Industrial Energy Data Exploration Interface

2024-2025 Master of Environmental Data Science Capstone Project Proposal

October 18, 2024

Authors

Oksana Protsukha | oprotsukha@bren.ucsb.edu | +1 415-670-0620 |Master of Environmental Data Science Student, Bren School of Environmental Science and Management

Client

Eric Masanet | Professor and Mellichamp Chair in Sustainability Science for Emerging Technologies | Bren School of Environmental Science and Management | emasanet@ucsb.edu

Proposed Project

Objective

The project aims to develop an expanded dataset and an interactive dashboard that builds upon the Industrial Assessment Centers (IAC) database by integrating additional relevant data sources. This enhanced dataset will enable researchers to conduct more comprehensive analyses of energy consumption, air pollutant emissions, and decarbonization opportunities through energy efficiency across small to mid-size manufacturing plants in the United States.

Environmental Motivation

The client, Dr. Eric Masanet, leads a research group at the Industrial Sustainability Analysis Laboratory at the University of California, Santa Barbara, focusing on industrial decarbonization pathways through energy efficiency. This project will support their research by integrating missing information into the IAC database and developing a data exploration interface to provide a comprehensive view of industrial energy consumption and emissions. Additionally, by eliminating manual data integration and extensive Excel-based formulas, researchers can dedicate more time to analyzing energy efficiency solutions. This is environmentally important, as the industrial sector contributes 30% of U.S. primary energy-related CO2 emissions, totaling 1,360 million metric tons in 2020 [1]—equivalent to the annual energy use of about 124 million homes [2]. With the U.S. targeting net-zero emissions by 2050 under the Paris Agreement [3], it is essential to accelerate research into decarbonizing this sector.

While the project focuses on technical solutions for climate change mitigation, it may also positively impact environmental justice by accelerating the decarbonization of industrial facilities in specific locations, thereby improving air quality in fenceline communities, which are disproportionately low-income and minority populations [4].

Data Science Need

The primary data science need is to create an expanded, integrated dataset and an interactive dashboard that enhances the IAC database with additional relevant information and provides data exploration capabilities. The Industrial Assessment Centers (IAC) program, founded in 1976, accumulated more than forty years of data from thousands of audits and tens of

thousands of structured recommendations on how to optimize energy consumption within manufacturing processes [5]. Though vast amounts of IAC data enables researchers to analyze energy consumption patterns and electricity savings associated with recommended efficiency measures in small to mid-size manufacturing plants, it has several limitations.

- Lack of greenhouse gas emissions data to calculate CO2 emissions associated with the energy consumption and recommended efficiency measures at a given industrial facility.
- Limited data exploration and visualization capabilities of the IAC website.
- Suboptimal data structure which is time-consuming to analyze.
- Lack of automated updates to the dataset requiring manual updates of the analysis results to keep the analysis current.
- No inherent way to convert investment data from prior years into present-day cost equivalents.

Deliverables

Project Deliverables

The project will produce the following deliverables:

- Integrated Dataset: Expanded IAC database incorporating power grid, air pollutant emissions data, economic data, and financial metrics.
- Dashboard: Interactive platform for data exploration and visualizations of the integrated dataset.
- User Documentation: Comprehensive guides covering data structure, integration methods, and dashboard usage.

Refer to Table 2 in the Supporting Materials section for a detailed breakdown of each project deliverable.

Stretch goals

Automated Data Maintenance and Scalability: Develop a process to update the database when new data becomes available.

Audience

The primary audience for these deliverables are researchers analyzing energy demand and decarbonization opportunities through efficiency in the industrial sector.

Ongoing support and maintenance

The database and data exploration dashboard will be used by researchers to derive insights into energy consumption and decarbonization efforts within the industrial sector. The project deliverables will be maintained by the client on the client's website, The Industrial Sustainability Analysis Lab (ISAL), after project completion.

Data Access and Availability

The project will utilize publicly available datasets maintained by government agencies and educational institutions. All data can be downloaded through the provided links with some preprocessing expected as part of the project deliverables.

1. Industrial energy audits: IAC database (primary dataset)
2. Air pollutants emissions data for the electric power sector:
    a. The Emissions & Generation Resource Integrated Database (eGRID)
    b. Electric Power Annual
3. Air pollutants emissions factors from stationary sources: AP-42
4. Electricity generation data: Electric Power Annual
5. Economic data: Bureau of Labor Statistics Present Price Index data: BLS PPI

Additionally, the client will provide qualitative attributes to classify assessment recommendation codes (ARC), such as system type, equipment type, and Avoid-Shift-Improve (ASI) measures.

For a detailed overview of the data refer to Table 1 in the Data section under the Supporting Materials.

Project Requirements

The project does not have a predefined implementation approach or technical requirements. Students are encouraged to design an approach and apply data science and software engineering tools that are best suited to meet the project objectives. All tools should be open-source and easy to maintain to ensure reproducibility and ongoing maintenance.

SUPPORTING MATERIALS

References

[1] Energy.gov. “DOE Industrial Decarbonization Roadmap.” Accessed September 23, 2024. https://www.energy.gov/industrial-technologies/doe-industrial-decarbonization-roadmap.
[2] US EPA, OAR. “Frequent Questions: EPA’s Greenhouse Gas Equivalencies Calculator.” Overviews and Factsheets, December 12, 2016. https://www.epa.gov/energy/frequent-questions-epas-greenhouse-gas-equivalencies-calculator
[3] Kerry, John. “The Long-Term Strategy of the United States, Pathways to Net-Zero Greenhouse Gas Emissions by 2050,” n.d.
[4] “Health Consequences of Air Pollution.” Accessed September 28, 2024. https://www.who.int/news/item/25-06-2024-what-are-health-consequences-of-air-pollution-on-populations.
[5] “Industrial Assessment Centers.” Accessed October 3, 2024. https://iac.university/#database.

Budget and justification

The project does not require additional budget beyond $250 from the Bren School to cover the group’s basic operations and printing.

Client letter of support

Click here to view the PDF document.

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

BREN SCHOOL OF ENVIRONMENTAL SCIENCE & MANAGEMENT

SANTA BARBARA, CA 93106
http://www.bren.ucsb.edu/

October 9th, 2024

MEDS Project Selection Committee
Bren School of Environmental Science & Management
2400 Bren Hall
UC Santa Barbara, CA 93106

Dear MEDS Project Selection Committee,

I strongly support the proposed Capstone Project "Industrial Energy Data Exploration Interface". This project addresses an important need in industrial decarbonization research by developing an interface that expands upon the existing Industrial Assessment Centers (IAC) database from the U.S. Department of Energy with important additional metrics related to local air pollution and resource use.

The proposed interface will enable my team of researchers to more effectively analyze the data collected through IACs, advancing our research on industrial decarbonization through energy efficiency. It will also provide a valuable public resource for the industrial decarbonization research community.

As the client for this project, I commit to at least monthly meetings and to being responsive via email throughout the project duration (January - June 2025) to ensure consistent communication and guidance. I anticipate that no additional funding will be required to complete the project.

The Industrial Sustainability Analysis Lab (ISAL) will provide continued support and maintenance for the integrated dataset and associated dashboard through our website at http://industrial-sustainability.org/, which is currently under development.

All necessary data are publicly available via the links listed in the proposal. I am available to answer questions and provide additional data as needed.

This Capstone Project offers a valuable opportunity to contribute to the research on industrial energy efficiency. Thank you for your consideration. Please contact me if you require any further information.

Sincerely,

Eric R. Masanet

Eric Masanet

Professor, Bren School of Environmental Science and Management
Professor (by courtesy), Department of Mechanical Engineering
Mellichamp Chair of Sustainability Science for Emerging Technologies
Faculty Scientist, Lawrence Berkeley National Laboratory

PHONE: (312) 792-9776 • EMASANET@UCSB.EDU

Data

The table below provides the list of data sources that the students are expected to work with to complete the project.

Table 1

Dataset	Description	Data Format and Delivery Method
Industrial energy audits	Database containing the technical information about more than 13,000 industrial energy audits done by all the industrial assessment centers. This includes information on the type of facility assessed (size, industry, energy usage, etc.) and details of resulting recommendations (type, energy & dollars savings etc.). Technical details on the database and the description of all data points are available on the IAC website: https://iac.university/#database	Excel 2003 (xls) Download Link
The Emissions & Generation Resource Integrated Database (eGRID)	The dataset provides information on environmental attributes, including air pollutant emissions, from all electricity-generating plants in the United States that supply power to the national grid and report their data to the government. The data will be used to determine air pollutant emission quantities (CO2, SOx, NOx, etc.) avoided by ARCs that lead to electricity savings.	Excel (xlsx) Download Link
Electric Power Annual: Emissions from Energy Consumption (Table 9.5)	National and regional data on electricity generating capacity, electricity generation and useful thermal output,	Excel (xlsx) Download Link

	fuel receipts, consumption, and emissions. The dataset will be used to determine air pollutant emission quantities (CO2, SOx, NOx, etc.) avoided by ARCs that lead to electricity savings.	
Electric Power Annual: Utility Scale Facility Net Generation (EIA-923)	This dataset provides a detailed view of electric power operations in the United States such as electricity generation, fuel consumption, fossil fuel stocks, and fuel receipts. 1. Fuel consumption 2. Fossil fuel stocks 3. Fuel receipts The data will be used to calculate upstream fuel input (coal, natural gas, etc.) quantities avoided by the ARCs.	Excel (xlsx) Download Link
Compilation of Air Pollutant Emissions Factors from Stationary Sources (AP-42)	The dataset will be used to determine air pollutant emission quantities (CO2, SOx, NOx, etc.) avoided by ARCs that lead to onsite fuel savings.	PDF Download Link
Bureau of labor statistics present price index data (BLS PPI)	The dataset measures the average change over time in the selling prices received by domestic producers for their output. The dataset will be used to adjust the investments and savings to present day values in order to determine their investment cost, energy cost, and payback.	Excel (xlsx) Download Link Public Data API

Deliverable Components and Tasks

The table provides a detailed outline of the tasks expected for each project deliverable.

Table 2

Integrated Dataset	● Design a data structure that allows for easy incorporation of new data. ● Integrate IAC dataset with state-level power grid data to estimate air pollutant emissions associated with electricity use (GHG, SO2, NOx, and etc.) ● Integrate IAC dataset with fuel-specific emissions factors to estimate air pollutant emissions associated with onsite fuel use. ● Incorporate power plant fuel data to estimate upstream resource use. ● Integrate economic data to normalize historical costs into present-day values. ● Implement equations to convert investments into financial metrics such as NPV and levelized costs. ● Enhance tagging of assessment recommendation codes (ARC) with additional information on intervention types.
Dashboard	● Develop a user-friendly interface for exploring the integrated dataset. ● Create interactive features for custom queries and data filtering. ● Implement visualization tools for data analysis and presentation.
User Documentation	● Create a comprehensive data dictionary explaining all fields in the expanded dataset. ● Document the data integration process and methodologies used.

	• Provide user documentation on how to navigate and use the interface. • Include sample queries or tutorials explaining dashboard capabilities.

Glossary

ARC: Assessment Recommendation Codes provided by IAC as part of the energy audit.
Air Emission Factor: A representative value used to link the quantity of a pollutant released into the atmosphere to a specific activity related to that pollutant's emission.
GHG (Greenhouse Gas): Gasses in Earth's atmosphere that trap heat, contributing to the greenhouse effect and global warming.
IAC: Industrial Assessment Centers based at accredited engineering schools and staffed by faculty and students who perform energy, waste, and productivity audits for small and medium-sized manufacturers.
NPV: Net Present Value (NPV) is a financial metric used to evaluate the profitability of an investment or project.