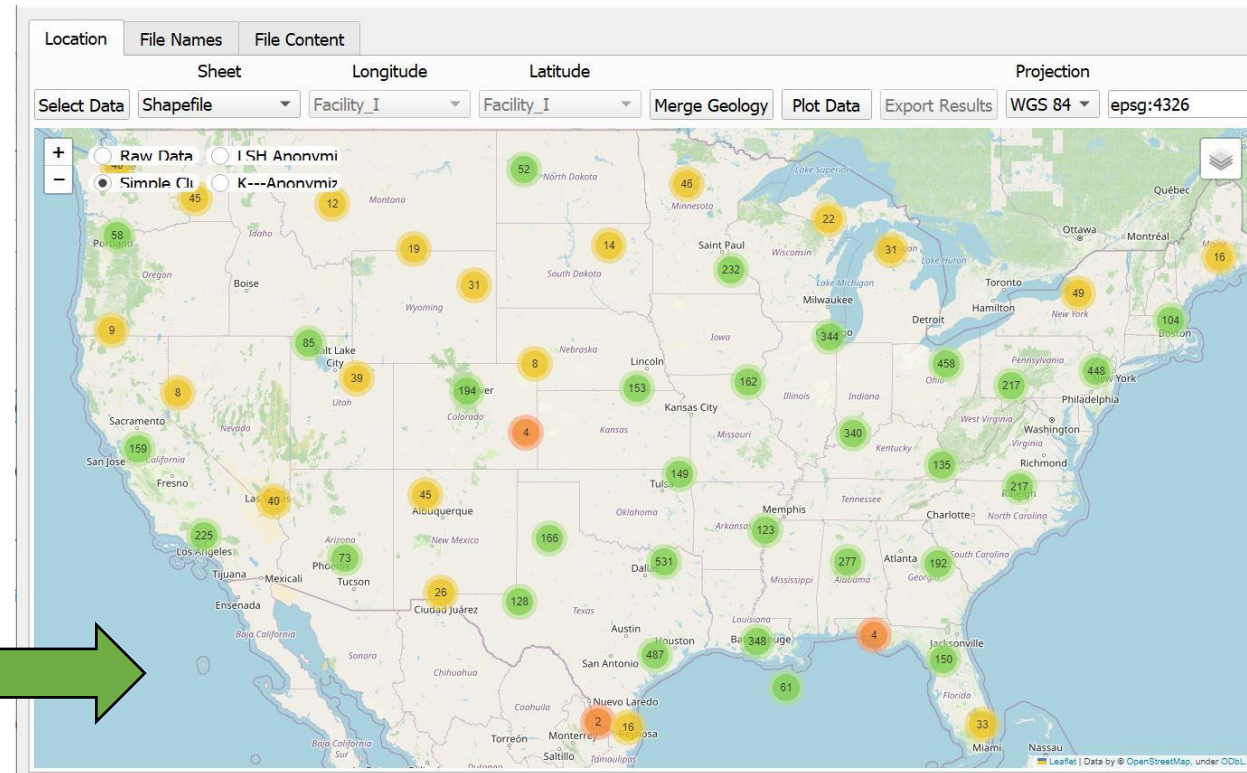
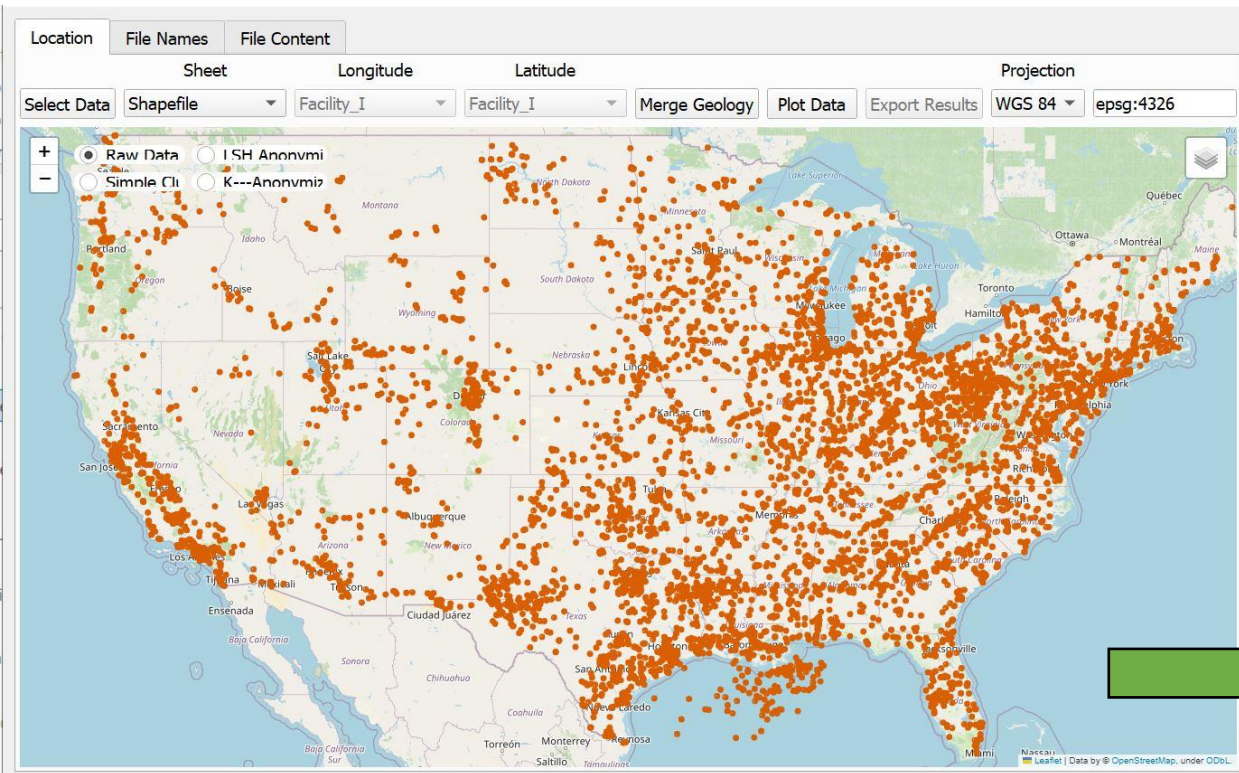


Anonymizing Sensitive Carbon Storage Data Tool

The Geospatial and Information Substitution and Anonymization Tool (GISA)

Paige Morkner
Geo-data Scientist



2023 Carbon Management Research Project Review Meeting
Aug. 29, 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



**Paige Morkner^{1,2}, Jennifer Bauer¹, Patrick Wingo^{1,2}, Michael Gao^{1,2}, Maneesh Sharma^{3,4},
Brendan Hoover⁵, Cedric Neumann⁶, Christopher Johnson⁶, Jared Schuetter⁶, Kelly Rose¹**

¹National Energy Technology Laboratory, 1450 Queen Avenue SW, Albany, OR 97321, USA

²NETL Support Contractor, 1450 Queen Avenue SW, Albany, OR 97321, USA

³National Energy Technology Laboratory, 3610 Collins Ferry Road, Morgantown, WV 26507, USA

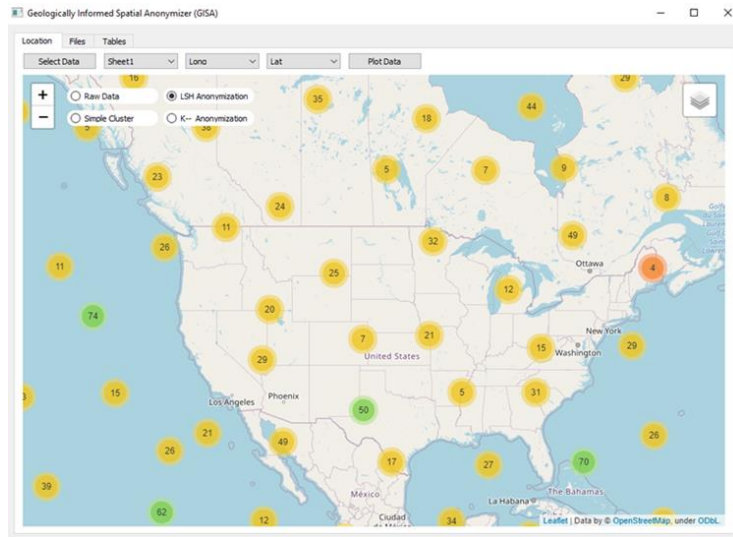
⁴NETL Support Contractor, 3610 Collins Ferry Road, Morgantown, WV 26507, USA

⁵ U.S. Army Engineer Research and Development Center

⁶Battelle, 505 King Avenue, Columbus, OH 43201, USA

The Geospatial and Information Substitution and Anonymization (GISA) Tool

Challenge: Internal data and data shared by partners with DOE sometimes contains sensitive information that cannot be shared with the public or other entities.



Preliminary layout and interface for the anonymization tool demonstrated to anonymize spatially explicit data.

Objective:

- **Deploy an anonymization tool for spatial, text-based, and tabular data, leveraging previous work on the approach, to improve access to critical data** and important variables that can facilitate meaningful analysis and models **without sharing sensitive information.**

Ultimate Product:

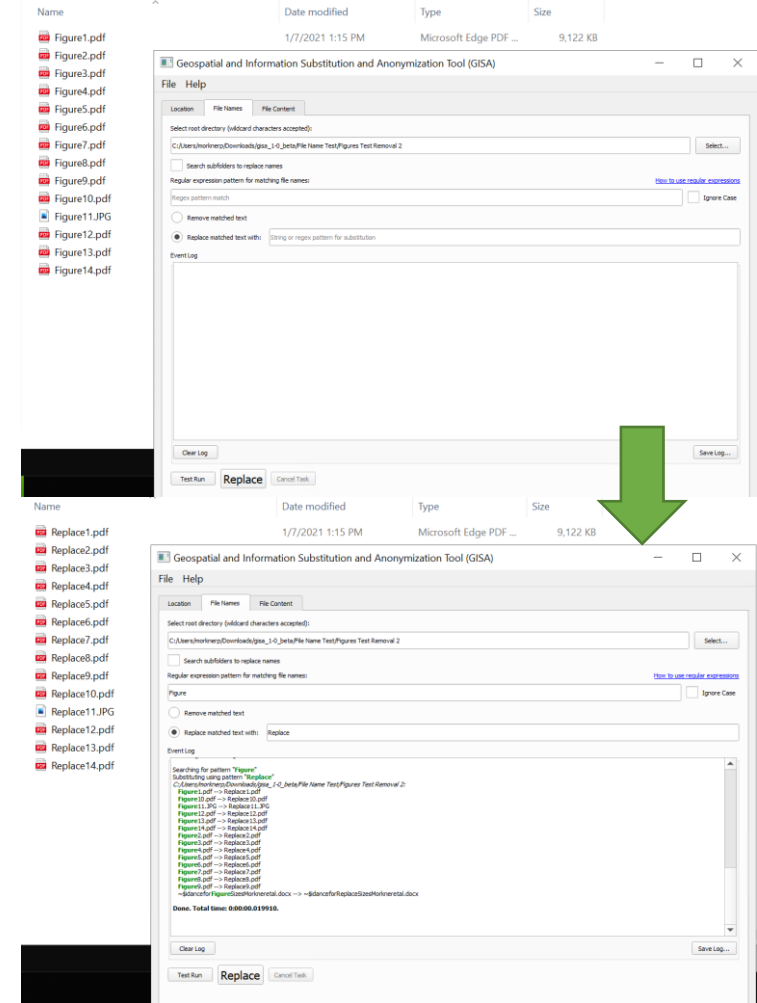
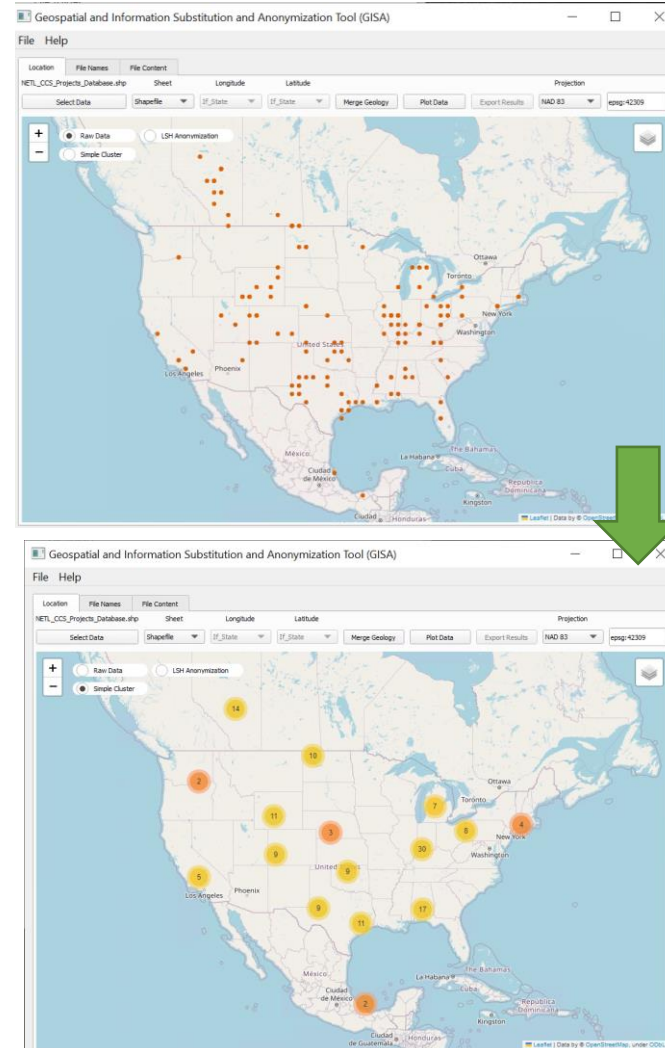
- Virtualized data-anonymization tool integrated for use within EDX DisCO₂ver Platform.
- The tool will enable users to prepare data for publishing on EDX, which promotes open-data sharing practices, while also:
- Not publishing sensitive information that otherwise would have prevented the data from being published.
 - Retaining essential information for data interpretation without publishing the original data collected.

Current Status

Geospatial and Information Substitution and Anonymization Tool

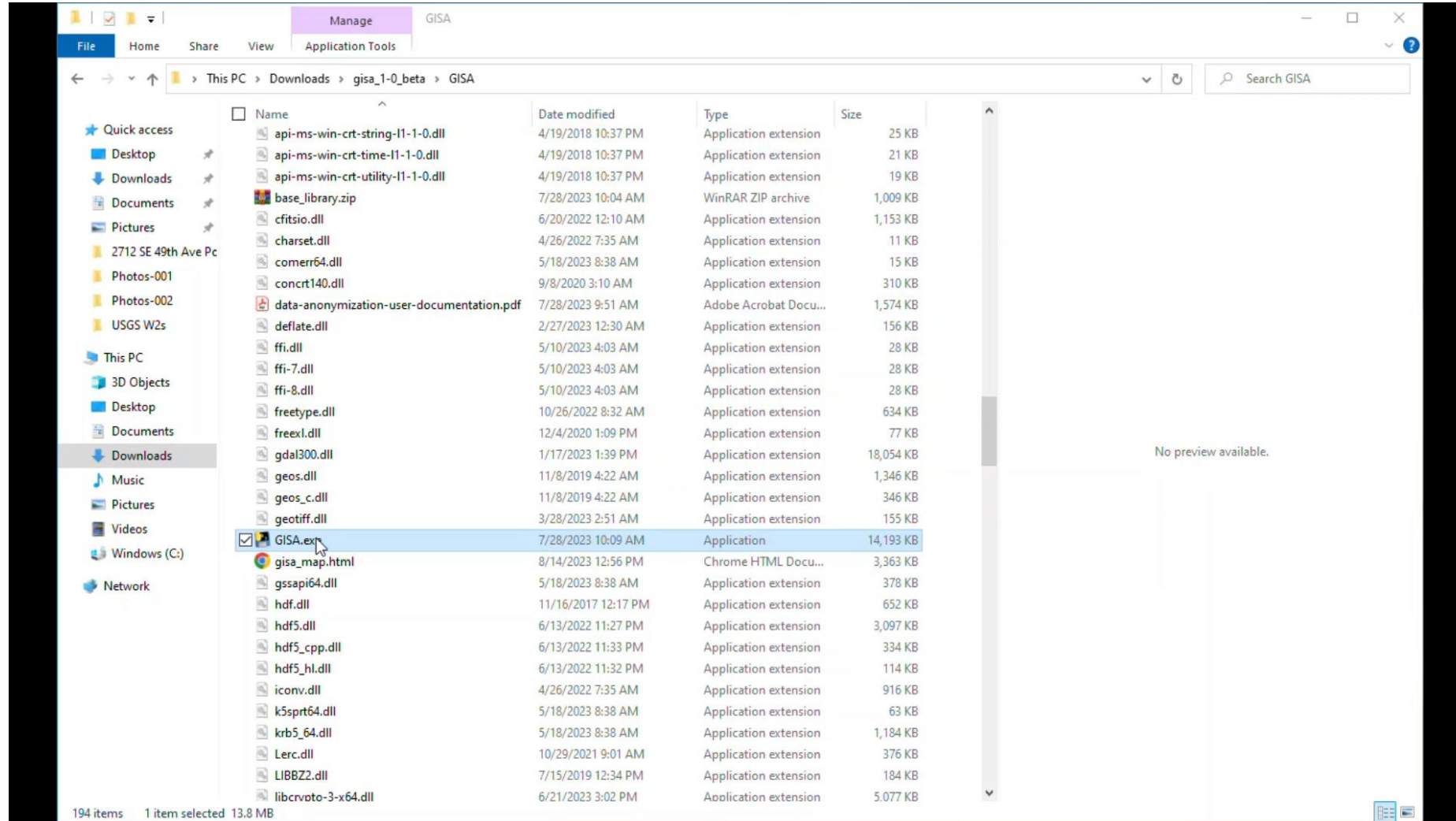
Current Capabilities:

- Geospatial data anonymization of point data
 - Point Shapefiles
 - CSVs/XLSX files
- Batch File Name replacement or removal of user-specified terms
 - All file types
- Batch removal of user-specified terms from text and tabular data formats
 - Text, CSV, XLSX, Doc, and more



GISA Tool Walk Through

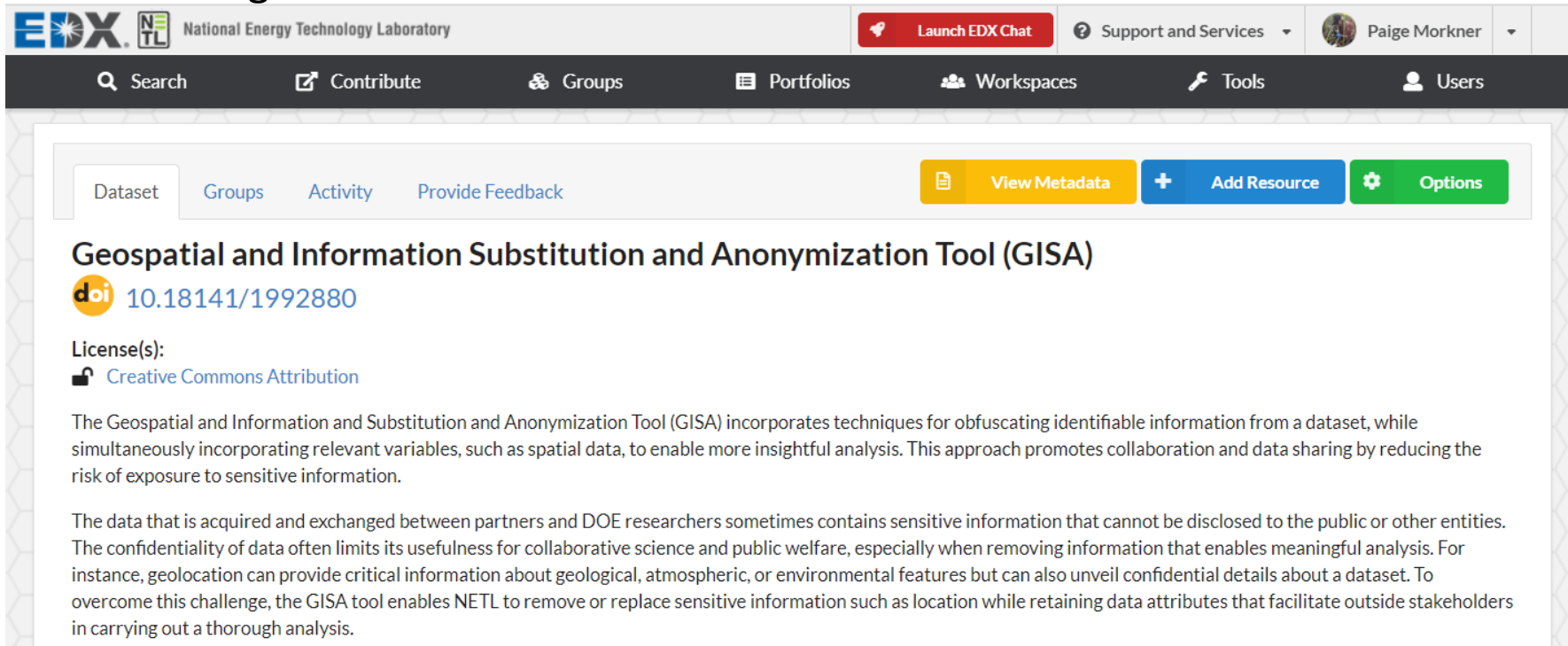
Geospatial and Information Substitution and Anonymization Tool



Beta Tool Availability on EDX

Geospatial and Information Substitution and Anonymization Tool

- GISA Tool was published on EDX in late July 2023
- Available for download and testing now
 - User documentation published alongside the desktop tool for download
 - Looking for user feedback from beta users



Access the GISA
Beta tool on EDX!



<https://edx.netl.doe.gov/dataset/geospatial-and-information-substitution-and-anonymization-tool-gisa>

Geospatial and Information Substitution and Anonymization Tool

New module development includes:

1. Module leveraging **natural language processing** to **produce recommendations for removal** including company names, American Petroleum Institute (API) numbers, and locations from PDF, Word Documents, tables, and LAS files

- **Image recognition in PDF documents (including logos) allowing for review and removal of images**

2. **Module to add additional anonymization capability for point data, based on underlying boundary information such as:**

- Land cover, geologic formations, aquifers, field site boundaries, and more

Use cases for EY23 capabilities:

- **Recommending removal of all locations and company names from data prior to publishing for a CCS site currently going through the permitting process**
- **Relocation of data representing infrastructure such as wells and CO₂ sources**
- **Removal of sensitive logos and images**

Summary

Geospatial and Information Substitution and Anonymization Tool

- **The GISA tool was developed to:**
 - Help users generalize locations in point data (shapefiles and tabular)
 - Help users batch remove or replace terms within file names (all file types)
 - Help users batch remove or replace terms from within file content (text, docs, tabular, csv, LAS)
- **The GISA tool is available for download as a Beta tool**
 - All user feedback is welcome for the Beta tool
- **Additional work to integrate code developed into this tool in EY23**
 - The production tool is expected to be release in March 2024

Access the GISA Beta tool on EDX!



<https://edx.netl.doe.gov/dataset/geospatial-and-information-substitution-and-anonymization-tool-gisa>

Tuesday Evening – Live Tool Demos!

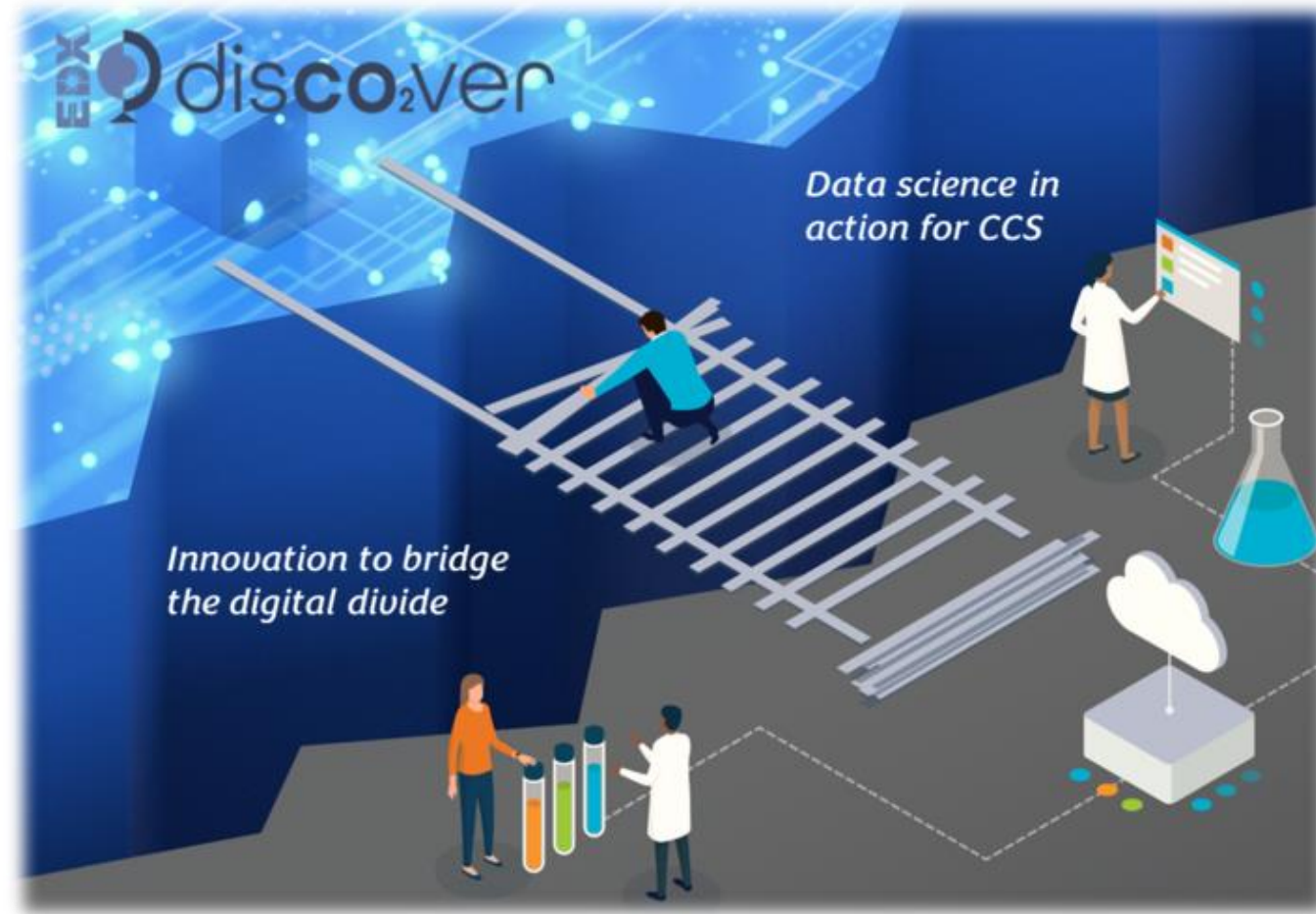
When: 5:45 - 7:45 p.m.

Where: The Ballroom Foyer and East/West Atriums

What:

- **Environmental Justice and Social Justice** for CS Systems
- The **international offshore CS** and web-database and tool
- RokBase, Virtualizing **CS Rock Property Data** platform
- **Class VI Data Support Tool** for regulatory requirements
- **CO2 Pipeline Routing Smart Tool**
- Co2Locate - **Class II Well Reuse and Regional Evaluation Tool**
- **Carbon Storage Planning Framework Dashboard**
- **3D Data Viewer** and Preview Capability
- AIIM Model, **Assessing Infrastructure Reuse Potential** for CS
- **EDX disCO₂ver**, a one-stop tool for CO₂ digital resources

 In demo "theater room"  **support team will offer in person demos & Q&A**



Additional Information



Contact:

Paige Morkner, Paige.Morkner@netl.doe.gov

Data resources and important URLs:

Hoover, B., Wingo, P., Gao, M., Morkner, P., Sharma, M., Bauer, J., and Rose, K. Geospatial and Information Substitution and Anonymization Tool (GISA) Beta version 1.0. National Energy Technology Laboratory, 8/1/2023. www.edx.netl.doe.gov/dataset/geospatial-and-information-substitution-and-anonymization-tool-gisa, DOI: 10.18141/1992880

<https://edx.netl.doe.gov/>

Acknowledgments



This work was performed in support of the U.S. Department of Energy's (DOE) Fossil Energy and Carbon Management's Carbon Storage Program and executed through the National Energy Technology Laboratory (NETL) Research & Innovation Center's Bi-Partisan Infrastructure Law Funded EDX4CCS FWP DE-FE 1025007.

NETL

RESOURCES

VISIT US AT: www.NETL.DOE.gov

 @NETL_DOE

 @NETL_DOE

 @NationalEnergyTechnologyLaboratory



U.S. DEPARTMENT OF
ENERGY