

## **Abstract for CRADA between Environmental Defense Fund and NETL**

The production, transportation, storage, and consumption of oil and gas resources rely on a global network of infrastructure, which is maintained and regulated by numerous commercial, non-profit, academia, and local to national government entities. Thus, information about this infrastructure is often highly variable in the quality, quantity and accessibility of the data. However, integrating and standardizing these disparate data sets into a global, open-access assessment and inventory of oil and gas infrastructure can provide a framework to support advanced analyses, including predictions and assessments of global methane infrastructure and emissions risk, identify information gaps, evaluate economic costs, and support a range of decision making needs. At present, there is no single point of access to authoritative, publicly available oil and gas infrastructure data. In addition, conventional online search approaches are increasingly inefficient as the volume and variety of open-data online continues to evolve and increase. Consequently, there exists a great deal of uncertainty surrounding the extent, gaps, and quality of infrastructure data across the world. This six-month effort used a combination of expert driven and machine learning techniques to build a virtual taxonomy of global oil and gas infrastructure to provide an open data foundation for addressing a range of stakeholder questions and needs.

Here, we summarize an effort to develop a global oil and gas infrastructure (GOGI) taxonomy and geodatabase, using a combination of big data computing, custom search and data integration algorithms, and expert driven spatio-temporal analytics to identify, access, and evaluate open oil and gas data resources and uncertainty trends worldwide. This approach leveraged custom National Energy Technology Laboratory (NETL) tools and capabilities in collaboration with Environmental Defense Fund (EDF) and Carbon Limits subject matter expertise, to identify over 380 datasets and integrate more than 4.8 million features into the GOGI database. In addition to acquisition of open oil and gas infrastructure data, information was collected and analyzed to assess the spatial, temporal, and source quality of these resources, and estimate their completeness relative to the top 40 hydrocarbon producing and consuming countries.

Ultimately, the GOGI database provides a single, unified source for open oil and gas infrastructure data that can be leveraged by a range of users. The resulting geodatabase is now hosted for public access and download via the U.S. Department of Energy's, Energy Data eXchange (EDX) at <https://edx.netl.doe.gov>. Within the EDX platform, users of the GOGI database will have access to metadata supporting the data resources housed within the database and this technical report describing how the database was assembled. Finally, through the EDX platform, users will have the ability to visualize the spatial information within GOGI using EDX's open, online web mapping tool, Geocube, or download the entire database to support individual needs and analyses.