The Foundational Industrial Energy Dataset (FIED): Open-Source Data on Industrial Facilities

Colin A. McMillan, Carrie Schoeneberger, Sarang Supekar, and David Thierry

1 National Renewable Energy Laboratory
2 Argonne National Laboratory

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-6A20-90442
September 2024

The Foundational Industrial Energy Dataset (FIED): Open-Source Data on Industrial Facilities

Colin A. McMillan, Carrie Schoeneberger,
Sarang Supekar, and David Thierry

1 National Renewable Energy Laboratory
2 Argonne National Laboratory

Suggested Citation
McMillan, Colin A., Carrie Schoeneberger, Sarang Supekar, and David Thierry. 2024. *The Foundational Industrial Energy Dataset (FIED): Open-Source Data on Industrial Facilities*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-90442. https://www.nrel.gov/docs/fy24osti/90442.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-6A20-90442
September 2024

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Strategic Analysis Team. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via www.OSTI.gov.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Acknowledgments

The authors acknowledge Gale Boyd (Duke University), Administrator Joseph DeCarolis (U.S. Energy Information Administration), Ozge Kaplan (U.S. Environmental Protection Agency), and Eric Masanet (University of California, Santa Barbara) as panelists in the American Center for Life Cycle Assessment 2022 Conference special session on datasets for industrial modeling, as well as Heather Liddell (Purdue University) for her help organizing and moderating the session. Thank you to Ashna Aggarwal, Patrick Bryant, and Will Dean (U.S. Department of Energy Office of Energy Efficiency and Renewable Energy [EERE] Strategic Analysis); Steve Capanna, Tomy Granzier-Nakajima, Nicole Ryan, and Jun Shepard (U.S. Department of Energy Office of Policy); Amy Jordan (Carbon Solutions, LLC); Dan Bilello, Elaine Hale, Danny Inman, and Adrienne Lowney (National Renewable Energy Laboratory) for providing reviews of the report. Thank you to Liz Breazeale for editing the report and to Sam Baldwin (EERE Strategic Analysis) and Ookie Ma (Energetics) for their early support of the project.

List of Acronyms

CBP	County Business Patterns
DOE	U.S. Department of Energy
EF	emission factor
EIA	U.S. Energy Information Administration
EIS	Emissions Inventory System
EPA	U.S. Environmental Protection Agency
FIED	foundational industry energy dataset
FRS	Facility Registry Service
GHG	greenhouse gas
GHGRP	U.S. Environmental Protection Agency Greenhouse Gas Reporting Program
MECS	Manufacturing Energy Consumption Survey
NAICS	North American Industrial Classification System
NEI	National Emissions Inventory
SCC	source classification code
SEDS	State Energy Data System

Executive Summary

The state of data on industrial energy use has co-evolved over several decades with the demands of industrial energy analysis. The most recent development—analysis in support of decarbonizing the industrial sector—has changed the characteristics of industrial data that are useful for analysts and model developers. Although data and its collection processes may be cast from a conventional viewpoint as objective and free from the influence of social dynamics, this provides an incomplete picture of not only the processes by which information is generated, but also the limitations and opportunities of data to be useful for analysis.

The foundational industry energy dataset (FIED) is a result of the confluence of trends in open data and the demand for higher resolution industrial energy analysis. The general approach to compiling the FIED involves accessing, filtering, and formatting data published by federal organizations on the internet for public use. Unlike most industrial energy datasets, which are published by the U.S. Energy Information Administration (EIA), the FIED relies on core datasets from the U.S. Environmental Protection Agency (EPA). Whereas data collection by the EIA has been opposed by industry out of concern for potential regulation, data collection by the EPA is expressly for regulating industry.

The FIED addresses several of the areas of growing disconnect between the demands of industrial energy analysis and the state of industrial energy data by providing unit-level characterization—including estimates of energy use, greenhouse gas (GHG) emissions, and design capacities—for facilities that are identified by latitude and longitude. This enables local-level analysis of existing combustion equipment, as well as regional comparisons with traditional industrial energy data estimates. The report summarizes the general logic behind compiling the FIED. The FIED itself and the Python code used to generate it are available from the Open Energy Data Initiative (https://doi.org/10.25984/2437657) and GitHub (https://github.com/NREL/foundational-industry-energy-data/), respectively.

Industrial energy data in the United States has been shaped by several decades of competing efforts to collect and withhold information about how industries use energy. After gaining momentum in the 1970s during the energy crises and the formation of the U.S. Department of Energy (DOE), efforts to systematically collect and publish detailed industrial energy data were dramatically scaled back in the early 1980s due to changing priorities and resistance from industry. The data products that emerged from these contentions not only offer less detail than the industrial datasets from the late 1970s, but are also far less detailed than what is being demanded by decarbonization analysis.

In addition to a continuation of prior attempts to overcome the limitations of conventional sources of industrial energy data, the FIED is also intended to be a building block for analysis and modeling of the U.S. industrial sector. For these reasons, the FIED should be considered part of an ongoing process of improving the understanding of industrial energy use. Despite best current efforts to provide a comprehensive facility- and unit-level dataset on industrial energy use, notable gaps remain, including data on electricity purchases and production/output. Lessons learned from other industrial energy datasets indicate that a robust constituency is needed to promote use of the FIED (potentially for industrial decarbonization analysis at the local level), as well as to refine and expand its underlying methods.

Table of Contents

List of Figures

List of Tables

1 Introduction

Analysis and data collection are mutually dependent processes. The types of data collected and the methods of collection are determined by the scope of the analysis. In a similar way, the scope of the analysis may incorporate the actual or perceived data characteristics of quality and availability. Analysis and data collection are not only dependent processes, but they are also historically contingent and shaped by the dynamic interests of policymakers and analysts.

Energy data are invariably used in energy policy analysis. Although this relationship may be framed as "independent and impartial information" that can "promote sound policymaking" (U.S. Energy Information Administration 2024a), an alternative view sees data as "socially produced and opaque" (Lave, Nost, and Goldstein 2022) and that political judgements are "implicit in the choice of what to measure, how to measure it, how often to measure, and how to present and interpret the results" (Alonso and Starr 1987). From this perspective, to cast data collection as a completely objective endeavor, free from the influence of social dynamics, provides an incomplete picture of the processes by which information is generated, as well as the limitations and opportunities of data to be useful for analysis. In actuality, "data is the hybrid offspring of technology and social practice" (Lave, Nost, and Goldstein 2022).

Calls for open data, or data that is free for anyone "to access, use, modify, and share it—subject, at most, to measures that preserve provenance and openness" (Open Knowledge Foundation n.d.), have been made in order to improve the quality of science; make collaboration between scientists and policymakers more effective; increase productivity through collaboration; and more transparently support public debate about the energy system (Pfenninger et al. 2017). For the U.S. government's open data initiative, Data.gov, open data can "inform decisions by the public and policymakers, drive innovation and economic activity, achieve agency missions, and strengthen the foundation of an open and transparent government" (U.S. General Services Administration 2024).

In the same way researchers have pursued higher-resolution data to better analyze renewable energy generation (e.g., Buster et al. 2024), analysis in support of decarbonizing the industrial sector has changed the types of industrial data that are useful for analysts and model developers. For example, the prospect of electrifying industrial processes with variable renewable electricity generation introduces the importance of having data at a much higher geographic and temporal resolution to enable analysis of grid interactions. Likewise, knowing process temperatures, in addition to having greater geographic and temporal resolution, is significant for evaluating the potential use of solar thermal technologies (Schoeneberger et al. 2020).

This report introduces and describes the foundational industry energy dataset (FIED), which was created in response to the demands for not only open data, but also for decarbonization analysis and policymaking. As we describe shortly, the openness of traditional sources of industrial energy data is constrained by its data collection procedures. These collection procedures also constrain the level of spatial, temporal, and operational resolution for analysis. These constraints have created a growing disconnect between the traditional sources of industrial energy data and what is demanded by decarbonizing the industrial sector. This disconnect either does not exist or is less significant for other areas of energy analysis, most notably for the electricity sector.

The FIED provides a blueprint for how to overcome many of the existing limitations for greater openness and increased resolution of traditional sources of industrial energy data. By not simply viewing data through a technical lens and instead treating data as the result of social practices, we expand the goals of developing the FIED beyond identifying these limitations and their potential solutions. Equally importantly, we also wanted to understand why these limitations exist in the first place.

What has been missing in much of the previous work on industrial energy data is analyzing the social, economic, and political context in which those data came into being. The state of information on industrial energy use in the United States has co-evolved with the demands of industrial energy analysis over several decades. For instance, we can trace our current collection of industrial energy data back even before the creation of the U.S. Energy Information Administration (EIA) and the U.S. Department of Energy (DOE) to the early 1970s, when the first oil crisis, subsequent changes in the natural gas and electricity markets, and new environmental and energy regulations were among the catalysts that drove the federal government to collect and study data on energy use of energy-intensive industries (Isser and Limaye 1982).

The collection of energy information is shaped not only by legislation and the organizations responsible for collection and dissemination, but also by the actors who may or may not be subject to providing information. Data openness and overall transparency are in direct conflict with private industry claims of proprietary information and the protection of trade secrets. Schnaiberg (1977) discusses claims like these as part of larger social and economic control over research on the impacts of technologies. Morin (1989) describes how the environments of different centralized statistical agencies are shaped by the level of conflict (e.g., the newness of an agency or a survey; the importance of the larger political and policy environment; and a change in information demands) and the heterogeneity of network actors.

The remainder of Section 1 provides a brief history of industrial energy data, then describes the resolution gap(s) and summarizes current perceptions of industrial energy data. Section 2 details how the FIED is constructed and provides various descriptions of its characterization of industrial energy use, including unit-level descriptions and comparisons to existing data sources. Additional details on the logic underlying various aspects of the FIED are provided in the Appendices. Section 3 concludes the report and identifies several paths of future activities to improve and socialize the FIED.

1.1 A Brief History of Industrial Energy Data

While it may be attractive to ascribe the lack of detail in publicly available industrial energy data to technical characteristics of the sector (e.g., heterogeneity of industrial processes and equipment), or the proprietary nature of data collected from industrial firms, the issue is much more complex. A deep discussion of how we got here for industrial energy data is introduced by this report, although it is not the primary focus. It is important to establish that industrial energy data collection processes have a history, and this history has involved contested processes that have shaped what types of data are permissible to collect and to publish.

The EIA was established with DOE in 1977, with staff from the Federal Energy Administration, the Bureau of Mines, and the Federal Power Commission (U.S. Energy Information

Administration 1980). Along with personnel, the EIA came into existence with files, collection instruments, and statistical reports from these predecessor organizations. As Lincoln E. Moses, the first EIA Administrator, put it in an annual report to Congress, "[n]ew data collections of considerable interest and importance have been started, but the historical files, that is to say, the earlier years of many EIA's data series still bear, **as they must always, the stamp of their origins** [emphasis added]" (U.S. Energy Information Administration 1980).

Before detailing the existing sources of industrial energy data and how they are collected in Section 2.1, it is instructive to briefly review the origins of industrial energy data and the efforts over the decades to improve industrial energy data. We frame our short discussion in terms of two historical periods. The first ran on an initial period of interest and development of the first national estimates of industrial energy use. Until recently, this period has represented the high-water mark of momentum to increase not only the coverage of industrial energy estimates, but also the level of detail of industrial energy data. This period ran roughly from 1973 until the beginning of 1981, with the most substantial activities occurring in the second half of the period. However, the many examples of datasets and analyses developed in the period of rapid expansion became ad hoc contributions that were never updated. This expansion was followed immediately by a period of contraction, when EIA and DOE budgets were reduced dramatically,[1] and stabilization with the first Manufacturing Energy Consumption Survey (MECS) published in 1986. Around 2013, efforts to use non-EIA data to estimate facility-level energy use became established.

[1] DOE’s solar thermal energy program drove many of the efforts in the late 1970s to improve the characterization of industrial energy use. Beginning in Fiscal Year 1981, the Solar Energy Research Institute’s mission and activities were redirected from demonstration and commercialization activities to high-risk R&D activities to align with the new administration's priorities. This redirection came with a 25% reduction in staff and a nearly a 53% reduction in budget from the previous fiscal year (U.S. General Accounting Office 1983).

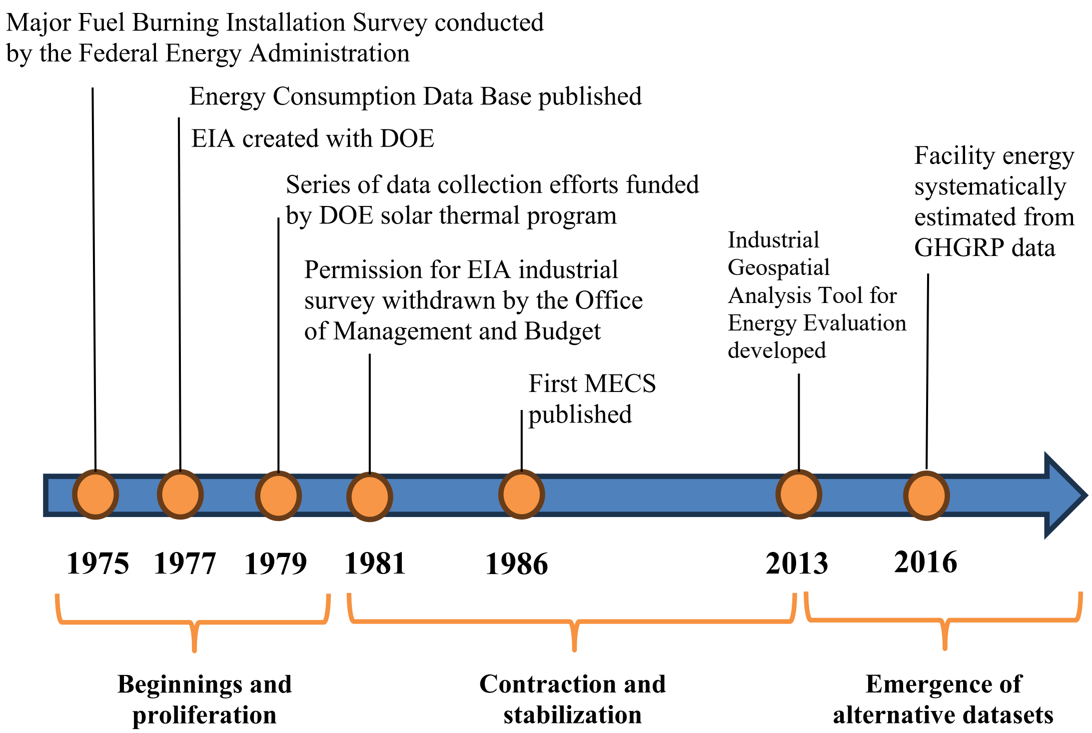


Figure 1. Timeline of significant events in the history of U.S. industrial energy data discussed in this report

1.1.1 Industrial Energy Data: Beginnings and Proliferation (1975–1981)

Development of industrial energy datasets and related analyses[2] proliferated in the mid to late 1970s (see reviews by Krawiec et al. 1981; Isser and Limaye 1982). These projects tended to be funded by DOE and its precursor agencies, as well as private companies. The Energy Consumption Data Base[3] was the first comprehensive compilation of industrial energy data, as well as data from other end use sectors. The overall result was the culmination of a series of projects that began in 1975 with the Federal Energy Administration, included the Energy

[2] The DOE Office of Science's Office of Scientific and Technical Information database (U.S. Department of Energy 2024) contains the abstracts and reference information of many of these reports. However, not all these records include the full-text documents.

[3] Like other forms of data, the energy data that underlie energy models are rooted in institutional and technological histories that have shaped their development and use. For example, the Energy Consumption Data Base of the mid-1970s can be linked to the 2022 Industrial Demand Module of the EIA's National Energy Modeling System. National Energy Modeling System documentation identifies the Consolidated Impacts Modeling System (CIMS) as a source of technology data for the aluminum, iron and steel, glass, pulp and paper, and cement and lime industries (U.S. Energy Information Administration 2022a). CIMS originally referred to the Canadian Integrated Modeling System, which was developed at Simon Frasier University in the mid-2000s using modules from the Industrial Sector Technology Use Model (Bataille 2005). The Industrial Sector Technology Use Model came to Simon Frasier University in the mid-1980s (Jaccard and Roop 1990) from Pacific Northwest National Laboratory, who had developed a personal computer version (Roop and Kaplan 1984) from the original mainframe program first developed in the late 1970s by Energy and Environmental Analysis for DOE (Bohn and Herod 1978). The Industrial Sector Technology Use Model's underlying data were based on the Energy Consumption Data Base (Isser and Limaye 1982).

Research and Development Administration, and concluded with DOE (Isser and Limaye 1982). The Energy Consumption Data Base was developed not as a survey, but as a compilation of existing sources from federal agencies (e.g., Census of Manufacturers and the Annual Survey of Manufacturers), as well as industry trade associations (Isser and Limaye 1982).

It is notable that the Energy Consumption Data Base estimated energy for non-manufacturing sectors (i.e., agriculture and mining; the EIA does not currently have end-use surveys for these sectors, although some energy use data is collected by the U.S. Department of Agriculture and the Census Bureau), as well as the manufacturing, but also included end uses (referred to as "functional uses") for all three sectors separately. These end uses included direct heating applications for at least three temperature ranges, as well as process steam, for the manufacturing and mining sectors (Isser and Limaye 1982). Many additional resources were used to estimate disaggregated energy by end use (e.g., Energy and Environmental Analysis 1978; Energy and Environmental Analysis, Inc 1977; Energy and Environmental Analysis, Inc. 1977a; 1977b), in part because of the aggregated nature of data obtained from the Bureau of the Census and other federal sources.

Likely due to a combination of policy interests at the time, which included fuel switching and energy conservation measures, and the limitations of existing industrial energy datasets from federal agencies (i.e., lack of operational and geographic detail and disclosure restrictions), much of the initial work to develop industrial energy datasets was focused on developing more detailed estimates of industrial energy use, particularly at the end use and unit process level. Several notable examples of these detailed studies include Drexel studies (e.g., Brown et al. 1976; 1977).

One of the first large-scale efforts to specifically collect industrial energy data directly from facilities was the Major Fuel Burning Installation Survey, conducted in 1975 by the Federal Energy Administration. The survey was sent to facilities with combustion equipment larger than 99 MMBtu/hour and was motivated by the Energy Supply and Environmental Coordination Act, under which the Federal Energy Administration was authorized to force these facilities to switch from burning oil or natural gas to burning coal (Energy and Environmental Analysis, Inc. 1979).

During this time, several solar research organizations recognized the growing research activity related to industrial energy use and proposed a coordinated effort to manage data collection (Green 1979). This effort also described many limitations and challenges to data collection at the time that we still face today. For instance, in order to evaluate a new industrial technology, it is helpful to have process data on the operational schedule, process temperature, and peak electric demand, as well as information on company investment criteria, local regulations, and management perceptions of new technology. Green (1979) also discussed how data collection activities should be centralized and coordinated to avoid duplicative efforts and overburdening industrial respondents, as well as the limitations of government surveys that must be approved by the Office of Management and Budget.

The result was the Industrial Energy Data Collection, a voluntary effort among the organizations that were involved with industrial energy research. At the time of the report, there were 30 members of the Industrial Energy Data Collection who were tracking contacts at 39 industrial trade associations and 150 individual industrial facilities. The Industrial Energy Data Collection proposed a comprehensive set of data collection fields that not only provided details on how major processes were operated (e.g., data fields for supply and return process temperatures, scheduled and unscheduled downtime, and equipment rated capacities), but also economic factors (e.g., discount rate, payback period, age of plant, rural or urban location, and value of shipments) and institutional factors (e.g., perspectives on fuel costs, preferred incentives, existence of an energy management program, and persons responsible for energy-related investments) (Green 1979).

1.1.2 Contraction and Stabilization (1981–2015)

The ambition and momentum that had been building in the late 1970s crashed headlong into the priorities of a new executive administration and coordinated resistance from individual industrial corporations (notably General Motors), as well as industrial trade organizations (Morin 1989). The beginning of this period reflects an overall shift in federal statistical policy, driven by the Office of Management and Budget, that involved: (1) reducing the Information Collection Budget; (2) increasing the number of agencies and their forms that must comply with the Office of Management and Budget regulations; and (3) reducing the scope of federal information collection overall (Morin 1989).

Authority to collect data can be granted, rescinded, or otherwise changed, as was the case of the EIA's Manufacturing Industries Energy Consumption Study and Survey of Large Combustors (U.S. Energy Information Administration 1983).[4] The survey, which began in 1980, also represents a turning point in industrial energy data collection. The survey was scoped with input from many other federal agencies (e.g., the U.S. Environmental Protection Agency [EPA]; the Department of Commerce Bureau of the Census and Bureau of Economic Analysis; the Department of Labor's Bureau of Labor Statistics; the General Accounting Office; and the Office of Technology Assessment of the U.S. Congress). However, the survey continued into a new presidential administration and in 1981, the Office of Management and Budget, led by a new director, withdrew permission for the survey, stopping it before it was completed (U.S. Energy Information Administration 1982).

During this period, the EIA was also forced to discontinue funding the Annual Survey of Manufacturers, which was, and still is, conducted by the Census Bureau. From 1978 to 1981, the EIA used the Annual Survey of Manufacturers to collect industrial energy consumption data. This time, the change was brought about from budget cuts: between the 1981 and 1984 fiscal years, EIA's budget and staff were reduced by 38% and 41%, respectively (U.S. General Accounting Office 1984).

---

[4] More recent events related to attempts to collect data from cryptocurrency mining facilities demonstrate how opposition from industry groups continues to play a role in the social processes of energy data collection. The Office of Management and Budget approved at the beginning of 2024 an EIA survey to collect data from bitcoin miners under the emergency approval provisions of the Paperwork Reduction Act (U.S. Energy Information Administration 2024b). By the end of February, a federal judge in Texas issued a temporary restraining order against DOE (Texas Blockchain Council v. Department of Energy 2024) based on a lawsuit filed against DOE and the EIA by the Texas Blockchain Council, a nonprofit industry association based in Texas, and Riot Platforms, whose Rockdale facility has been pegged as the most electricity-intensive Bitcoin mining facility in the United States (Dance et al. 2023). On March 7, 2024, the EIA withdrew the information collection notice, discontinuing the intended survey (U.S. Energy Information Administration 2024c).

The Manufacturing Industries Energy Consumption Study and Survey of Large Combustors is also important because the experiences gained by individual companies, industrial trade groups, and other industry associations in opposing the survey enabled them to become better coordinated in their responses to future data collection requests. This, in turn, enabled them to shape the MECS survey (Morin 1989). As early as 1984, the EIA began planning a new industrial energy consumption survey that would be conducted in 1986 (U.S. General Accounting Office 1984). At this point, the EIA had conducted the Residential Energy Consumption Survey annually from 1978 through 1982 and the Nonresidential Buildings Energy Consumption Survey (later to become the Commercial Buildings Energy Consumption Survey) in 1979 and 1983 (U.S. General Accounting Office 1984).

The first year of MECS data collection was 1985, following a pilot survey conducted in 1984. Although it was not clear during the planning stage if the EIA or Census Bureau would conduct the actual survey, Census was ultimately chosen because of the benefits of consistency and use of the Annual Survey of Manufacturers sample frame, as well as the strict confidentiality protection provided under Title 13 USC § 9 (U.S. General Accounting Office 1984). Industries were concerned that data collection would lead to new regulations by the EIA; the confidentiality protection, however, limited the use of facility-level data to agents of the Census Bureau (Morin 1989). Stanford University's Energy Modeling Forum (1987) later had the following to say regarding the relationship between industry and MECS:

> Nearly all industrial users and industrial trade associations have opposed the MECS survey and would certainly oppose any expansion of that survey…The cost of preparation, the proprietary nature of energy data, and the lack of offsetting benefits to the companies involved are the key reasons for industrial opposition. (p. xiii – ix)

1.1.3 Emergence of Alternative Datasets (2013–Present)

Since 2013,[5] interest and activities related to industrial energy data collection and reporting have focused on using non-EIA data to characterize facility-level energy use. The Industrial Geospatial Analysis Tool for Energy Evaluation estimates facility energy use based on data collected through the Industrial Assessment Center program (U.S. Department of Energy 2024a), proprietary industrial dataset, and MECS (Alkadi et al. 2013; Hale et al. 2018). Other efforts have been the result of increasing focus on collecting data on GHG emissions from industrial facilities and decarbonizing industrial process heat (e.g., McMillan et al. 2016; McMillan and Ruth 2019).

1.2 The Industrial Energy Resolution Gap(s)

Although energy analysis interests have evolved over time, apart from MECS, the EIA has by and large collected energy data from a supply perspective through fuel-based surveys. In 1992, the Committee on the National Energy Modeling System advised DOE and the EIA that a significant effort was needed to expand information collection, especially related to the use of energy (National Research Council 1992). The Committee explicitly called out the need for disaggregated data on the current stock of technologies, as well as data that characterized

[5] A report detailing an Industrial Large Energy Users Database was published in 2006 (McKane et al. 2006), but this appears to be an ad hoc effort that was not subsequently used for analysis, or updated after its original publication.

existing and new energy-using technologies. The Committee also distinguished data requirements of renewable energy technologies, including the need for more site- and technology-specific representations. Nearly 30 years later, the 2021 Infrastructure Investment and Jobs Act (commonly known as the Bipartisan Infrastructure Law) laid out improvements to industrial energy data. The Bipartisan Infrastructure Law directed the EIA Administrator to expand the scope and increase the frequency of its end-use surveys, as well as use new methods and tools to collect more comprehensive data (*Infrastructure Investment and Jobs Act*, 42 U.S.C. § 18773 2021).

The Bipartisan Infrastructure Law directions to the EIA (e.g., collecting information on demand response, disaggregating process heat demand by temperature, and expanding data resolution overall) reflect the disconnect between the demands of analysis for industrial decarbonization and the current state of industrial energy data collection and publication. These demands have been growing for some time, as evidenced by previous calls to expand data collection on the use of energy. The disconnect between the current state of industrial energy data and the demands of decarbonization analysis can be characterized by three categories[6]:

1. **Spatial:** facility location at the sub-state level, either at the county- or ZIP-code level, or locating with latitude and longitude
2. **Temporal:** sub annual level, such as hourly electricity load data, that captures the variation in energy use and production
3. **Operational:** characterizing industrial processes and associated equipment, including material and energy inputs and outputs (i.e., production), process temperature, equipment vintage, and equipment lifetime.

There is possibly no greater juxtaposition to the resolution of industrial energy data than electricity generation data. This juxtaposition is a result of differences in data collection and dissemination processes, which can be captured using several indicators. First, surveys of electricity generation have allocated to them a disproportionate amount of time, represented in terms of respondent burden hours,[7] compared to the surveys of industrial energy use. The industrial sector accounts for 24% of total primary energy,[8] yet only 5% of EIA's Office of Management and Budget-approved information collection activities' total burden hours (34,565

---

[6] Although these data features have emerged in the context of decarbonization analysis, they are by no means new interests. For example, Isser and Layme's (1982) review of industrial energy databases for electric utilities was motivated by new process technologies, industrial heat pumps, and load management systems. Electric utilities saw then the need for data on type and size of loads; technological changes and their effects on loads; process demand profiles; and firm ability and willingness to adopt new technologies.

[7] Burden hours represent the time required to generate, maintain, or provide information for a collection. The Paperwork Reduction Acts of 1980 and 1995 give the Office of Management and Budget authority over federal agency information collection (U.S. Office of Personnel Management 2011). Burden hours are available from https://www.reginfo.gov/public/do/PRAMain.

[8] Based on 2023 data for the industrial sector from the Monthly Energy Review (U.S. Energy Information Administration 2024e). With purchased electricity and electricity losses included, the industrial sector accounted for 33% of U.S. primary energy. The manufacturing sector accounts for the majority of primary energy used by industry (agriculture, construction, mining, and manufacturing).

of 611,194 hours as of June 20, 2024) was devoted MECS. Electricity generation, or 34% of total primary energy, was the focus of 33% (202,322 hours) of EIA's information collection through various electric power and renewable electricity surveys. The three power plant operations surveys alone (EIA-923, EIA-923 [Monthly], and EIA-923 [Supplemental]) accounted for 16% (100,302) of EIA's total Office of Management and Budget -approved burden hours.

The second set of indicators of the juxtaposition of industrial energy data and electricity generation are in terms of how frequently data are collected and published, and the resolution of data. While MECS currently provides quadrennial estimates of annual manufacturing energy use, the Hourly Electric Grid Monitor (U.S. Energy Information Administration 2024) provides hourly operating data of the high-voltage bulk electric power grid across the 64 balancing authorities of the contiguous United States. Additionally, unlike electricity generation, which is subject to facility and even unit-level reporting that includes data on generation, revenues, and fuel purchases (i.e., the EIA-860, EIA-861, and EIA-923 surveys), results of MECS cannot be released publicly at a level of detail that would identify an institution, organization, or business, as specified by Title 13 U.S.C. § 9 (U.S. Energy Information Administration 2020). This disclosure law sets a limit of data openness not only for MECS, but for all information collected by the Census Bureau. The implications for MECS data resolution are not even the withholding of facility-level detail; energy estimates are reported at no higher geographic resolution than census region (i.e., Northeast, South, Midwest, and West)[9] and energy use is reported at the 6-digit North American Industrial Classification System (NAICS) code level for only 46 of the 360 NAICS codes[10].

1.3 Expert Perceptions of Current Industrial Energy Data

Engaging with experts in the industrial energy systems field was part of the FIED development process. Four experts from DOE, EPA, EIA, and academia participated in a special session on developing datasets for industrial modeling that was held at the American Center for Life Cycle Assessment Conference in November 2022. The session consisted of short presentations by panelists, followed by a panel discussion related to data gaps, data quality, and relationships between federal databases.

Common themes on the types of demands and limitations for industrial energy data emerged from the session. These themes were largely consistent with our framing of the features of industrial energy data demanded by decarbonization analysis. The experts in the session

---

[9] The 1994 MECS is an anomaly in that its sample size was expanded by 40% compared to the 1991 MECS, which enabled EIA to provide results at the 9-unit census division level (U.S. Energy Information Administration 1997). Note that this was also the last time that MECS was conducted as a triennial survey and not on its current quadrennial schedule. That said, the Energy Policy Act of 1992 had specified that MECS was to be conducted "at least once every two years" (Energy Policy Act of 1992, Pub L. No. 102-486 § 171, 106 Stat. 2864). Later, the Consolidated Appropriations Act of 2014 changed MECS to "once every four years" (Consolidated Appropriations Act of 2014, Pub L. No. 113-76 § 315, 218 Stat. 177).

[10] Although these 46 industries are undoubtedly responsible for the majority of manufacturing energy use (MECS methodology states that the most-energy intensive industries are represented at the 5- and 6-digit NAICS [U.S. Energy Information Administration 2020]), the argument for comprehensive coverage of the manufacturing sector is one of increased operational resolution, as well as data openness.

emphasized that data to support decarbonization roadmaps and decarbonization modeling should be technology-rich, including granular process detail, technology readiness level values over time, and cost accounting. For example, modeling new technology adoption requires a focus on their integration with existing processes at a systems-level. Additionally, while technology readiness level values are often included in technology characterizations, there is a gap in understanding and analysis on what it takes for technologies to achieve higher technology readiness levels. Similarly, it was noted that many industrial modeling efforts stress the precision of baseline data, whereas there should be increased efforts in quantifying the implications of technology transformations. Obtaining best estimates of emerging technology cost data and future cost projections is especially important for modeling these transformations and identifying the cost factors that will speed the transition.

The second common theme of the session centered around limitations of federal, publicly available databases and the lack of connectivity between them. Multiple federal agencies (EPA, EIA, Bureau of Economic Analysis, Census Bureau) report data that are used by industrial energy modelers, but the coverage of industrial data in each database varies according to reporting requirements and is updated in different years, making it difficult to calibrate results among models that use the data. While some public databases report industrial facility-level data for a portion of facilities, other databases aggregate surveyed data into relevant categories. No public database of every industrial facility exists, but the Census Bureau provides economic micro-data on the production and energy use of industrial products at the facility-level that can be through an application process (U.S. Census Bureau 2024a). Furthermore, because databases are maintained under different agencies and are designed to report data specific to the agency's purpose (e.g., emissions of air pollutants, energy consumption, gross output by industry), there is a lack of connectivity between them. This issue adds a significant amount of extra analysis and uncertainty for modelers using data from multiple databases in tandem. The issue also reveals a need to create a standardized industrial facility dataset that links data from existing sources.

The frequency of reporting and publication was highlighted as a specific limitation of existing industrial energy data. For example, MECS data is released every 4 years, and in the year it is released (most recently published in 2021), the data refers to a calendar year several years prior (most recent data is for the year 2018). National Emissions Inventory (NEI) data is released every 3 years (with the most recent data published in 2023 and referring to the year 2020). The different timelines for reporting not only make reconciling data difficult, but the long timespans between releases due to processing do not allow for timely analyses. It was noted that frequent releases of more current data would allow for better trending of new technology adoption. In fact, EIA is already planning to conduct supplemental surveys in between main survey years to monitor key trends and inform future questions that could be added to surveys. The additional questions could target the level of growth in emerging technologies, such as hydrogen generation and use and renewable energy purchases.

2 FIED

2.1 FIED Data Sources and Compilation

The FIED is a continuation of previous efforts to overcome limitations of existing publicly available sources of industrial energy data by developing facility- and county-level estimates (Wachs et al. 2022; McMillan and Ruth 2019; McMillan 2019; McMillan and Narwade 2018; McMillan et al. 2016). The FIED also builds on the approach used to create an inventory of industrial boilers in the United States (Schoeneberger et al. 2022) and to develop a decarbonization analysis framework for industrial process heat (Schoeneberger 2023).

Unlike other efforts to assemble facility-level industrial energy data (e.g., Modak et al. 2015), the FIED relies exclusively on publicly available data. The FIED is available for download from the Open Energy Data Initiative (https://doi.org/10.25984/2437657); the Python code used to assemble the FIED is available in a public GitHub repository (https://github.com/NREL/foundational-industry-energy-data/). At this time, the FIED is a historical dataset for 2017 and does not include projections of energy use, or cost or performance of technologies

The majority of the unit-level estimates of energy use presented by the FIED exist not because they were explicitly collected as energy data, but were instead collected in the NEI as part of regulatory requirements for reporting air emissions from fuel combustion and other processes. Information about the fuel used and the unit type, along with conversion factors in the form of emissions factors (e.g., kilogram CO_2 equivalents [kg CO2e] emitted per megajoule [MJ], or natural gas combusted kg sulfur dioxides [SOx] emitted per MJ diesel fuel combusted), have made it possible to systematically estimate energy use at the unit level. The use of these alternative data sources in lieu of traditional sources of energy data highlights how legislation and other social factors shape information collection.

In addition to a continuation of prior attempts to overcome the limitations of conventional sources of industrial energy data, the FIED is intended to be a building block (Pfenninger 2024) for analysis and modeling of the U.S. industrial sector. For these reasons, FIED should be considered part of an ongoing process, and not a final outcome, of improving the understanding of industrial energy use. We have attempted to clearly document the underlying data sources, code, and basic logic of developing energy data on the levels of individual units within individual facilities.

The general approach to compiling the FIED involves accessing, filtering, and formatting data published by federal organizations for public use (see Appendix B for an overview of the compilation process). The FIED relies on three core datasets from the EPA, as summarized in Table 1: the Facility Registry Service (FRS), the Greenhouse Gas Reporting Program (GHGRP), and NEI. The FRS was created as a way of integrating various federal and state environmental regulations for air, water, and waste into a single record (U.S. Environmental Protection Agency 2011). Each entity reporting to the EPA is assigned an FRS identification that can be cross-referenced to other EPA program databases, including those for compliance history and enforcement actions (i.e., U.S. Environmental Protection Agency 2024a). The GHGRP refers to EPA's Mandatory Greenhouse Gas Reporting (40 C.F.R § 98) from facilities with direct GHG emissions that exceed 25,000 metric tonnes per year, or certain production processes that emit GHGs (e.g., lime and cement production, aluminum production), as well as suppliers of fossil fuels and GHGs (e.g., carbon dioxide, sulfur hexafluoride). The NEI is the EPA's triennial compilation of air emissions data on criteria pollutants, criteria precursors, and hazardous air pollutants provided by state, local, and Tribal agencies for point, nonpoint, on-road, and nonroad sources.

Table 1. Core Datasets Used To Compile the FIED

Name	Publisher	Purpose	Use in FIED
FRS	EPA	Integrate facility environmental regulations reporting into a single record	Facility location, facility name, facility NAICS
GHGRP	EPA	Report GHG emissions for direct emitters of at least 25,000 MTCO2e annually	Unit characterization, derivation of energy use, reported GHG emissions
NEI	EPA	Compile air emissions (i.e., criteria pollutants, criteria precursors, and hazardous air pollutants) reported to state, local, and Tribal agencies	Unit characterization, derivation of energy use, derivation of GHG emissions

Of the three core datasets used to compile the FIED, the FRS provides the most comprehensive coverage of industrial facilities in terms of location and industrial classification. The FRS has also benefitted from efforts to map facility locations and standardize data fields. Other EPA databases, such as the GHGRP and NEI, contain industrial site information but only within the bounds of their emissions reporting. However, they each contain useful facility, process, and unit information, as described below. The appendices provide more detailed information on how the FIED develops energy and throughput estimates from the NEI, as well as the GHGRP.

Table 2 provides an additional summary of the datasets used to construct the FIED, as well as a comparison with existing conventional sources of industrial energy data. The Quarterly Survey of Plant Capacity Utilization also appears in the table based on its use in the FIED for estimating average weekly operating hours.

Table 2. Comparison of FIED Underlying Data and Conventional Sources of Historical Industry Energy Data for the United States

Source Data	FIED			Traditional				
	FRS	GHGRP	NEI	Quarterly Survey of Plant Capacity Utilization	MECS	State Energy Data System (SEDS)	Annual Survey of Manufacturers[a]	Economic Census
Owner	EPA	EPA	EPA	Census Bureau	EIA	EIA	Census Bureau	Census Bureau
Primary Data Purpose	Identify facilities, sites, or places subject to environmental regulations or of environmental interest	Report GHG emissions for direct emitters of at least 25,000 $MTCO_2e$ annually	Estimate air emissions	Provide statistics on the rates of capacity utilization	Report manufacturing energy consumption and related information	Report energy supply, consumption, prices	Provide key intercensal measures of manufacturing activity, products, and location	Describe the structure and functioning of the U.S. economy
Reporting Frequency	N/a	Annual	Triennial	Quarterly	Quadrennial	Annual	Annual	Quinquennial
Industry Resolution	Up to 6-digit NAICS (11, 21, 23, 31–33)	6-digit NAICS (11, 21, 23, 31–33)	Up to 6-digit NAICS (11, 21, 23, 31–33)	Up to 6-digit NAICS (31-33)	Up to 6-digit NAICS (31–33)	Aggregated industry	Up to 6-digit NAICS (31–33)[b]	Up to 6-digit NAICS (11, 21, 23, 31–33)
Temporal Resolution	N/a	Annual	Annual	Weekly, quarterly	Annual	Annual	Annual	Annual
Geographic Resolution	Latitude, longitude	Latitude, longitude	Latitude, longitude	National	Census region	State	County[c]	County[c]
Technology Resolution	Facility	Unit	Unit	N/a	Energy end use (e.g., machine drive, conventional boilers)	N/a	N/a	N/a
Data Disclosure Rules	N/a	40 C.F.R. § 2, subpart B[d]	N/a	13 U.S.C. § 9	13 U.S.C. § 9	1 0 C.F.R § 1004.11 and 18 U.S.C. § 1905	13 U.S.C. § 9	13 U.S.C. § 9
Office of Management and Budget Burden Hours in 2017 (annual hours per response)[e]	N/a	66.7[f]	N/a[g]	8.3	6.9[h]	N/a[i]	3.7	1.3

[a] The Annual Survey of Manufacturers has been integrated with and replaced by the Annual Integrated Economic Surveys. Data collection for the full-scale implementation of Annual Integrated Economic Surveys is beginning in calendar year 2024.

[b] Data may be withheld at more disaggregated NAICS codes to avoid disclosure.

[c] Data may be withheld at the county level to avoid disclosure.

[d] Determination of Confidential Business Information has been made by Subpart and data element (U.S. Environmental Protection Agency 2018).

[e] Burden hours represent the time required to generate, maintain, or provide information for a collection. The Paperwork Reduction Acts of 1980 and 1995 give the Office of Management and Budget authority over federal agency information collection. Federal agencies must publish 60-day Notices in the Federal Register of estimates of public burden hours for information collection (U.S. Office of Personnel Management 2011).

[f] Represents calculation using 2016 GHGRP information (U.S. Environmental Protection Agency 2015). No information collection requests were found in the Federal Register for the 2017 or 2018 GHGRP.

[g] The NEI is based on information reported to the Emissions Inventory System (EIS) by state, local, and Tribal air agencies and not emissions information collected directly by the EPA (U.S. Environmental Protection Agency 2024b).

[h] Burden hours reported for 2018 MECS (U.S. Energy Information Administration 2017).

[i] SEDS is a compilation of EIA surveys conducted by individual fuel.

In addition to the steps for acquiring and formatting the FIED's underlying data, a portion of this source data is used to derive estimates of unit-level energy use and throughput. The FIED expands previous efforts to use publicly available EPA datasets to fill gaps in industrial energy data for the United States. These include the estimation of facility-level combustion energy demand (McMillan et al. 2016; McMillan and Ruth 2019; McMillan et al. 2021) and county-level industrial energy demand (McMillan and Narwade 2018). The FIED adds to these efforts by refining existing methods and expanding their scope by characterizing individual units within facilities.

The overall logic for deriving energy and throughput (e.g., metric tonnes of material processed) estimates in the FIED is that air emissions and consumption of fuels are symmetric. In other words, as shown in Figure 2, the emission factors (EFs) that are used by GHGRP and NEI reporters to convert fuel consumption into emissions can be used in reverse, to convert emissions to fuel consumption and throughput (see Appendix E for details on the handling of disparate NEI EFs). How well this theory applies in practice depends on the unit or process producing emissions, whether a custom conversion factor is used, and the extent of supporting data reported. McMillan and Ruth (2019) estimated the uncertainty in energy estimates derived from GHGRP emissions associated with variations in reported higher heating value and carbon content. They found uncertainties for energy estimates as high as ±39% for the combustion of wood and wood residuals and as low as ±5 for the combustion of bituminous coal. At least in terms of the GHGRP, this logic has been shown to result in relatively similar energy estimates as conventional sources of industrial energy data estimates, such as MECS (Wachs et al. 2022).

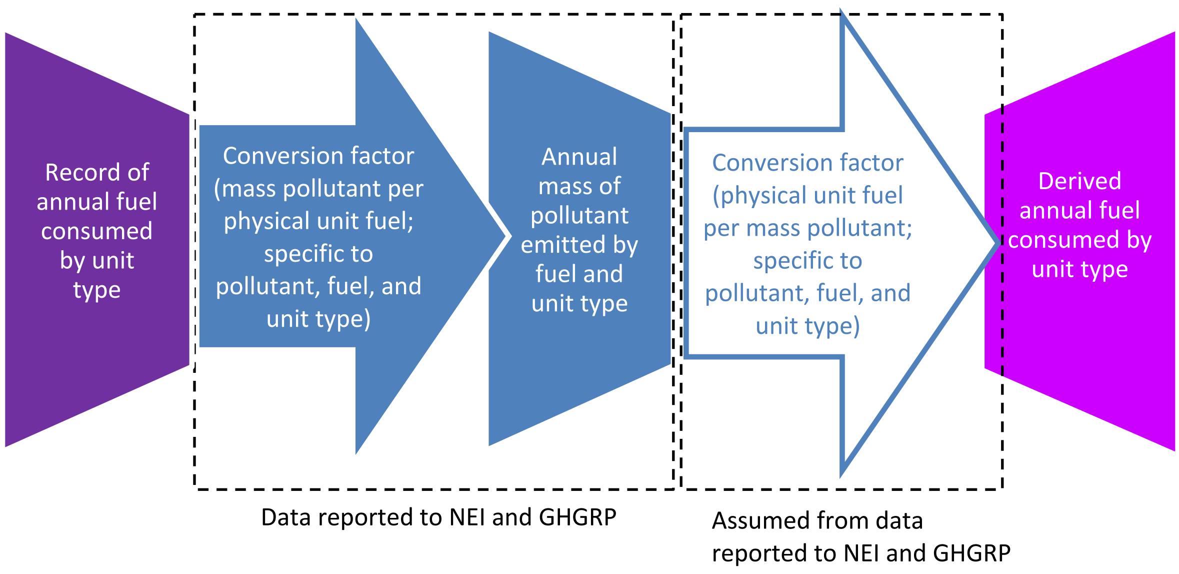


Figure 2. Generalized logic underlying FIED's energy and throughput estimates derived from reported emissions data

The NEI, in conjunction with the GHGRP and other data sources, has been used to characterize the industrial boiler population by their industrial subsector, county, heat capacity, and fuel type (Schoeneberger et al. 2022), as well as to estimate local air quality benefits from industrial decarbonization technologies (Jordan et al. 2024). The NEI is a compilation of state, local, and

Tribal air emissions inventories. These inventories may have varying levels of uncertainty based on the pollutant and its source (Miller et al. 2006).

Deriving energy and throughput estimates from the NEI is complicated by the ways that information on emissions factors, unit types, and fuel types is reported. In terms of emissions factors, there are 23 different calculation methods for reporting emissions, and these range in data collection rigor and unit specificity from continuous emissions monitoring to engineering judgement. With far fewer pollutants, a well-defined reporting tier system, and a single-point source of data collection, the GHGRP data provide a much more straightforward approach for estimating energy use than the NEI, as shown in Appendix D, Appendix F, and Appendix G. The method for deriving energy estimates from GHGRP-reported emissions has also benefitted from several years of accumulated knowledge about the data.

When we applied the equivalent approach to derive energy estimates from NEI data, we had expected that the different pollutants reported in the NEI would yield roughly an equivalent amount of energy for a given combustion unit. However, this was often not the case, as indicated by two example units shown in Table 3. In the first example, the amount of natural gas derived for SP Fiber Technologies Southeast LLC's combined cycle turbine ranges from 7.4587E+5 MMBtu to 9.382E+10. This facility also reports to the GHGRP; the derived amount of natural gas using that approach is 8.375E+5 MMBtu.

The range of derived energy is even larger for the second example in Table 3. Although the estimates using carbon monoxide and sulfur dioxide emissions are relatively close, equal to approximately 1.25E+8 gallons of distillate oil, the volatile organic compounds EF results in an estimate of only 117 gallons. Note that these emissions were calculated using engineering judgement and that the Anchor Oxford Plant is not a GHGRP reporter.

We developed various approaches to identify and adjust emissions factors that appeared to be incorrect or resulted in unreasonable energy estimates, as described in Appendix E. Additionally, the FIED reports energy and throughput estimates derived from NEI data as a range of values (minimum, median, and third quartile) in order to demonstrate the uncertainty that may be reflected in the estimates.

Table 3. Sample 2017 NEI Records Demonstrating the Inconsistencies That Can Result From Estimating Energy Use From Reported EFs of Different Pollutants

Site Name	Unit Type	Fuel Type	Pollutant	Reported Total Emissions (tons)	Calculation Method	Reported EF	Back-Calculated Fuel Amount
SP Fiber Tech-nologies Southeast LLC	Combined cycle (boiler/gas turbine)	Natural gas	Carbon monoxide	30.58	EPA EF (no Control Efficiency used)	0.08 lbs / MMBtu	7.645E+5 MMBtu
			Nitrogen oxides	137.24	State / Local / Tribal (S/L/T) EF (no Control Efficiency used)	0.368 lbs / MMBtu	7.4587E+5 MMBtu
			Sulfur dioxide	74.59	EPA EF (no Control Efficiency used)	1.59E-4 lbs / MMBtu	9.382E+10 MMBtu
Anchor Oxford Plant	Direct-fired dryer	Distillate oil	Carbon monoxide	3.24	Engineering judgment	5.00E-5 lbs / 1,000 gal	1.296E+8 gallons
			Nitrogen oxides	1.63	Engineering judgment	2.00E-5 lbs / 1,000 gal	1.63E+8 gallons
			Sulfur dioxide	0.25	Engineering judgment	4.00E-6 lbs / 1,000 gal	1.25E+8 gallons
			Volatile organic compounds	0.02	Engineering judgment	0.34 lbs / 1,000 gal	117 gallons

2.2 FIED data fields

The FIED is built on a structure of data fields, as shown in Figure 3. At the highest level, the site information encompasses locational detail of industrial facilities at varying degrees, from latitude and longitude to census region. This situates geographical information of industrial facilities in a single comprehensive dataset. Site data can also be cross-referenced to other relevant geographical data, such as renewable resources, land availability, and environmental justice areas.

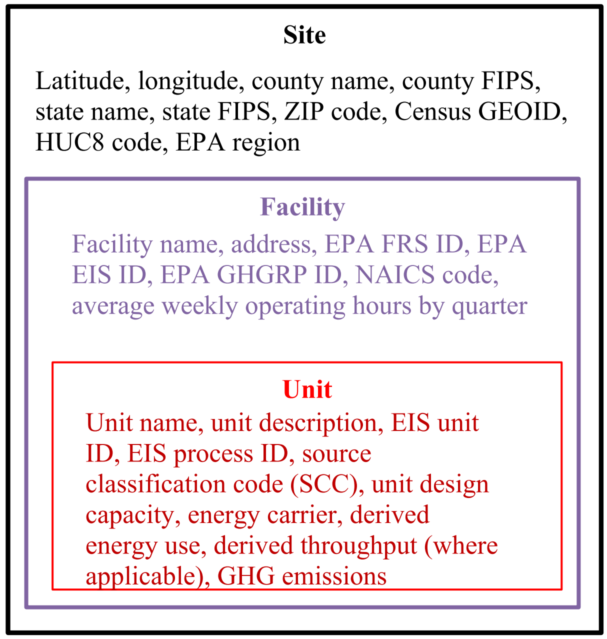


Figure 3. Nested structure of FIED data fields.

A complete list and description of data fields is provided in Appendix A.

At the next level, facility information contains descriptions of the specific facility, including its name, NAICS code and description, ID codes linking it to relevant EPA databases (e.g., Emissions Inventory System [EIS] facility, unit, and process IDs for air emissions data), employment size, and high-level operational data such as capacity, production, and hours of operation. This level is meant to classify a facility according to industry and size, which is useful for analyses that target specific industries or industry-wide technology adoption.

Within a facility, there are major processes that can be represented by the main material inputs and outputs and total energy per material output. For example, in an ammonia production facility, major processes could include syngas production and ammonia synthesis. Within each major process, there can be multiple unit operations, such as reformers, reactors, heaters, and coolers. Unit operations information contains the equipment type or unit description, energy input, energy carrier, capacity, efficiency, vintage, and costs. Unit operation data has been historically lacking due to much of it being inherently non-public proprietary operational data. However, it is necessary for accurate assessments of technologies, energy usage, and emissions accounting.

Unit type is a reported data element in the NEI, but the classification of unit types varies broadly. This is primarily due to the way the NEI is constructed, which relies on integrating data from the numerous state, local, and Tribal agencies across the country. In some state agencies, unit type is not collected at all, whereas in others, the same physical unit may be referred to as a boiler or an incinerator (U.S. Environmental Protection Agency 2021). In terms of fuel type, the NEI does not explicitly include a fuel type data element, but fuel information can be extracted from other data elements, including unit description and process description.

Conversely, the GHGRP provides standardized unit types and fuel types. This does not, however, ensure that identifying the type of unit and type of fuel is straightforward. As described in Section 2.3.2, there are several instances in which a reporter can indicate that a unit is an "other combustion source." Reporters can also input custom fuel types as "blends" or "other."

2.3 FIED Summary

2.3.1 Facility Identification

A facility[11] appears in the in the FIED if it has been assigned an FRS Registry ID by the EPA and if it has been categorized with an industrial NAICS code, including non-manufacturing sectors (i.e., NAICS 11, 21, and 23) and manufacturing sectors (i.e., NAICS 31–33). As has been highlighted previously in the context of industrial energy use, NAICS is an economic classification system that delineates groups based on their primary production activities and processes (Wachs et al. 2022). Some facilities produce multiple types of products or use more than one type of production process. In these cases, the NAICS code is assigned based on the final activity or the activity that represents the highest portion of facility costs. In terms of the FRS, a facility may be categorized using different NAICS codes by different regulatory programs, as well as by the facilities themselves; the FIED has incorporated any additional NAICS codes along with the primary NAICS code used to represent a facility.

The U.S. Census Bureau's County Business Patterns (CBP) is an annual data product that provides economic information on establishments[12] with paid employees (U.S. Census Bureau 2023a). The number of establishments reported in CBP data serves as a useful comparison for the coverage of facilities in the FRS data. Because the Census Bureau provides subnational economic data by industry each year, CBP estimates can be considered a comprehensive source of the current state of industrial facilities. Based on the CBP, there were 4,222,569 industrial establishments, of which 2,862,558 were construction (NAICS 23), 1,163,741 were manufacturing (NAICS 31 – 33), 102,916 were mining (NAICS 21), and 93,444 were agriculture, forestry, fishing and hunting (NAICS 11) (U.S. Census Bureau 2024b).

For the whole industrial sector, there are fewer facilities reported in FRS data compared to the assumed total count, as reported in CBP data. However, in the manufacturing sector specifically, the facilities listed in the FRS are very close in count to establishments in the CBP. Given that most industrial energy use occurs in manufacturing, this supports using FRS facility data as the basis for site information.

As shown in Figure 4, there is a close match between the number of manufacturing facilities in the FIED and manufacturing establishments in the CBP. In the cases of agriculture and mining,

---

[11] For the EPA, a facility could be defined as "a place where regulatory activities of interest to the agency occur or have occurred in the past" (Ross & Associates Environmental Consulting 2000).

[12] According to the Census Bureau, "[a]n establishment is a single physical location at which business is conducted or services or industrial operations are performed. It is not necessarily identical with a company or enterprise, which may consist of one or more establishments. When two or more activities are carried on at a single location under a single ownership, all activities generally are grouped together as a single establishment. The entire establishment is classified on the basis of its major activity and all data are included in that classification" (U.S. Census Bureau 2023b).

however, the FIED includes a much larger count. This is explained in part for agriculture by the fact that the CBP does not include NAICS codes associated with crop production or animal husbandry. For mining, it could be the case that, unlike the FIED, the CBP may group together individual drilling rigs that fall under NAICS 2111. The CBP, conversely, provides a larger count for construction than the FIED. This could be because many construction operations are not subject to the types of environmental regulations that would result in them being assigned an FRS ID. We also note that both sources may assign a different NAICS code to a facility, which may also affect the comparison of facility counts.

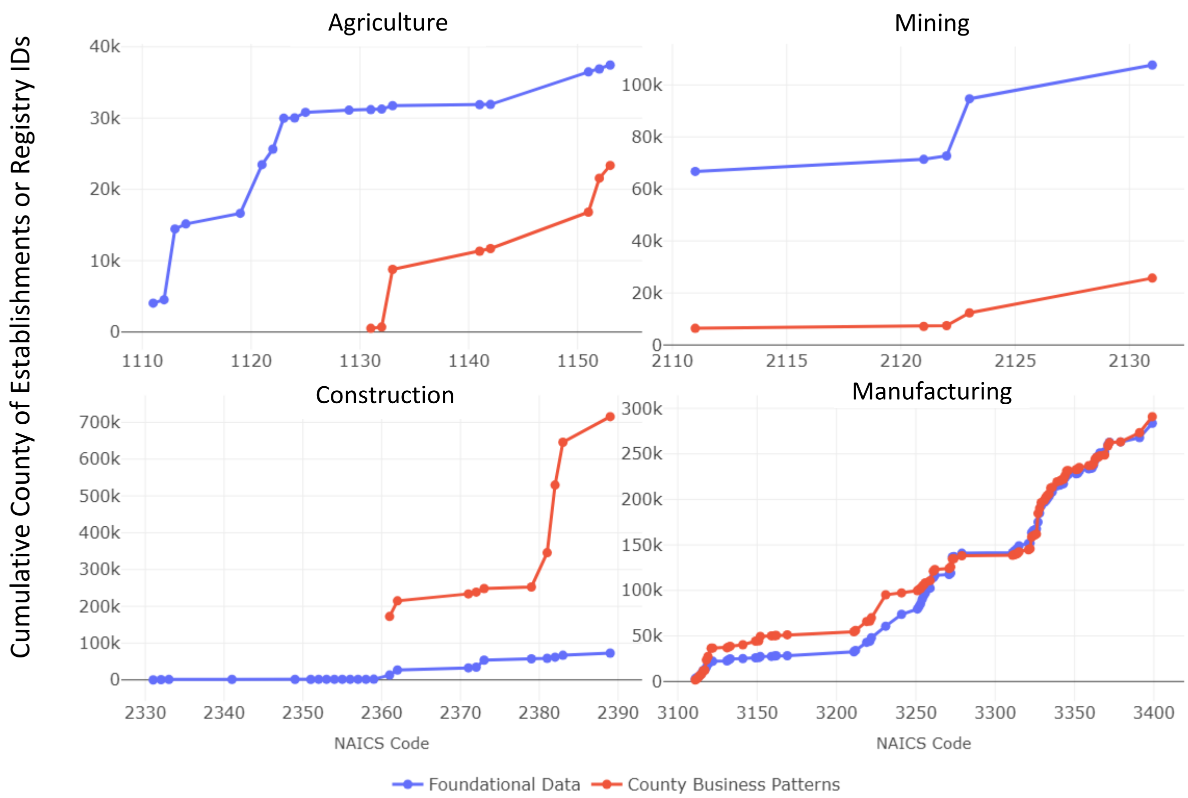


Figure 4. Comparison of cumulative facility counts between the FIED (number of unique Registry IDs) and Census County Business Patterns (number of establishments) by industrial sector for 2017

2.3.2 Unit Characterization

This first version of the FIED contains 644,881 rows of data points that represent 502,165 unique registry IDs. The 644,881 rows of data points contains 165,298 values for standardized unit types. This means that the fraction of data points missing for standardized unit type is about 0.74. Similar descriptions for other data fields are shown in Table 4.

Because the FRS includes environmental regulations that are less directly related to energy use than others (e.g., National Pollution Discharge Elimination System permitting for water discharges and Toxic Release Inventory reporting), it is understandable that many facilities would not have unit or even facility-level unit information associated with them. Additionally, given the relatively small number of industrial facilities that report to the GHGRP we would expect that large portion of the data points in the FIED are missing GHGRP-derived energy estimates. Table 4 shows that the fraction of data points missing GHGRP-derived energy estimates is in fact 0.98.

Table 4. Summary Description of FIED Data Field Coverage

	Registry ID	Standardized Unit Type	Standardized Fuel Type	Design Capacity	Energy (NEI-Derived)	Energy (GHGRP-Derived)
Number of Data Points	644,881[a]	165,298	62,492	48,388	46,227	12,939
Number of Missing Data Points	0	479,583	582,389	596,493	598,654	631,942
Fraction of Data Points Missing	0	0.74	0.90	0.92	0.93	0.98

[a] Of these, 502,165 are unique registry IDs.

Overall, however, facilities that do report to the GHGRP are much more likely to report information on, at minimum, the type of unit than facilities appearing in the NEI. Figure 5 shows that in the FIED, nearly 60% of the facilities reporting under the GHGRP include unit-level data (see Appendix F for the unit characterization process); however, this number is only about 25% for facilities reporting under the NEI (see Appendix C for detail on the unit characterization process).

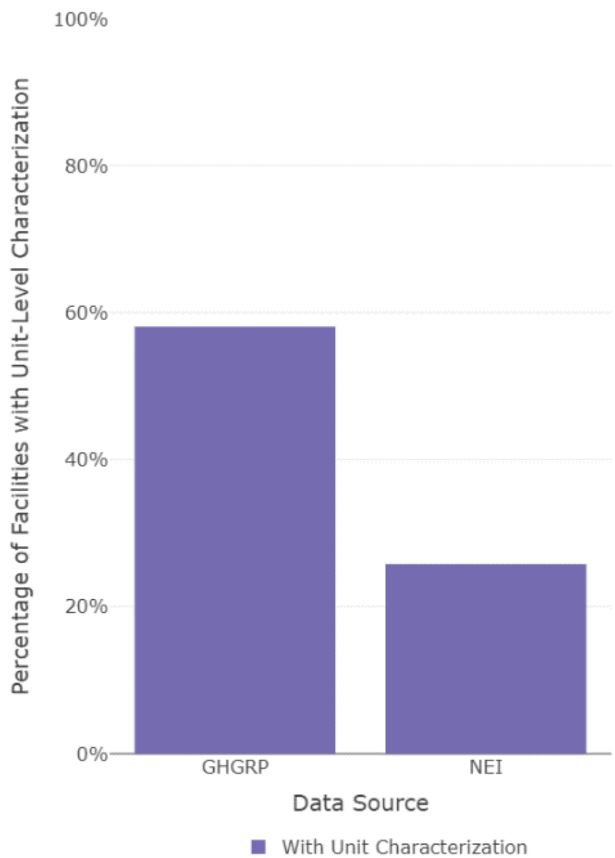


Figure 5. Comparison of the percentage of GHGRP- and NEI-reporting facilities with unit characterization (i.e., an obtainable named unit type)

The differences in unit characterization between the NEI and GHGRP are associated with the scope and purpose of each dataset. As described previously, the NEI represents a compilation of air emissions reporting from states and other sub-national entities. We assume it is likely that states do not share the same reporting requirements for data fields that may be of secondary importance to emissions data themselves, such as information on unit type. Conversely, GHGRP data are reported by emitters directly to the EPA. The NEI also includes a higher proportion of non-manufacturing facilities, in addition to covering much smaller facilities and many more facilities, than the GHGRP.

Figure 6 shows the breakdown of NEI-reporting facilities with and without unit-level characterization across the manufacturing and non-manufacturing sectors of industry. Nearly 20,000 of the approximately 51,000 manufacturing facilities have named unit types. Named unit types are much less common in the non-manufacturing industries.

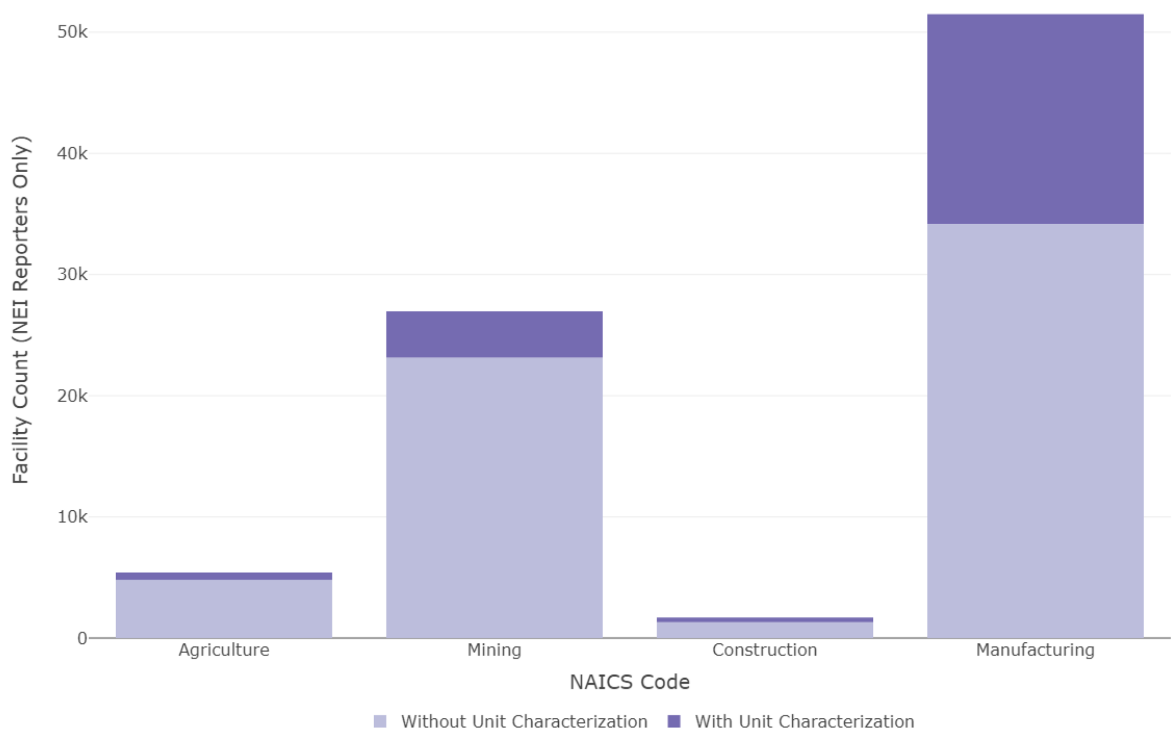


Figure 6. Comparison of the number of NEI-reporting facilities with and without unit characterization (i.e., an obtainable named unit type)

As shown in Figure 7, for GHGRP reporters, more than half of the manufacturing facilities report a unit type. One of the complications of identifying unit type in GHGRP data is the option for reporters to simplify unit-level reporting by aggregating either units that each have a maximum rated heat input capacity of equal to or less than 250 million British thermal units (MMBtu) per hour; units whose emissions are monitored from a common stack or duct; units that combust the same type of liquid or gaseous fuel and share a supply pipe; or small units that share a supply pipe with larger units, such as boilers and combustion turbines (40 C.F.R. § 98.36(c)). This can not only affect the usefulness of the description of the unit type (e.g., a unit name of CP-1 identified as an other combustion source), but also the reported design capacity (e.g., a unit whose design capacity represents an aggregation of units).

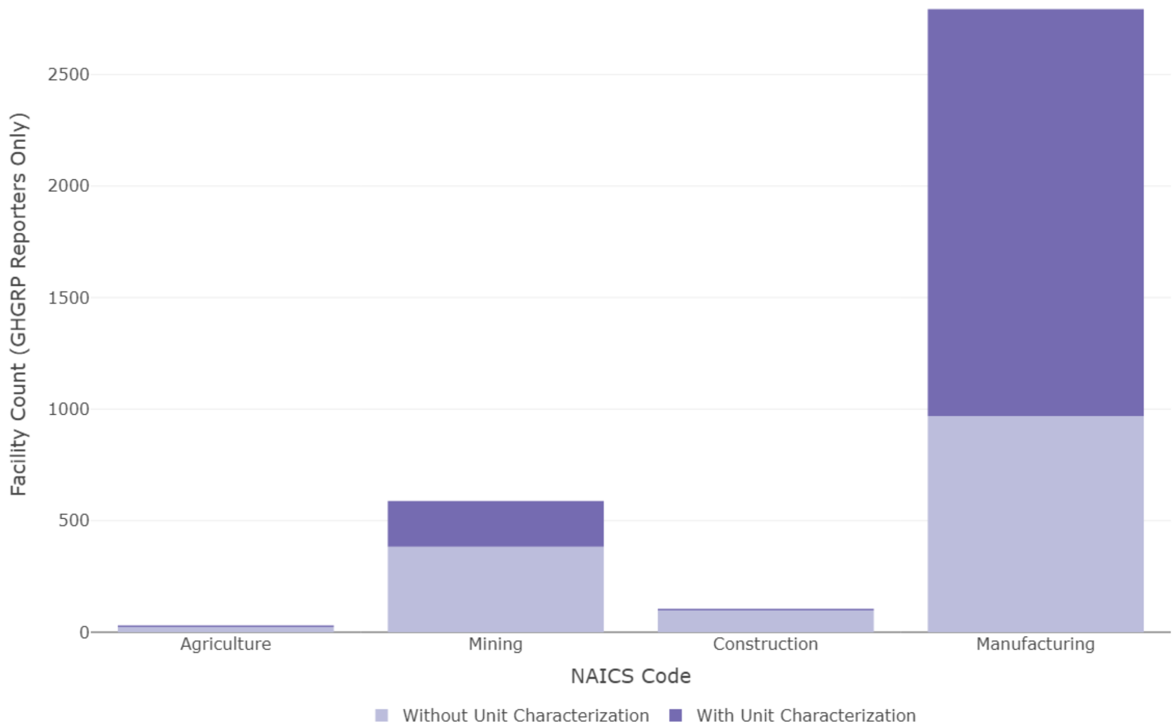


Figure 7. Comparison of the number of GHGRP-reporting facilities with and without unit characterization (i.e., an obtainable named unit type)

As shown in Figure 8 through Figure 10, the FIED assigns units to one of 19 standard unit types. The "other combustion" type captures units that were identified as combustion units, but did not necessarily fall neatly into another category (e.g., roasters, units described as both furnaces and heaters). The "other" type captures units whose type was not possible to characterize as combustion units based on available data sources, as well as units that were identified as non-combustion (e.g., hammermills, crushers, mixers, and conveyers) and had sufficient data to estimate mass throughput. This category of unit type contains the largest number of named units in the FIED and also represents a significant portion of the unit design capacity and estimated energy use. Boilers also represent a significant number of units, design capacity, and estimated energy use.

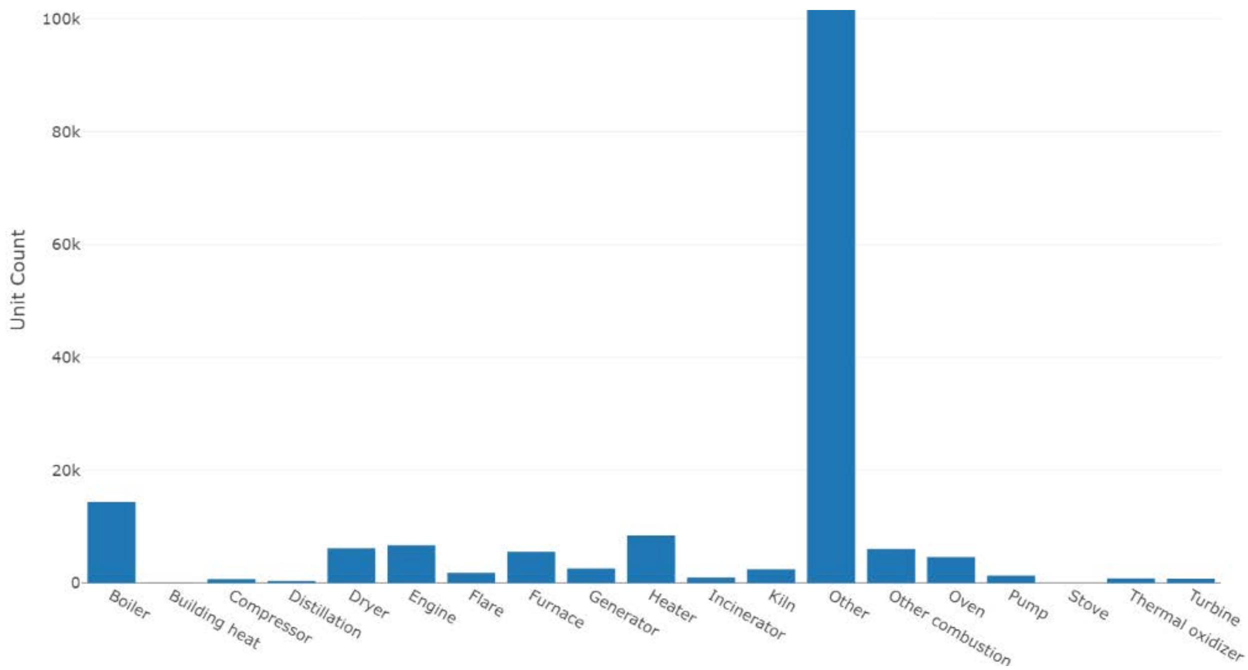


Figure 8. Total count by standardized unit type

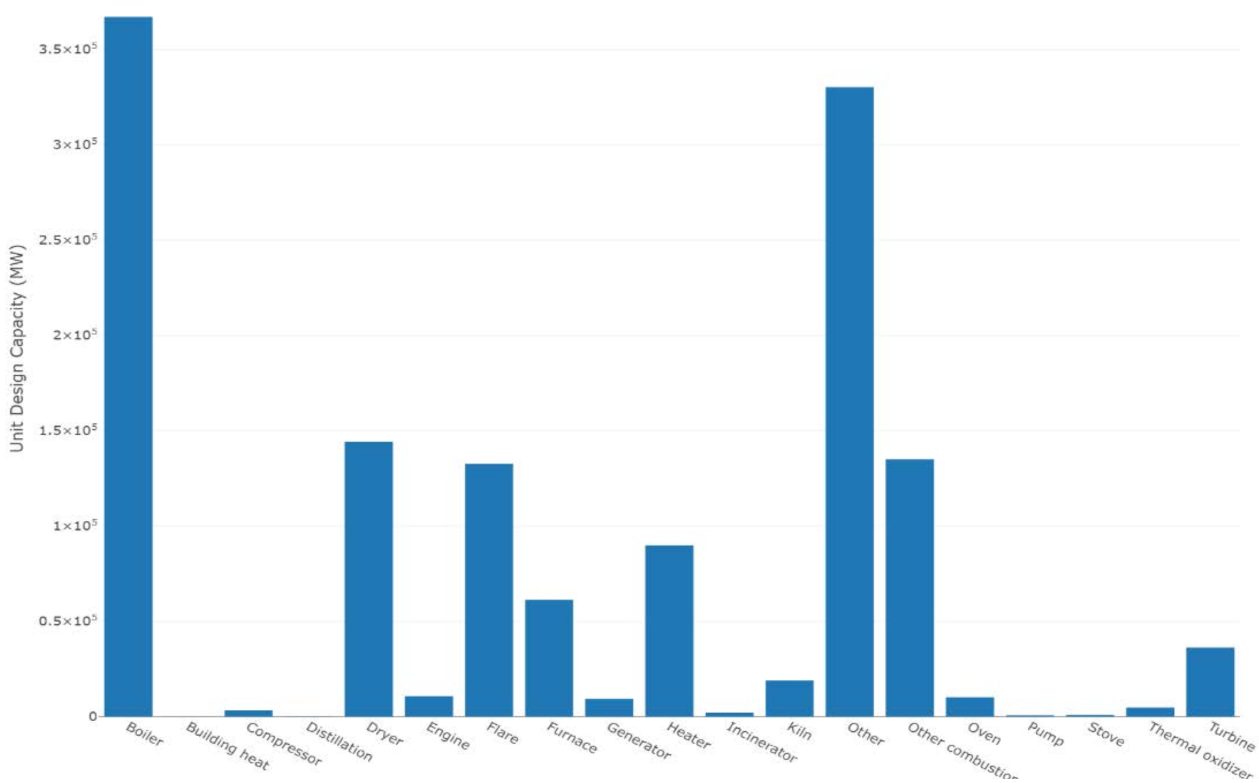


Figure 9. Design capacity (in MW) by standardized unit type

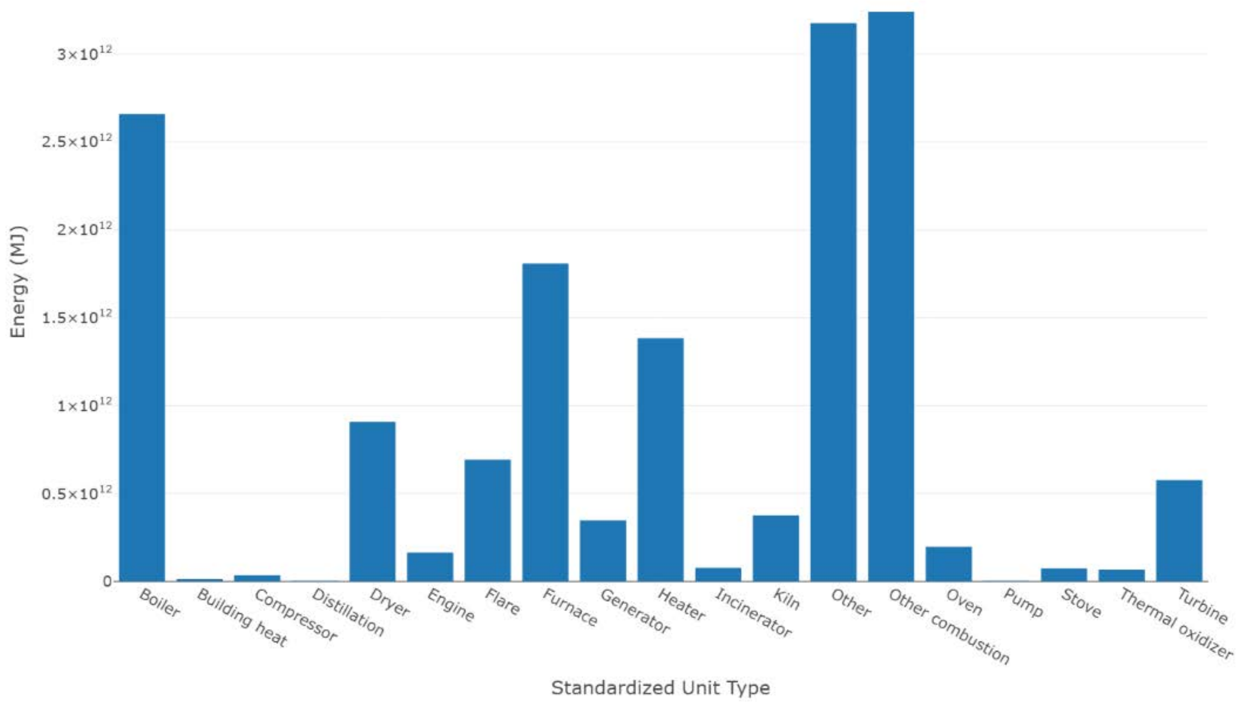
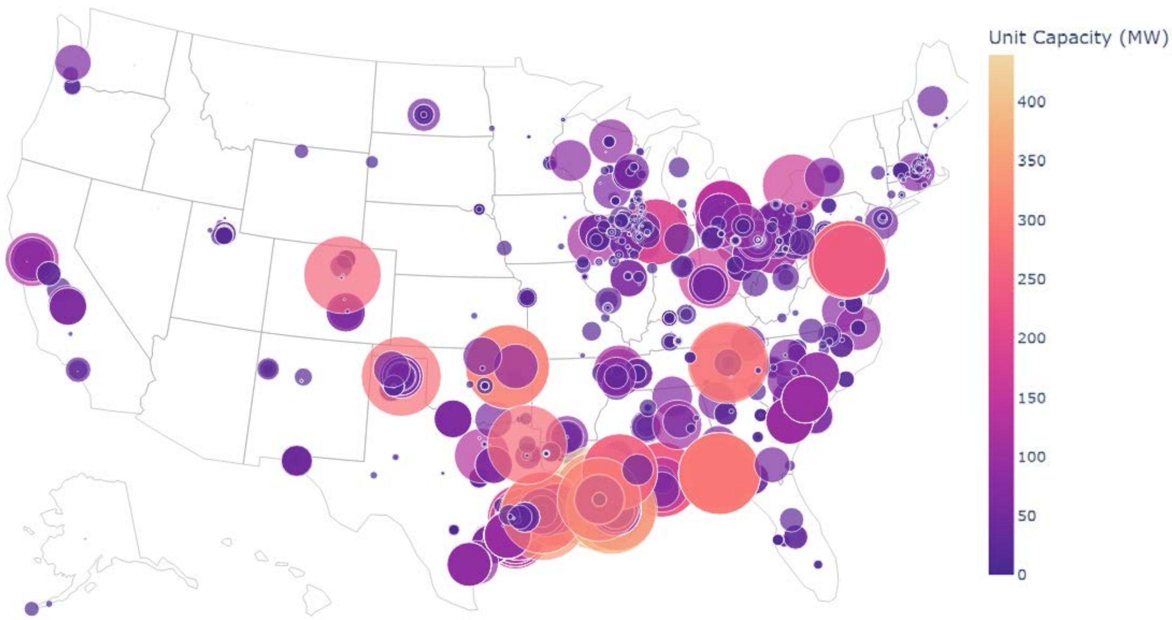


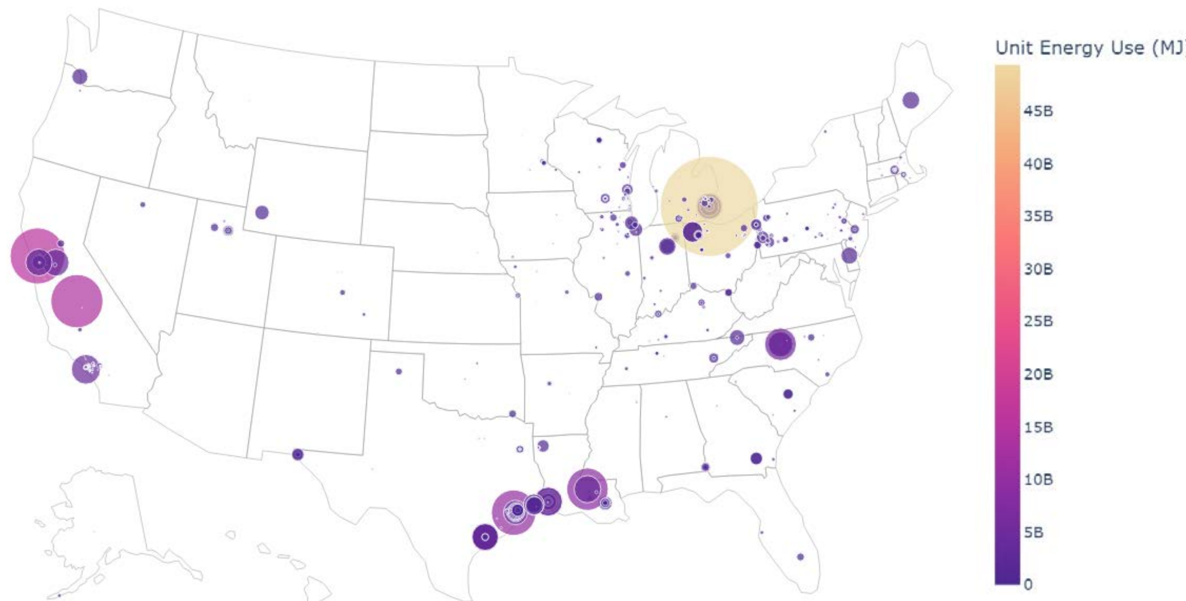
Figure 10. Estimated energy use (in MJ) by standardized unit type

One of the strengths of the FIED is its combined geographic resolution of facilities and unit-level characterization of units within facilities. Because of their associated latitude and longitude coordinates, it is possible to use the FIED to conduct analysis at the community level. The following series of figures identifies the locations, as well as relative estimated capacity and energy use, of several different unit types to indicate their distribution across the United States. It may be the case that not enough information was provided to estimate both the capacity and energy use for a given unit. As a result, the figures may show a location that has an estimated capacity, but not an estimated energy use, or vice versa. The following figures provide examples of the geographic distribution of unit capacity and unit energy use for furnaces and kilns.

Furnaces are a significant source of industrial energy use. The standardized unit type includes chemical recovery furnaces used in papermills, cracking furnaces used for ethylene production, as well as various furnaces used for producing and forming metals. As Figure 11 shows, these units are concentrated in the states around the Great Lakes and states in the Southeast.



(a)

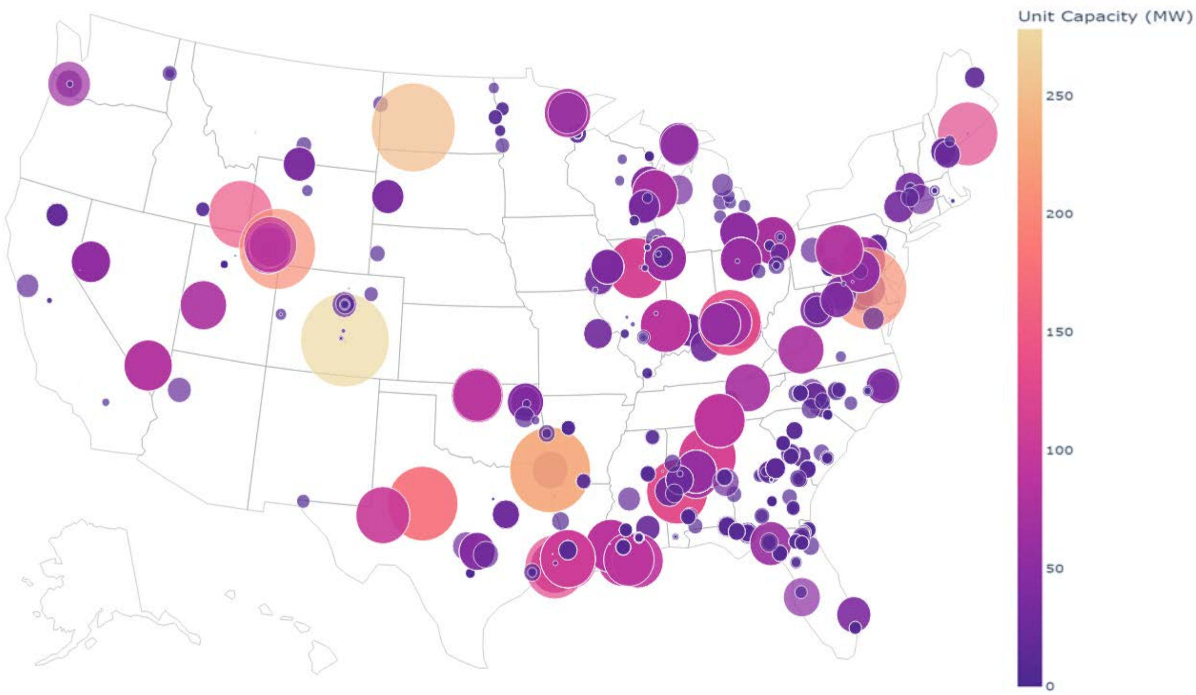


(b)

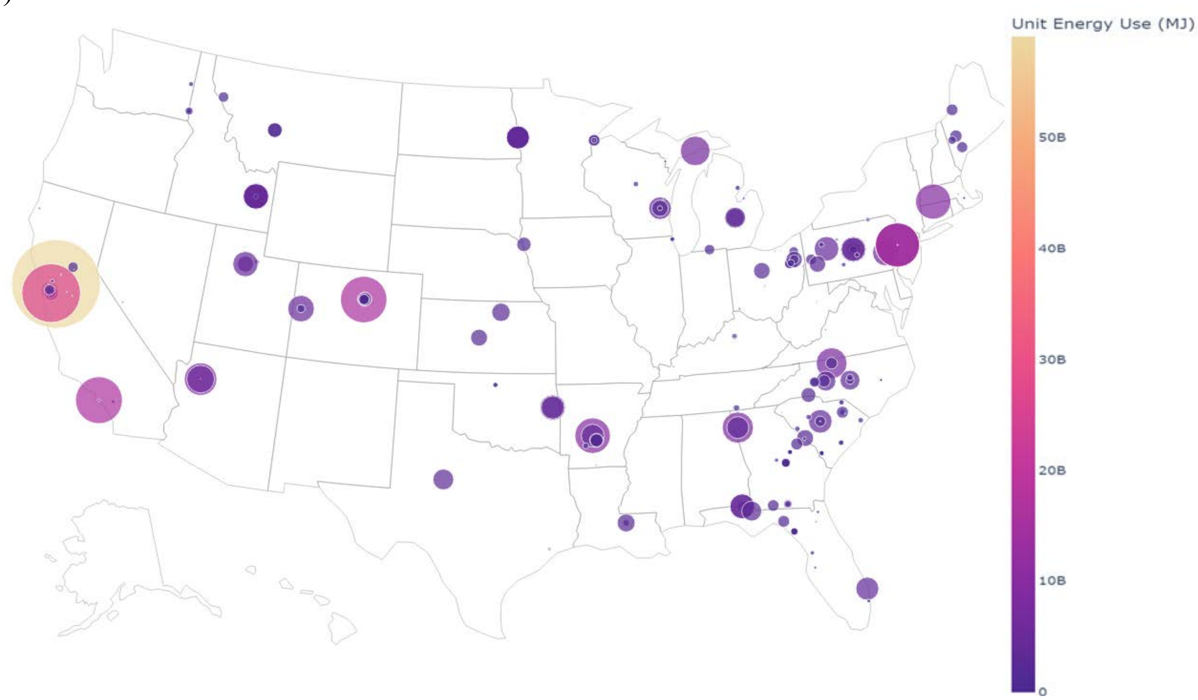
Figure 11. Location and relative size of identified furnace units by estimated capacity (a) and estimated energy use (b)

The standardized unit type category of kilns includes cement kilns, lime kilns in manufacturing and mining industries, and sawmill kilns. Figure 12 shows the distribution of kiln capacities and estimated energy use, which includes slightly more units in the Western states than furnaces. The map for unit capacity (Figure 12(a)) shows more units than the map for energy estimates (Figure

12(b)). Deriving energy estimates is only possible when the unit type and fuel type can be identified in the raw NEI data. In the case of kilns, it was possible to obtain unit capacity data for more units than it was possible to derive energy estimates.



(a)



(b)

Figure 12. Location and relative size of identified kiln units by estimated capacity (a) and estimated energy use (b)

2.3.3 Energy Estimates

Comparing energy estimates from the FIED (see Appendix D, Appendix F, and Appendix G for details on the derivation of energy estimates) to traditional sources of industrial energy estimates is complicated by the mismatched geographic and operational resolution. The lower resolution of traditional data sources—especially in terms of geographic resolution—is related to the disclosure policies associated with their data collection procedures. Likewise, the FIED's higher geographic and operational resolution is the result of unit-level reporting specified by the EPA's reporting requirements. These EPA data collection procedures are not protected by the same disclosure requirements as energy estimates developed by the EIA from surveys of industrial energy users and their energy suppliers.

As we discussed in Section 1, MECS is conducted by the Census Bureau and subject to disclosure rules that prohibit the release of data that could identify a single establishment. Consequently, MECS results published by the EIA are highly geographically aggregated. Results are often published at the census region by NAICS code, although some results may only be published at the national level by NAICS code (e.g., energy consumption disaggregated by end use). As previously discussed for the expansion of MECS 1994 to census regions, this geographic resolution is a function of the number of survey respondents, which is in part a function of survey funding, and is not fixed.

However, it should be noted that, although MECS and SEDS collect data on energy use instead of emissions from energy use, these energy estimates may not be substantially closer to the "true" amount of energy used by industrial facilities than EPA emissions reporting. In some cases, the sources of energy estimates may rely on very similar methods and the same starting data, with the only difference being a different type of conversion factor. For MECS, respondents can use a default conversion factor to translate the mass or volume of fuel purchases to an estimate of the amount of energy used, whereas the FIED energy estimates may rely on default conversion factors used by GHGRP respondents to translate the mass or volume of fuel purchases to energy and then to GHG emissions (i.e., Tier 1 emissions reporting for Subpart C).[13]

In both cases of this example, the underlying source of energy use is inferred from the annual amount of fuel purchased (recorded in mass or volume units, not energy units) and not the direct measurement of the amount of fuel used. Unlike the GHGRP, however, it is not possible to determine using public data what portion of reported energy use was estimated using generic or establishment specific conversion factors. This information is not published with MECS data, an indication of a lack of transparency relative to the FIED.

To further indicate how close these two approaches to estimating energy use can be, the 2018 and proposed 2022 MECS survey forms recommended that respondents use the "more precise conversion factors used for the Green House Gas (GHG) Reporting rule" in place of the generic conversion factors where available (Bureau of the Census 2018; 2022). Respondents record how much natural gas was purchased and delivered to their establishment, as well as how much was

[13] Reporters of emissions from direct combustion (Subpart C) may also report using Tier 2, Tier 3, or Tier 4 methodologies. See U.S. Environmental Protection Agency (2017) for more information.

consumed as a fuel and for any purpose other than a fuel (U.S. Census Bureau 2022). MECS respondents can report their natural gas use in units of therms, cubic feet, or MMBtu. The survey form provides generic factors by fuel to convert from mass or volume to energy units, but cautions they should be used only in circumstances in which respondents do not know the specific conversion factor of fuels used by their establishment.

In terms of geography, the FIED provides energy estimates that are geographically located by latitude and longitude, allowing aggregation to any spatial grouping such as county or ZIP code. Conversely, industrial energy estimates from the EIA are not published below the state level. In order to provide an initial comparison energy estimates between the FIED and EIA sources, we have aggregated FIED estimates to the census region and state level, as shown in Figure 13 and Figure 14, respectively. These figures depict a simple ratio between the FIED and EIA estimates, and because the FIED at this time primarily captures the use of combustion energy, the MECS estimates exclude electricity use.

Figure 13 shows that the FIED estimates of combustion energy use in the West and Northeast Census Region are about 2.3 and 2.1 times greater, respectively, than combustion energy estimated by MECS for the same region. The relative values for the Midwest and South Census Regions are much closer to MECS. Closer examination of the differences between the two datasets has been left for future work. A reasonable first step would be to aggregate the manufacturing NAICS codes in the FIED to match MECS and then compare the energy use by fuel type for each census region.

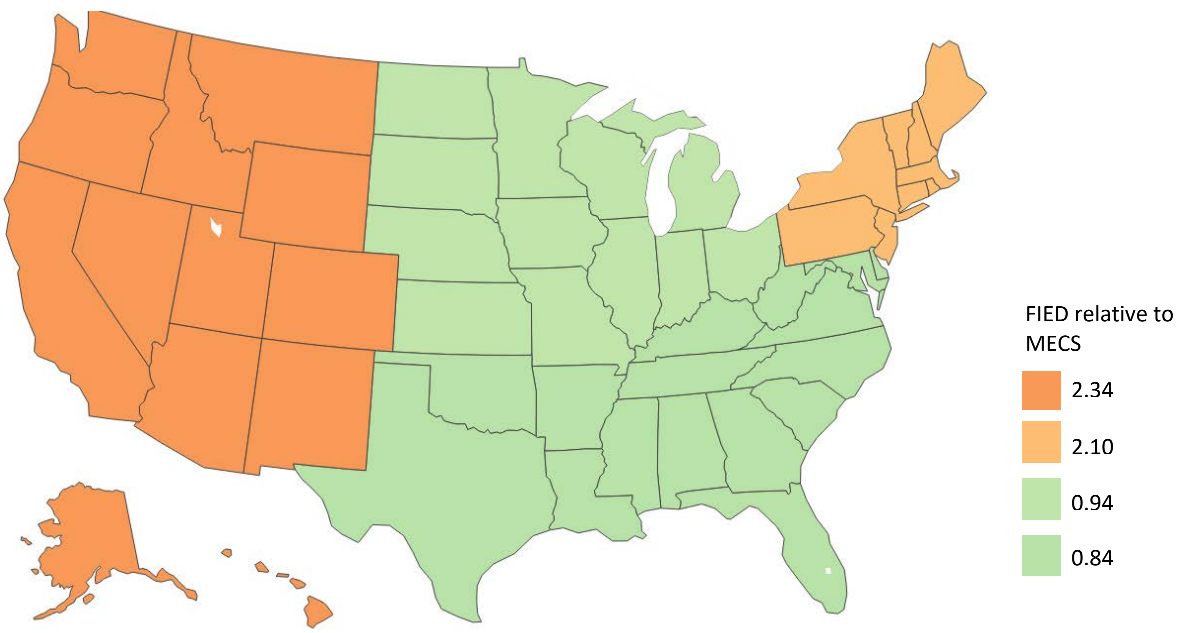


Figure 13. Relative comparison of manufacturing combustion energy estimated by FIED for 2017 and by MECS for 2018

Unit- and facility-level estimates for FIED have been aggregated by census region to match MECS geographic resolution.

The other EIA source of industrial energy data is SEDS. SEDS is a state-level compilation of data collected by various fuel surveys carried out by the EIA. With the exception of MECS, the EIA does not publish sub-national industry energy estimates by NAICS sector (i.e., NAICS 11 Agriculture, Forestry, Fishing and Hunting; NAICS 21 Mining, Quarrying, and Oil and Gas Extraction; NAICS 23 Construction; and NAICS 31-33 Manufacturing). Instead, the EIA aggregates these NAICS sectors as “industry” for SEDS. We note that industrial energy consumption is primarily collected from energy producers (e.g., electric utilities, natural gas processing plants) and not from industrial facilities themselves.[14]

Like MECS, we use a relative comparison between the FIED and SEDS, as shown in Figure 14. FIED estimates are the largest relative to SEDS for Colorado and California, which are about 2.5 times the combustion energy estimates from SEDS. Overall, however, the FIED’s combustion energy estimates tend to be less than estimates from SEDS. The FIED is lowest relative to SEDS for several states in New England, including Vermont and New Hampshire. The FIED combustion energy estimates for these states are about 85% less than SEDS.

Although the disparity between the FIED and SEDS may be large, it is difficult to examine the reasons why, without more geographically and operationally disaggregated data from the EIA. Even without more disaggregated data from EIA, given the lack apparent patterns of discrepancies by state, it is not clear what aspects of the datasets might be responsible for the differences shown in Figure 14.

[14] Since 1990, the EIA-861Annual Electric Power Industry Report has been the source of industrial electricity use (U.S. Energy Information Administration 2022b). Electric utilities, energy service providers, and distribution companies are required to submit EIA-861 annually. EIA’s *Annual Coal Report* has been the source of coal industrial consumption data since 2001 (U.S. Energy Information Administration 2022c). This report is based off Form EIA-3, the Quarterly Survey of Industrial, Commercial & Institutional Coal Users and Form EIA-7A Annual Survey of Coal Production and Preparation (U.S. Energy Information Administration 2024f). Facilities that consume more than 1,000 short tons of coal are required to submit Form EIA-3 quarterly. SEDS industrial natural gas consumption is the sum of natural gas delivered to the industrial sector, natural gas consumed as lease fuel, and natural gas consumed as plant fuel (U.S. Energy Information Administration 2022d). Since 1997, the EIA’s *Natural Gas Annual* has been the source of all three estimates. Form EIA-176 Annual Report of Natural and Supplemental Gas Supply and Disposition is the primary source of data for the report. The respondents of Form EIA-176 are the 2,000 natural gas processing plant operators. For industrial wood and waste consumption, SEDS uses Form EIA-923 Power Plant Operations report for combined heat and power and electricity-only facilities. For other industrial facilities, SEDS uses data from MECS and the Economic Census (U.S. Energy Information Administration 2022e).

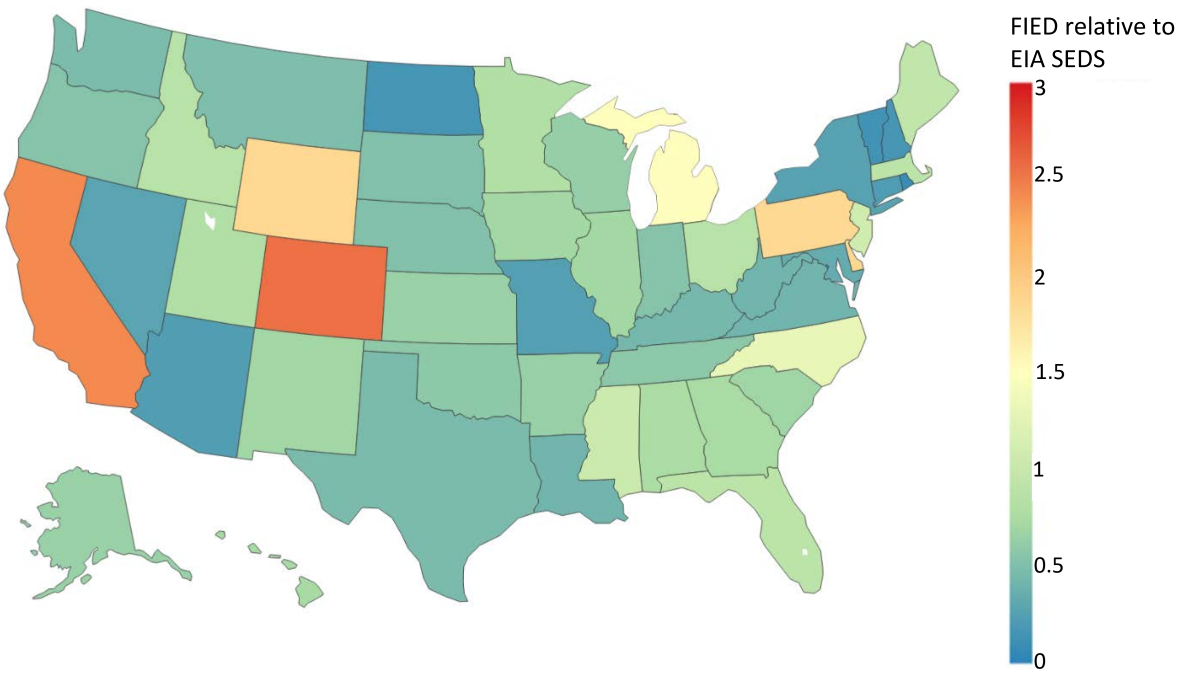


Figure 14. Relative comparison of industrial combustion energy estimated by FIED and by SEDS for 2017

Unit- and facility-level estimates for FIED have been aggregated by state to match SEDS geographic resolution.

2.3.4 Remaining Gaps

Despite best current efforts to provide a comprehensive facility- and unit-level dataset on industrial energy use, several data gaps remain. For one, many units remain uncharacterized, even after attempts to glean as much information as possible from data fields in the NEI and GHGRP. This is particularly the case for facilities reporting only to the NEI. Additionally, electricity end uses are underrepresented in FIED, similar to previous efforts to estimate energy use from facility emissions data. Because emissions from purchased electricity occur off-site, these are not included in the scope of emissions reporting requirements for the NEI and GHGRP. It has been possible in certain limited circumstances to estimate electricity use based on information provided by the NEI. Stolaroff et al. (2009) argued that the GHGRP should include not only emissions from purchased electricity, but also the amount of electricity purchased in energy units. The authors also recommended that the EPA require facilities in certain industries to report emissions per unit output.

The FIED provides only very limited information on facility production. In certain circumstances, it is possible to estimate throughput values by unit, but we continue to lack a systematic way of estimating facility production. Without production estimates or another way of estimating relative energy use, it is difficult to make inter-facility comparisons of energy intensity.

3 Conclusions

The FIED was created as a response to related demands for open data and more highly resolved industrial energy data to support decarbonization analysis. The conventional sources of industrial energy data that exist today have been shaped by histories that ultimately constrain the levels of transparency and resolution available to analysts and the public at large. These histories reflect the social processes that underpin the planning and implementation of data collection and dissemination activities. These processes may be contested and reflect the tension between private industry's desired control over their operational data and societal desire to analyze the impacts of technology. Additionally, the ability of an organization to collect information can be revoked or otherwise changed by other actors that control budgets or have other forms of authority. These processes are not unidirectional, however, and data collection efforts could expand in their scope and detail if society places a greater emphasis on the need to analyze the impacts of technology.

Industry concerns about potential regulation were part of historical processes that have limited the level of detail available in data collected and reported by EIA today. Yet, the authority granted to EPA to collect emissions data for its regulations has enabled the FIED and other related efforts to estimate industrial energy use at much higher resolutions than EIA data. Estimating energy use from emissions is not inconsistent with how a conventional industrial data source, such as MECS, allows respondents to report their energy use. With far fewer pollutants, a well-defined reporting tier system, and several years of revising the energy derivation methodology, the GHGRP data provide a much more straightforward approach for estimating energy use than the NEI, as shown in Appendix D, Appendix F, and Appendix G. We had expected that the different pollutants reported in the NEI would yield roughly an equivalent amount of energy for a given combustion unit. However, this was often not the case, indicating potential errors in the mass of reported emissions, the emissions factors used, or a combination of the two.

The challenges presented by NEI data also indicate potential paths to improve the FIED and expand its use:

1. **Increase coordination with state and local organizations.** The organizations responsible for providing NEI data may also be able to shed light on the discrepancies of reported emissions and emissions factors and identify additional sources of information that could be used to improve unit characterization.

2. **Develop a FIED constituency.** Recognizing data collection and dissemination as social processes highlights the importance of actors, networks, institutions, and practices in promoting or inhibiting the use of a dataset. Several of the early industrial energy datasets were developed in parallel with new modeling capabilities. Molina (1990) introduced the concept of sociotechnical constituencies to explain the "dynamic ensembles of technical constituents (e.g., machines, instruments) and social constituents (e.g., institutions, interest groups) which interact and shape each other in the course of the creation, production and diffusion of specific technologies such as formal methods." A FIED constituency could be grown from coordinating the different groups involved with

collecting underlying data and using the data through open source data platforms such as Open Energy Data Initiative and GitHub.

3. **Support local analysis.** The strongest use cases for FIED involve local analysis of facility- and unit-level energy use. The FIED's geographic resolution allows it to be easily paired with EJScreen (U.S. Environmental Protection Agency 2024c) and corporate data, such as the Corporate Toxics Information Project (Political Economy Research Institute 2024). By using the EPA FRS IDs as the unique facility identifiers, FIED can also be used in conjunction with other emissions and waste inventories (e.g., Young et al. 2022).

4. **Update to 2020 and 2014.** The FIED could be updated for 2020 and 2014, based on the NEI and GHGRP data for those years. This would enable additional analysis of trends in energy use and emergence and retirement of units.

References

Alkadi, Nasr E., Michael R. Starke, Ookie Ma, Sachin U. Nimbalkar, and Daryl Cox. 2013. "Industrial Geospatial Analysis Tool for Energy Evaluation (IGATE-E)." Oak Ridge, TN: Oak Ridge National Laboratory (ORNL). https://www.osti.gov/biblio/1082107-industrial-geospatial-analysis-tool-energy-evaluation-igate.

Alonso, William, and Paul Starr, eds. 1987. *The Politics of Numbers*. New York : Russell Sage Foundation. http://archive.org/details/politicsofnumber00alon_0.

Bataille, Christopher George Frederic. 2005. "Design and Application of a Technologically Explicit Hybrid Energy-Economy Policy Model with Micro and Macro Economic Dynamics." Doctoral thesis, Burnaby, B.C.: Simon Fraser University. https://www.researchgate.net/profile/Chris-Bataille/publication/252530941_Design_and_application_of_a_technologically_explicit_hybrid_energy-economy_policy_model_with_micro_and_macro_economic_dynamics/links/56154bca08ae4ce3cc652bea/Design-and-application-of-a-technologically-explicit-hybrid-energy-economy-policy-model-with-micro-and-macro-economic-dynamics.pdf.

Bohn, R. E., and J. S. Herod. 1978. "Industrial Sector Technology Use Model (ISTUM): Industrial Energy Use in the United States, 1974-2000. Volume 1: Primary Model Documentation, Book 2, Chapters IV, V, VI, and VII." DOE/EIA/2344-2. Energy and Environmental Analysis, Inc., Arlington, VA (USA). https://doi.org/10.2172/6056181.

Brown, Harry L., Bernard B. Hamel, Som Karamchetty, William H. Steigelmann, Birur C. Gajanana, Anil P. Agarwal, Lawrence W. Klock, et al. 1976. *Industrial Applications Study. Volume II. Industrial Data Base. Final Report. [Waste Heat Recovery and Utilization]*. COO/2862-2/2. Drexel Univ., Philadelphia, Pa. (USA); United Technologies Research Center, East Hartford, Conn. (USA); Mathematica, Princeton, N.J. (USA). https://doi.org/10.2172/5366400.

———. 1977. *Industrial Applications Study. Volume III. Technology Data Base Evaluation of Waste Recovery Systems. Final Report.* COO/2862-3. Drexel Univ., Philadelphia, Pa. (USA); United Technologies Research Center, East Hartford, Conn. (USA); Mathematica, Princeton, N.J. (USA). https://doi.org/10.2172/5283606.

Bureau of the Census. 2018. "Form EIA-846 2018 Manufacturing Energy Consumption Survey." U.S. Department of Commerce. https://www.eia.gov/survey/form/eia_846/form_a.pdf.

———. 2022. "Form EIA-846 2022 Manufacturing Energy Consumption Survey." U.S. Department of Commerce. https://www.eia.gov/survey/form/eia_846/proposed/2022/form.pdf.

Buster, Grant, Brandon N. Benton, Andrew Glaws, and Ryan N. King. 2024. "High-Resolution Meteorology with Climate Change Impacts from Global Climate Model Data Using Generative Machine Learning." *Nature Energy*, April, 1–13. https://doi.org/10.1038/s41560-024-01507-9.

Dance, Gabriel J. X., Tim Wallace, and Zach Levitt. 2023. "The Real-World Costs of the Digital Race for Bitcoin." *The New York Times*, April 10, 2023, sec. Business. https://www.nytimes.com/2023/04/09/business/bitcoin-mining-electricity-pollution.html.

Energy and Environmental Analysis. 1978. *Industrial Sector Technology Use Model (ISTUM): Industrial Energy Use in the United States, 1974-2000. Volume 4. Technology Appendix*. DOE/EIA/2344-5. Energy and Environmental Analysis, Inc., Arlington, VA (USA). https://doi.org/10.2172/5952168.

Energy and Environmental Analysis, Inc. 1977. *Energy Consumption Data Base. Volume III, Chapter 4, Part 1. Small Industries SIC 20-25, 27, 30-32, 34-39. Final Report. [1967, 1971, and 1974]*. DOE/TIC-10779(Vol.3)(Ch.4)(Pt.1). Washington, D.C.: Federal Energy Agency, Office of Energy Information and Analysis, Consumption Studies Division. https://doi.org/10.2172/6233453.

Energy and Environmental Analysis, Inc. 1977a. *Energy Consumption Data Base. Volume III, Chapter 4, Part 2. PAPER AND ALLIED PRODUCTS. SIC 26. Final Report.* DOE/TIC-10779Vol3Ch4Pt2. Washington, D.C.: Federal Energy Agency, Office of Energy Information and Analysis, Consumption Studies Division. https://doi.org/10.2172/12235942.

———. 1977b. *Energy Consumption Data Base. Volume III, Chapter 4, Part 3. The Chemical Industry. SIC 28. Final Report.* DOE/TIC-10779Vol3Ch4Pt3. Washington, D.C.: Federal Energy Agency, Office of Energy Information and Analysis, Consumption Studies Division. https://doi.org/10.2172/12235943.

———. 1979. *Industrial Fuel Choice Analysis Model. Volume II. Appendices to Model Documentation*. DOE/EIA/2832-2. Washington, D.C.: Department of Energy, Office of Policy and Evaluation. https://doi.org/10.2172/5953220.

Energy Modeling Forum. 1987. *Final Summary Report for EMF 8: Industrial Energy Demand*. Energy Modeling Forum. Stanford, CA: Standford University. https://web.stanford.edu/group/emf-research/docs/emf8/8v1.pdf.

Green, H. J. 1979. *Cooperative Effort for Industrial Energy Data Collection (IEDC).* SERI/RR-333-422. Solar Energy Research Inst., Golden, CO (USA). https://doi.org/10.2172/5807882.

Hale, Elaine, Henry Horsey, Brandon Johnson, Matteo Muratori, Eric Wilson, Brennan Borlaug, Craig Christensen, et al. 2018. *The Demand-Side Grid (Dsgrid) Model Documentation.* NREL/TP-6A20-71492. Golden, CO: National Renewable Energy Lab (NREL). https://doi.org/10.2172/1465659.

Isser, S., and D. R. Limaye. 1982. *Review of Industrial Energy Data Bases. Final Report*. EPRI-EM-2647. Synergic Resources Corp., Bala-Cynwyd, PA (USA). https://doi.org/10.2172/6622701.

Jaccard, Mark, and Joe Roop. 1990. "The ISTUM-PC Model: Trial Application to the British Columbia Pulp and Paper Industry." *Energy Economics* 12 (3): 185–96. https://doi.org/10.1016/0140-9883(90)90030-J.

Jordan, Amy B., Daniel S. Rodriguez, Jeffrey A. Bennett, Kat Sale, and Christopher Gilhooley. 2024. “Quantifying Air Quality Co-Benefits to Industrial Decarbonization: The Local Air Emissions Tracking Atlas.” *Frontiers in Public Health* 12 (May). https://doi.org/10.3389/fpubh.2024.1394678.

Krawiec, Frank, Dilip Limaye, Roy Beatty, Glenn Colville, and Karen Lang. 1981. “Energy End use Requirements in Manufacturing.” SERI/TR-733-790R. Golden, CO: Solar Energy Research Institute. https://books.google.com/books/download/Energy_end_use_requirements_in_manufactu.pdf?id=31n18A5CXpYC.

Lave, Rebecca, Eric Nost, and Jeny Goldstein. 2022. “Toward a Political Ecology of Data.” In *The Nature of Data: Infrastructure, Environments, Politics*, edited by Jeny Goldstein and Eric Nost, 303–9. Lincoln, NE: University of Nebraska Press. https://muse.jhu.edu/book/100261.

McKane, Aimee, Lynn Price, Scott Murtishaw, Stephane de la Rue du Can, Andrea Carroll, and Craig Cheney. 2006. “Development of the Industrial Large Energy Users Database.” Berkeley, CA: Lawrence Berkeley National Laboratory.

McMillan, Colin. 2019. “Manufacturing Thermal Energy Use in 2014.” Manufacturing Thermal Energy Use in 2014. September 27, 2019. https://dx.doi.org/10.7799/1570008.

McMillan, Colin A., and Vinayak Narwade. 2018. *The Industry Energy Tool (IET): Documentation.* NREL/TP-6A20-71990. Golden, CO: National Renewable Energy Lab (NREL). https://doi.org/10.2172/1484348.

McMillan, Colin A., and Mark Ruth. 2019. “Using Facility-Level Emissions Data to Estimate the Technical Potential of Alternative Thermal Sources to Meet Industrial Heat Demand.” *Applied Energy* 239 (April):1077–90. https://doi.org/10.1016/j.apenergy.2019.01.077.

McMillan, Colin A., Carrie Schoeneberger, Jingyi Zhang, Eric Masanet, Parthiv Kurup, Steven Meyers, Robert Margolis, and William Xi. 2021. *Opportunities for Solar Industrial Process Heat in the United States.* NREL/TP-6A20-77760. Golden, CO: NREL. https://doi.org/10.2172/1762440.

McMillan, Colin, Richard Boardman, Michael McKellar, Piyush Sabharwall, Mark Ruth, and Shannon Bragg-Sitton. 2016. *Generation and Use of Thermal Energy in the U.S. Industrial Sector and Opportunities to Reduce Its Carbon Emissions*. NREL/TP-6A50-66763; INL/EXT-16-39680. Golden, CO: National Renewable Energy Laboratory. https://doi.org/10.2172/1334495.

Miller, C. Andrew, George Hidy, Jeremy Hales, Charles E. Kolb, Arthur S. Werner, Bernd Haneke, David Parrish, et al. 2006. “Air Emission Inventories in North America: A Critical Assessment.” *Journal of the Air & Waste Management Association* 56 (8): 1115–29. https://doi.org/10.1080/10473289.2006.10464540.

Modak, Nabanita, Kelley Spence, Saloni Sood, and Jacky Ann Rosati. 2015. "Building a Comprehensive Mill-Level Database for the Industrial Sectors Integrated Solutions (ISIS) Model of the U.S. Pulp and Paper Sector." *PLOS ONE* 10 (3): e0120954. https://doi.org/10.1371/journal.pone.0120954.

Molina, A. H. 1990. "Transputers and Transputer-Based Parallel Computers : Sociotechnical Constituencies and the Build-up of British-European Capabilities in Information Technologies." *Research Policy* 19 (4): 309–33. https://doi.org/10.1016/0048-7333(90)90016-Y.

Morin, Arthur Lyman. 1989. "Dynamics of Environments of Centralized Statistical Agencies (CESTAs)." Ph.D., United States -- California: University of California, Santa Barbara. https://www.proquest.com/pqdtglobal/docview/303675472/FBBA87FD9EE74419PQ/1.

National Research Council. 1992. *The National Energy Modeling System*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/1997.

Open Knowledge Foundation. n.d. "Open Definition 2.1." Open Definition - Defining Open in Open Data, Open Content and Open Knowledge. Accessed May 21, 2024. https://opendefinition.org/od/2.1/en/.

Pfenninger, Stefan. 2024. "Open Code and Data Are Not Enough: Understandability as Design Goal for Energy System Models." *Progress in Energy* 6 (3): 033002. https://doi.org/10.1088/2516-1083/ad371e.

Pfenninger, Stefan, Joseph DeCarolis, Lion Hirth, Sylvain Quoilin, and Iain Staffell. 2017. "The Importance of Open Data and Software: Is Energy Research Lagging Behind?" *Energy Policy* 101 (February):211–15. https://doi.org/10.1016/j.enpol.2016.11.046.

Political Economy Research Institute. 2024. "PERI - Corporate Toxics Information Project." Corporate Toxics Information Project, 2024. https://peri.umass.edu/corporate-toxics-information-project.

Roop, J. M., and D. T. Kaplan. 1984. *ISTUM PC: Industrial Sector Technology Use Model for the IBM-PC.* PNL-5225. Richland, WA: Pacific Northwest Lab. https://doi.org/10.2172/6791153.

Ross & Associates Environmental Consulting. 2000. *Facility Identification Template for States: Version 2 : Working Guidelines for Integrating Facility Identification Information.* https://www.epa.gov/sites/default/files/2015-09/documents/facility_identification_template_for_states.pdf.

Schnaiberg, Allan. 1977. "Obstacles to Environmental Research by Scientists and Technologists: A Social Structural Analysis." *Social Problems* 24 (5): 500–520. https://doi.org/10.2307/800121.

Schoeneberger, Carrie. 2023. "Energy Systems Analysis Framework for the Decarbonization of Industrial Process Heat." Evanston, IL: Northwestern University. https://doi.org/10.21985/n2-tzef-r956.

Schoeneberger, Carrie, Colin McMillan, Parthiv Kurup, Sertac Akar, Robert Margolis, and Eric Masanet. 2020. "Solar for Industrial Process Heat: A Review of Technologies, Analysis Approaches, and Potential Applications in the United States." *Energy*, June, 118083. https://doi.org/10.1016/j.energy.2020.118083.

Schoeneberger, Carrie, Jingyi Zhang, Colin McMillan, Jennifer B. Dunn, and Eric Masanet. 2022. "Electrification Potential of U.S. Industrial Boilers and Assessment of the GHG Emissions Impact." *Advances in Applied Energy* 5 (February):100089. https://doi.org/10.1016/j.adapen.2022.100089.

Stolaroff, Joshuah K., Christopher L. Weber, and H. Scott Matthews. 2009. "Design Issues in a Mandatory Greenhouse Gas Emissions Registry for the United States." *Energy Policy*, New Zealand Energy Strategy, 37 (9): 3463–66. https://doi.org/10.1016/j.enpol.2009.04.028.

Texas Blockchain Council v. Department of Energy. 2024. 6:24-cv-00099, (W.D. Tex. Feb 23, 2024) ECF No. 13. https://www.courtlistener.com/docket/68276281/13/texas-blockchain-council-v-department-of-energy/.

U.S. Office of Personnel Management. 2011. *Paperwork Reduction Act (PRA) Guide: Version 2.0.* Washington, D.C. https://www.opm.gov/about-us/open-government/digital-government-strategy/fitara/paperwork-reduction-act-guide.pdf.

U.S. Census Bureau. 2022. "Form EIA-846 2022 Manufacturing Energy Consumption Survey," 2022. https://www.eia.gov/survey/form/eia_846/proposed/2022/form.pdf.

———.2023a. "County Business Patterns." Census.gov, October 16, 2023. https://www.census.gov/programs-surveys/cbp.html.

———. 2023b. "Statistics of U.S. Businesses Glossary." Census.gov, October 12, 2023. https://www.census.gov/programs-surveys/susb/about/glossary.html.

———. 2024a. "Restricted-Use Data Application Process." Census.gov, July 31, 2024. https://www.census.gov/StandardApp.

———. 2024b. "All Sectors: County Business Patterns by Legal Form of Organization and Employment Size Class for U.S., States, and Selected Geographies: 2017." 2024. https://data.census.gov/table/CBP2017.CB1700CBP?q=CBP2017.CB1700CBP&n=N0200.00.

U.S. Department of Energy. 2024a. "Industrial Assessment Centers." Energy and Cost Saving Assessments for Small and Medium-sized US Manufacturers, 2024. https://iac.university/.

———. 2024b. "U.S. Department of Energy Office of Scientific and Technical Information." OSTI.GOV, 2024. https://www.osti.gov/frontpage.

U.S. Energy Information Administration. 1980. *Annual Report to Congress, 1979 (Volume 1: Activities).* DOE/EIA-0173(79)/1. Washington, D.C.: U.S. Energy Information Administration. https://doi.org/10.2172/6894180.

———. 1982. *Methodological Report on the 1980 Manufacturing Industries Survey of Large Combustors (EIA-463).* DOE/EIA-0306. Washington, D.C. https://www.eia.gov/consumption/manufacturing/reports/archive/establishments/hp_pdf/DOE%20EIA-0306.pdf.

———. 1983. *Report on the 1980 Manufacturing Industries Energy Consumption Study and Survey of Large Combustors.* DOE/EIA-0358. Washington, D.C.: Energy Information Administration. https://www.eia.gov/consumption/manufacturing/reports/archive/establishments/hp_pdf/DOE%20EIA-0358.pdf.

———. 1997. *Manufacturing Consumption of Energy 1994*. DOE/EIA-0512(94). Washington, D.C.: Energy Information Administration. https://www.eia.gov/consumption/archive/manufacturing/051294.pdf.

———. 2017. Agency Information Collection Extension. *Federal Register* 82, no. 245 (December 22, 2017): 60720. https://www.federalregister.gov/documents/2017/12/22/2017-27596/agency-information-collection-extension

———. 2020. *2014 Manufacturing Energy Consumption Survey Methodology and Data Quality: Survey Design, Implementation, and Estimates.* Washington, D.C.: U.S. Energy Information Administration. https://www.eia.gov/consumption/manufacturing/data/2014/pdf/methodology.pdf.

———. 2022a. *Model Documentation Report: Industrial Demand Module of the National Energy Modeling System.* Washington, D.C.: Energy Information Administration. https://www.eia.gov/outlooks/aeo/nems/documentation/industrial/pdf/IDM_2022.pdf.

———. 2022b. "Section 6: Electricity." State Energy Data System 2022: Consumption. Washington, D.C.: U.S. Energy Information Administration, 2022. https://www.eia.gov/state/seds/sep_use/notes/use_elec.pdf.

———. 2022c. "Section 2: Coal." State Energy Data System 2022: Consumption. Washington, D.C.: U.S. Energy Information Administration, 2022. https://www.eia.gov/state/seds/sep_use/notes/use_coal.pdf.

———. 2022d. "Section 3: Natural Gas." State Energy Data System 2022: Consumption. Washington, D.C.: U.S. Energy Information Administration, 2022. https://www.eia.gov/state/seds/sep_fuel/notes/use_natgas.pdf.

———. 2022e. U.S. Energy Information Administration. "Section 5: Renewable Energy." State Energy Data System 2022: Consumption. Washington, D.C.: U.S. Energy Information Administration, 2022. https://www.eia.gov/state/seds/sep_fuel/notes/use_renew.pdf.

———. 2024a. "About EIA." About EIA. 2024. https://www.eia.gov/about/.

———. 2024b. “Tracking Electricity Consumption from U.S. Cryptocurrency Mining Operations.” Today in Energy, February 1, 2024. https://www.eia.gov/todayinenergy/detail.php?id=61364.

———. 2024c. Agency Information Collection Proposed Extension; Withdrawal. *Federal Register* 89, no. 51 (March 14, 2024): 18630. https://www.federalregister.gov/documents/2024/03/14/2024-05411/agency-information-collection-proposed-extension-withdrawal.

———. 2024d. “Hourly Electric Grid Monitor,” 2024. https://www.eia.gov/electricity/gridmonitor/index.php.

———. 2024e. “Total Energy Monthly Data.” May 28, 2024. https://www.eia.gov/totalenergy/data/monthly/index.php.

———. 2024f. “Survey Forms.” Survey, 2024. https://www.eia.gov//Survey/index.php.

U.S. Environmental Protection Agency. 2011. “Facility Registry Services.” Facility Registry Services. 2011. https://sor.epa.gov/sor_internet/registry/facilreg/home/basicinformation/.

———. 2015. Agency Information Collection Activities; Proposed Collection; Comment Request; Information Collection Request for the Greenhouse Gas Reporting Program. *Federal Register* 80, no. 214 (November 5, 2015): 68534-68536. https://www.federalregister.gov/documents/2015/11/05/2015-28275/agency-information-collection-activities-proposed-collection-comment-request-information-collection.

———. 2017. “Greenhouse Gas Reporting Program: Subpart C Methodologies.” U.S. Environmental Protection Agency. https://www.epa.gov/sites/default/files/2017-12/documents/subpartcmethodologiesfactsheet.pdf.

———. 2018. “Greenhouse Gas Reporting Program: Data Reported by Facilities Subject to the Direct Emitter Subparts C-V, X-II, and RR-TT,” February 8, 2018. https://www.epa.gov/sites/default/files/2020-04/documents/direct_emitters_cbi_table.pdf.

———. 2021. “Standardizing Reporting Codes for The Combined Air Emissions Reporting System (CAERS): Team Report.” U.S. Environmental Protection Agency. 2021. https://www.epa.gov/sites/default/files/2021-02/documents/standardize-code-tables-final-report.pdf

———. 2024a. “Enforcement and Compliance History Online.” https://echo.epa.gov/.

———. 2024b. “National Emissions Inventory (NEI).” Other Policies and Guidance, May 6, 2024. https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei.

———. 2024c. “EJScreen.” EPA’s Environmental Justice Screening and Mapping Tool (Version 2.3), August 9, 2024. https://ejscreen.epa.gov/mapper/.

U.S. General Accounting Office. 1983. *Selected Aspects of The Department of Energy's Operating Contract for the Solar Energy Research Institute.* GAO/RCED-83-66. U.S. General Accounting Office. https://www.gao.gov/assets/rced-83-66.pdf.

U.S. General Accounting Office. 1984. "Information on Budget Reductions in Energy Information Administration Programs." Washington, D.C.: General Accounting Office. https://www.gao.gov/assets/123707.pdf.

U.S. General Services Administration. 2024. "Data.Gov Home." Data.Gov. 2024. https://data.gov/.

Wachs, Liz, Colin McMillan, Gale Boyd, and Matt Doolin. 2022. *Exploring New Ways to Classify Industries for Energy Analysis and Modeling.* NREL/TP-6A20-82957. Golden, CO: National Renewable Energy Lab. https://doi.org/10.2172/1893357.

Young, Ben, Wesley W. Ingwersen, Matthew Bergmann, Jose D. Hernandez-Betancur, Tapajyoti Ghosh, Eric Bell, and Sarah Cashman. 2022. "A System for Standardizing and Combining U.S. Environmental Protection Agency Emissions and Waste Inventory Data." *Applied Sciences* 12 (7): 3447. https://doi.org/10.3390/app12073447.

Appendix A. FIED Data Fields

- Facility Identifiers
  - registryID
    - dtype: float
    - description: >
      The identification number assigned by the EPA Facility Registry Service to uniquely identify a facility site
    - source: EPA Facility Registry Service (FRS)
  - name
    - dtype: object
    - description:>
      The public or commercial name of a facility site (i.e., the full name that commonly appears on invoices, signs, or other business documents, or as assigned by the state when the name is ambiguous).
    - source: EPA Facility Registry Service (FRS)
  - eisFacilityID
    - dtype: float
    - description: first EPA Emissions Inventory System (EIS) ID associated with registryID
    - source: >
      EPA Facility Registry Service (FRS) and
      EPA National Emissions Inventory (NEI)
  - eisFacilityIDAdditional
    - dtype: float
    - description: first EPA Emissions Inventory System (EIS) ID associated with registryID
    - source: >
      EPA Facility Registry Service (FRS) and
      EPA National Emissions Inventory (NEI)
  - ghgrpID
    - dtype: float
    - description: first EPA Greenhouse Gas Reporting Program (GHGRP) ID associated with registryID
    - source: EPA Facility Registry Service (FRS)
  - ghgrpIDAdditional
    - dtype: float
    - description: additional Greenhouse Gas Reporting Program (GHGRP) IDs associated with registryID
    - source: EPA Facility Registry Service (FRS)
  - naicsCode
    - dtype: float
    - description: >
      first North American Industrial Classification System (NAICS) codes associated with registryID.
      See documentation for more information on NAICS codes and their descriptions.

```
    - source: <a href="https://www.epa.gov/frs">EPA Facility Registry Service (FRS)</a>
  - naicsCodeAdditional
    - dtype: float
    - description: >
        additional North American Industrial Classification System (NAICS) codes
        associated with registryID.
        See <a href="https://www.census.gov/naics/"> documentation </a> for more information
        on NAICS codes and their descriptions.
    - source: <a href="https://www.epa.gov/frs">EPA Facility Registry Service (FRS)</a>
  - siteTypeName
    - dtype: object
    - description: >
      descriptive name for the type of site. EPA values are currently
      'STATIONARY', 'POTENTIALLY CONTAMINATED SITE', 'FACILITY',
      'PORTABLE', 'CONTAMINATED SITE', 'BROWNFIELDS SITE', 'MOBILE',
      'CONTAMINATION ADDRESSED', and 'WATER SYSTEM'.
    - source: <a href="https://www.epa.gov/frs">EPA Facility Registry Service (FRS)</a>
  - sensitiveInd
    - dtype: float
    - description: Indicates whether or not the associated data is enforcement sensitive.
    - source: <a href="https://www.epa.gov/frs">EPA Facility Registry Service (FRS)</a>
  - smallBusInd
    - dtype: float
    - description: >
        Code indicating whether or not a business is requesting relief under EPA’s Small Business
Policy,
        which applies to businesses having less than 100 employees.
    - source: <a href="https://www.epa.gov/frs">EPA Facility Registry Service (FRS)</a>
  - envJusticeCode
    - dtype: float
    - description: >
        The code that identifies the type of environmental justice concern affecting the
        facility or enforcement action.
    - source: <a href="https://www.epa.gov/frs">EPA Facility Registry Service (FRS)</a>

- Unit Identifiers
  - eisUnitID
    - dtype: float
    - description: >
      unique Emissions Inventory System (EIS) identifier for unit associated with a emissions-
producing process.
      More than one eisUnitID may be associated with a eisProcessID.
    - source: <a href="https://www.epa.gov/air-emissions-inventories/national-emissions-
inventory-nei">EPA National Emissions Inventory (NEI)</a>
  - unitName
    - dtype: object
```

- description: reported name of unit.
- source: >
EPA Facility Registry Service (FRS) or
 EPA GHGRP.
- unitType
- dtype: float
- description: >
Identified unit type. May be taken from a reported unit type, unit description, or other data field.
- source: >
EPA National Emissions Inventory (NEI)
 EPA GHGRP.
- unitTypeStd
- dtype: object
- description: >
standardized unit type. Current types are 'dryer', 'other combustion', 'kiln', 'boiler', 'turbine', 'pump',
'generator', 'other' [non-combustion unit], 'heater', 'engine', 'furnace', 'oven', 'incinerator', 'flare',
'thermal oxidizer', 'compressor', 'distillation', 'building heat', and 'stove'.
- source: EPA National Emissions Inventory (NEI)
- unitDescription
- dtype: object
- description: description of the unit
- source: EPA National Emissions Inventory (NEI)
- designCapacity
- dtype: float
- description: design capacity of the unit. Directly reported, or obtained from unit description or other data fields.
- source: >
EPA National Emissions Inventory (NEI) or
 EPA GHGRP.
- designCapacityUOM
- dtype: object
- description: >
unit of measurement (UOM) for design capaicty. Currently includes 'MW' [megawatts], 'TON/DAY',
'E3LB/HR', 'LB/HR', 'TON/HR', 'GAL', 'FT3/DAY', 'E3GAL/HR', 'DATAMIGR', 'E3FT2/HR', 'TON/YR',
'E6FT2/YR', 'E3BDFT/YR', 'GAL/HR', 'LB/YR', 'GAL/DAY', 'E6BDFT/YR',
'GAL/YR', 'DATAMIGRATION', 'FT3/MIN', 'GAL/MIN', 'FT2/HR', 'AMP-HR/HR', 'E3FT3/DAY',

'FT3SD/HR', 'FT2/YR', and 'BBL'.
  - source: EPA National Emissions Inventory (NEI)

- Geographic Identifiers
  - latitude
    - dtype: float
    - description: latutide associated with registryID
    - source: EPA Facility Registry Service (FRS)
  - longitude
    - dtype: float
    - description: longitude associated with registryID
    - source: EPA Facility Registry Service (FRS)
  - geoID
    - dtype: float
    - description: >
      Census Geographic Identifier. Numeric codes that uniquely identify all administrative/legal and
      statistical geographic areas for which the Census Bureau tabulates data. See
       GEOID overview
      for more information.
    - source: EPA Facility Registry Service (FRS)
  - locationAddress
    - dtype: object
    - description: >
      The address that describes the physical (geographic) location of the front door or
      main entrance of a facility site, including
      urban-style street address or rural address.
    - source: EPA Facility Registry Service (FRS)
  - postalCode
    - dtype: float
    - description: >
      The combination of the five digit ZoneImprovement Plan (ZIP) code and the four
      digit extension code (if available) that represents the geographic segment that is
      a subunit of the ZIP Code, assigned by the U.S. Postal Service to a geographic
      location, where the facility site is located.
    - source: EPA Facility Registry Service (FRS)
  - cityName
    - dtype: object
    - description: >
      The name of the city, town, village or other locality, when identifiable, within whose
      boundaries (the majority of) the facility site is located. This is not always the same as
      the city used for USPS mail delivery.
    - source: EPA Facility Registry Service (FRS)
  - countyName

- dtype: object
- description: county name of facility.
- source: EPA Facility Registry Service (FRS)

- stateCode
  - dtype: object
  - description: two-letter state abbreviation (e.g., "AL") of facility
  - source: EPA Facility Registry Service (FRS)
- countyFIPS
  - dtype: float
  - description: >
    The Federal Informationl Processing Standard (FIPS) code that represents the county or county equivalent and the state or state equivalent of the United States.
  - source: EPA Facility Registry Service (FRS)
- legislativeDistrictNumber
  - dtype: object
  - description: The number that represents a Legislative District within a state.
  - source: EPA Facility Registry Service (FRS)
- epaRegionCode
  - dtype: float
  - description: EPA Region Code associated with registryID
  - source: EPA Facility Registry Service (FRS)
- hucCode8
  - dtype: float
  - description: >
    Hydrolic Unit Code (HUC) from the USGS. Accoring to the USGS, hydrologic unitsare classified
    into four levels: regions, sub-regions, accounting units, and cataloging units.
    Each unit is identified by a unique hydrologic unit code (HUC) consisting of two to eight digits based on its classification.
  - source: EPA Facility Registry Service (FRS)

- Energy and Emissions Identifiers
  - fuelType
    - dtype: float
    - description: >
      fuel type that is directly reported, or is derived from SCC codes, unit or process descriptions, or other data fields.
    - source: >
      EPA National Emissions Inventory (NEI) or
       EPA GHGRP.
  - fuelTypeStd
    - dtype: object
    - description: >
      standardized fuel type. Current types are 'diesel', 'naturalGas', 'resFuelOil' [residual fuel oil], 'biomass', 'lpgHGL', 'gasoline', 'other', 'coal', 'coke', 'jetA'

- source: EPA National Emissions Inventory (NEI)
- SCC
  - dtype: float
  - description: >
    Source Classification Codes are a standardized hierarchical classification of the
    different types of activities that generate air emissions. See additional information from
    this documentation.
  - source: EPA National Emissions Inventory (NEI)
- eisProcessID
  - dtype: float
  - description: Emissions Inventory System (EIS) identifier for a emissions-producing process.
  - source: EPA Facility Registry Service (FRS)
- processDescription
  - dtype: float
  - description: description of an emissions-producing process (eisProcessID)
  - source: EPA Facility Registry Service (FRS)
- energyMJq0
  - dtype: float
  - description: minimum estimated energy use by unit in megajoules (MJ). Derived from reported EPA NEI emissions.
  - source: EPA National Emissions Inventory (NEI)
- energyMJq2
  - dtype: float
  - description: median estimated energy use by unit in megajoules (MJ). Derived from reported EPA NEI emissions.
  - source: EPA National Emissions Inventory (NEI)
- energyMJq3
  - dtype: float
  - description: third quartile estimated energy use by unit in megajoules (MJ). Derived from reported EPA NEI emissions.
  - source: EPA National Emissions Inventory (NEI)
- throughputTonneQ0
  - dtype: float
  - description: minimum estimated throughput by unit in metric tons. Derived from reported EPA NEI emissions.
  - source: EPA National Emissions Inventory (NEI)
- throughputTonneQ2
  - dtype: float

- description: median estimated throughput by unit in metric tons. Derived from reported EPA NEI emissions.
- source: EPA National Emissions Inventory (NEI)
- throughputTonneQ3
  - dtype: float
  - description: third quartile estimated throughput by unit in metric tons. Derived from reported EPA NEI emissions.
  - source: EPA National Emissions Inventory (NEI)
- ghgsTonneCO2eQ0
  - dtype: float
  - description: >
    minimum greenhouse gas emissions by unit in metric tons CO2 equivalent (TonneCO2e),
    calculated from either data reported by EPA NEI, derived from data reported by EPA NEI.
  - source: EPA National Emissions Inventory (NEI)
- ghgsTonneCO2eQ2
  - dtype: float
  - description: >
    median greenhouse gas emissions by unit in metric tons CO2 equivalent (TonneCO2e),
    calculated from either data reported by EPA NEI, derived from data reported by EPA NEI.
  - source: EPA National Emissions Inventory (NEI)
- ghgsTonneCO2eQ3
  - dtype: float
  - description: >
    third quartile of greenhouse gas emissions by unit in metric tons CO2 equivalent (TonneCO2e),
    calculated from either data reported by EPA NEI, derived from data reported by EPA NEI.
  - source: EPA National Emissions Inventory (NEI)
- energyMJ
  - dtype: float
  - description: energy use by unit in MJ, either derived from reported EPA GHGRP emissions, or taken directly from reported data.
  - source: EPA GHGRP.
- ghgsTonneCO2e
  - dtype: float
  - description: >
    emissions of greenhouse gases (GHGs) in metric tons carbon dioxide
    equivalents (TonneCO2e) reported by the EPA GHGRP.
  - source: EPA GHGRP.
- energyEstimateSource
  - dtype: float
  - description: Source of energy estimate. Value is either 'nei' or 'ghgrp'.

- source: not applicable
- ghgsEstimateSource
  - dtype: float
  - description: Source of ghg emissions estimate. Value is either 'nei' or 'ghgrp'.
  - source: not applicable

- Other Identifiers
- weeklyOpHoursLow_q1
  - dtype: float
  - description: >
    lower bound of 95% confidence interval for average
    weekly operating hours first quarter. Calculated using reported data.
  - source: Census QPC
- weeklyOpHoursLow_q2
  - dtype: float
  - description: >
    lower bound of 95% confidence interval for average
    weekly operating hours second quarter. Calculated using reported data.
  - source: Census QPC
- weeklyOpHoursLow_q3
  - dtype: float
  - description: >
    lower bound of 95% confidence interval for average
    weekly operating hours thrid quarter. Calculated using reported data.
  - source: Census QPC
- weeklyOpHoursLow_q4
  - dtype: float
  - description: >
    lower bound of 95% confidence interval for average weekly operating hours fourth quarter.
    Calculated using reported data.
  - source: Census QPC
- weeklyOpHours_q1
  - dtype: float
  - description: reported average weekly operating hours first quarter.
  - source: Census QPC
- weeklyOpHours_q2
  - dtype: float
  - description: reported average weekly operating hours second quarter
  - source: Census QPC
- weeklyOpHours_q3
  - dtype: float
  - description: reported average weekly operating hours third quarter
  - source: Census QPC
- weeklyOpHours_q4
  - dtype: float
  - description: reported average weekly operating hours fourth quarter

```
    - source: <a href="https://www.census.gov/programs-surveys/qpc.html"> Census QPC </a>
- weeklyOpHoursHigh_q1
    - dtype: float
    - description: >
        upper bound of 95% confidence interval for average weekly operating hours first quarter.
        Calculated using reported data.
    - source: <a href="https://www.census.gov/programs-surveys/qpc.html"> Census QPC </a>
- weeklyOpHoursHigh_q2
    - dtype: float
    - description: >
        upper bound of 95% confidence interval for average weekly operating hours second
quarter.
        Calculated using reported data.
    - source: <a href="https://www.census.gov/programs-surveys/qpc.html"> Census QPC </a>
- weeklyOpHoursHigh_q3
    - dtype: float
    - description: >
        upper bound of 95% confidence interval for average
        weekly operating hours third quarter. Calculated using reported data.
    - source: <a href="https://www.census.gov/programs-surveys/qpc.html"> Census QPC </a>
- weeklyOpHoursHigh_q4
    - dtype: float
    - description: >
        upper bound of 95% confidence interval for average
        weekly operating hours fourth quarter. Calculated using reported data.
    - source: <a href="https://www.census.gov/programs-surveys/qpc.html"> Census QPC </a>
```

Appendix B. FIED Compilation

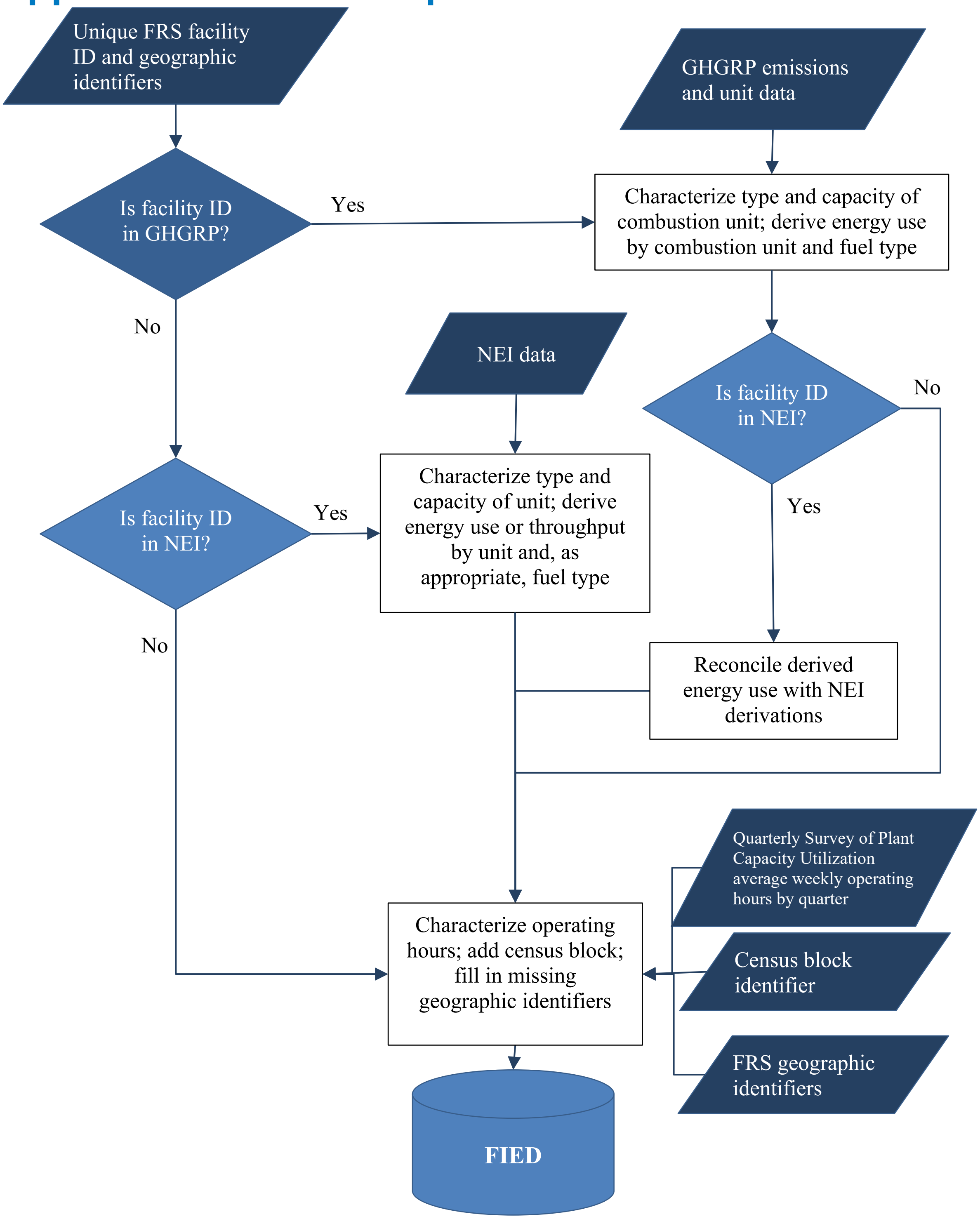


Figure B- 1. Summary flowchart for compiling the FIED

Appendix C. NEI Unit Characterization

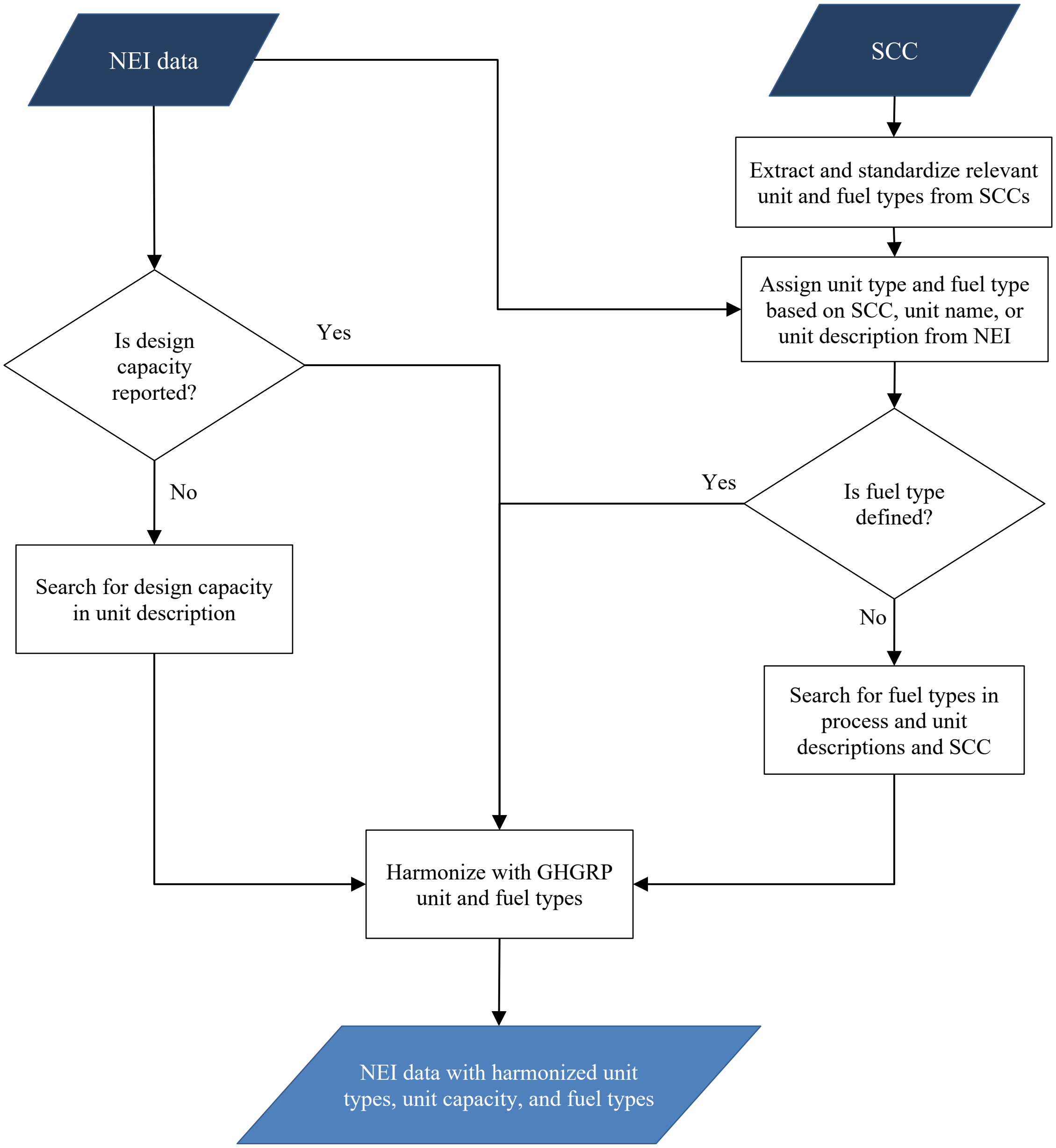
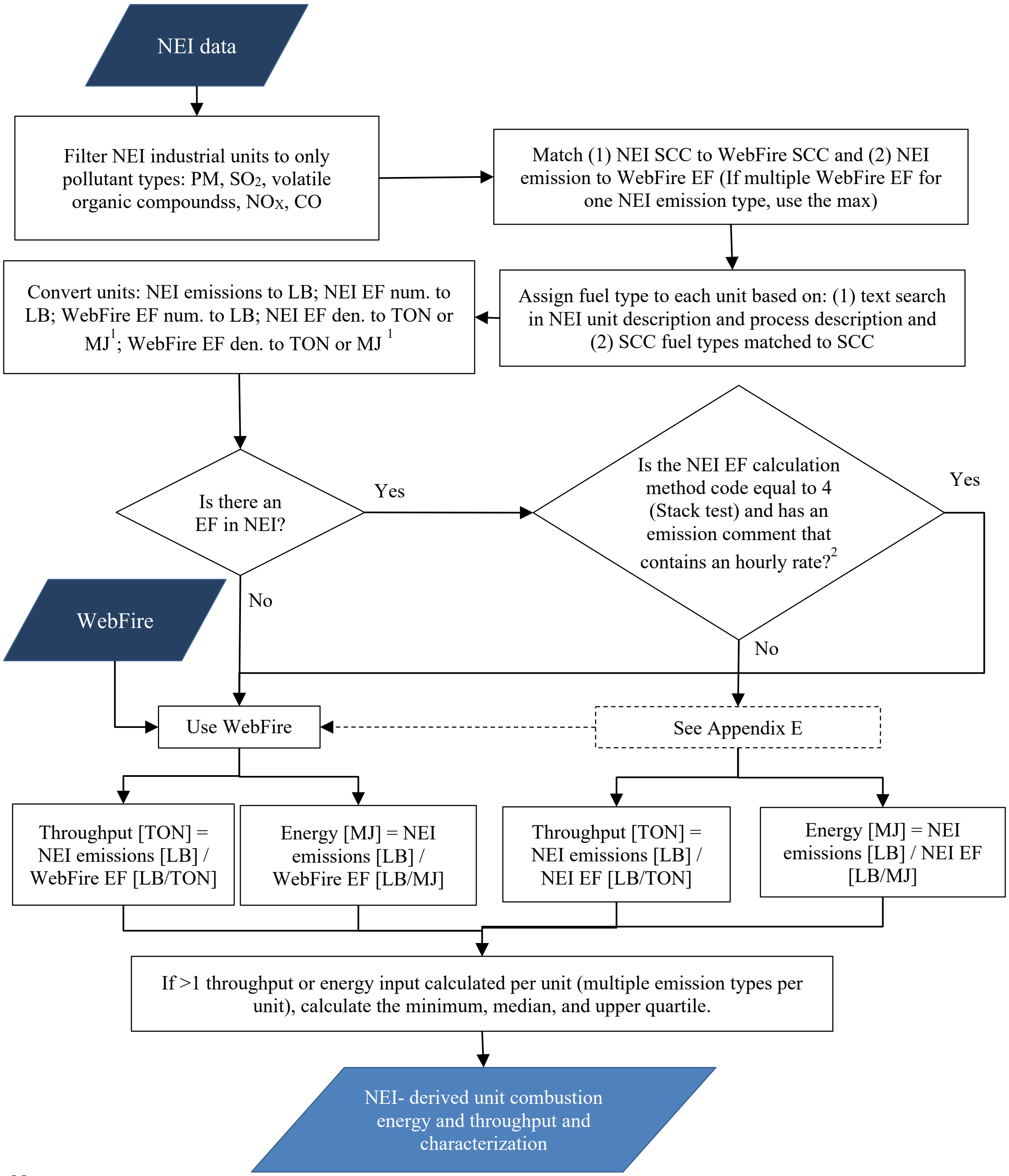


Figure C- 1. Summary flowchart for characterizing NEI units

Appendix D. NEI Energy and Throughput Derivations



Notes:
[1]Convert units of fuel energy that are volumetric or mass-based to MJ based on fuel type.
[2]If units have an "emission comment" of text indicating an hourly emission rate (e.g., LB/HR) that matches the reported EF and does not match reported EF units, NEI EFs are not used in this case.

Figure D- 1. Summary flowchart for deriving energy and throughput estimates from NEI data

Appendix E. Addressing Disparate NEI EFs

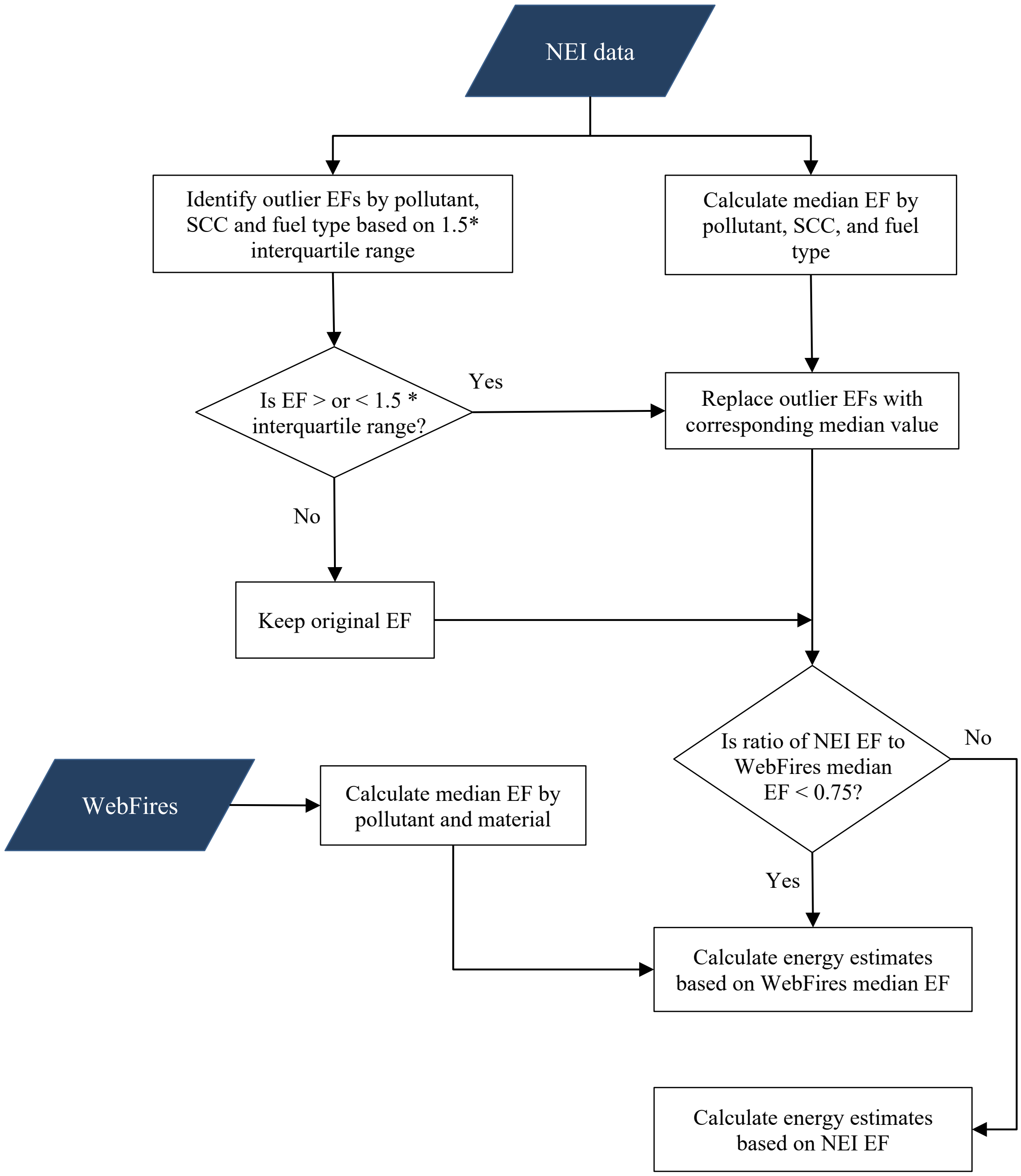


Figure E- 1. Summary flowchart for addressing disparate NEI EFs

Appendix F. GHGRP Energy Derivations and Unit Characterization

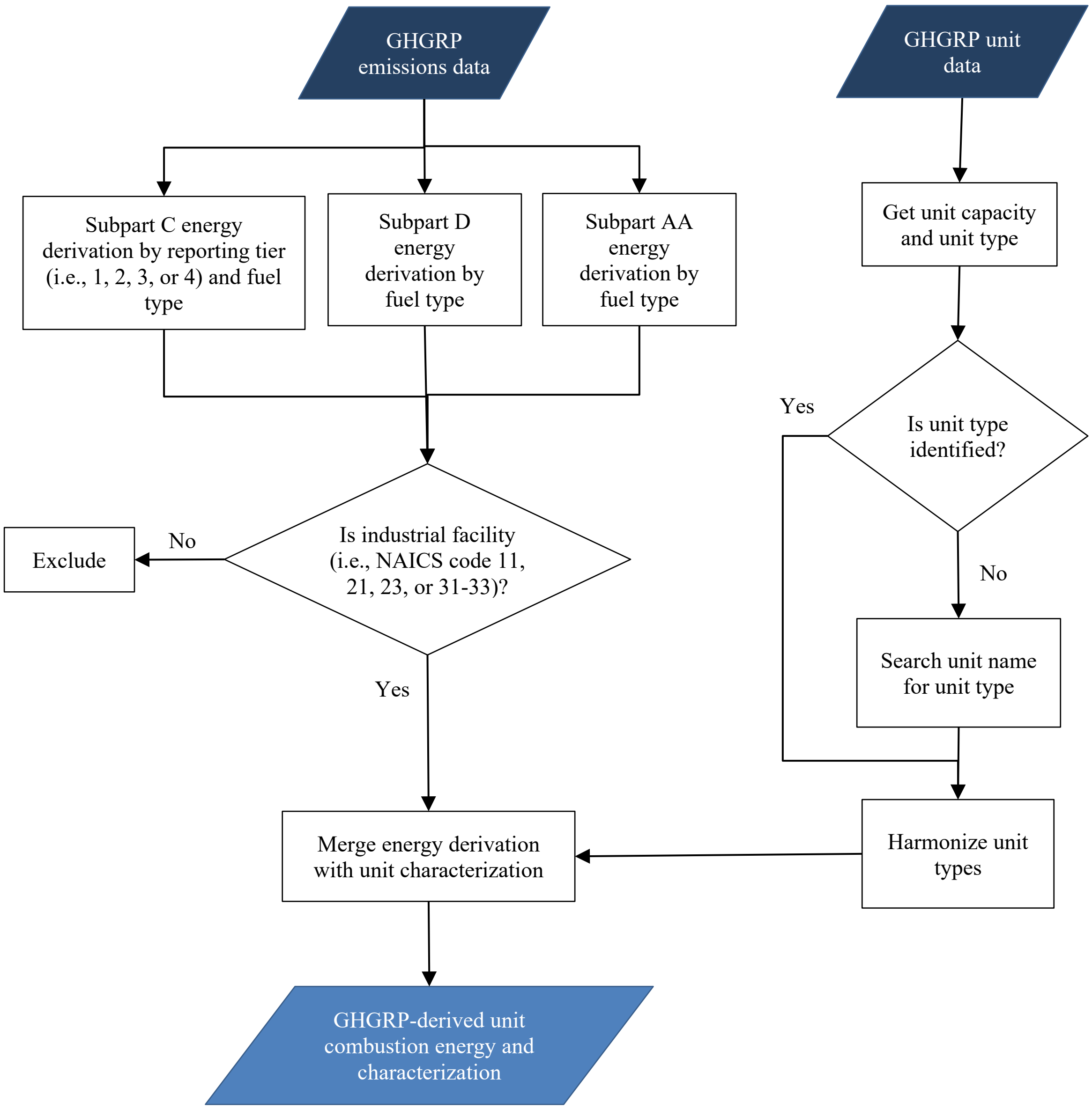


Figure F- 1. Summary flowchart for characterizing units and deriving energy estimates from GHGRP data

Appendix G. Blending NEI and GHGRP Energy Derivations

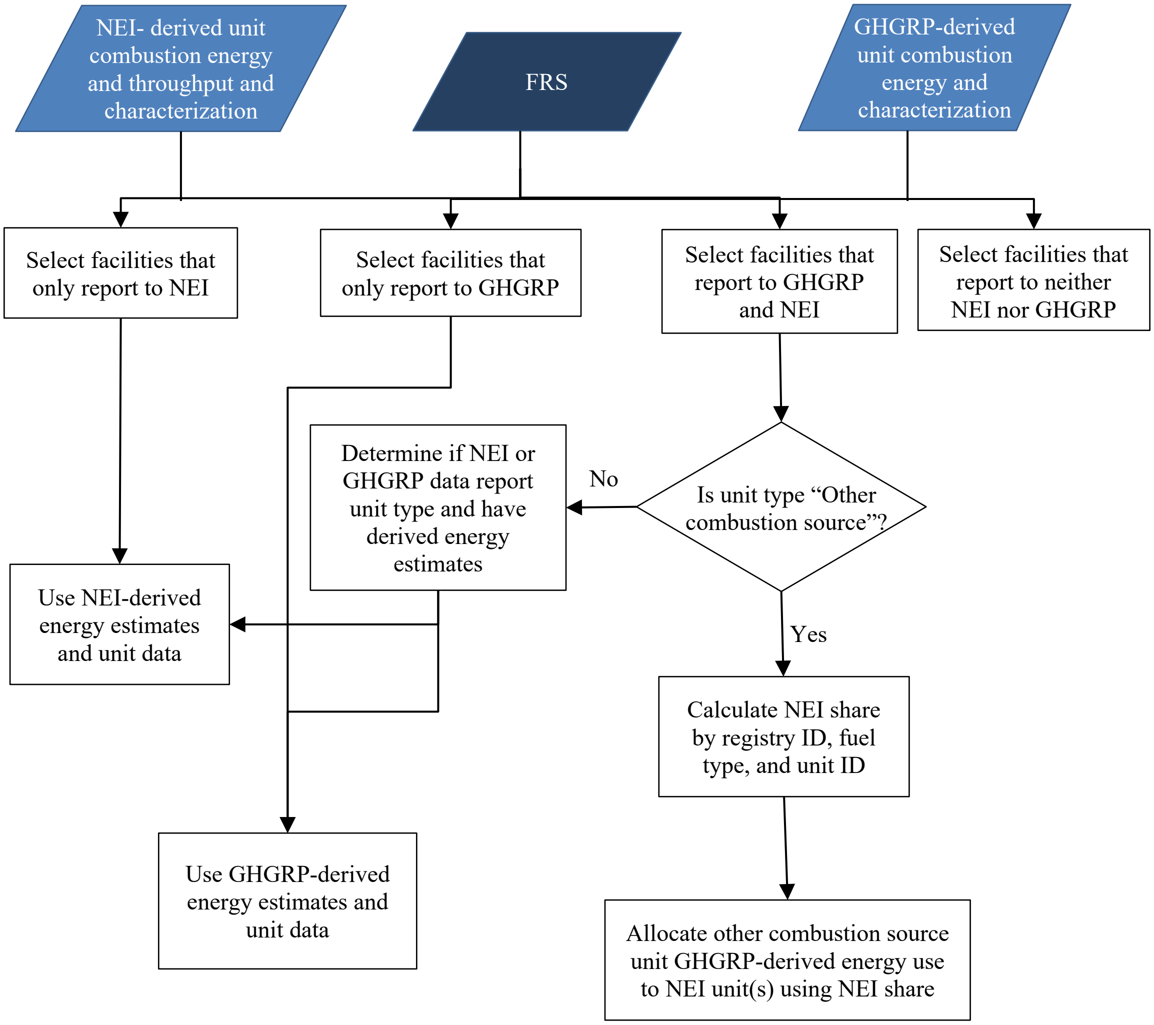


Figure G- 1. Summary flowchart for blending derived energy estimates from NEI and GHGRP data