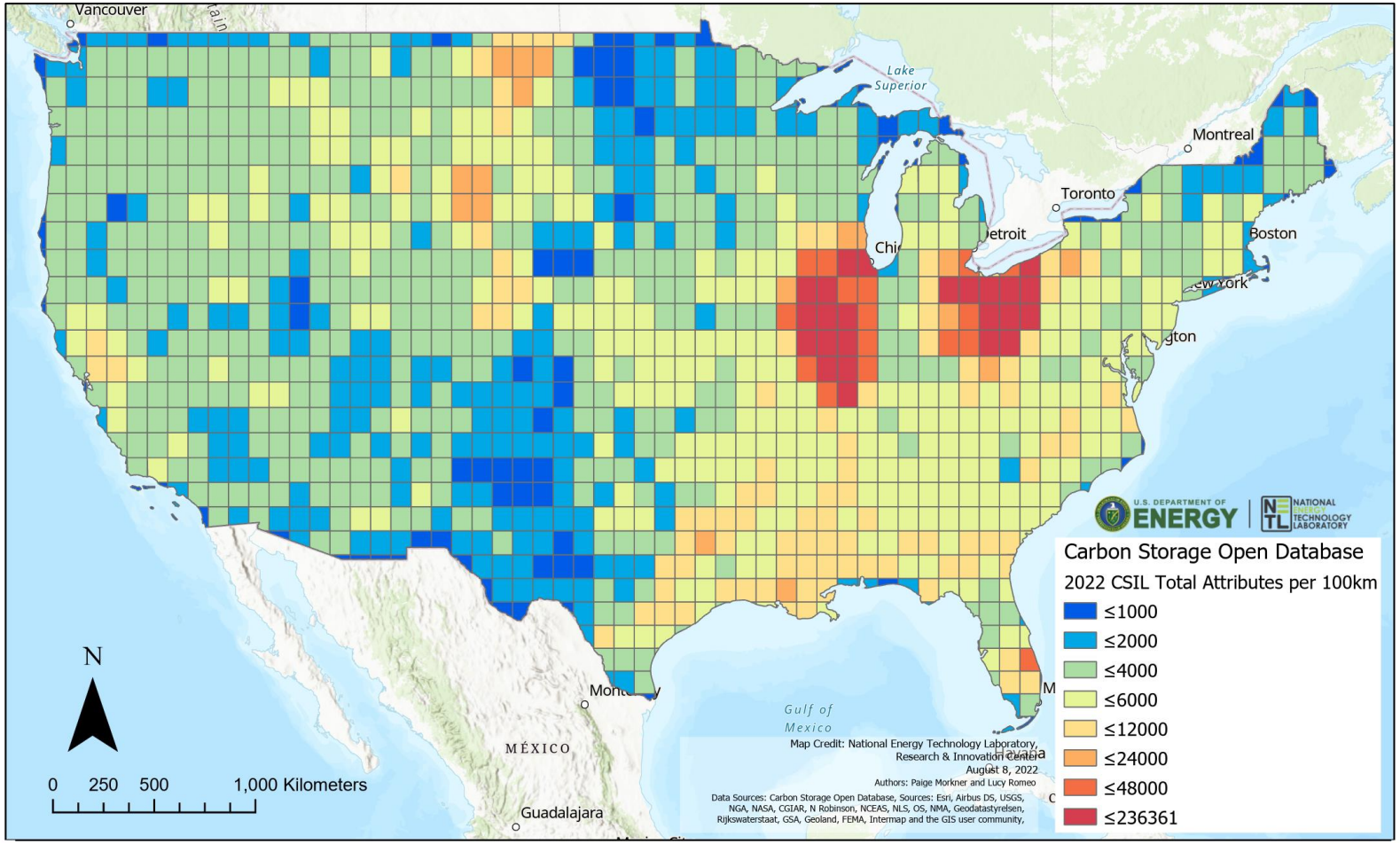
Updating NATCARB and Carbon Storage Geospatial Resources via EDX Cloud

Paige Morkner
NETL Support Contract, Research Innovation Center



U.S. Department of Energy,
National Energy Technology Laboratory
Carbon Management Project Review Meeting
Pittsburgh, PA

Aug. 16, 2022

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

Paige Morkner[1,2], Jennifer Bauer[1], Scott Pantaleone[1,2], Jacob Shay[1,2], Chad Rowan[1,3], Vic Baker[1,3] , Joseph Obradovich[1,3] , and Kelly Rose[1]

[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

[3] NETL Support Contractor, 3610 Collins Ferry Road, Morgantown, WV 26505, USA

A History of DOE Program Product Curation

DOE Carbon Storage

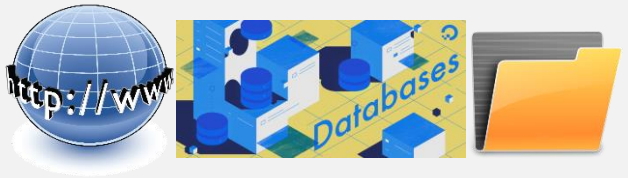
2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022

- 2015 DOE Program Managers **initiated requirement** to contribute and curate carbon storage (CS) data products from CS program in EDX
- Included requirements in FOAs and data management plans for individual projects implemented across the US
- Funded NETL RIC EDX team to provide customized support to the CS-affiliated performers, as well as development of data science-informed optimizations in EDX for the CS Program
- EDX team also hosted in-person and web-based training and informational webinars for extramural partners

NETL SmartSearch automates data discovery using AI/ML by …

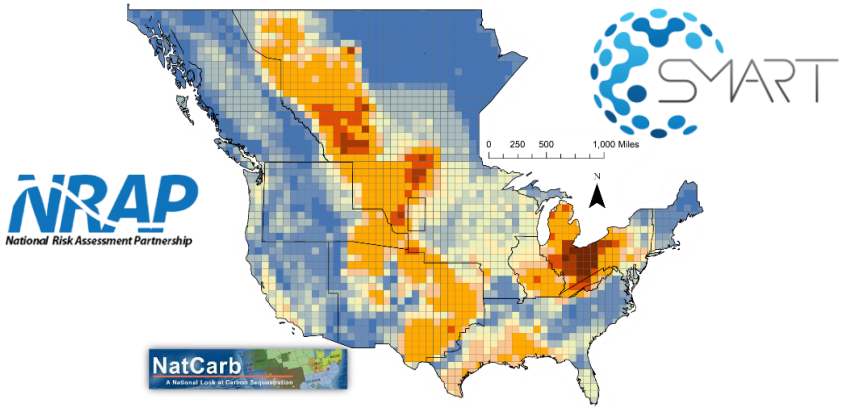
1) **Analyzing** content you like

2) **Finding** new, targeted content



- 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

- Expanded to CarbonSafe and additional extramural programs
- EDX used by each CS Program extramural team (**51 individual Private Workspaces**) in their ongoing work, staging space for products. New awardees initiate their workspace at beginning of each project
- NETL geo-data science and EDX teams launched development of AI and data-science enhanced efforts to acquire, tag, organize, refine, and improve virtualization of CS-program products as they were accumulating



EDX ++ Connecting Data to Resources for Analysis and Computing

Driving Next-Gen R&D…Scaling the Pyramid

NATIONAL ENERGY TECHNOLOGY LABORATORY

Off-Prem Cloud Compute

OSTI.GOV

OpenEI Open Energy Information

Other systems & resources

NATIONAL ENERGY TECHNOLOGY LABORATORY

JOULE NETL SUPERCOMPUTER

WATT The Power of Ai

On-Prem Compute

ESnet

EDX++ FRAMEWORK

…ensuring compliance with Federal/DOE regulations

EDX® Energy Data eXchange

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

Big Data

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

Carbon Storage Data on the Energy Data eXchange (EDX)

Carbon Storage Data on EDX

- More than 2,500 submissions on EDX
  - *1.569TB of Published EDX resources*
  - *3,185 open data resources*
- **100+ TB of seismic data and counting on NETL's WATT**
- ***Groups specific to: RCSPs, CarbonSAFE, FutureGen, Illinois Basin Decatur Project, NRAP, and more***

NATCARB Database 2015

- Reservoir data for deep saline, unminable coal, and depleted oil and gas reservoirs
- Brine produced waters data
- Geochemical data
- Major sedimentary basins for USA

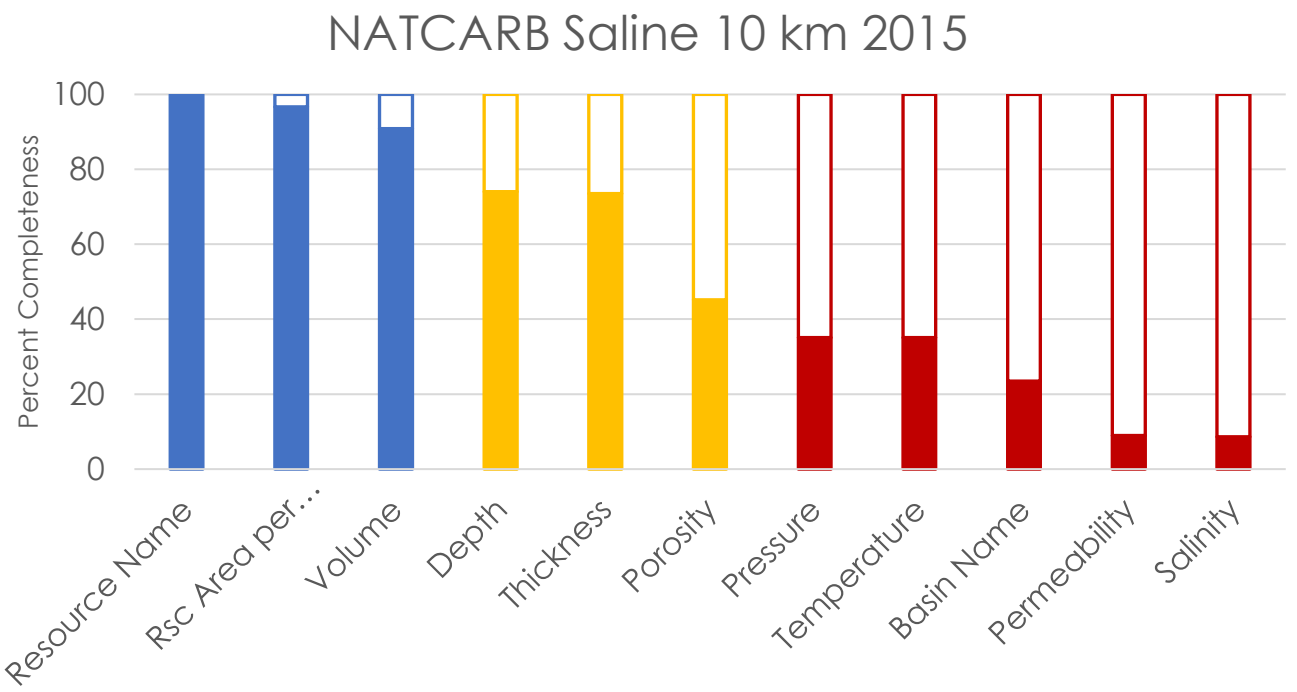
Carbon Storage Open Database:

- **Group on EDX for targeted data resources from FECM research related to Carbon Storage**
- **Data collection scraped from public websites and ArcREST servers**
- **896 Shapefiles and rasters available on EDX's GeoCube**
- **1,846 text-based documents on EDX**

NATCARB Data Gaps

Why Do These Need to be Addressed?

- Gap analysis of NATCARB database conducted in 2019 showed gaps in carbon storage property values reported by the RCSPs including salinity, permeability, basin name, temperature, pressure, thickness, porosity, and depth
- Goal in EY22 was to mitigate gaps leveraging AI/ML tools developed through CS Program



Discovering and Transforming Information with AI/ML Tools at NETL

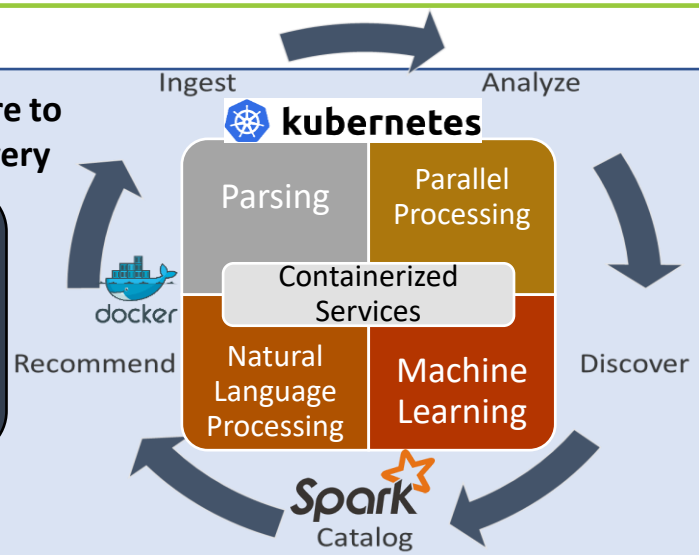
NATIONAL ENERGY TECHNOLOGY LABORATORY

AI informed approach

Challenge: data infrastructure to AI/ML enhanced data discovery

Employing AI/ML tools to find open resources

NETL SmartSearch



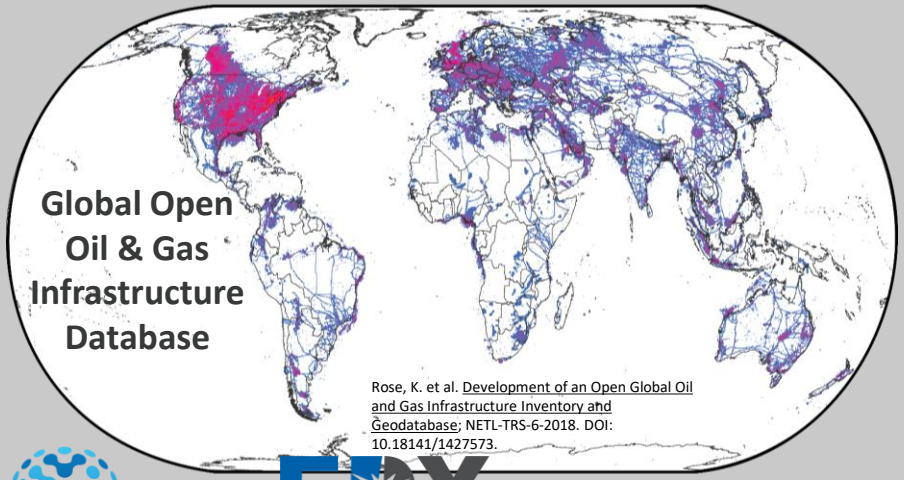
SmartSearch leverages ML+NLP to:

1) **Analyzing** content you like
2) **Finding** new content via www, local, enterprise data stores
3) Telling you **how relevant** the new data is to what you like

Opportunity:

Infinitely scalable to return text, graphical, tabular, image, html, spatial, etc., result

Example applications to date



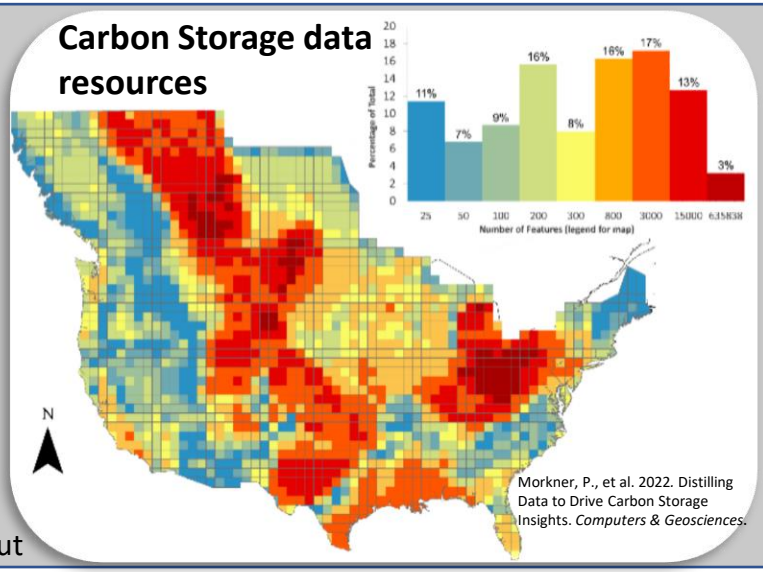
Rose, K. et al. Development of an Open Global Oil and Gas Infrastructure Inventory and Geodatabase; NETL-TRS-6-2018. DOI: 10.18141/1427573.

SMART

EDX Energy Data eXchange

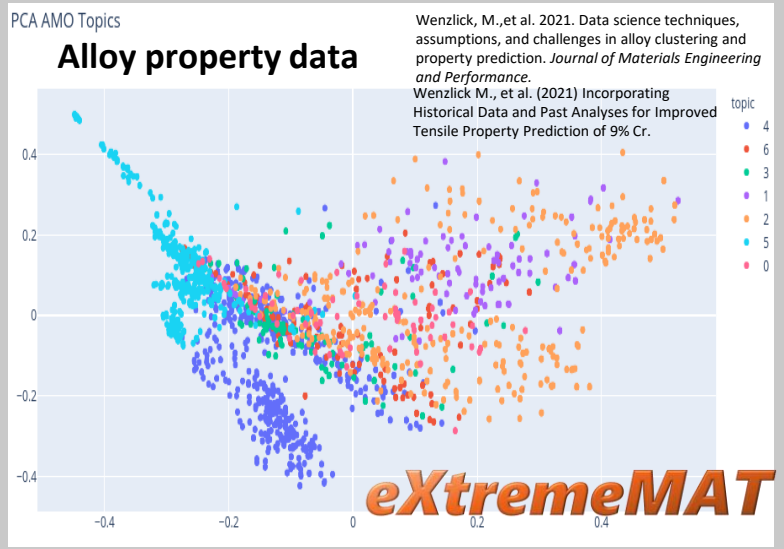
https://edx.netl.doe.gov/about

Carbon Storage data resources



Morkner, P., et al. 2022. Distilling Data to Drive Carbon Storage Insights. *Computers & Geosciences.*

Alloy property data



Wenzlick, M.,et al. 2021. Data science techniques, assumptions, and challenges in alloy clustering and property prediction. *Journal of Materials Engineering and Performance.*

Wenzlick M., et al. (2021) Incorporating Historical Data and Past Analyses for Improved Tensile Property Prediction of 9% Cr.

Leveraging SmartSearch to Update NATCARB

Leveraging information from gap analysis, SmartSearch was used to look for data that would help fill gaps

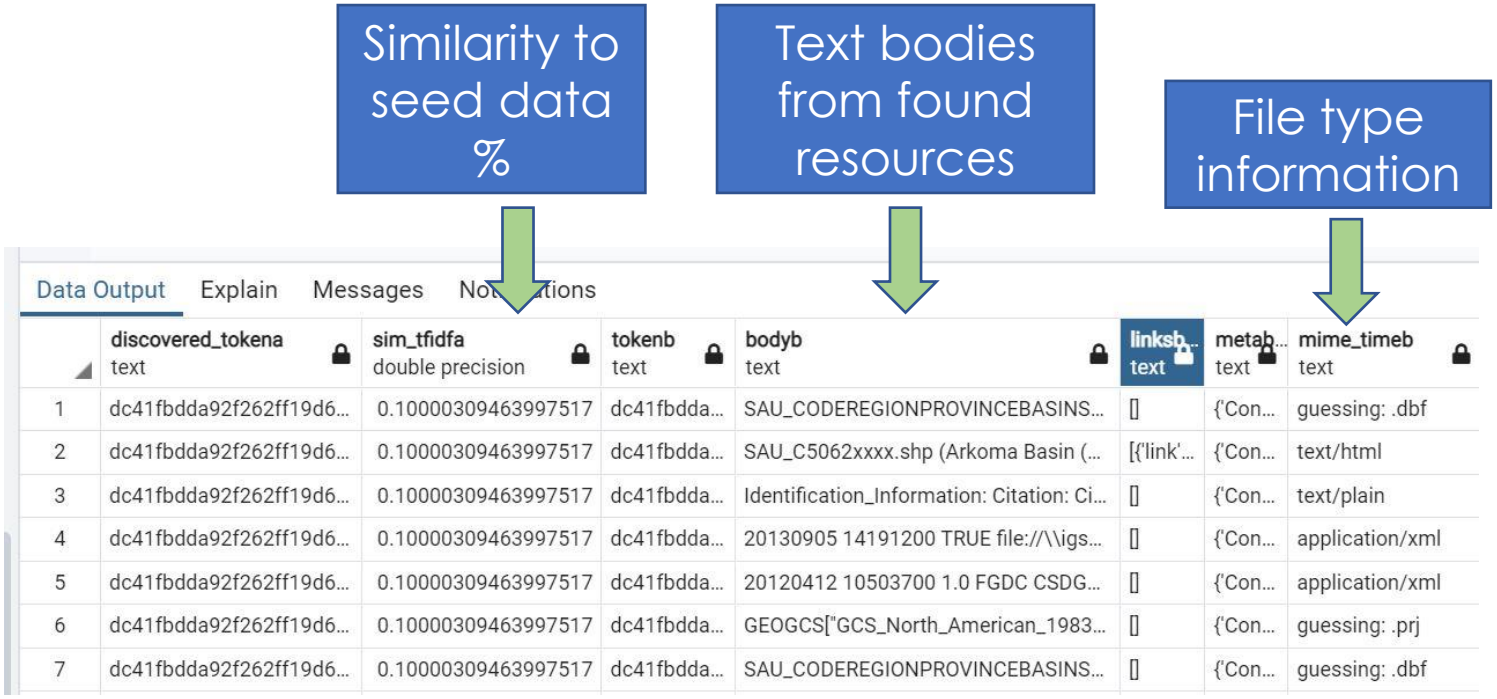
Used NATCARB database as seed data

Seed parsed terms are then used to search the WWW for other relevant results

Results are returned of webpages that likely contain relevant results

Output Data from SmartSearch and Processing into PostgreSQL

- 10,000+ results of text-based documents, shapefiles, and other data were collected using SmartSearch
- The data text bodies were inputted into a PostgreSQL database to make them easily quarriable and searchable via keywords
- Data was integrated from SmartSearch results, manual research for information, and from the USGS Saline Resource Assessment data



	discovered_tokena text	sim_tfidfa double precision	tokenb text	bodyb text	linksb... text	metab... text	mime_timeb text
1	dc41fbdda92f262ff19d6...	0.10000309463997517	dc41fbdda...	SAU_CODEREGIONPROVINCEBASINS...	[]	{'Con...	guessing: .dbf
2	dc41fbdda92f262ff19d6...	0.10000309463997517	dc41fbdda...	SAU_C5062xxxx.shp (Arkoma Basin (...	[{'link'...	{'Con...	text/html
3	dc41fbdda92f262ff19d6...	0.10000309463997517	dc41fbdda...	Identification_Information: Citation: Ci...	[]	{'Con...	text/plain
4	dc41fbdda92f262ff19d6...	0.10000309463997517	dc41fbdda...	20130905 14191200 TRUE file://\\igs...	[]	{'Con...	application/xml
5	dc41fbdda92f262ff19d6...	0.10000309463997517	dc41fbdda...	20120412 10503700 1.0 FGDC CSDG...	[]	{'Con...	application/xml
6	dc41fbdda92f262ff19d6...	0.10000309463997517	dc41fbdda...	GEOGCS["GCS_North_American_1983...	[]	{'Con...	guessing: .prj
7	dc41fbdda92f262ff19d6...	0.10000309463997517	dc41fbdda...	SAU_CODEREGIONPROVINCEBASINS...	[]	{'Con...	guessing: .dbf

Results

What Attributes were Contributed to NATCARB?

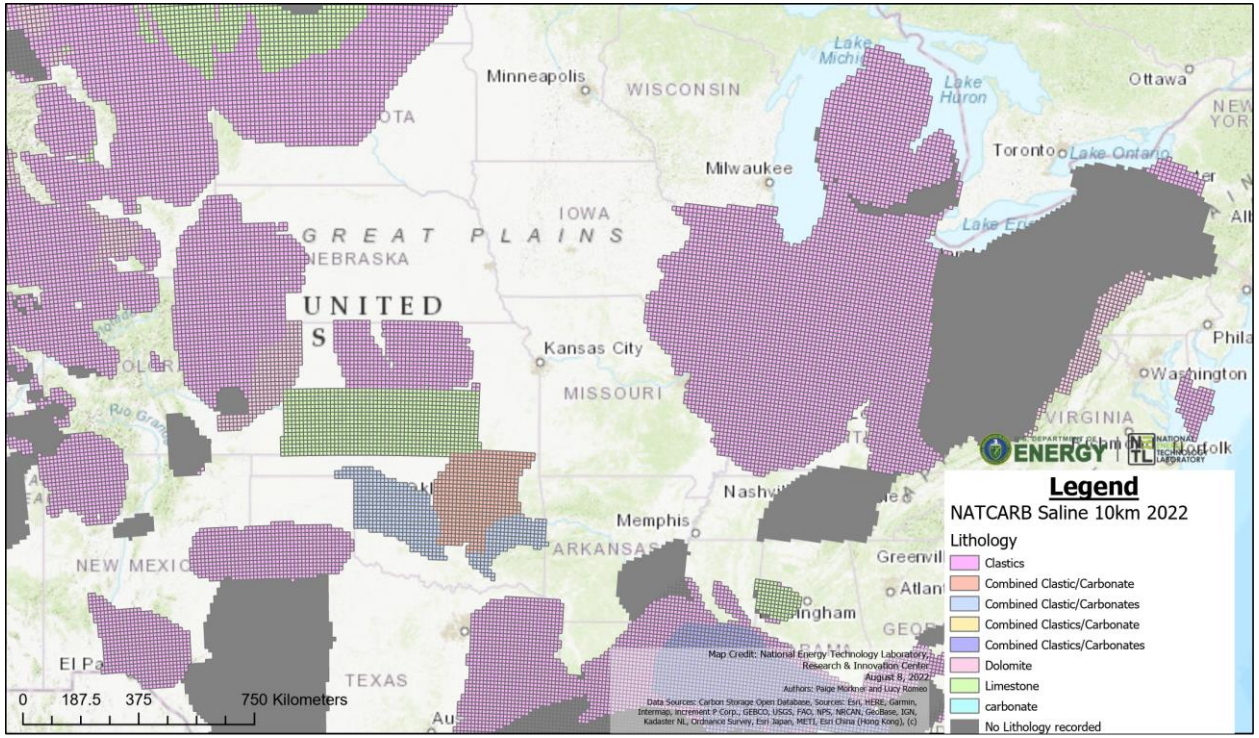
New data was added to new columns within Saline 10 km layer within NATCARB

Added additional data to:

- Basin Names
- Porosity, Permeability, Depth (top and bottom), Thickness

New columns of data added:

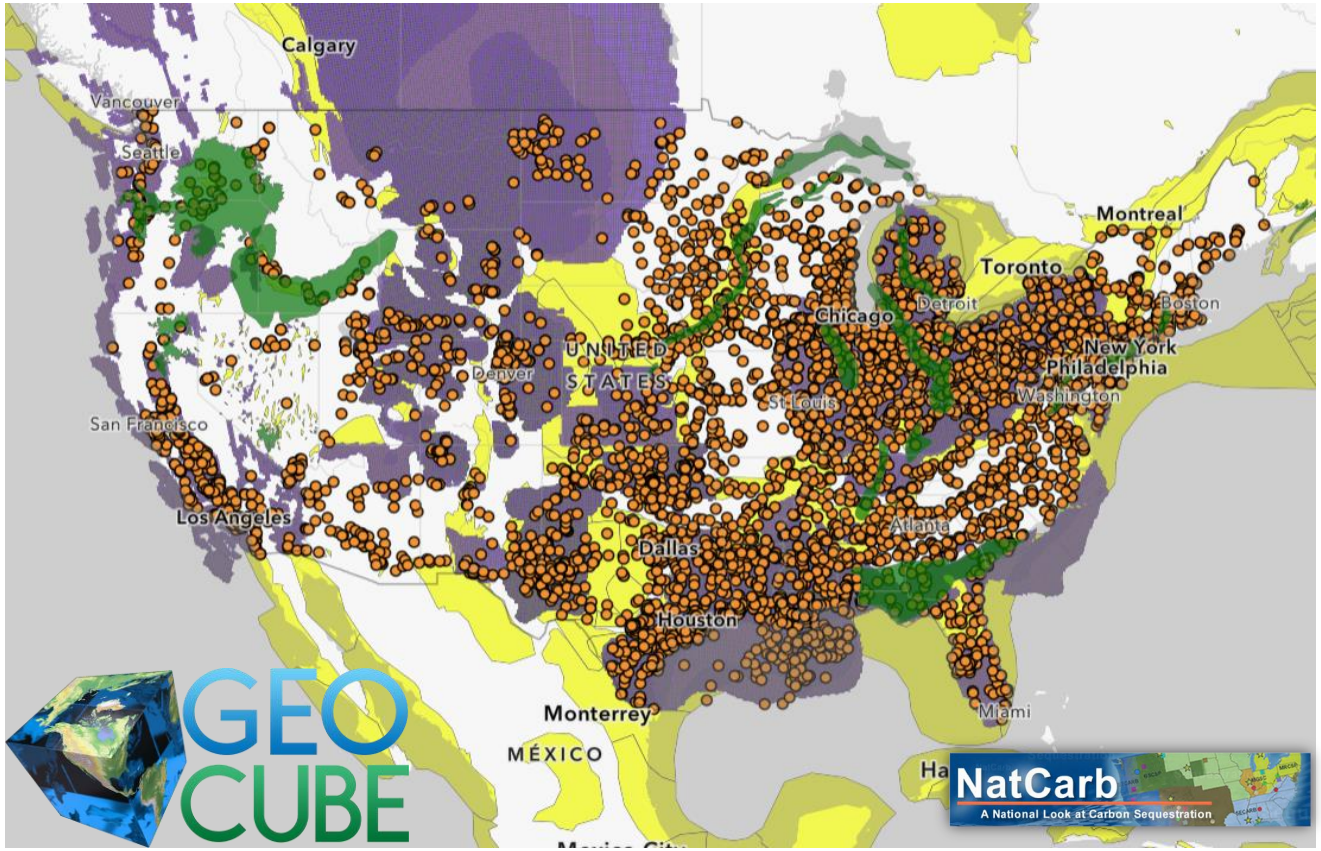
- Formation Name (more granular level than Resource in NATCARB v5)
- Lithology
- Depositional Environment
- Minor Lithology and Depositional Environment
- Citations for all new data



Example of lithologies identified for NATCARB Saline 10 km reservoir units

GeoCube and EDX Geospatial

- Data added to the NATCARB Viewer on GeoCube:
  - EPA 2020 Large CO2 emitter point sources
  - Basalt Basins
  - Updated Saline 10 km Grid Shapefile
    - Also can be downloaded directly with documentation from EDX:
    - https://edx.netl.doe.gov/dataset/natcarb-atlas-saline-basin-10km-grid
- Additional 586 Shapefiles added to the Carbon Storage Open Database – see talk tomorrow for more details
- New GeoCube release leveraging EDX++
  - See demo during Tools demo Tuesday evening!



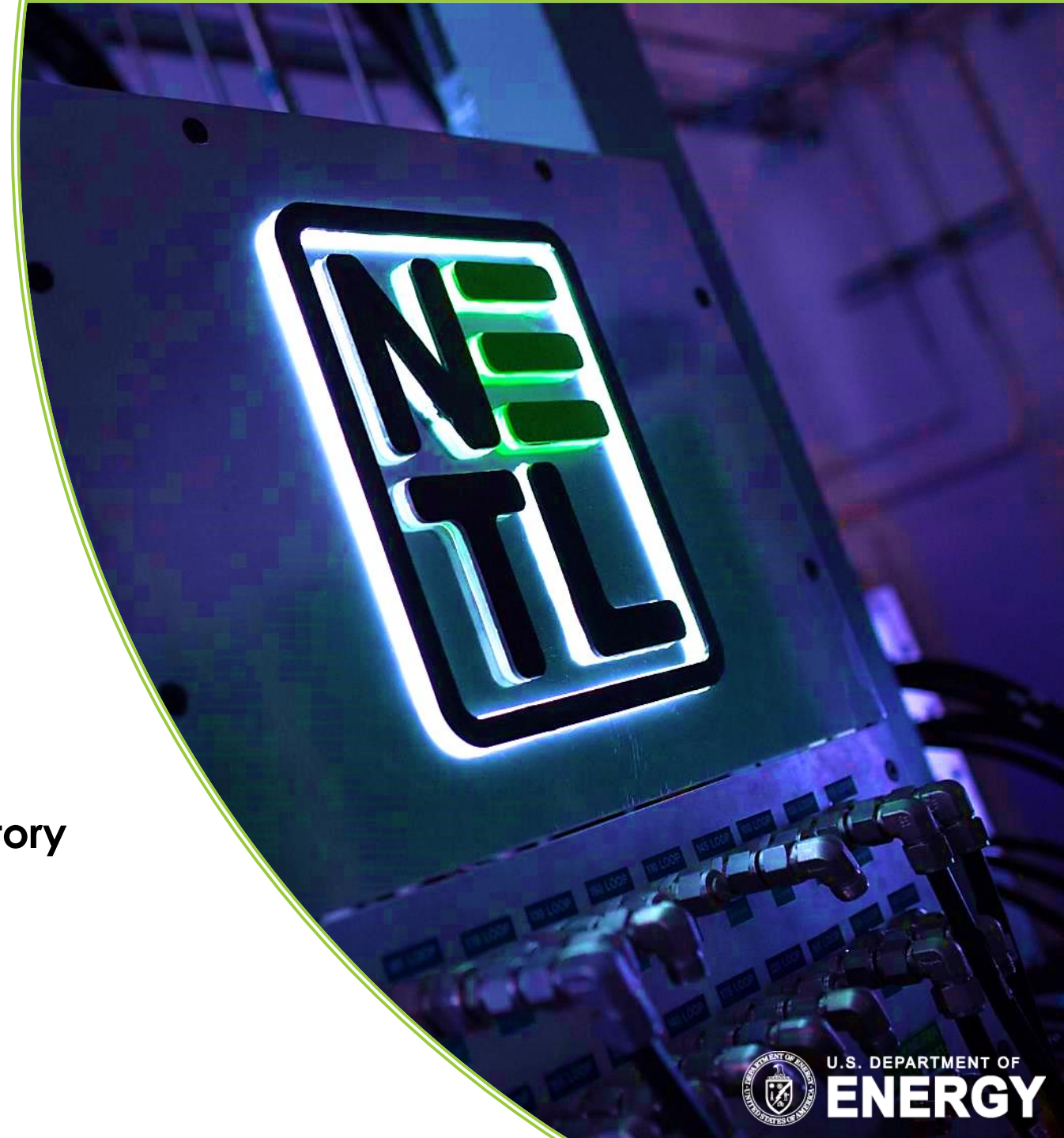
NETL RESOURCES

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

@NETL_DOE

@NETL_DOE

@NationalEnergyTechnologyLaboratory



U.S. DEPARTMENT OF ENERGY

Organization Chart

Project Partners
DOE
NETL
RCSPs – Big Sky Carbon Sequestration Partnership, Southwest Partnership, Southeast Regional Carbon Sequestration Partnership, Midwest Regional Carbon Sequestration Partnership, Midwest Geological Sequestration Consortium, Plains CO2 Reduction Partnership
CarbonSAFE projects
SMART
National Risk Assessment Partnership

Lead Organization
NETL

Principal Investigators
Kelly Rose, Jennifer Bauer

Task 28
Curation of Carbon Storage R&D Products Through Advanced Data Computing Solutions

Lead: Jennifer Bauer
Team: Kelly Rose, Chad Rowan, Michael Sabbatino, Paige Morkner, Lucy Romeo, TJ Jones, Aaron Barkhurst, Vic Baker, and other Matric Software Engineers and Developers

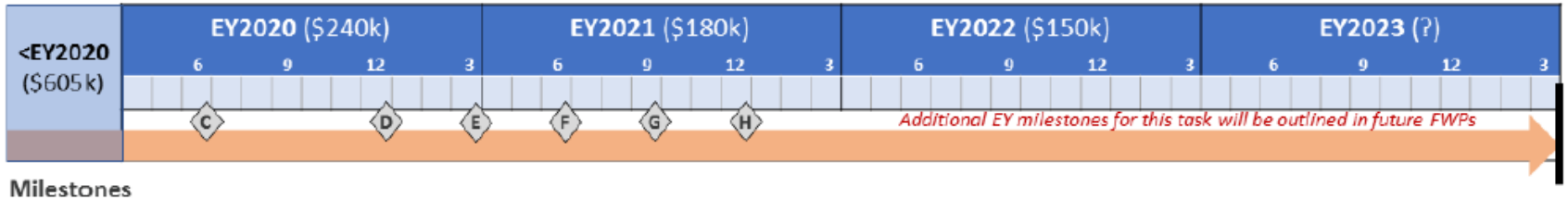
Task 27.0
Next Generation Development, Deployment, and Modernization of Database, Tools, Online Viewer, and Atlas

Lead: Jennifer Bauer
Team: Kelly Rose, Paige Morkner, Michael Sabbatino, Patrick Wingo, Andrew Bean, Aaron Barkhurst, and other Matric Software Engineers and Developers

Gantt Chart

Task 27.0: Project Timeline Overview

Natcarb - Next Generation Development, Deployment, and Modernization of Database, Tools, Online Viewer, and Atlas (PIs: Paige Morkner, Jennifer Bauer)



Milestones

Number	Expected Completion Date	Milestone Description
EY20.27.C	06/30/2020	Identify tools and models that will be targeted for integration and inclusion within the Natcarb Viewer.
EY20.27.D	12/31/2020	Outline report/manuscript on updated technical capabilities of Natcarb Viewer.
EY20.27.E	03/31/2021	Release update of Natcarb Viewer and Natcarb Database to EDX.
EY21.27.F	06/30/2021	Catalog additional datasets, models, and text-based resources on EDX for future integration into the Natcarb Viewer and GeoCube.
EY21.27.G	09/30/2021	Catalog datasets returned from SmartSearch results targeting known data gaps in existing Natcarb and Open Carbon Storage Databases.
EY21.27.H	12/31/2021	Document capabilities to be incorporated in advanced spatial search capability for discovering spatial data from EDX and GeoCube.

Chart Key
- Milestone
- Project Completion
- Go/No-Go Timeframe

Impact

Key Accomplishments/Deliverables	Value Delivered
• 2018, Enhanced interface and updates to Natcarb Viewer and release through EDX (Barkhurst et al., 2018; Bauer et al., 2018) • 2019, Integration of advanced data use tools in Natcarb Viewer & GeoCube to improve data access and use • 2020, Integration of open-source data to develop Open Carbon Storage Database (Morkner et al., 2020) • 2021, Manuscript detailing innovative data integration strategies used to aggregate Natcarb, RCSP, and open CS data sources (Morkner et al., in review) • 2022, Support updates to Natcarb database and CS estimates	• Produce a robust subsurface data framework that provides **improved data access, data discoverability, and ease of use within the CS community.** • Integrate **online, advanced analytics and models to help facilitate research** across the CS community. • Support development of content and materials for Carbon Storage Atlas updates.

U.S. DEPARTMENT OF ENERGY

* Task 27.0 is updating content into an existing tool with no development of a technology. Therefore, no TRL is assigned.