# Dealing with the data deluge to build the first open global oil and gas infrastructure (GOGI) database

**DiGiulio, J.[1,2], Romeo, L.[1,2], Rose, K.[1], Bauer, J.[1], Baker, V.[1,4], Sabbatino, M.[1,2], Justman, D.[1,2], Jones, K.[1,3]**

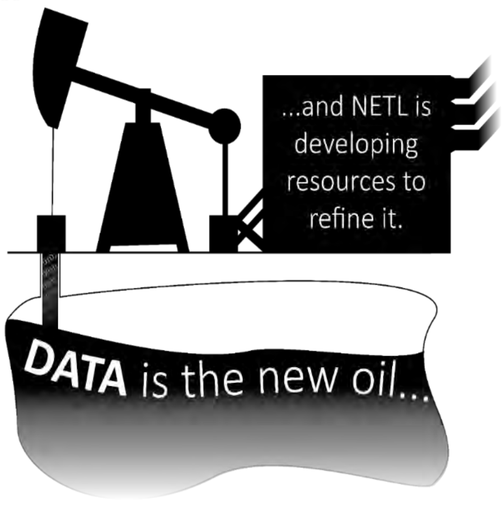
[1]U.S. Department of Energy, National Energy Technology Laboratory, Albany, OR; [2]AECOM, Albany, OR; [3]ORISE, Albany, OR; [4]MATRIC, Morgantown, WV

NATIONAL ENERGY TECHNOLOGY LABORATORY

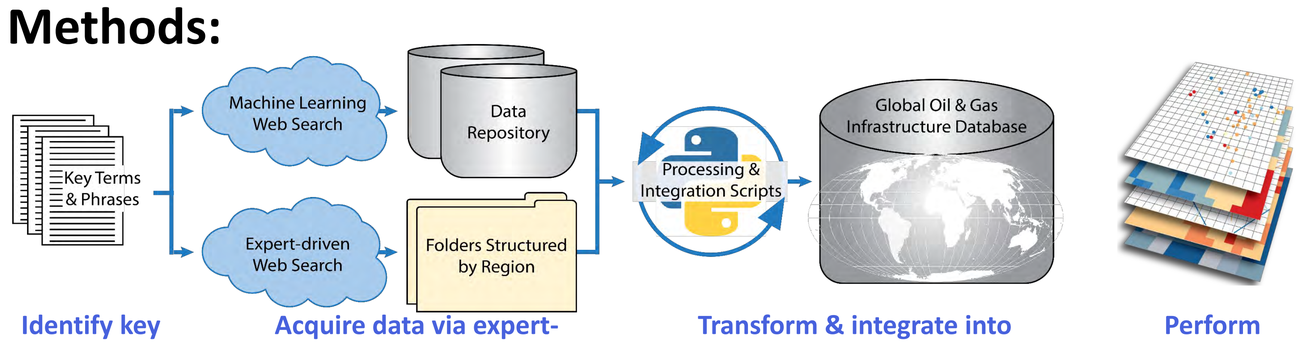
Research & Innovation Center

## Introduction:

- Reliable, accessible data is the foundation of scientific inquiry and basis for all empirical discovery
- The volume of open data is growing rapidly, unlocking huge potential to implement big data analytics and resolve energy-related and economic challenges
- However, rapid data growth comes with challenges; an immense amount of researcher's time is dedicated to finding, accessing, and integrating disparate data before beginning analytics
- To reduce this effort, NETL researchers are developing advanced data science tools to expedite data discovery, accessibility, and processing times for utilizing new data
- Here we demonstrate how NETL researchers used custom data science tools and capabilities to rapidly (~4 months) acquire and analyze open oil and gas infrastructure data across the globe using expert-driven and machine learning search strategies
- **The resulting spatial database demonstrates the first-ever footprint of global oil and gas infrastructure (GOGI)**




## Methods:




**Identify key search terms & phrases**

**Acquire data via expert-driven & machine-learning search strategies**

**Transform & integrate into geodatabase using custom tools & scripts**

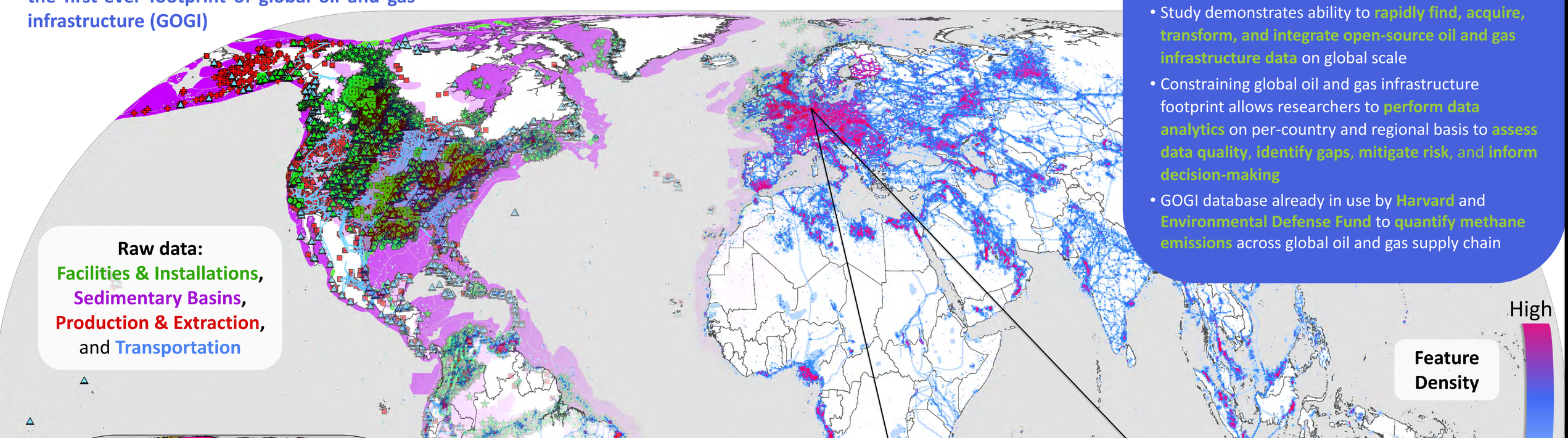
**Perform advanced data analytics**

Access & visualize GOGI through **GeoCube** – NETL's interactive web mapping platform!

## Key Takeaways:

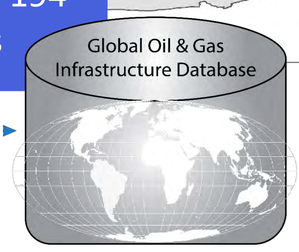
- GOGI database includes **4.8 million features**, **700+ sources**, and **380+ datasets** spanning **194 countries**
- Researchers used **expert-driven and machine learning** search strategies to acquire spatial datasets
- Study demonstrates ability to **rapidly find, acquire, transform, and integrate open-source oil and gas infrastructure data** on global scale
- Constraining global oil and gas infrastructure footprint allows researchers to **perform data analytics** on per-country and regional basis to **assess data quality, identify gaps, mitigate risk**, and **inform decision-making**
- GOGI database already in use by **Harvard** and **Environmental Defense Fund** to **quantify methane emissions** across global oil and gas supply chain

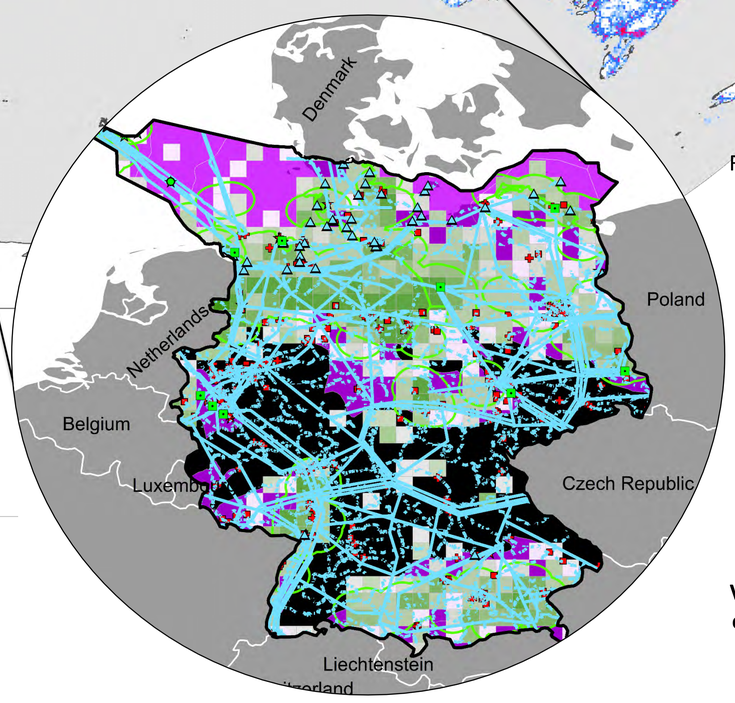
**Raw data:**
**Facilities & Installations**,
**Sedimentary Basins**,
**Production & Extraction**,
and **Transportation**





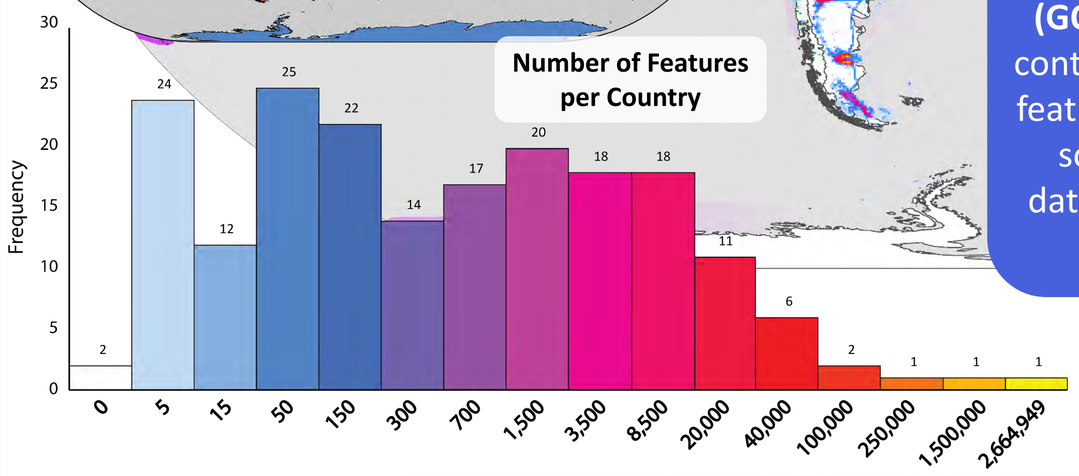

The open **Global Oil and Gas Infrastructure (GOGI) database** contains 4.8 million features from 700+ sources, 380+ datasets, and 194 countries



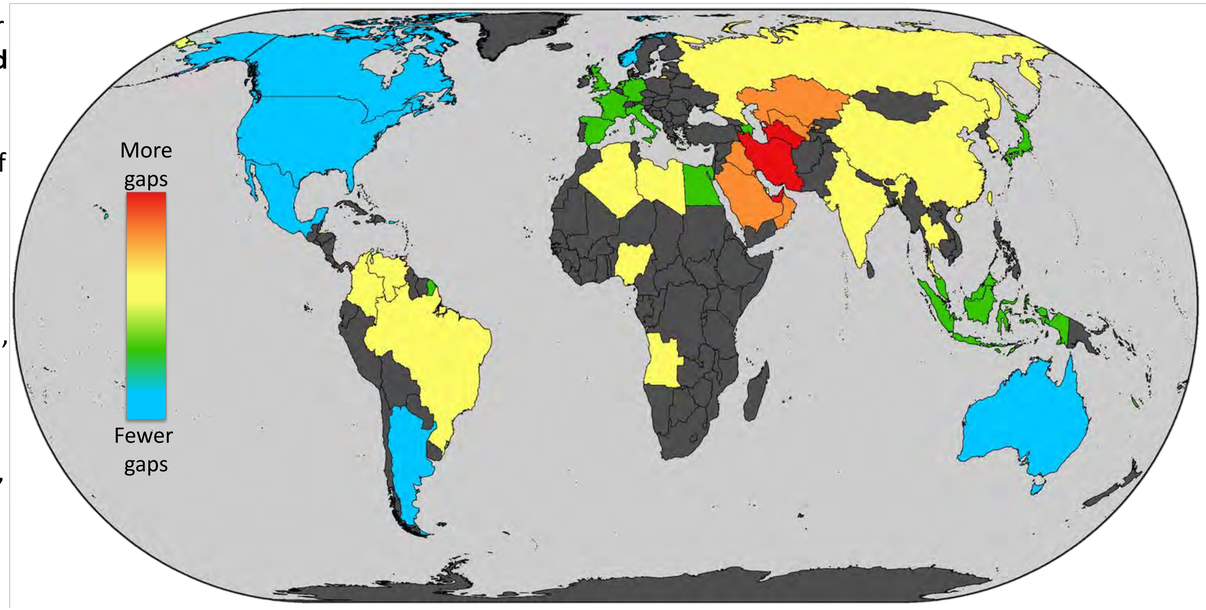



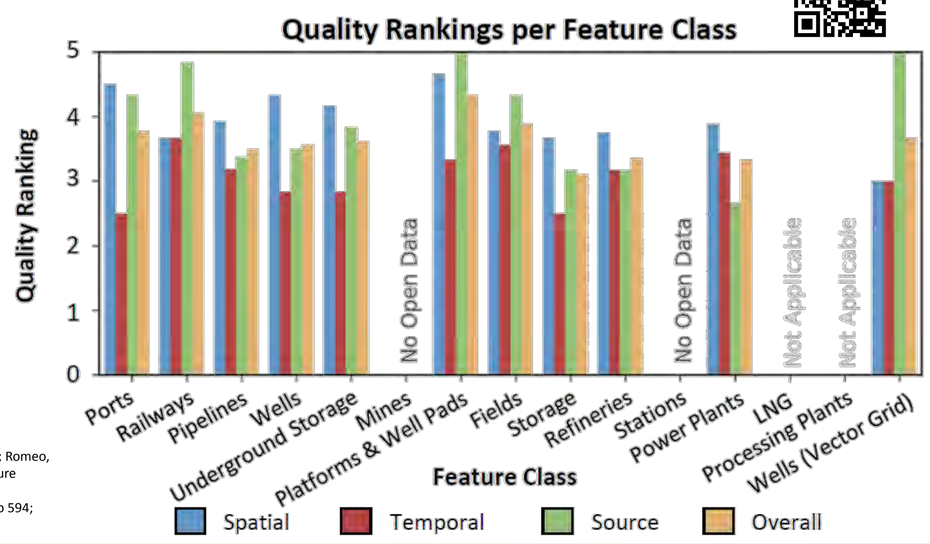

Feature class map (left) displaying all oil and gas infrastructure features in **Germany.** Colors indicate: **Facilities & Installations**, **Sedimentary Basins**, **Production & Extraction**, and **Transportation**. Corresponding dataset's quality rankings shown in graph (below).

**View details for all 194 countries via QR code.**




**Data Gap Analysis** calculated for **top 40 oil and gas producing and consuming countries**. This gap analysis helps evaluate uncertainty and completeness of data obtained for each country in the GOGI database. Cool colors demonstrate countries expected to have fewer data gaps (e.g., United States, Canada, Argentina, Norway), while hot colors demonstrate countries expected to have more gaps (e.g., Iran, United Arab Emirates, Turkmenistan).




EDX Energy Data eXchange

*Download GOGI geodatabase here*

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

*View full technical report here*

Rose, K.; Bauer, J.; Baker, V.; Bean, A.; DiGiulio, J.; Jones, K.; Justman, D.; Miller, R. M.; Romeo, L.; Sabbatino, M.; Tong, A. Development of an Open Global Oil and Gas Infrastructure Inventory and Geodatabase; NETL-TRS-6-2018; NETL Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Albany, OR, 2018; p 594; DOI: 10.18141/1427573.

U.S. DEPARTMENT OF ENERGY

Mid-Atlantic Technology, Research & Innovation Center MATRIC

AECOM

**Jennifer DiGiulio,** Geologist
**jennifer.digiulio@netl.doe.gov**

Acknowledgment: This work was completed by the U.S. Department of Energy's (DOE) National Energy Technology Laboratory (NETL) in support of Cooperative Research and Development Agreement 5655 for the Environmental Defense Fund. The authors wish to thank Daniel Zavala-Araiza and Drew Nelson with EDF (Environmental Defense Fund), Ann-Cathrin Voege of Statoil, and Stephanie Saunier and her colleagues at Carbon Limits for their input and feedback that helped with formulation and execution of this study.



Science & Engineering To Power Our Future