New Mexico Comprehensive Energy Transition Strategy

Data Management Strategy

CHARTING NEW MEXICO'S SUSTAINABLE FUTURE A COMPREHENSIVE ROADMAP FOR THE ENERGY TRANSITION AND CLIMATE RESILIENCE

October 2025

Authors: Jean-Lucien Fonquergne[1*], Marco Lopez[2], Martha Cather[1], Shari Kelley[3], Virginia McLemore[3], Van Romero[4] Robert Balch[1].

[1] New Mexico Tech, Petroleum Recovery Research Center
[2] Papermill LLC
[3] New Mexico Tech, New Mexico Bureau of Geology and Minerals Resources
[4] New Mexico Tech, Physics Department
[*]Corresponding author: jean-lucien.fonquergne@nmt.edu

Acknowledgments

This report was prepared by the New Mexico Institute of Mining and Technology (New Mexico Tech) in collaboration with the Energy, Minerals and Natural Resources Department (EMNRD) as part of the Comprehensive Energy Transition Strategy (CETS) initiative. The authors thank the staff and leadership at EMNRD for their guidance and partnership throughout this project, as well as the EFI Foundation (EFIF) and the Climate Equity Foundation (CEF) for their outstanding support and collaboration throughout the CETS phase 1 project.

Special appreciation is extended to all agency representatives, researchers, and stakeholders who contributed data, insights, and feedback through interviews, surveys and correspondences that informed the development of the DETECT (*Data for the Energy Transition, Economy, Climate Change, and Technologies*) platform and the accompanying data management strategy.

This work was supported by **EMNRD Contract No. 25-521-0100-0226**.

Executive Summary

This report presents the Data Management Strategy for the Comprehensive Energy Transition Strategy (CETS) led by EMNRD, and outlines the foundation for developing **DETECT: Data for the Energy Transition, Economy, Climate Change, and Technologies**. The goal is to establish a unified, state-level data system that consolidates energy, emissions, and infrastructure information across agencies and stakeholders, enabling New Mexico to plan and manage its energy transition more effectively.

New Mexico's energy landscape is complex, encompassing oil and gas production, renewables, complex grid infrastructure, and potential for technologies development like carbon sequestration and enhanced geothermal energy. However, existing datasets are fragmented across state, federal, and private sources, often inconsistent in format, and not regularly updated nor easily accessible. This limits the state's ability to make data-driven policy decisions or track progress toward climate goals.

In this report, the **DETECT platform** is proposed as a centralized, phased built, and modular solution. First, it will act as a "database of databases" to integrate these diverse datasets with their raw data. Then, it will integrate and consolidate the data to make it more accessible. Finally, it will enable standardized reporting, interagency collaboration, and public access through dashboards and maps similar to those used by organizations such as the California Independent System Operator (CAISO) or the U.S Environmental Protection Agency Greenhouse Gas Reporting Program (EPA GHGRP). The platform will be developed in stages:

1. **Phase 1:** Integrate existing public data from EIA, EPA, state agencies and other sources;
2. **Phase 2:** Expand to include more data that could be regulated or proprietary data;
3. **Phase 3:** Develop advanced tools for planning, regulation, and public transparency.

Another key component of DETECT will be education and outreach, using storytelling and regional data insights to communicate energy and climate trends to policymakers, industry, and the public. This approach supports community understanding of regional disparities, from the oil-producing southeast and the Four Corners, to the renewable-rich central and eastern regions.

A comparable model for DETECT already exists in New Mexico through the Water Data Portal, developed by the New Mexico Bureau of Geology following the Water Data Act passed by the state legislature in 2019. That initiative established authority, governance, and collaboration across agencies to centralize and share critical water data. This demonstrates how a similar legislative and institutional framework could guide the development of DETECT, providing a tested template for building an integrated, state-managed energy and climate data platform.

Finally, the report emphasizes that achieving this vision depends on both resources and governance. Data platform maintenance, sustainable staffing, data-sharing agreements, and enabling legislation are needed to ensure consistent data reporting. A fully resourced DETECT plan would give New Mexico the capacity to manage, update, and use data effectively. In more constrained scenarios, a phased implementation ensures progress.

In summary, this strategy aims to transform fragmented datasets into a coherent system that strengthens New Mexico's energy governance, supports transparency, and positions the state as a national leader in evidence-based climate and energy planning.

Table of Contents

List of Figures and Tables

Acronyms

AQB: Air Quality Bureau (New Mexico Environment Department)
ASAI: Average Service Availability Index
API: Application Programming Interface
CAIDI: Customer Average Interruption Duration Index
CAISO: California Independent System Operator
CAP: Climate Action Plan
CEF: Climate Equity Foundation
CETS: Comprehensive Energy Transition Strategy
CPUC: California Public Utilities Commission
CUSP: Carbon Utilization and Storage Partnership (of the Western U.S.)
DER: Distributed Energy Resource
DETECT: Data for the Energy Transition, Economy, Climate Change, and Technologies
DHS: U.S. Department of Homeland Security
DOE: U.S. Department of Energy
ECAM: Energy, Conservation and Management
EIA: U.S. Energy Information Administration
EIA-860: Annual Electric Generator Report (EIA)
EIA-923: Power Plant Operations Report (EIA)
EIA-930: Hourly Electricity Operations (EIA)
EFIF: EFI Foundation
EMNRD: Energy, Minerals and Natural Resources Department (New Mexico)
EPA: U.S. Environmental Protection Agency
EPE: El Paso Electric
ERCOT: Electric Reliability Council of Texas
FERC: Federal Energy Regulatory Commission
FLIGHT: Facility Level Information on GreenHouse gases Tool (EPA)
GHG: Greenhouse Gas
GHGRP: Greenhouse Gas Reporting Program (EPA, 40 CFR Part 98)
GIO: Geographic Information Officer (New Mexico)
GIS: Geographic Information System
HIFLD: Homeland Infrastructure Foundation-Level Data (DHS)
IBEW: International Brotherhood of Electrical Workers
ICA: Integration Capacity Analysis (distribution hosting maps)
IOU: Investor-Owned Utility
IRP: Integrated Resource Plan
ISO: Independent System Operator
LANL: Los Alamos National Laboratory
MECS: Manufacturing Energy Consumption Survey (EIA)
MISO: Midcontinent Independent System Operator
MOUs: Memoranda of Understanding
MRV: Measurement, Reporting, and Verification
NDA: Non-Disclosure Agreement
NMBGMR: New Mexico Bureau of Geology & Mineral Resources
NMED: New Mexico Environment Department

NMEI / NMEIC: New Mexico Energy Initiatives / Consortium
NMRETA: New Mexico Renewable Energy Transmission Authority
NMPRC / PRC: New Mexico Public Regulation Commission
NMSU: New Mexico State University
NMT: New Mexico Tech (New Mexico Institute of Mining & Technology)
NREL: National Renewable Energy Laboratory
NMOCD / OCD: (New Mexico) Oil Conservation Division (EMNRD)
NPMS: National Pipeline Mapping System (PHMSA)
NYSERDA: New York State Energy Research and Development Authority
OASIS: Open Access Same-Time Information System (FERC Order 889)
PG&E: Pacific Gas and Electric Company
PIES: Permian Integrated Energy Systems
PJM: PJM Interconnection (RTO)
PHMSA: Pipeline and Hazardous Materials Safety Administration (U.S. DOT)
RECA: New Mexico Rural Electric Cooperative Association
RECS: Residential Energy Consumption Survey (EIA)
ResStock / ComStock: NREL Residential/Commercial building stock models
RTO: Regional Transmission Organization
SAIDI: System Average Interruption Duration Index
SAIFI: System Average Interruption Frequency Index
SEDS: State Energy Data System (EIA)
SIC: State Investment Council (New Mexico)
SLO: State Land Office (New Mexico)
SNL: Sandia National Laboratories
SPP: Southwest Power Pool (RTO)
UNM: University of New Mexico
ZEV: Zero-Emission Vehic

1 DETECT Development Strategy

1.1 Why This Matters

This report establishes the baseline of New Mexico's energy and climate data and outlines the framework to develop **DETECT: Data for the Energy Transition, Economy, Climate Change, and Technologies**. DETECT will serve as the state's central platform to collect, manage, and share consistent, verified information on energy production, consumption, and greenhouse gases (GHG) emissions.

Reliable data is essential for informed decisions. It enables state agencies and lawmakers to design effective policies, track progress toward clean energy and emissions goals, and demonstrate accountability. For industry, access to clear and accurate data supports compliance, planning, and investment. For researchers and communities, it builds transparency and trust in how New Mexico manages its energy transition.

As one of the nation's top energy-producing states, New Mexico's ability to attract investment and manage its transition depends on its data capacity. DETECT provides the structure to meet that need. It is built to grow in phases, from organizing existing public data to integrating new reporting requirements as resources and legislation allow, ensuring meaningful progress even under limited budgets.

In short, the DETECT initiative turns data into a tool for better energy data literacy, governance, stronger policy, and smarter investment in New Mexico's energy future.

1.2 Methodology

To develop a state data management strategy and guide the creation of the DETECT platform, our team of energy experts conducted a comprehensive review of state and federal databases to assess the availability, quality, and structure of energy and emissions data. We evaluated both existing datasets and analytical tools used by New Mexico and other states to understand best practices for data integration and accessibility. Particular attention was given to metadata, which is essential for evaluating data accuracy, update frequency, and overall reliability. A detailed gap analysis was then performed to identify missing or incomplete datasets critical for policy planning, emissions tracking, and compliance with state and federal climate goals.

Building on this review, we defined a data management strategy for the CETS, prepared an inventory of existing data (Chapter 2, 3 and 4 – Data inventory), and developed actionable recommendations for new collection processes, digitization, and interagency coordination (Chapter 5 and 6). The resulting recommended approach is a 3-phase strategy for robust, scalable data management strategy that can be implemented step by step to guide New Mexico's long-term energy and climate policy planning.

In addition, our team developed a digital platform first prototype during this first phase of the CETS. This prototype has been developed to demonstrate the effectiveness and suitability of the proposed architecture and softwares proposed in this report.

For the future development of the platform, we recommend a three phases approach, the first one incorporates only publicly available data to ensure transparency and accessibility. In parallel, the framework anticipates a separate, secure component where proprietary or regulatory compliance data can be submitted and managed. This structure balances transparency with confidentiality, supporting both public insight and future compliance or regulatory needs.

1.2.1 Phase 1 - Short-Term (Publicly Available Data)

In the initial stage, the DETECT platform should leverage data that is already public, reliable, and readily accessible to quickly produce useful tools. This includes datasets from the U.S. Energy Information Administration (EIA), such as the State Energy Data System (SEDS) and EIA Application Programming Interfaces (APIs), as well as U.S. Environmental Protection Agency (EPA) greenhouse gas reporting data through the Greenhouse Gas Reporting Program (GHGRP) and Facility Level Information on Greenhouse Gases Tool (FLIGHT). Additional sources include New Mexico Oil Conservation Division (NMOCD) methane reporting data (Form C-115 and the Natural Gas Waste Dashboard) and other datasets from Energy, Conservation and Management Division (ECAM) and the New Mexico Environment Department (NMED).

These sources can immediately support dashboards, interactive maps, and automated reports for high level electricity generation, sectoral fuel use, GHG emissions, and methane tracking. Tools in this phase should demonstrate DETECT value to policymakers and the public by providing good User Interface (UI) with interactive visualizations and sectoral energy/GHG snapshots with storytelling, while establishing consistent metadata and data-quality checks.

For Phase 1, given the expected timeframe, the focus should be on defining a clear, achievable goal for the prototype. It should be centered on a few high-impact datasets and visualization products as described above, while simultaneously establishing the processes for data acquisition, management, and support in phase 2.

It will also be important to anticipate data variability across agencies, as formats, protocols, and rules for processing often differ or change over time, especially given the current uncertainties of some programs at the federal level. To ensure stability and long-term usability, the team should begin setting up communication protocols and contingency plans with agencies during this phase, laying the groundwork for data consistency and smoother integration in future phases.

1.2.2 Phase 2 - Mid-Term (Agreements and Modest Effort)

The second stage could expand the platform by incorporating data that is collected but not yet public, requiring targeted agreements or modest investment to access. Examples include GHG emissions data collected by NMED Air Quality Bureau (AQB), dashboards for ECAM programs data, IOU operational datasets (hourly generation, sub-sector load data, interconnection queues), and oil and gas reporting platforms in partnership with NMOCD or NMOGA that must be structured into databases (Updates on GHG estimations, support data reconciliation between OCD and NMED, and update Go-tech).

In addition, partnerships with utilities, co-ops, and trade associations could theoretically unlock more granular data such as infrastructure data (pipeline capacities, transmission congestion, transformer/substation loading) and demand-side details by sub-sectors. At this stage, the DETECT platform should evolve from historical dashboards toward **near real-time monitoring tools**, sub-sectoral demand analytics, and integrated compliance reporting modules and tools, creating mid-term value for regulators, planners, and researchers.

The agreements needed for Phase 2 can begin to be negotiated in parallel to Phase 1. Early engagement with utilities, co-ops, trade associations, and state agencies such as the ECAM, NMOCD, NMED AQB or the Public Regulation Commission (PRC) will help establish trust and the terms for accessing operational, infrastructure, and compliance datasets.

In addition, the DETECT platform should be developed in response to the needs of agencies, legislators, and the public, rather than being implemented without meaningful consultation and stakeholder input. Building an inclusive working group with these stakeholders and starting these negotiations during Phase 1 ensures that once the initial prototype is operational, more detailed and higher-value datasets can be integrated smoothly during phase 2, accelerating the transition from basic dashboards to more advanced tools that incorporate real-time monitoring, sub-sectoral demand, and infrastructure data.

1.2.3 Phase 3 - Long-Term (Legislative and Structural)

The final stage envisions an ideal, state-level data ecosystem enabled by legislative authority and expanded regulatory mandates. This could include a comprehensive, New Mexico-specific greenhouse gas reporting program modeled after EPA's GHGRP subparts targeting the energy systems but adapted for state needs; mandatory reporting of Scope 1, 2, and 3 emissions or reporting based on a specific framework defined by the state; systematic collection of industrial sub-sector consumption (oil and gas, mining, refining, manufacturing, data centers); and state-level energy surveys similar to EIA's national Manufacturing Energy Consumption Survey (MECS) or Residential Energy Consumption Survey (MECS), but statewide, to have a full picture of the New Mexico energy ecosystem.

Additional long-term goals for specific sectors could include integration of regulatory and reporting systems, harmonized facility identifiers between agencies, and the use of digital MRV (dMRV) technologies such as satellite monitoring, automated datasets anomaly detection, and smart-sensor reporting. At this phase, the DETECT platform would become similar to a statewide digital twin. The digital twin will be a virtual model of New Mexico's energy system that integrates real-time data on generation, transmission, consumption, and emissions to monitor, simulate, and plan future scenarios. It should be built on: (1) mandatory reporting with MRV for consistency and data quality, (2) surveys to capture harder-to-track data, and (3) dashboards and maps for public accessibility and transparency. Finally, the platform should also consolidate cross-sector energy, environment, water, land, and socio-economic data into easily accessible datasets.

1.2.4 Importance of Robust and Scalable Development

To conclude, the DETECT platform should be designed from phase 1 to be able to evolve through different development scenarios depending on available resources and needs. In the fully resourced scenario, New Mexico would establish a complete, state-managed digital twin of its energy and emissions system, supported by dedicated staff, legislative reporting requirements, and seamless data integration across agencies and industries. This version would provide near real-time dashboards, predictive modeling, and comprehensive public transparency. Under a moderate resource scenario, the platform could expand beyond the prototype to include key partnerships and data-sharing agreements with utilities, co-ops, and state agencies, supported by a small technical team to maintain data pipelines and analytics. In a limited resource scenario, DETECT would focus on maximizing existing public datasets to build targeted visualization tools that demonstrate value and guide future investment. Each scenario builds toward the same long-term goal, ensuring progress, strategical approach and impact even within budgetary and administrative constraints.

To that end, a robust and scalable development approach (Chapter 5) is essential to ensure that DETECT remains functional and relevant over time. Energy systems, technologies, and data requirements will continue to evolve, and the platform must be able to adapt without requiring complete redesigns. Building scalability from the start allows for seamless integration of new datasets, analytical tools, and policy frameworks as they emerge. It also ensures that investments made early in the project yield long-term value by creating a durable digital infrastructure capable of supporting both near-term decision-making and future innovation.

1.3 Building a DETECT Team

Starting in Phase 2, maintaining and advancing the DETECT platform databases and tools will require a dedicated full-time equivalent (FTE) team with complementary expertise. The ideal core team should include at least:

- 2 Data Engineers/Developers to manage data pipelines, APIs, and system architecture;
- 1 Front-End Developer to design and maintain UI, dashboards, and visualization tools;
- 1 Energy Analyst to guide development and integration of new tools;
- 1 GHG Analyst to handle emissions inventories, verification, and reporting alignment with state GHG accounting frameworks and Monitoring, Reporting and Verification (MRV);
- and 1 Data Steward responsible for overseeing data privacy, quality, integrity, security, and compliance with state rules.

Together, this small but multiskilled team would ensure the DETECT platform remains technically sound, valuable, and compliant as it scales across subsequent phases.

1.4 Coordination with State Agencies and other Stakeholders

A critical aspect of the DETECT platform's success will be the verification and coordination of data across state agencies and partners. Establishing consistent procedures for data validation ensures that the information used for decision-making is accurate, traceable, and credible.

Verification should be built into the data ingestion process from the start, using clear metadata standards, automated quality checks. For example, data from the NMED and the OCD should follow common reporting formats or include at least standardized fields for identification, timestamps, units, and data sources. Simply getting facilities identification numbers (IDs) that are consistent across agencies would improve data reconciliation.

Because equally important is interagency coordination. Each agency currently collects data for its own regulatory or operational purposes, but these datasets often remain siloed. To avoid duplication and inconsistency, DETECT should establish formal data-sharing agreements with agencies such as NMED, EMNRD, OCD, PRC. These agreements would clarify roles and responsibilities, define data submission schedules, and set expectations for review and validation.

Regular coordination meetings and a shared governance framework will be necessary to align data updates, ensure compliance with reporting standards, and address emerging data needs. Ultimately, the credibility and usefulness of DETECT will depend not only on the technology but on a collaborative data governance culture, where agencies and partners work together to maintain a reliable, transparent, and up-to-date picture of New Mexico's energy and climate landscape.

It is important to note, based on discussions with state agencies, that workforce capacity is already a significant challenge; therefore, the data-sharing efforts under DETECT should minimize additional workload for state employees. Instead, DETECT should be designed to support agencies, not burden them, with the DETECT team taking the lead on data integration, processing, and management, allowing agency staff to focus on their core missions while benefiting from improved data accessibility and coordination.

Therefore, the stakeholder working group for DETECT should remain focused and manageable, as a too large group risks slowing progress and diluting effectiveness. A phased approach is recommended, beginning with a core working group in Phase 1 composed of key stakeholders capable of setting up a strong framework and ensuring the platform is successfully launched. Then, the working group could extend and include:

- **State Agencies Point of Contact (POC):** Public Regulation Commission (PRC), Oil Conservation Division (OCD), New Mexico Environment Department (NMED), Energy, Minerals and Natural Resources Department (EMNRD), Indian Affairs Department (IAD), Department of Workforce Solutions (DWS), State Investment Council (SIC), State Land Office (SLO).

- **Investor-Owned Utilities (IOUs):** Public Service Company of New Mexico (PNM), El Paso Electric (EPE), Southwestern Public Service (SPS), and others.

- **Electric Cooperatives:** Tri-State, New Mexico Rural Electric Cooperative Association (RECA), Kit Carson Electric Cooperative, and additional member co-ops.

- **Tribal Governments and Organizations:** Navajo Nation Resource Development Committee and individual Pueblo governments.

- **Labor and Workforce Organizations:** Boilermakers, International Brotherhood of Electrical Workers (IBEW), Building Trades Councils.

- **Research and Technical Institutions:** New Mexico State University (NMSU), University of New Mexico (UNM), New Mexico Tech (NMT), Los Alamos National Laboratory, and Sandia National Laboratories.

This is an exhaustive list, but a comprehensive group of stakeholders that, if organized effectively in phases, and potentially different groups, can establish a strong foundation for DETECT and build momentum across subsequent phases with broader engagement as needed.

1.5 Integrating Outreach, Education, and Storytelling into DETECT

Another critical component of the DETECT platform's long-term success will be its capacity not only to manage and analyze data but also to communicate it effectively. Like the EIA, DETECT should be designed to generate public-facing reports, interactive stories, and visual summaries that transform technical data into accessible narratives. This outreach and storytelling function are essential for demonstrating the platform's value, translating energy and emissions data into fact-based insights that resonate with policymakers, educators, communities, and the general public.

For example, DETECT could publish annual "New Mexico Energy and Emissions Outlook" reports, such as the recent NMEIC New Mexico Energy Landscape report, summarizing key trends in electricity generation, renewable integration, and sectoral greenhouse gas emissions. Interactive dashboards and story maps could make the platform interactive and highlight case study on regional energy developments, such as growth in solar deployment, methane reduction progress, or the potential for geothermal heating in rural areas. Educational modules, fact sheets, infographics and videos could be developed, helping stakeholders but also the general public to understand New Mexico's energy systems through real data.

Education and outreach efforts through DETECT can also be tailored to address regional discrepancies across New Mexico, recognizing that each area faces distinct energy realities and opportunities. Energy production, resource potential, and societal and environmental impacts vary between regions such as the Permian Basin, the Four Corners area, and central New Mexico. By developing region-specific data stories and educational materials, DETECT can connect local challenges and strengths to statewide energy and climate goals, ensuring that all communities see themselves reflected in the project and the energy transition.

By embedding outreach and education into its core design, DETECT would become more than a data platform, it would be a communication tool that strengthens trust in public data and initiatives, supports transparent policy development, and fosters a data-literate culture across New Mexico. Through storytelling grounded in facts, DETECT can make complex energy systems understandable, actionable, and inspiring for all stakeholders.

1.6 Working Group and Legislator Champions

As described in the CETS Data Memo from the CETS phase 1 project, several legislators could serve as champions for supporting the DETECT project throughout its different phases. In the Senate, **Mimi Stewart (District 17)** has been an advocate for the Energy Transition Act and statewide renewable energy reporting frameworks, while **Mark Moores (District 21)** has engaged in clean energy deployment and modernization legislation. **Elizabeth "Liz" Stefanics (District 39)** has been active in environmental oversight and supports transparency in statewide energy reporting. In the House, **Gail Chasey (District 18)** brings experience from Appropriations & Finance, having supported environmental data management legislation, and **Jason Harper (District 6)** has advocated for transparency in emissions and utility oversight. Finally, **Kristina Ortez (District 42)** focuses on just transition and rural economic development, aligning with DETECT's goals of equitable access and transparent reporting. Together, these champions could provide bipartisan support for establishing a comprehensive, state-led data platform.

2 Data Inventory - Energy

Objective: Establish a comprehensive understanding of the current state of data availability, management, and data gaps relevant to New Mexico for effective policy planning, emissions tracking, and monitoring compliance with state climate goals. Data for the DETECT platform must be comprehensive, granular, and regularly updated.

Note to the reader: We have made every effort to identify and compile all relevant data currently available for New Mexico's energy and emissions systems. However, additional datasets surely exist across agencies and stakeholders. As this project advances, we will continue to expand coverage and work toward building a comprehensive "dataset of databases" that collect relevant information for New Mexico.

Below is a review of primary data needs and availability for planning DETECT development. Firstly, we focused on the energy supply, demand and infrastructure. Secondly, we search for GHG data, and finally all other energy technologies, including petroleum sector, geothermal, mineral and mining, and water data.

We began by reviewing energy generation, demand, and infrastructure because these datasets form the backbone of any energy transition strategy. Generation data show where energy is coming from (fossil fuels, renewables, imports), demand data show how it is being used across sectors (residential, commercial, industrial, transportation), and infrastructure data capture the capacity and constraints of the grid, storage, and pipelines.

Together, these elements determine both the baseline of New Mexico's current energy system and the physical and economic limits/opportunities of future change. Without this foundation, it is not possible to reliably measure progress toward climate goals, forecast system needs, or plan for new technologies deployment.

Energy Generation, Demand and Infrastructure

To reliably characterize New Mexico's energy system, we need baseline data on production, use, and infrastructure. Without this minimum information, it is impossible to set an accurate baseline, identify bottlenecks, or plan effectively for the state's energy transition. Therefore, we need at a minimum the following information:

Energy Production & Exchanges:
- Electricity generation by source (coal, natural gas, wind, solar, hydro…).
- Fuel production (oil, gas, coal, biofuels…).
- Renewable energy mix (wind, solar, geothermal, biomass).
- Energy imports/exports across state lines.

Energy Use:
- Final energy demand by sector and sub-sector (residential, commercial, industrial, transportation).
- Energy consumption by fuel type, sector and sub-sector.

- Time-of-use electricity demand data (hourly, seasonal load curves).

Energy Infrastructure:
- Power plant database (capacity, type, age).
- Grid infrastructure (transmission, substations, distribution networks).
- Storage assets (batteries, pumped hydro)
- Fuel pipelines, refineries…

2.1 Data Sources

2.1.1 EIA Data Review

A great portion of the data mentioned above is available on the EIA website and updated regularly (weekly/monthly/annually) based on mandatory reporting under Section 13(b) of the **Federal Energy Administration Act (FEAA)**. Several surveys are sent to energy stakeholders to collect the data under strict guidelines.

This data is reliable, with good granularity, good update frequencies and QA/QC. See **Appendix 1** for detailed information on EIA data coverage, FEAA forms, and also the Data Review Slide Deck accompanying this report. Therefore, EIA data is a good baseline but this is not enough to have the granularity and update frequency we need to build a comprehensive DETECT platform.

A lot of the EIA data collected through the FEAA forms is not easily accessible for New Mexico. Only part of the data can be found directly for the state using the **EIA State Energy Data System (SEDS)** and the **New Mexico State Energy Profile (SEP)** on EIA's website.

- **SEDS** offers a comprehensive, annual historical time series of production, consumption, prices, expenditures, emissions, and infrastructure metrics for every state (SEDS NM). SEDS provides high-quality data, but its annual reporting and Quality Assurance and Quality Control (QA/QC) process limits timeliness. As of August 2025, the most recent release (June 27, 2025) covers only 2023. While more frequent updates are needed, SEDS remains an excellent foundation for historical energy data.

- The **New Mexico SEP** provides current overviews based on FEAA data analysis, including energy supply, demand, production, consumption, trade, and key energy indicators for 2024 (EAI SEP). The SEP data can be more recent and provide a broad general analysis of the state energy system for the previous year. However, it is not possible to navigate to historical data or more recent (monthly) data.

- **EIA API**: It's a big advantage that EIA exposes its data by Application Programming Interface (API), this gives the opportunity to automate a comprehensive, end-to-end pipeline for the New Mexico data platform instead of web data scraping.

The EIA data provides a good foundation for the CETS data platform but additional data sources and collection are needed and discussed in this document.

2.1.2 Fuel Infrastructure

Below is a list of additional data sources to get access to New Mexico fuel infrastructure. However, similarly to electric grid data, this data is very protected for security reasons.

The **Pipeline and Hazardous Materials Safety Administration (PHMSA) National Pipeline Mapping System (NPMS)** contains maps for data review only. A request by a government organization needs to be sent in order to be able to download the data. The data (link) contains:

- Gas transmission pipelines (gas distribution and gathering pipelines are not included);
- Hazardous liquid transmission pipelines;
- Regulated rural hazardous liquid gathering pipelines;
- Liquefied Natural Gas (LNG) plants; and
- Breakout tanks.

The **New Mexico State Land Office (SLO)'s Oil, Gas, and Minerals Online Mapping Service** is a web-based GIS tool that allows users to explore energy and mineral resources on State Trust Lands. It includes interactive layers for unit agreements, wells, land status, leases, and infrastructure: https://mapservice.nmstatelands.org/OilGasMinerals/

The **NMOCD Geospatial Hub** is an interactive mapping platform by New Mexico's Oil Conservation Division that provides access to well locations, oil and gas infrastructure, and regulatory data. Users can explore, filter, and download spatial datasets: https://ocd-hub-nm-emnrd.hub.arcgis.com/

The **NMOCD Oil and Gas Map** is an interactive GIS web application that provides access to key oil and gas data across the state. Users can explore layers such as well locations**,** facilities**,** groundwater wells**,** seismicity zones**,** and leasing information**,** as well as USGS earthquake data. The platform supports API searches and township queries. https://nm-emnrd.maps.arcgis.com/apps/webappviewer/index.html?id=4d017f2306164de29fd2fb9f8f35ca75

The **NMOCD Brine Wells Map** is a web-based GIS tool to display and manage brine well locations across the state. It features color-coded status markers (Active, Plugged, Cancelled, etc.), searchable API numbers, and detailed well information: https://nm-emnrd.maps.arcgis.com/apps/webappviewer/index.html?id=5a610e64e6b1492d9f64449fccb6dcda

OpenEnviroMap, developed by NMED, is a GIS-based platform for visualizing environmental data. This version focuses on Petroleum Storage Tank Facilities and their risk scores, and associated release reports**.** Users can explore underground storage tank sites color-coded by risk (low to high), and view contamination threats. https://gis.web.env.nm.gov/oem/?map=gonm

2.1.3 Grid

To complement EIA-930 (see **Appendix 1** or attached data slide deck), a good source of data for grid information is the Open Access Same-Time Information System (OASIS). OASIS is a federally mandated, public web platform that U.S. transmission providers (like PNM, SPP, EPE etc.) must maintain under **FERC Order 889**. Its main purpose is to make wholesale electric transmission information available transparently and simultaneously to all market participants.

Key Functions of OASIS:

1. Transmission Service Information
    - Available transmission capacity (ATC)
    - Transmission service offerings (firm, non-firm, hourly, daily, monthly)
    - Transmission rates and tariffs
2. System Operations & Status
    - Outage schedules for transmission lines and major equipment
    - System constraints and congestion notices
    - Real-time status updates affecting grid availability
3. Interconnection & Queue Data
    - List of proposed generation projects requesting interconnection to the grid
    - Queue positions, sizes, and expected in-service dates
4. Standardized Data Exchange
    - All OASIS sites use a common set of templates and data fields, making it possible to programmatically query them (though they're not as API-friendly as EIA's Open Data API).

For New Mexico:

- PNM OASIS covers part of NM (NM WECC side) grid footprint. As a requirement of FERC Order 2023, PNM's Heatmap for points of interconnections is now live: Link
- SPP/SPS OASIS covers the eastern NM grid footprint under the Southwest Power Pool. SPS is the primary provider.
- EPE OASIS: Covers Southern NM, EPE also publishes transmission information on its website, with transmission notices, expansion plans and also real time data. The real time data are weekly and daily data from El Paso Electric's Energy Management System (EMS).
- Tri-State Generation & Transmission (serves NM co-ops): Tri-state OATI OASIS.

Some datasets include information on transmission capacity, outage notifications, and interconnection queues relevant to New Mexico's portion of the grid. However these resources are often not user-friendly (pdfs) nor easily accessible, and their formats vary significantly. As a result, the CETS team will need to conduct further analysis to understand their coverage and develop a process for retrieving, processing, and integrating the most relevant data to enhance the grid infrastructure component of the DETECT dataset.

Ultimately, there is a significant lack of publicly available grid data, and greater access to GIS information would be highly valuable for planning and development, but security concerns persists. Some datasets, such as those once available through the U.S. Department of Homeland Security's Homeland Infrastructure Foundation-Level Data (HIFLD), were previously accessible but have been restricted since September 2025, further limiting transparency and the ability to conduct comprehensive energy system analyses.

This aligns with findings from the Interwest Energy Alliance report, "Evaluating Transmission Opportunities in Integrated Resource Plans" (Interwest, 2025), which highlights the need for stronger integration of transmission planning into state energy strategies.

2.1.4 New Mexico Public Regulation Commission

Investor-owned utilities (IOUs) in New Mexico, such as PNM, are required to submit annual reports to the PRC covering system reliability indices, renewable energy compliance (procurement plans, Renewable Energy Certificate acquisitions…), and every 3 years Integrated Resource Plans (IRPs) that provide long-term resource and cost projections (Link). Key reliability metrics such as SAIFI (average outage frequency), SAIDI (average outage duration), CAIDI (average restoration time), and ASAI (percentage of service availability) are essential indicators used to assess and compare the reliability and performance of the electrical grid.

There is no regulatory requirement for daily or hourly generation reporting to the NM PRC. However, utilities routinely track such operational data internally through Supervisory Control and Data Acquisition (SCADA) systems, energy management systems, and smart meters for dispatch, forecasting, and reliability management.

In many regions, similar data is made public through market operators: for example, the California ISO (Link CAISO) and Electric Reliability Council of Texas (ERCOT) dashboards provide hourly generation by resource type in compliance with FERC Order 2000 that created the framework for Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs) and their reporting requirements.

These examples illustrate how utilities already collect and use fine-grained operational data even where state commissions like the PRC only mandate annual, aggregated reporting. IOUs in New Mexico could be required to provide more granular data to the DETECT team. The information could be complemented by CO_2 emission estimates like it is the case for CAISO or DETECT could estimate emissions, like the Electricity Map website.

2.1.5 Federal Energy Regulatory Commission

The Federal Energy Regulatory Commission (FERC) provides several datasets that can support New Mexico's DETECT, particularly for complementing and understanding grid infrastructure, market operations, and transmission planning. FERC's Open Data Plan is part of its commitment to the **Foundations for Evidence-Based Policymaking Act of 2018** (the "Evidence Act"). It outlines how the agency:

- Establishes a data governance framework, creating a structured approach to managing, sharing, and improving public data assets.
- Commits to annually updating this plan.
- Publishes data in machine readable, accessible formats to comply with open data requirements and improve usability across external platforms.

Therefore, this data is easily usable and trustworthy for the DETECT data platform. The datasets available are described in **Appendix 2**.

2.2 Missing or Incomplete Data for DETECT

Our preliminary analysis comparing the data requirements for the CETS (comprehensive, granular, and frequently updated) with the datasets available indicates that a substantial portion of the essential data for energy supply, demand, and infrastructure is already available, especially on EIA website. One of the key challenges, in addition to data quality, lies in identifying the source, extracting the data, and organizing the information relevant to New Mexico in a more accessible and usable format.

Now, despite the notable advantages and comprehensiveness of the EIA datasets, we identified several gaps.

Timeliness and Granularity of State-Level Energy Data: A major limitation of EIA datasets lies in their insufficient temporal resolution, granularity and reporting lags, particularly for state-level energy data. The SEDS is one of the most comprehensive sources for tracking production, consumption, and emissions at the state level. However, it is updated only annually and with a delay of 12 to 18 months.

In addition, while the **EIA-930** dataset provides near real-time electricity data, it is limited to balancing authority-level (BA) granularity, offering no insight into distribution-level loads, local generation variability, or community-scale trends within New Mexico. Sub-counties level data will be important for the DETECT effort.

While we identified some additional datasets available for visualization on the IOUs' websites, these resources remain limited and insufficient for comprehensive analysis. To support meaningful planning, modeling, and decision-making, the state should partner with IOUs to collect or have them provide more detailed, consistent, and accessible information beyond what is currently shared.

Such a partnership would not be a one-way effort, IOUs would also benefit from improved data coordination through better access to statewide analytics, enhanced planning capabilities, reduced reporting redundancies, and better understanding of opportunities for grid modernization and clean energy development.

Energy Infrastructure and System Operations Data: Available datasets provide only limited insight into physical and operational aspects of New Mexico's energy infrastructure, particularly in relation to grid functionality, system constraints, and infrastructure readiness for clean energy integration. Key missing or incomplete components include: transmission and distribution infrastructure data, energy storage data and interconnection data.

Transmission and distribution infrastructure data: We did not find datasets on electric lines capacities, congestion zones, transformer loading, or substation locations, critical elements for planning new power integration, connection planning, siting batteries, and enhancing grid resilience. The only available resources are maps that are online explorable, data cannot be extracted.

A good example of a comprehensive grid data platform is the RTE-France interactive map (Link). France Electric Grid network (RTE) has just released a new interactive map showing real-time energy flows across the French HTB2 grid. You can explore live electricity flows on the 400kV national grid and zoom in and see 225kV lines regionally, plus additional data on connections, evolution of the grid and commissionings. The advantage that France has is that the grid is nationally owned.

By comparison, states like California, with support from the California Public Utilities Commission (CPUC), have asked utilities to develop Distributed Resource Planning Portals (DRPP) and other interconnection resources that provide detailed information and data to support grid planning and integration of distributed resources. California's major IOUs (PG&E, SCE, SDG&E) each host Integration Capacity Analysis (ICA): CPUC, PGE, DRPP, SDGE.

At the federal level, FERC Form 715 contains transmission planning and evaluation data, but accessibility and usability remain limited. More data were available publicly at the Homeland Infrastructure Foundation-Level Data (HIFLD), but they have been removed from public access in September 2025.

Ideally, New Mexico could require utilities to file standardized, open datasets with the PRC or EMNRD as part of Integrated Resource Plans (IRPs) or grid modernization efforts, building on these established models, to enhance planning for new generation.

Energy storage data: While utility-scale battery and pumped storage facilities are reported in EIA-860 and EIA-923, the data is often not available or not disaggregated at the state level or by operational characteristics (duration, capacity factor, etc…). Therefore making it hard for project developer to consider New Mexico for storage projects.

On the contrary, good state storage data sharing can be found for California at the California Energy Storage System Survey (published via the California Energy Commission), which presents data on residential, commercial, and utility-scale storage installations (capacity, counts, trends), and allows data download (California Energy Commission). Additionally, California ISO publishes daily storage performance reports including metrics like state of charge, bid-in capacity, awards, and ancillary services usage (CAISO Energy Storage)

In New-York, NYSERDA maintains a data system that includes detailed data on distributed energy resources (including storage), with location, performance characteristics, and filtering by technology (NYSERDA DER). NYSERDA also publishes storage data showing project status and other storage metrics. NYSERDA

For New Mexico, an ideal framework would require utilities and storage developers to report standardized operational data, including information about location, duration and performance. This could fall under the PRC oversight.

Interconnection queue status: There is no centralized reporting in any databases on queued energy or storage projects awaiting grid access at the state level. These data are essential for modeling grid development, future generation integration and understanding bottlenecks.

In many organized wholesale markets, ISOs/RTOs maintain transparent interconnection queues that detail the status of generation and storage projects seeking grid access. For example, CAISO publishes a downloadable Public Queue Report, listing projects by point of interconnection, status, receive date, and study phase (CAISO Public Queue Report).

PJM (RTO) has also developed a Queue Scope tool within its Planning Center, which allows users to explore interconnection projects and assess potential grid impacts.

For New Mexico, adopting a similar approach would mean ensuring that IOUs and co-ops publish queue data in a consistent, machine-readable format (e.g. by interconnection point, MW size, status, study phase). In the absence of an ISO/RTO, it may be necessary for the state (via EMNRD or the regulatory commission) to require that all interconnection applicants report standardized queue metadata publicly. This transparency can help stakeholders model future generation build-outs, identify infrastructure bottlenecks, and prioritize grid investments and long-term planning.

Generally speaking, other incomplete data are due to the lack of data granularity.

Sectoral Disaggregation: EIA's datasets offer only broad sector-level energy use figures. It has minimal information on industrial, commercial and residential sub-sector data. Same goes for heat versus electric energy use, or the share of clean energy technology uptake within specific sectors. These are important information for the CETS, for example to be able to promote geothermal direct use or electrify low-heat industries and heating and cooling.

Some of the other specific deficiencies identified include:

- **Transportation sector detail:** EIA lacks data on electric vehicle (EV) adoption rates, charging infrastructure deployment, and transit trends.
- **Industrial subsector differentiation**: Like mentioned above, industrial energy consumption is aggregated, hiding distinctions between mining, manufacturing, refining, and oil and gas operations. Given the energy-intensive nature of these activities in New Mexico (rapidly rising source of CO_2 emission), disaggregation is critical for a state tailored decarbonization strategy.

California shows that it is possible to require more granular reporting than EIA provides, but it is only sub-divided by counties and still very limited. Examples: CEC Energy Almanac; CEC Electricity Consumption Dashboard.

A comprehensive disaggregated reporting should combine: (1) mandatory reporting by utilities and regulated industries; (2) surveys designed to capture subsector-specific energy patterns; (3) public dashboards to make the data easily usable for planners, researchers, and industry.

New Mexico would be a pioneer in the U.S. if sub-sectoral reporting is implemented, as most states only publish broad sector-level energy data, which hides important differences across industries. To support tailored decarbonization strategies, CETS could recommend that the state adopt new reporting requirements for large industrial operators (such as mines, refineries, and manufacturers) to disclose energy use by fuel type and process. This could be complemented with partnerships with industry associations (NMOGA) or IOUs to share anonymized data to model subsector splits. Doing so would place New Mexico at the forefront of state-level data transparency, filling a gap that even leading states like California and New York have yet to address.

Energy Heat vs Electricity: Energy is not only electricity but also heat, and we use a lot of heat in our industrial processes or we lose heat when there is a combustion to produce electricity. But EIA does not consistently or explicitly differentiate heat energy use or loss as a distinct end-use category, especially for non-electric thermal energy.

EIA datasets report final energy consumption by sector and fuel type, but they do not isolate end-use functions like space heating, water heating, or industrial process heat in a consistent, energy-agnostic way. This means there is no direct accounting of how much total energy is used specifically for heat or is lost in the processes, regardless of whether the source is electricity, natural gas, oil, biomass, or geothermal.

Even if EIA does track electricity used for heating in residential and commercial buildings through surveys (Residential Energy Consumption Survey and Commercial Building Energy Consumption Survey), these are intermittently released, and national or regional, not state-level, even less county level. In addition, the data lags by several years.

For the industrial sector, EIA's Manufacturing Energy Consumption Survey (MECS) provides data on process heating, but it is national only, not state-specific, and it is conducted only every 4 years and often has a publication delay.

Finally, there is no comprehensive tracking of heat pump deployment, geothermal direct use, or district heating systems at the state level in EIA's main datasets. So key heating transition pathways such as fuel switching, electrification of space/water heating, or geothermal heating systems are poorly captured in New Mexico.

Interviews with state energy stakeholders suggest that additional datasets exist within state-level agencies but much of this information remains underutilized. The lack of mechanisms for cross-agency sharing, and the even more limited accessibility for public entities, creates barriers to fully analyzing existing data sources and leveraging these resources for planning and decision-making.

2.3 Actionable Recommendations

2.3.1 Phased Approach to Complex Data Strategy Needs

We recommend establishing a cross-agency working group with representatives from state agencies, universities, and other stakeholders to guide the development of the DETECT data management strategy and platform. Each agency should designate a Point of Contact (POC) to act as a liaison, ensuring streamlined communication and integration. Addressing existing data gaps, particularly in energy supply, demand, and infrastructure but also GHG emissions (Chapter 3), will require strong collaboration and a phased, sector-by-sector approach, supported by ongoing coordination and trust building. Finally, discussions on data governance, including access, sharing, and standardization protocols, are critical to building a robust and sustainable data ecosystem.

Priority elements include for the energy supply, demand and infrastructure sectors:

- Transmission and distribution system assets (lines, substations, interconnections).
- Fuel production and supply chain data.
- Final demand by sub-sector.
- End-use data, including hourly and seasonal load profiles.
- Fuel consumption trends by type and sub-sector.

Proposed Data Sources:

- Federal: EIA, FERC, EPA.
- State: EMNRD, NMED, PRC, OCD, BLM (regional datasets), NMBGMR.
- Utilities: Annual reports, filings to PRC + potentially additional agreements.
- Academic/Research: NMT, UNM, NMSU, NTU, national lab collaborations.

Timelines:

Timelines are outlined in more detail in Chapter 5 and 6, but depend heavily on (1) the volume of data to be collected, (2) the tools and platforms to be developed, and (3) the resources and capacities available to DETECT and state agencies. Similar to Chapter 1, recommend a phased approach:

- **Phase 1:** Focus on high-priority sectors and publicly available data for electricity generation, transmission, fuel production, GHG. This phase should establish standardized formats and secure data-sharing agreements for next phase.
- **Phase 2:** Expand to sub-sectors, demand-side data, emissions inventories, and compliance monitoring datasets.
- **Phase 3:** Integrate cross-sectoral datasets (water, land, socio-economic indicators) and advanced tools for monitoring, modeling, forecasting.

Finally, we recommend that the future DETECT digital platform for CETS be designed to:

- Systematically retrieve data via the EIA API and supplemental data scraping from the SEDS and state, and consolidate it into a New Mexico-specific database with robust search and query capabilities. We can ensure up-to-date access across electricity, petroleum, and gas data on the New Mexico platform, with hourly operations data for power sector (by jurisdiction), weekly/monthly for petroleum and gas, and annual for historical context, all from one interoperable interface.
- Generate dynamic visualizations on dashboards, including interactive graphs, to support near-real-time (hours to week) analysis and trend analysis. For example, we've worked with the New Mexico Energy Initiatives Consortium (NMEIC) students to make the SEDS data for New Mexico more accessible (NMEIC SEDS), notably with interactive graphs.

- Produce geospatial outputs, such as interactive maps, using the underlying data to visualize energy infrastructure, production, emissions, and other key indicators across the state, in coordination with ongoing statewide GIS efforts such as those outlined in the New Mexico Geographic Information Office (GIO) Report

Important note for platform developers: EIA surveys and reporting forms are occasionally updated, renumbered, or retired. Before integrating or automating data pipelines, always verify the current status of relevant surveys and forms through the EIA Survey Portal.

2.3.2 Leveraging State Agency Data

Insights gained through state employee engagement during Phase 1 of the CETS revealed that valuable data exists within New Mexico state agencies**,** including, to only cite some of the most important for this first phase effort: ECAM, NMED, and the OCD. However, this data is often siloed, stored in disparate formats, or not accessible. Moreover, most of the agencies do not have sufficient staffing dedicated to data management to undertake a comprehensive effort of data consolidation and management. Due to the scale of the efforts required to identify, retrieve, and integrate all state agencies information, a dedicated, agency-by-agency data assessment and recovery process is essential.

A strong model for this type of effort could be Project Velocity, currently underway at NMED. This project aims to consolidate decades of legacy environmental data into a centralized, digitized, and easily searchable database. It employs modern data management practices, including AI-enabled search functionality, to transform static records into an actionable information system.

Project Velocity could serve as a proof of concept with a proven framework for what is possible when data modernization and digital transformation are prioritized at the agency level. More analysis of the project and collaboration with the Velocity team would give us an idea of the budget and timeline required to undertake such an endeavor for each agency. We suggest leveraging synergies between the CETS and the Velocity projects.

To build on this example, we recommend that similar initiatives be launched within each state agency involved in energy, environmental management, permitting, and infrastructure planning. These efforts should follow a shared data management framework, informed by lessons learned, governance model, and technical specifications so they can be easily integrated to the future DETECT platform. Applying a unified structure across agencies will ensure interoperability and data consistency. This inter-agency effort would yield several benefits:

- It would surface hidden or underutilized datasets relevant to energy production, emissions tracking, and infrastructure planning.
- It would improve institutional databases and transparency, making it easier for agencies to collaborate and respond to public inquiries or federal reporting requirements.
- Most importantly, it would allow the state to have comprehensive, granular databases that would feed the DETECT platform.

These efforts should be considered in parallel to the DETECT data platform and would represent only the foundation for establishing strong, reliable, and comprehensive data repositories. These efforts at each agency would create data sources for the DETECT platform. A data platform is only as powerful as the quality of the data it holds.

Chapters 5 and 6 describe how the DETECT data platform could be structured to integrate diverse sources of information into a unified, interoperable architecture. The sources of data could range from federal agencies (EIA, FERC, EPA) to state departments (which should maintain and contribute their own databases), as well as utilities (through partnerships or formal agreements) and other partners.

The CETS proposed platform would function as a "**dataset of databases**", consolidating structured and unstructured data into a centralized system capable of supporting regular analysis, reporting, and visualization. For this reason, we have recommended the development of a robust and adaptable tool, one that is capable of evolving with future data sources, data governance needs, policy objectives, and technological advances.

3 Data inventory - Greenhouse Gases

During Phase 1 of the CETS, we started engaging with NMED, EMRND and the Eastern Research Group (ERG) staff to learn more about past GHG emissions inventory (GHGEI) work at the state level.

3.1 Data Source

The primary data sources used for New Mexico GHGEIs are:

1. U.S. EPA State GHGEIs**:** State GHG Emissions and Removals
2. U.S. EPA State Inventory Tool (SIT): SIT Tool
3. EPA Greenhouse Gas Reporting Program (GHGRP)**:** GHGRP Data Sets
4. U.S. Energy Information Administration (EIA): State Energy Data System (SEDS)
5. New Mexico Oil and Gas Activity Data from NMOCD
6. New Mexico Oil and Gas GHGEI for 2020 (ERG)
7. Population and Employment Projections**:** U.S. Census Bureau and University of New Mexico Bureau of Business and Economic Research (BBER).
8. Waste Management Sector Data: Based on state-level solid waste reports and EPA factors.
9. Agriculture Data: Derived from USDA Census of Agriculture and National Agricultural Statistics Service (NASS) reports.
10. Land Use, Land Use Change, and Forestry (LULUCF)**:** Derived from EPA State Inventory Tool and other forestry datasets.

The inventories incorporate default emission factors published by:

- U.S. EPA (e.g. SIT modules, EPA Air Emissions Inventories, Modeling, Monitoring and Emissions Factors)
- IPCC (Intergovernmental Panel on Climate Change GHGEI guidelines)
- EIA (for fuel-specific emissions)

A review of the principal available GHG data sources for New Mexico is presented in **Table 1**.

Table 1: Greenhouse Emissions Primary Data Sources

Source Name	Description	Granularity	Regulation/ Authority	Public Access	Update Frequency
NMED AQB Emissions Inventory	State-inventory GHG, part of criteria pollutants	Facility + County	NM AQB regulates for compliance	Partial	Variable
NMED Climate Change Bureau	State-reporting	Sector wide inv and forecast		Yes	Annual (Third party)
EPA GHGRP	Facility-level GHG emissions by sector	Facility	Clean Air Act (CAA) §114	Yes	Annual
EPA FLIGHT	Public interface for GHGRP	Facility	Same as GHGRP	Yes	Annual
EPA Facility Registry System (FRS)	Cross-references facility data	Facility	Cross-agency	Yes	Variable
EIA GHG CO_2 by State & Sector	CO_2 emissions from fossil fuels	State + Sector	N/A (EIA estimate)	Yes	Annual
SEDS (State Energy Data System)	Estimated state-level emissions and energy consumption	State	None (EIA-modeled)	Yes	Annual
NM OCD	Flare volumes & production data	Well, basin	NM state rules	Partial (public search)	Monthly
BLM Methane Waste Rule Reporting	Methane venting/flaring on federal and tribal lands	Well pad	BLM Rule 3179	No (aggregated data only)	Annual

The EPA GHGRP is one of the most accurate and reliable process in place for GHG data collection. Each subpart of the GHGRP (Appendix 4) covers different topics and has a MRV in place. The only limitation is the update frequency.

3.1.1 State-Level GHG Emissions Inventory

NMED, with support from EMNRD and third parties, published state-level GHG inventories that track emissions across all sectors of the economy. Here is the list of published GHGEI from the most recent to the older:

1. 2021 GHGEI + 2005 baseline + projections through 2050, released in December 2024 (E3).
2. 2005 NM Oil and Gas GHGEI, released in December 2024 (ERG).
3. 2020 NM Oil and Gas GHGEI, published in August 2022 (ERD).

4. A New Mexico GHGEI for 2018, and 2005 baseline in October 2020 (E3).

They are developed with support from independent consultants such as Energy + Environmental Economics (E3) or Eastern Research Group (ERG).

3.1.2 Oil and Gas Specific Emissions Data

1.1.1.1 CO_2 from Oil and Gas

Several federal and state datasets provide information on GHG emissions specific for oil and gas, with emphasis on methane. These data sources complement production and injection data by linking volumes to actual emissions, and by supporting comparisons across facilities, basins, and states.

The EPA GHGRP - Subpart W and EPA FLIGHT Facility-level are reporting CO_2, CH_4, and N_2O from petroleum and natural gas systems, including production, processing, gathering, boosting, transmission, storage, and injection.

➢ **Pros:** Comprehensive, standardized annual reporting; facility-level resolution; directly tied to regulatory compliance.

➢ **Cons:** Reporting thresholds exclude smaller facilities; data typically lags by one year; methane estimates rely partly on emission factors rather than direct measurement.

1.1.1.2 Methane Emissions

Methane emissions data in New Mexico is primarily available through the OCD, which collects extensive compliance reports from oil and gas operators based on the Methane Waste Rule. The most significant dataset is the **C1.15b Monthly Venting and Flaring Report**, which requires each active well to report on 13 methane-related categories. When methane releases exceed 500 mcf, operators must also report an incident to OCD within 24 hours, although enforcement capacity remains limited.

While OCD verifies some submissions, the data is industry-reported. There is a lack of staff for enforcement based on this data. The public can access this information through OCD's **Geospatial Hub**, or via the **Natural Gas Waste Dashboard**, which provides near real-time updates as operators submit data. However, because each operator maintains their own proprietary system, fully integrating or linking these datasets remains challenging, and a live data API is currently unrealistic.

Key access points:

- OCD Geospatial Hub
- OCD Natural Gas Waste Dashboard
- Operator-specific reports can be found under the "Natural Gas Waste Reports" section in the data tools.

3.1.3 Tools for GHG Emissions Tracking

- EPA Methane Emissions Data Explorer **(MEDE)** Interactive tool compiling methane estimates by basin, sector, and source. Includes Permian and San Juan basins.

➤ **Pros:** Basin-level methane intensity metrics; allows comparison across producing regions.

➢ **Cons:** Based on modeled data; updates less frequent; not directly tied to individual wells or facilities.

- NMOCD Statistical Summaries Includes venting, flaring, spills, and incident reports, largely derived from C-115 filings.

➤ **Pros:** State-reported activity data; directly links to NM production records; updated monthly to annually.

➢ **Cons:** Quality depends on operator self-reporting; no independent verification; limited facility-level emissions detail.

- EIA Flaring and Venting Data State-level venting and flaring volumes reported through EIA Form 914.

➤ **Pros:** Provides federal consistency across states; useful for comparing NM to other producing states.

➢ **Cons:** Less granular than NMOCD data; potential inconsistencies with state records; limited facility detail.

- VIDRIS Nightfire Satellite observations of gas flaring, identifying flare locations and intensity globally.

➤ **Pros:** Independent, observation-based dataset; provides check on reported flare volumes; global coverage.

➢ **Cons:** Cannot measure non-flaring emissions; nighttime detection only; requires post-processing to integrate with production data.

- IEA Methane Tracker: It provides global estimates of methane emissions from the energy sector, including oil, gas, and coal. It also highlights abatement opportunities and policy progress by country.
- Climate TRACE is a coalition that uses satellites, remote sensing, and AI to provide near real-time, independent estimates of greenhouse gas emissions at the facility, sector, and country level.
- TCCON Network (Total Carbon Column Observing Network): The **TCCON** is a global network of ground-based spectrometers that measure atmospheric concentrations of CO_2, CH_4, and other gases with high precision. It can provide validation for satellite observations.
- MethaneSAT: Developed by the Environmental Defense Fund (EDF), is a dedicated satellite mission to detect, quantify, and attribute methane emissions worldwide with high spatial resolution.

- Other tools include platforms like GHGSat, Carbon Mapper, EPA GHG map, EPA which provide targeted monitoring.

3.2 Missing or Incomplete Data for DETECT

One of the issues identified in the GHGEI is that the 2021state-wide GHGEI incorporates updated practices and data sources, so it is not directly comparable to earlier versions such as the 2018 inventory. It includes revised 2005 baseline values and long-range emissions projections through 2050 that supersede previous estimates.

These updates illustrate progress in GHG accounting practices but also reveal the absence of a unified and agreed upon framework. To ensure that GHGEIs are consistent, reliable, and comparable over time, there is a critical need to develop and implement a standardized accounting program for the state. The methodologies and QA/QC procedures used for the 2021 GHGEI are detailed in the New Mexico Climate Action Plan: PCAP Appendix K.1 – Inventory and Quantification Methodology. Even though these documents provide some guidelines on how the GHGEIs were produced, there is a lack of transparency and clarity to enable an easy replicability.

3.2.1 Analysis of Legal Authority for GHG Monitoring

Existing legal authority for GHG monitoring, reporting and verification (MRV) are the Air Quality Bureau (AQB) at NMED and OCD at EMNRD.

1. Air Quality Control Act & NMED Authority**:** The Air Quality Control Act empowers the Environmental Improvement Board (EIB) and the AQB to require reporting of GHG emissions. Specific regulations include **20.2.87 NMAC** (GHG Emissions Reporting) and **20.2.300 NMAC** (Reporting of GHG Emissions). These establish mandatory annual reporting for certain large facilities (e.g., power plants ≥25 MW, petroleum refineries, cement plants, and other covered sectors) and align with EPA's 40 CFR Part 98 reporting framework (EPA GHGRP).
2. Methane Waste Rule: OCD regulations of methane leakage.

The Priority Climate Action Plan (PCAP, 2024) confirms that NMED, through its Climate Change Bureau, and EMNRD jointly lead statewide GHG inventory development. The plan also highlights gaps in authority and coordination across agencies, particularly in methane response and cross-agency data integration. AQB data focus on environmental violation and OCD data from Methane Waster Rule focus on avoiding waster of natural gas. It would be interesting for the DETECT platform to consolidate the data from each agency.

Currently, organizations submitting emissions inventories (mainly for air quality control) to NMED are also required to report their GHG emissions. For oil and gas operators, only the total GHG emissions across all facilities within New Mexico need to be reported, but sometimes they even include NM and Texas for the Permian basin.

For data collection, all GHG data must be submitted through the Air Emissions Inventory Reporting (AEIR) system, NMED's online reporting platform. AEIR contains a dedicated GHG reporting branch. The following three greenhouse gases are specifically tracked:

- Carbon Dioxide (CO_2)
- Methane (CH_4)
- Nitrous Oxide (N_2O)

Current Limitations:

- The AQB has statutory authority under the Air Quality Control Act, but its regulations primarily cover large stationary sources and limit to yearly data acquisition. Many smaller sources, indirect emissions, and some sectoral data fall outside direct reporting mandates.
- Verification requirements exist under **20.2.301 NMAC**, but depends on third-party and AQB capacity, which is limited.
- OCD (oil and gas), EMNRD (energy/mining), and NMED (air, water, waste) each maintain separate datasets. These are not fully integrated, which makes building a comprehensive digital platform challenging.

The PCAP (Section 7, "Review of Authority") explicitly notes the need to clarify and potentially expand agency authority to collect and integrate emissions data. To be able to have a comprehensive digital platform, we need (1) improved and uniform GHG data accounting and MRV systems that integrate data across NMED, EMNRD, and OCD; (2) legislative or regulatory action to expand mandatory reporting and incorporate modern technologies and methodologies; and (3) automation and interoperability, moving from compliance-driven reporting to real-time or near-real-time data flows.

To summarize, state agencies and primarily NMED under the Air Quality Control Act and Clean Future Act, have some authority to require GHG reporting from certain sectors, but this framework is not yet comprehensive enough for a full coverage of all GHG emissions, which would contain all the information from **Table 2**. Current rules focus on large stationary sources. The need for exploring these authority gaps and integration pathways is already recognized in the PCAP and should be considered also for the CETS.

Table 2: Theoretical Key Data Categories for GHG Reporting

Category	**Subcategories**
Emissions Data	CO_2, CH_4, NOx by source (stationary, mobile, fugitive)
Scope 1, 2, 3 Emissions	Direct, indirect, and value chain emissions
Facility-level Data	Power plants, oil & gas sites, industrial facilities
Sectoral Breakdown	Power, buildings, transportation, industry, agriculture
Temporal Resolution	Annual (most common), limited monthly/hourly data
Spatial Resolution	State, county, facility-specific, basin-specific
MRV Methodologies	Quantification methods, QA/QC, uncertainty analysis

3.2.2 GHG Data Quality and Availability

Table 3 summarizes the current data quality for GHG emissions data. Most of the data lack either granularity, regular update with good frequency, MRVs or QA/QC.

Table 3: Current Data Quality and Availability Assessment for GHG Emissions in New Mexico

Criterion	Status	Source	Comments
Facility-Level CO_2 Emissions	⚠ Available but	NMED, GHGRP	Long lead time
Facility-Level CH_4 Emissions	⚠ Partial	NM OCD	Partially reported by oil & gas, lack MRV and QA/QC
Scope 2 Emissions	✕ Not standardized		No tracking
Scope 3 Emissions	✕ Not available		
Sectoral CO_2 Emissions	⚠ Available but	EIA state/sector data	Mainly power sector and some other sectors, but uses estimates and no information on sub-sectors.
Emissions by Fuel Type	⚠ Available but	EIA & GHGRP	Bad lead time, and need sub-sectors
Monthly/Hourly Resolution	⚠ Limited		Mostly annual; some real-time electric emissions (PNM, EPE, SPS)
Geospatial Mapping	☑ Good		Many maps, need integration
MRV Methodologies	⚠ Limited		EPA guidance, but limited deployment and NM-specific rules
Consistency Across Datasets	⚠ Needs harmonization		Different scopes, assumptions, formats
Reporting	⚠ Limited and need harmonization	NMED, EMNRD, & EPA	AEIR AQB, NEI, NMOCD or GHGRP

3.3 Actionable Recommendations

Based on information collected through state agencies stakeholder engagement, there is a critical need to implement automation and advanced data analytics to improve GHGEI, GHG emissions oversight and regulatory effectiveness, especially for OCD methane data collection. Current manual processes, such as individual data reviews and field inspections, are not scalable given the volume and complexity of the data being reported. Agencies are faced with the challenge of having "too much data to use," highlighting the importance of investing in automated anomaly detection, intelligent dashboards, and targeted compliance triggers.

The DETECT platform should prioritize full utilization, integration and automatization of currently available datasets. Once foundational analytics and quality assurance processes are established, the system can evolve to incorporate more features, for example near-live data streams.

To summarize and suggest action items, key limitations preventing a reliable and comprehensive GHG data platform include:

Methane Emissions Reporting: Reliance on operator self-reporting, limited direct measurement, and no reconciliation between OCD and NMED data nor satellite data.

- Recommended action: Establish a standardized dMRV (digital MRV) framework for methane emissions. First the DETECT platform could ingest and analyze existing methane emission data automatically. But in a second time, we could require new operators to supplement self-reported emissions with automation and/or digital sensors implementation that can automatically report flaring/venting. Then we can reconcile these against independent satellite and aerial datasets.
- Impact: Improves accuracy of data, captures super-emitters, and aligns with best practices for transparent emissions reporting.

Digitalization and Automation: Many filings still rely on flat files, PDFs, or manual data entry.

- Recommended action: Promote digitalization, modernize data submissions by requiring machine-readable formats with standardized templates (CSV, JSON, API-based reporting) for AEIR, C-115, and other filings at state agencies. Create a unified data management process to ensure consistency.
- Impact: Reduces processing time, reduce errors, and enables automated integration into dashboards and DETECT digital platform.

Cross-Agency Integration: GHG datasets remain siloed between NMED (criteria pollutants, GHG), EMNRD/OCD (venting/flaring), PRC (utility filings), and federal systems (EIA, EPA).

- Recommended action: Create a Data Working Group (NMED, EMNRD/OCD, PRC, DOT, SLO…) to harmonize agencies data and create identifiers across agencies (statewide ID for facilities or other data to facilitate management).
- Impact: Breaks down silos, enables consistent analysis, and improves traceability of data and emissions across sectors.

Granularity and Timeliness: GHG datasets are often aggregated or have long delays, limiting their utility for relevant analysis.

- Recommended action: Build partnerships with utilities, trade associations and other stakeholders to collect granular near-real-time data on GHG emissions and trends for important sectors and sub-sectors.
- Impact: Narrows the current 12 to 18 months lag in EIA and EPA GHG datasets, enabling more granular and frequent analysis, support GHGEI.

Scope 1, 2 and 3 Emissions: Lack of mandatory reporting frameworks leaves Scope 1 sub-sectors, consumption-based emissions and value-chain emissions unaccounted for.

- Recommended action: Develop a mandatory reporting framework for GHG reporting that is consistent over time. If updates are made, make sure that GHGEI are comparable over time. Begin to full proof concepts with utilities, state agencies, and major industrial operators.
- Impact: Expands the emissions footprint beyond direct sources, enabling more accurate statewide accounting and compliance with statewide reduction targets.

4 Data Inventory - Energy Technologies

It is important to consider data sources for specific energy technologies because each technology, such as oil and gas, geothermal, CCS, renewables or mining, relies on distinct datasets (e.g., subsurface geology, water use, grid interconnections) that are often collected by different agencies or institutions. Without integrating these sector-specific data into the CETS platform, critical information remains siloed, limiting the state's ability to evaluate technology potential, ensure regulatory compliance, and support informed policy and investment decisions.

4.1 Petroleum Sector Data

New Mexico is very reliant on oil and gas activities. Therefore, comprehensive data on oil, gas, produced water, and CO_2 production and injection are essential, like highlighted by the state GHGEI that has a specific oil and gas GHGEI. Good oil and gas data can also support life-cycle analyses and help to forecast the scale and duration of future production. Information on CO_2 volumes and disposition, whether used in enhanced oil recovery (EOR) or injected for long-term storage, along with GHG flaring, venting, and facility-specific emissions, provides critical inputs for emissions modeling and carbon management. In addition, oil and gas revenues are currently paramount for the state and that type of data will be important for DETECT as well.

Together, these datasets form the foundation for roadmaps to near-zero emissions, like E3 report, and can guide infrastructure planning, regulatory strategies, and the evaluation of technologies such as CO_2 storage, methane capture, and produced water management.

4.1.1 Oil and Gas

NMOCD data can be accessed in three principal ways:

1. NMOCD Permitting is a web-based application for searching on wells, facilities, hearings, spills and much more

➢ Pros: Flexible, user-friendly searches; results are up-to-date and link to additional files. Results include a large variety of data including production and injection volumes, well information, history, and much more.
➢ Cons: Not all data are downloadable; bulk exports limited to basic well info (CSV). Expanded data such as production volumes cannot be downloaded in bulk.

2. NMOCD Map Hub Interactive GIS map + some data downloads

➢ Pros: Well data download provides well locations (lat/long), spatial context, and hyperlinks to additional well information.
➢ Cons: No direct download of production or injection data; links connect only to individual wells.

3. NMOCD FTP public-facing ftp server that provides access to bulk well and volume data

➢ Pros: Contains full production and injection volumes, well history, and completion data; the official "data of record" for the state; updated daily.

➢ Cons: Data are provided as large XML tables that must be structured into a relational database to take full advantage of the data. The entire data set may exceed 15 GB in size. Documentation is limited, and business rules are not provided, so the learning curve is steep.

Additional OCD resources include:
OCD Online Imaging A substantial archive of digitized paper files: well files, well logs, administrative and environmental orders, hearing orders, and case files.

NMOCD Statistics. Statistical summary files of production, injection, well activity, and other information (drawn from C-115 forms and other reports on flaring, venting, spills, and incidents). These are updated on variable schedules, monthly, annually, or as needed.

GO-TECH Petroleum data web access maintained by the Petroleum Recovery Research Center at NMT.

➢ Pros: Provides a less complex interface than OCD; allows single-well downloads and smaller annual or custom datasets (county or basin scale). Preserves some historical information (e.g., original well names, operators, drilling dates, or depths) that is missing or degraded in OCD records.

➢ Cons: Cannot download complete datasets at once; updated monthly; error-checking less rigorous than OCD data; overall less complete. Hardware/software upgrades can lag due to resource constraints.

FracFocus Registry National registry for hydraulic fracturing disclosures, required by New Mexico law.

➢ Pros: national database of chemicals used in hydraulic fracturing operations; data downloadable as an SQL database or per well.

➢ Cons: large database not directly tied to emissions, it may be of value for estimating emissions related to well completion and for tracking water use.

New Mexico State Land Office (SLO): NM SLO GIS Data

➢ Pros: Provides shapefiles of trust lands, lease boundaries, and rights-of-way, Includes an interactive oil and gas map with links back to ONGARD.

➢ Cons: Not specifically tied to emissions but may be of value in mapping

New Mexico Bureau of Geology & Mineral Resources (NMBGMR): NMBGMR Subsurface Library

➢ Pros: Archives thousands of geophysical logs, drillers' logs, cores, cuttings, sample descriptions, and related publications. Data from these records are essential for evaluating geologic context, reservoirs, and potential CO_2 storage

➢ Cons: Much of the subsurface data is not digitized due to complexity.

We have also collected significant amount of information on tax revenues from oil and gas that we will incorporate in future phases.

4.1.2 Produced Water

Produced water data are critical for life-cycle emissions analysis, regulatory compliance, and long-term water sustainability planning. Volumes, chemistry, disposition, and reuse information help connect production data to management pathways. Important produced water data needs include:

- Volumes produced and injected
- Water quality information and related geologic context
- Sources and disposition of water used for hydraulic fracturing
- Means of water transport and handling

Produced water affects emissions reduction in two main ways:

1. **Production burden:** It is an energy-intensive byproduct of hydrocarbon extraction. In the Permian Basin, water-to-oil ratios of 4:1 to 8:1 (or more) make handling a major operational cost and emissions source.
2. **Treatment and reuse:** Cost-effective treatment for industrial or agricultural applications would reduce freshwater demand while offsetting some costs and emissions of water management. But nowadays, disposal through Salt Water Disposal wells is cheaper.

Fully assessing these opportunities requires integrating detailed volume and geologic data with water quality information, a key gap that currently limits large-scale reuse planning.

Data Sources:

New Mexico already has a solid foundation of produced water data through several existing sources. The NMOCD C-115 provides official well-level records of production, injection, and disposal volumes, while GO-TECH offers a simplified interface for single-well or county-level queries. Chemistry data are available through the USGS Produced Waters Database and the New Mexico Produced Water Database, and visualization tools such as WaterStar provide map-based access to select water quality and volume information. Additional context is supplied by GWPC reports, EPA's UIC well inventories, FracFocus disclosures, and academic or DOE-funded pilot studies. Together, these sources capture volumes, chemistry, spatial patterns, and regulatory framing.

Despite these resources, important gaps remain. Water chemistry data are sparse, dated, and uneven across basins, with no routine or standardized sampling. There is no systematic tracking of treatment or reuse pathways beyond disposal in saltwater disposal wells (SWD), and economic data on costs of handling, treatment, and transport are limited to case studies. Integration across datasets is weak, with little capacity to join production, chemistry, and completion records at the well level. Finally, reporting is often monthly or annual, leaving no real-time or near-real-time visibility for planners or regulators.

Other states provide useful models for improvement. Texas requires detailed reporting of produced water injection and is advancing efforts to track reuse. Colorado maintains transparent portals linking produced water data to water management initiatives, and Wyoming and North Dakota integrate produced water reporting with beneficial use and spill tracking. In contrast, New Mexico does not yet collect or share reuse volumes, treatment methods, or ongoing water quality monitoring from areas where best practices could be adopted.

For CETS, a robust produced water module should build on existing datasets while expanding collection to capture treatment, reuse, and cross-border flows. A lot of the produced water is sent to Texas for ease of disposal, but if this is not feasible anymore the state will have to deal with enormous volumes in the Permian basin.

Standardized and regularly updated water quality data, better integration of production and water management datasets, and GIS-enabled linkages across basins would strengthen the state's ability to manage produced water sustainably.

Below is a list of the sources mentioned above with their data content, advantages and limitations.

NMOCD C-115 Production/Injection Reporting Monthly reported volumes of oil, gas, and produced water; includes injection/disposal and well history via related files.

➢ Pros: Official data of record for NM volumes; updated daily at source. Well-level granularity and time series suitable for trend and intensity analysis.
➢ Cons: Large XML flat files; requires joining multiple tables (well, completion, volumes). Data dictionary and business rules not well documented.

GO-TECH Petroleum data web access maintained by the PRRC at NMT

➢ Pros: Provides a less complex interface than NMOCD; allows single-well downloads and smaller annual or custom datasets (county or basin scale). Preserves some historical information (e.g., original well names, operators, drilling dates, or depths) that is missing or degraded in OCD records.
➢ Cons: Cannot download complete datasets at once; updated monthly; error-checking less rigorous than OCD data; overall less complete. Hardware/software upgrades can lag due to resource constraints.

USGS Produced Waters Database (PWDB) National produced-water chemistry (TDS, major/trace constituents, isotopes) with historical coverage.

➢ Pros: Broad spatial coverage and standardized attributes. Useful for treatment design, reuse feasibility, and environmental screening.
➢ Cons: Sparse and uneven temporal coverage; not comprehensive for recent NM production. Sample metadata can vary; requires curation for basin-level analyses.

New Mexico Produced Water Database Searchable water quality database specific to NM.

➢ Pros: NM-specific chemistry and management data; bridges volumes (OCD) and chemistry (USGS). Useful for mapping treatment/reuse opportunities by basin/formation.
➢ Cons: Web structure/links can change; no recent water quality updates; coverage depends on partnerships. Interface stability and completeness can vary over time.

New Mexico WaterStar map-based interface to water quality and volume data.

➢ Pros: New Mexico based data with a well-designed user interface that generates maps and graphics
➢ Cons: water quantity and quality data is grouped by ¼ Township and only data on produced water identified as being disposed through Saltwater Disposal Wells (SWD) are displayed.

Groundwater Protection Council (GWPC) Produced Water Reports National and regional assessments of produced water management, reuse, and policy.

➢ Pros: Synthesizes practices, costs, and regulatory context; actionable for planning. Helpful framing for reuse pathways and LCA boundary conditions.
➢ Cons: Primarily qualitative/summary reports; not a raw data repository. Coverage by state and year varies.

EPA Underground Injection Control (UIC) Program Class II injection well inventory and program information (disposal/EOR) and Class VI (CO_2 storage).

➢ Pros: Federal oversight view of injection activity; complements NMOCD records. Useful for cross-checking disposal capacity and spatial patterns.
➢ Cons: Program-level data; accessing well-level records may require multiple sources. Not a produced-water volume dataset per se; pairs best with NMOCD C-115.

FracFocus Registry National registry for hydraulic fracturing disclosures

➢ Pros: Useful for water-use accounting and completion-related emissions assumptions. Downloadable at well-level or as database for regional aggregation.
➢ Cons: Covers injected frac fluids, not produced-water disposition. Linkage to produced-water volumes requires well-level joins.

Academic & Regional Initiatives (NMSU WRRI, NMT PRRC, UNM) and DOE PWRC: Project-level datasets and reports on treatment pilots, reuse feasibility, and costs relevant to NM.

➢ Pros: Rich technical detail for process modeling and LCA assumptions. Can provide up-to-date case studies and pilot-scale data.
➢ Cons: Fragmented and project-specific; not standardized across studies. Data availability may be in PDFs/appendices rather than structured tables.

4.1.3 Carbon Management: Injection and CO_2 Management

New Mexico produces significant volumes of natural CO_2, primarily for EOR in the Permian Basin. In addition, anthropogenic CO_2 from industrial sources such as gas processing plants is injected for EOR and, in some cases, for disposal through acid gas injection (AGI) wells. Key data needs include CO_2 production, EOR utilization, acid gas disposal wells, and long-term geologic storage resources. The sources below provide permitting and inventory records, facility-level injection data, and geologic capacity assessments to support planning and documentation of CO_2 management pathways in New Mexico.

Monthly production and injection volume data for CO_2 are sourced the same as oil and gas data reported in Section 1.1, with the same constraints. For purposes of brevity, only the links are provided in the following list;

- NMOCD Permitting
- NMOCD Map Hub
- NMOCD FTP
- OCD Online Imaging
- NMOCD Statistics
- GO-TECH
- NMBGMR Subsurface Library

Other useful resources include:

- EPA Underground Injection Control (UIC) Program: Program oversight and data for Class II (oilfield fluids: produced water disposal and EOR, including CO_2 for EOR or AGI) and Class VI (geologic sequestration of CO_2) wells.

➢ Pros: Federal regulatory framework for siting, construction, operation, and monitoring of injection wells. Complements NMOCD records and distinguishes Class II EOR vs. Class VI storage.

➢ Cons: Detailed well-level data may require stitching from multiple sources. Not a unified volume dataset; best used alongside state production/injection records.

- EPA GHGRP: Subparts RR & UU (CO_2 Injection & Storage Reporting) Subpart UU covers CO_2 injection (e.g., EOR). Subpart RR covers geologic sequestration, including MRV (monitoring, reporting, verification) plans and annual injected/retained CO_2.

➤ Pros: Facility-level annual reporting distinguishes EOR vs. storage. MRV documentation supports accounting of injected and retained CO_2.

➢ Cons: Reporting thresholds and ~1-year lag limit immediacy. Public extracts may not include all site-specific details.

- NETL Carbon Storage Atlas and Natcarb Viewer Apps: Regional assessments of storage capacity in saline formations, depleted oil/gas fields, and unmineable coal seams; includes play/formation-level summaries relevant to New Mexico.

➤ Pros: Consistent national framework for high-level capacity screening. Useful for prioritizing basins/formations for feasibility work.

➢ Cons: Not a siting tool; estimates require local geologic and engineering validation. Resolution is regional; project design needs site-specific data.

- USGS Carbon Sequestration Assessments Independent assessments of technically accessible CO_2 storage with geologic detail by region/formation.

➤ Pros: Geoscience-forward methodologies and uncertainty framing. Complements NETL with alternative estimates and formation detail.

➢ Cons: Assessment vintages vary and may predate recent subsurface insights. Still requires site-specific characterization for projects.

- Regional & Academic Research Efforts (CUSP, PRRC-NMT, National Labs) Regional DOE-funded consortium (CUSP) and in-state research (NMT; Sandia and Los Alamos National Laboratories) providing geologic characterization, storage assessments, pilot studies, and monitoring technology development.

➤ Pros: Formation-specific data and pilot-scale results beyond national assessments. Local expertise and technology development applicable to NM basins.

➢ Cons: Project-specific and dispersed across reports/papers; not standardized. Data accessibility and update cadence depend on funding cycles.

4.2 Geothermal Data

New Mexico's abundant but still-underdeveloped geothermal resources, especially the blind systems hidden within the Rio Grande Rift, could supply baseload heat and power, support industrial decarbonization, and create rural jobs. A robust data-management framework is essential to:

- **Locate hidden resources.** Deep temperature, structural, and permeability data are prerequisites for derisking exploration.
- **Streamline permitting & investment.** Transparent, high-quality data lowers exploration costs, risks and attracts private capital.
- **Enable cross-sector uses.** Accurate demand, water-use, and market data guide coupling of geothermal heat with industries, agriculture, district energy, energy needs in buildings, and rising data center demand.

The management of geothermal data is not covered by current rules. Currently, most of the publicly available geothermal data is housed at the New Mexico Bureau of Geology and Mineral Resources (NMBGMR), both as a digital database and as paper records in the Subsurface Library and in the Geologic Information Center. The database includes petroleum-well bottom-hole temperatures, temperature-depth data, heat flow, geothermal gradient, and thermal conductivity data, drill-stem test data, groundwater chemistry, and porosity data. Cores and cuttings from geothermal wells are also stored at NMBGMR. Well logs from recently drilled geothermal wells by Eavor and Zanskar are housed at ECAM. Ensuring the geothermal community knows the location of all geothermal data is a top priority and will require collaboration between the industry, academia, and state and federal regulators.

Table 4: Overview of Geothermal Data Assets and Quality

Data Asset	Custodian	Coverage & Format	Accessibility	Notes
Legacy geothermal temp logs, chemistry, and gradient data	NMBGMR (digital DB + paper in Subsurface Library & GIC)	~600 shallow wells statewide	Public (partial digitization)	Many pre-1980 records were scanned but not georeferenced.
Core & cuttings from historic wells	NMBGMR Core Library	~250 boxes, varied depth	On-site only	Limited metadata; needs bar-coding.
Recent well logs (Eavor, Zanskar, others)	EMNRD – ECAM	4 deep exploration wells	Restricted	Data-release MOU in progress.
Oil & gas well & seismic data	NM OCD / BLM / USGS	Tens of thousands of wells and 2-D/3-D seismic in Permian, Raton, & San Juan Basins	Public (some w/ fees)	Sparse coverage in rift basins.
MT, gravity, magnetic surveys	Multiple private firms, USGS	Patchy regional data	Mostly proprietary	Key gap for Rio Grande Rift.

Critical Data Gaps & Research Needs:

- **Deep Rift Basin Control:** Drill-stem tests, down-hole pressure/temperature arrays, 2-D/3-D seismic, and magnetotelluric (MT) soundings.
- **Ground-Source Heat Pump (GSHP) Deployment:** Create statewide GSHP registry; survey HVAC contractors to geolocate existing installs and characterize performance.
- **Water-Resource Accounting:** Quantify water withdrawals, reinjection volumes, and produced-water compatibility for geothermal operations; coordinate with BLM and ECAM databases.
- **Industrial Heat-Demand Mapping:** Match basin-scale heat potential with process-heat users (dairies, dehydrators, chemistry, pulp/paper, mining). Pilot industry surveys through EMNRD would help.
- **Ag-Tech Applications:** Catalog livestock barns, greenhouses, and crop-drying sites suitable for geothermal space heating/cooling (Kuravi & Winingham 2025).

4.3 Mineral Data and Tools

The mineral data and data sources are included in the **Critical Mineral Technical Document, by Dr Virginia McLemore** associated with this report. Below are some additional tools that display mining activities (past and present) in the state.

The New Mexico **Mining and Minerals Division (MMD) GIS tool** to locate registered mines across the state. Link here.

Former Uranium Mine and Mill Sites Map: Developed by NMED and EMNRD, the interactive GIS dashboard displaying legacy uranium mines and former uranium mills across New Mexico. Users can view detailed site data, navigate by county or mine name, and explore regulatory programs.
https://nmenv.maps.arcgis.com/apps/dashboards/690621694d4e4906b2ae2886f528eec1

Coal Mine Resources and Land Ownership Map Explorer: Developed by New Mexico's MMD, the interactive GIS tool visualizes coal permit boundaries, mine locations, and coal field classifications across the state. https://catalog.newmexicowaterdata.org/application/new-mexico-coal-mine-resources-and-land-ownership-map-explorer

4.4 Water Data

Water data collected by a variety of state and federal agencies in New Mexico can be found at https://newmexicowaterdata.org/. The water data initiative develops a comprehensive database of all water data available in New Mexico. The organization of their data management system is a good example of how to manage different datasets using CKAN (see data platform chapters).

4.5 Other Data

While during the time period we had in this phase we were able to focus on oil and gas, geothermal, mineral, and water data, the next phase must expand to include additional datasets such as **energy economic impacts, workforce and societal statistics, and environmental**

indicators. These areas are critical to provide a more complete picture of New Mexico's energy landscape, linking technical resource data with the economic, social, and environmental dimensions needed to guide effective policy and investment decisions. Together, these datasets will enable CETS to serve as a truly comprehensive platform for understanding and guiding the state's energy transition.

5 DETECT Platform Development Strategy

Objective: Develop a strategy to develop the DETECT data platform and improve data collection, reporting, management, and integration. The overreaching goals are to:

- Provide a detailed, actionable recommendations on new data collection processes that align with the needs identified above (data elements and data sources), and a timeline for data gathering efforts.
- Develop a plan to develop the data platform and enhance data accessibility and management practices across agencies, including recommendations for digitization, data visualization platforms, and ongoing monitoring, reporting, and verification strategies.

5.1 Evaluation of Data Volumes

Before being able to develop a data strategy and a plan for the data platform, we need to have a better understanding of the volume of data that needs to be collected. To estimate the volume of data required to manage and upload all the data described previously in the data sources, we need to look at the reporting frequency, level of granularity (geographic, facility-level, fuel-type, time intervals, etc.), and typical data file sizes from all datasets. The following is a preliminary estimate based on EIA datasets and using an AI-assisted analysis (**Appendix 5**) of each form's characteristics and publicly available dataset sizes.

Assumptions for EIA Volume Estimation:

- Monthly forms: ~1–10 MB/month depending on granularity.
- Weekly forms: ~0.5–2 MB/week.
- Hourly forms: ~20–100 MB/day depending on resolution.
- Annual forms: ~1–5 MB/year.
- Historical data (~10 years) and ongoing uploads are considered.
- Uncompressed CSV format (conservative size).

Total Estimated Volume:

- Annual Data Upload Need (Current + Ongoing) ≈ 15–20 GB/year (dominated by EIA-930 hourly electric grid data)
- 10-Year Archive Estimate (if downloaded historically): ≈ 170–200 GB total

This estimate reflects only EIA data; additional datasets will need to be incorporated over time, but this provides a necessary starting point.

Notes:

- These are uncompressed file sizes (CSV or Excel format). Using optimized databases or compressed formats could significantly reduce volume.
- The EIA-930 alone makes up over 90% of the total data volume, due to its high-frequency (hourly) reporting. It highlights that the more frequent the reporting, the more volume of data we will need to manage.
- Based on the evolution of the needs, the data platform should be robust and highly adaptable.

5.2 Platform Architecture & Tools Recommendations

Based on the data assessment and the current needs expressed, we propose a comprehensive, adaptable (tool and size) and modern (latest software, ideally open-source) approach to building a data platform based on best practices. The proposed approach is a solution capable of collecting data from multiple sources, in various formats, and at different frequencies. This platform will centralize, store, transform, and make data available for analysis and visualization.

A key design principle of the DETECT platform is that it will be robust and scalable, allowing data and features to be integrated step by step, starting with core essentials and expanding over time into a comprehensive tool covering all energy and GHG data and tools needed for the CETS.

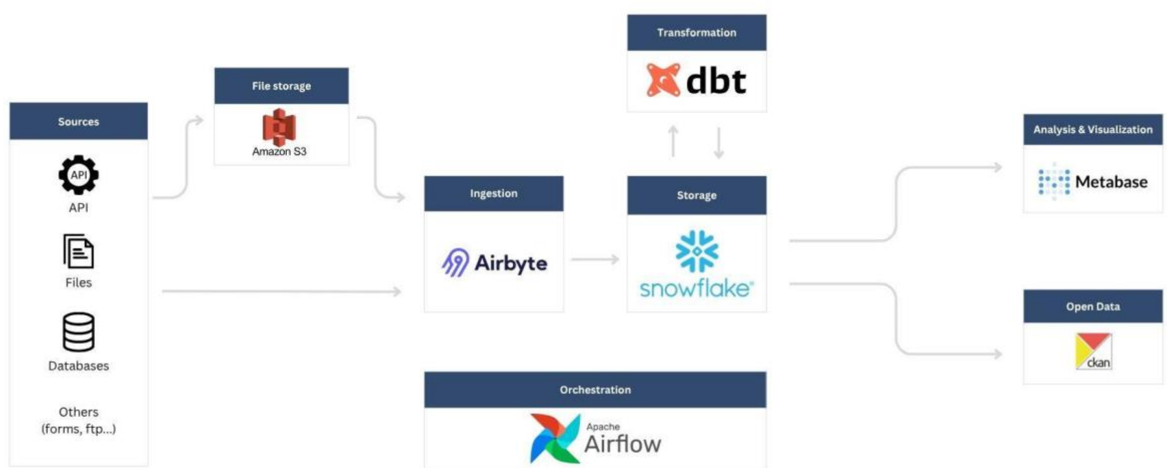
We also address key points around data privacy, security, and traceability, which are critical in the context of this project.

Based on conversation with stakeholders managing data in department agencies, we understand that a cloud-based platform is preferable, however, we outlined two implementation options:

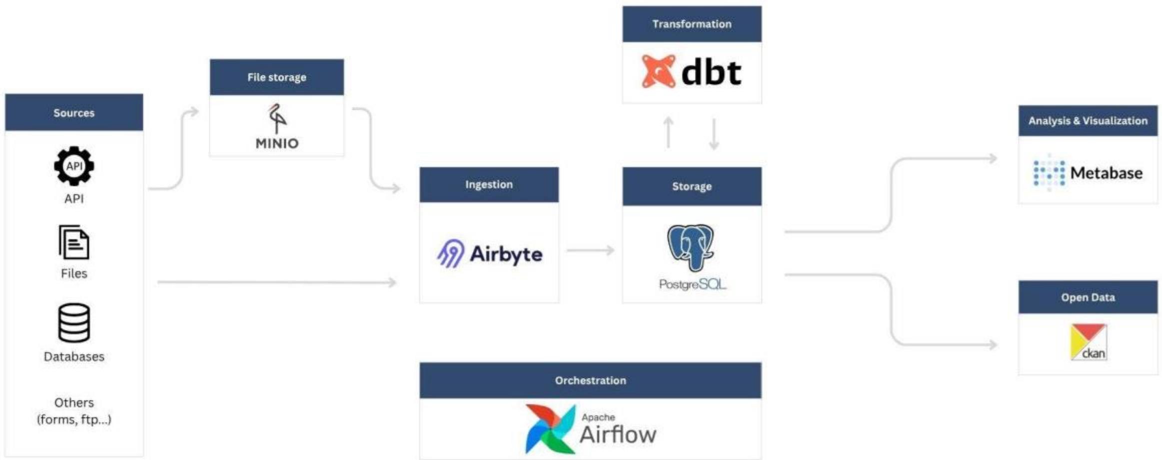
- A 100% private, on-premises solution, fully hosted on local infrastructure.
- A cloud-based or hybrid solution, leveraging managed services for scalability and ease of maintenance.

Each approach will be evaluated in terms of cost, scalability, operational complexity, and compliance with governance requirements.

Figure 1: Overview of the recommended data platform architectures (cloud and on-premises):



DATA PLATFORM ARCHITECTURE
(CLOUD)



DATA PLATFORM ARCHITECTURE
(ON-PREMISE)

Below is a description of the role of each section of the data workflow in the platform architecture.

5.2.1 Ingestion

Data ingestion refers to the process of collecting and integrating data from various heterogeneous sources, which may differ in structure, format, or storage technology. Our data platform must be capable of aggregating data from these multiple systems and accept external data contributions from third parties. Below, we present the different ingestion scenarios we anticipate supporting.

We recommend using Airbyte, a widely adopted open-source data integration tool that has become a standard in the data engineering ecosystem. Airbyte simplifies data ingestion by offering hundreds of pre-built connectors, allowing seamless integration with most common data sources. It can be deployed either on-premises or via a managed cloud service, depending on infrastructure choices.

We also recommend integrating a file storage service (such as Azure, Amazon S3 or Google Cloud Storage), which will be used in several of the ingestion workflows.

Ingestion Scenarios:

1. **Data Available via API**

When data is accessible via an API (e.g., data from https://www.eia.gov/opendata), we can use an existing Airbyte connector or develop a custom one if needed. This allows us to query the

API, extract the data, and load it into our data warehouse. The frequency of extraction (daily, weekly, monthly, etc.) can be adjusted according to update requirements.

2. **Data Available as a Published File (Excel, PDF, XML, etc.)**

If data is published in file formats on a website or FTP server, we propose a two-step ingestion process:

- File backup to a file storage system (e.g., Amazon S3).
- Ingestion via Airbyte, using a connector capable of parsing and importing the data into our database.

Importance of this intermediate step: Keeping the original file archived in storage ensures traceability, reproducibility, and auditability. It also protects us in case the source file is modified or removed by a third party. The cost of cloud file storage is minimal when using services like S3 or GCS.

- **Data Stored in External or Internal Databases**

For data already stored in third-party relational databases, we can leverage Airbyte's dedicated connectors. We will configure the appropriate extraction frequency depending on the use case and load the data into our platform's database accordingly.

- **Data Pushed to the Platform by a Third Party**

In some cases, third parties may need to push data to our platform rather than us pulling it. Several implementation options can be considered within our proposed architecture:

- **Receiving data via a custom API**: We can expose tailored API endpoints that allow third-party systems to push structured data directly into our platform securely (e.g., from industry-specific software)
- **Receiving data via a form**: We can provide web-based interfaces or upload forms where users can manually submit files or input data.
- **Other formats or methods**: Depending on the partner's system constraints, additional ingestion mechanisms (e.g., email-based dropboxes, SFTP servers) can be evaluated and integrated into the pipeline.

5.2.2 Storage

Storage is the central component of our data architecture, it is where all collected data will be recorded during the ingestion phase, whether in raw form or already cleaned and transformed. This layer serves as the single source of truth for further analysis and visualization.

We recommend using Snowflake as the primary data storage engine. Snowflake is a widely adopted cloud-based solution, recognized in the industry for its strong performance, flexibility, and ease of use. Snowflake automatically adjusts its resources based on data volume and workload, making it easy to scale without manual intervention.

In terms of security, Snowflake offers interesting built-in features such as end-to-end encryption, multi-factor authentication, fine-grained access control, and compliance with leading industry standards and regulations. It also provides strong data traceability, with detailed query history, change tracking, and auditing tools that ensure transparency, reproducibility, and compliance throughout the entire lifecycle of the data pipelines.

For the on-premise option, we recommend using PostgreSQL as the primary database engine. PostgreSQL is a widely used, open-source, enterprise-grade relational database system known for its stability. PostgreSQL integrates seamlessly with most data management tools.

Compared to Snowflake, PostgreSQL offers full control over the environment and data residency. However, it requires significantly more operational overhead, including infrastructure management, scaling decisions, security hardening, and therefore demands more dedicated manpower and expertise to maintain and optimize.

Data Warehouse Structure

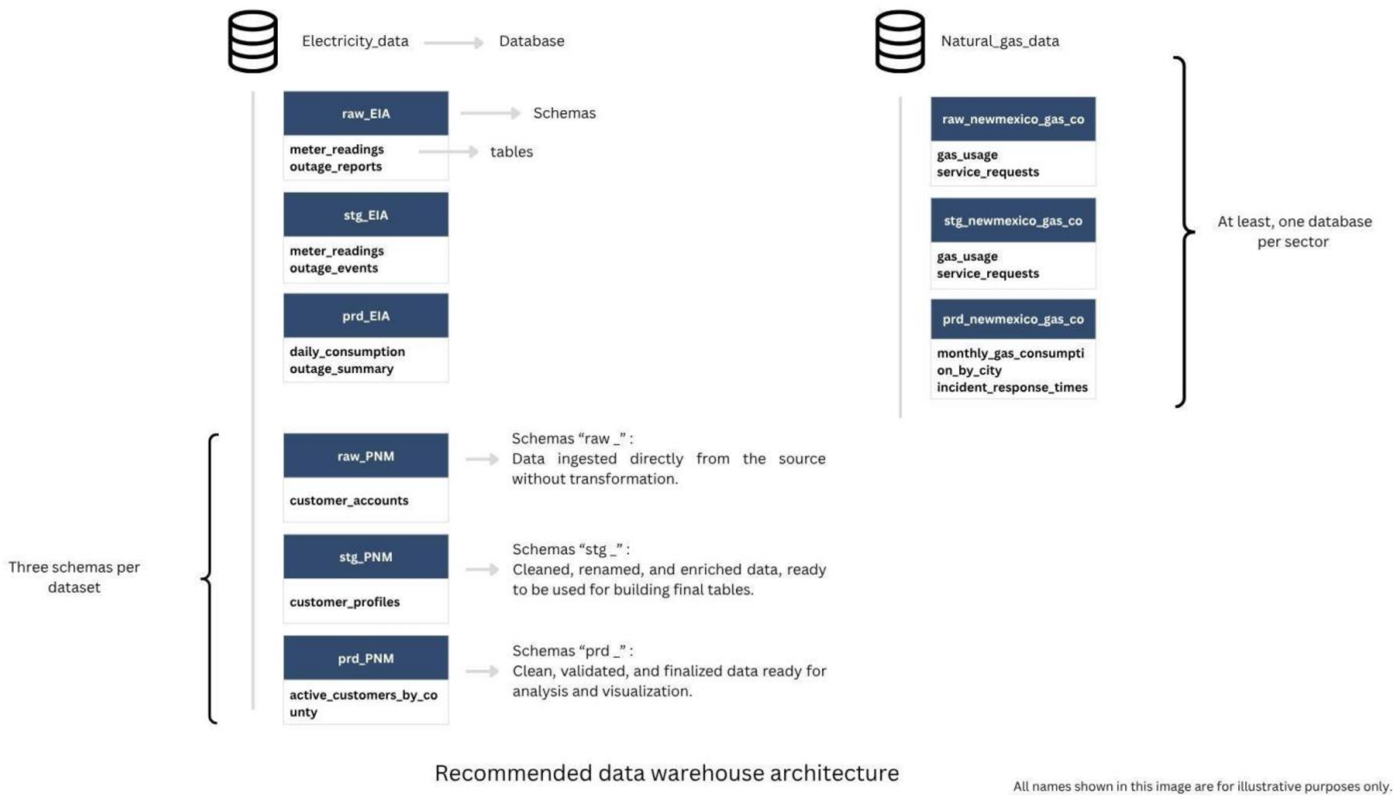
The organization of the data warehouse is a critical aspect of the project. We will need to store data from dozens of different sources, each potentially requiring its own cleaning, transformation, access policy and validation processes. It is therefore essential to implement a database structure that remains clear and scalable as new sources are added, while also being secure and compatible with traceability and data governance requirements.

We propose the following generic structure highlighted in **Figure 2** (which can be adapted as needed):

- One database per sector (e.g., electricity, gas…),
- Within each database, each data source is organized into three schemas:
    - **raw** – unprocessed data, ingested directly from the source without transformation,
    - **stg** – cleaned, renamed, and enriched data, ready to be used for building final tables,
    - **prd** – clean, validated, and finalized data ready for analysis and visualization.

This structure ensures long-term maintainability, facilitates auditability, and supports clear separation of responsibilities across the data lifecycle.

Figure 2: Recommended Data Warehouse Architecture



Recommended data warehouse architecture

5.2.3 Transformation

The transformation step is the process of converting raw data into a structured, clean, and meaningful format that can be reliably used for analysis and reporting. After ingestion, raw data is often inconsistent, incomplete, or difficult to interpret. During transformation, this data undergoes a series of operations such as:

- **Cleaning**: removing duplicates, correcting errors, handling missing values.
- **Standardization**: applying consistent naming conventions, formats, and units across datasets.
- **Enrichment**: adding context by joining with reference data or calculating derived fields.
- **Validation**: ensuring the data meets expected rules, constraints, and quality checks.

This step is essential to ensure that downstream users, whether analysts, data scientists, or public-facing dashboards, can trust and efficiently use the data.

To manage the transformation layer of the data pipeline in a structured, scalable, and maintainable way, we recommend using dbt (data build tool). dbt is an open-source framework that enables analysts and engineers to transform data in the warehouse using simple SQL statements, while applying software engineering best practices such as version control, testing, modularity, and documentation. It integrates seamlessly with Snowflake and other modern data warehouses, allowing transformations to run directly where the data lives, without requiring a separate processing layer.

Given the complexity and diversity of the data sources in this project, dbt will help ensure transparency, traceability, and consistency in all transformation logic, while making the pipeline easier to maintain and evolve over time to meet the CETS goals.

5.2.4 Orchestration

Orchestration is a critical step in a data platform, responsible for automating, scheduling, monitoring, and managing the entire set of tasks that make up the data pipeline, from ingestion to final delivery. It ensures complex workflows run in the correct sequence with fine-grained dependency handling, error management, and automatic retries.

To manage orchestration in a robust and scalable manner, we recommend using Apache Airflow, a widely adopted open-source platform. Airflow enables defining workflows using Python, providing great flexibility to model complex chains of operations. Airflow can be deployed both on-premises and in cloud environments, offering flexibility depending on infrastructure preferences or regulatory requirements.

5.2.5 Analysis, Visualization and Open Data

Once data is cleaned, transformed, and orchestrated, we want to make it accessible and usable both for internal users and the public.

For interactive analysis and visualization needs, we recommend Metabase, an open-source, user-friendly solution that enables business users and analysts to build dashboards, reports, and charts without coding. Metabase connects directly to databases or data warehouses (e.g., Snowflake) and facilitates dynamic query creation and visualizations via a modern web interface. This encourages data democratization and self-service analytics.

For open data publishing, dataset cataloging, and secure dissemination, we suggest **CKAN** (Comprehensive Knowledge Archive Network). CKAN is an open-source platform widely used by governments and institutions to publish and organize public datasets in a standardized way, complete with metadata, access controls, and APIs. Its adoption will structure access to energy data, facilitate reuse by third parties (researchers, companies, citizens), and ensure data traceability and quality in published datasets.

CKAN is used by the US government to share data (https://data.gov) and also by energy data service companies (for example https://www.energidataservice.dk/datasets).

For spatial visualization of energy and environmental data within the CETS platform, it may be necessary to incorporate ESRI GIS tools or other open-source alternatives.

5.3 Data Quality

5.3.1 Quality Metrics

For each dataset, it is essential to define specific quality criteria to ensure the reliability and relevance of the data. Some of these criteria will be common across all datasets, such as duplicate detection, identification of incomplete or missing data, verification of the last update date, and detection of outliers or inconsistent values.

But some criteria will be industry-specific, tailored to operational requirements. These may include compliance with expected formats, adherence to domain-specific calculation rules, or consistency with internal and external reference data.

Table 5: General Data Quality Metrics

Metric Name	Definition	Unit	Example Threshold
Completeness	Share of non-missing hourly production values for each plant.	%	≥ 98% per day
Range Check	Production must stay within 0 and the plant’s maximum installed capacity.	MW	Within 0 – Max Capacity
Sudden Spike Detection	Detection of unexplained abrupt changes in production.	% change	≤ ±20% in one hour
Uniqueness	Absence of duplicate records for the same plant, date, and time.	Count	0 duplicates

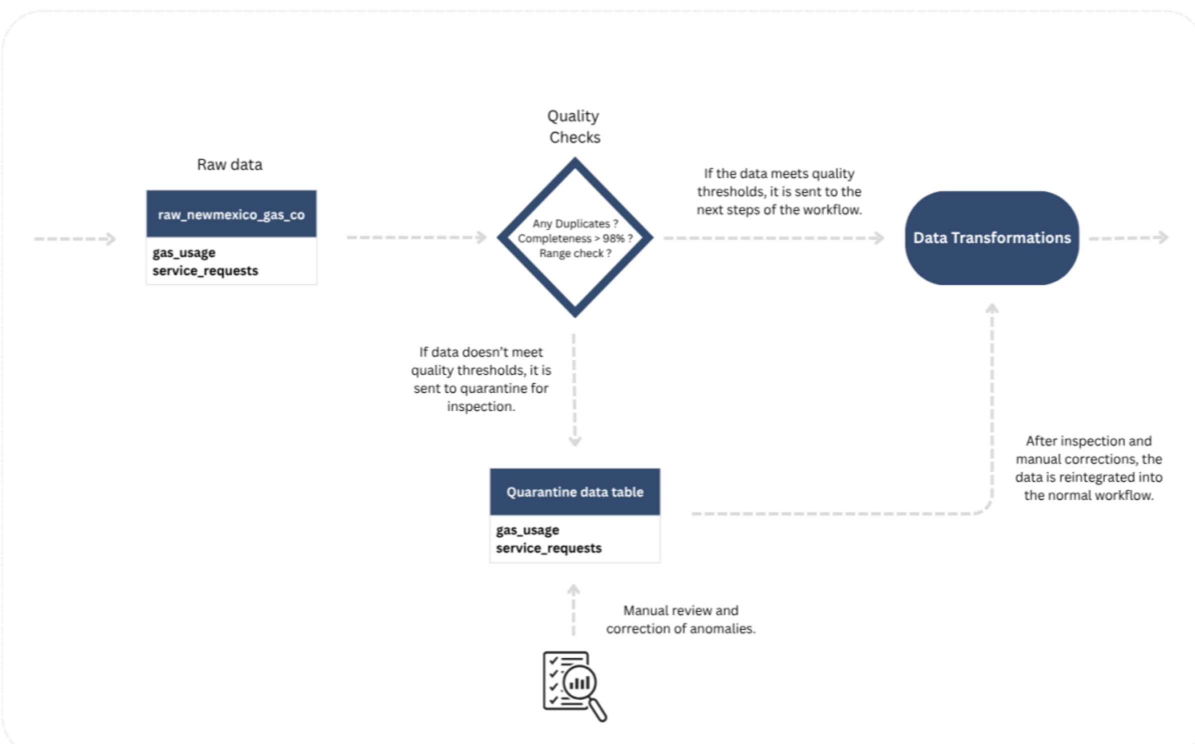
5.3.2 Automated Process for Quality Testing

Based on the previously defined quality metrics, we can implement automated workflows to verify that any new data meets the established quality criteria before it is published. This process can rely on a system of gates or checkpoints, where data that does not meet the defined standards is automatically quarantined in a dedicated database. These quarantined records must then be reviewed, corrected, and validated before being reintroduced into the main data platform. This approach not only prevents the release of inaccurate information but also ensures traceability of detected issues and corrective actions, fostering continuous improvement in data quality.

5.3.3 Data Quality Dashboards

Data quality dashboards provide a central, visual interface to monitor and track the health of datasets. By consolidating quality metrics such as completeness, timeliness, accuracy, and consistency, these dashboards offer both a global overview and the ability to drill down into specific datasets or issues. They enable stakeholders to quickly identify anomalies, track trends over time, and assess the effectiveness of corrective actions. Integrating automated alerts ensures that deviations from expected thresholds are promptly flagged, allowing for proactive remediation.

Figure 3: Example of Automated Quality Testing Workflow



Example of automated quality testing workflow

5.4 Data Governance

Data quality ensures that information is accurate, consistent, and reliable, while data governance provides the framework to manage, protect, document, and ensure the traceability of data across the organization. These two dimensions are essential and must be embedded in the design of the data platform we aim to build. By integrating them from the outset, we will be able to publish and deliver data that is not only trustworthy and actionable, but also compliant with regulatory requirements and aligned with operational needs. Therefore, Phase 2 of CETS should focus on establishing a clear framework for data governance.

5.4.1 Governance Roles and Data Access Control

Data governance relies on clearly defined roles and responsibilities for all stakeholders involved. Each team or agency point of contact must be identified and we must have a coordinated approach to ensure smooth data management. Access rights can be granted based on the principle of least privilege, ensuring that only authorized individuals can view or modify sensitive data, while promoting secure collaboration between agencies.

5.4.2 Documentation and Traceability

Comprehensive documentation is essential to guarantee data traceability and understanding. We also recommend using a data lineage tool that will precisely track the origin, transformations, and flow of data throughout the system, ensuring transparency.

5.4.3 Validation Step

If certain datasets require approval from an agency or a third party before publication, we can implement a quarantine system similar to the automated data quality checks (**Figure 3**). These datasets will be placed in a dedicated quarantine table and will not be published until they have received formal approval from the responsible agency or authorized personnel. This process ensures that sensitive or critical data undergoes necessary validation and oversight before being made publicly available.

5.5 Security & Compliance

5.5.1 Security

Cloud-based platform: The tools we recommend for building the platform, particularly Snowflake, offer comprehensive security features to protect data integrity and confidentiality. These include fine-grained access control policies, multi-factor authentication (MFA), and data encryption both at rest and in transit, role-based access control (RBAC), and audit logging.

Additionally, the platform is widely used by major enterprises and regulated industries, which demonstrates its robustness and compliance with industry standards.

On-premise deployment: If the data platform needs to be deployed on-premise, appropriate infrastructure and security measures must be planned. This includes secure physical access controls, network segmentation, firewalls, intrusion detection systems, and regular security audits. Data encryption and access management should be enforced consistently, and compliance with internal IT security policies must be guaranteed. On-premises deployment offers more direct control but requires dedicated resources for maintenance and security monitoring.

5.5.2 Compliance

Certain datasets will require certification or attestation to meet regulatory or internal standards. We recommend leveraging the data platform's file storage capabilities to securely store these certificates, for example, as PDF documents. These certificates can be linked permanently to their corresponding datasets, ensuring traceability and easy verification during audits. This integration supports ongoing compliance by maintaining a transparent connection between data and its certification status. Additionally, metadata fields can document certification dates, validity periods, and responsible parties, facilitating automated compliance checks and reporting.

6 Next Steps to Develop the DETECT Data Platform

6.1 Host Institution and Sustainability

The CETS data platform will require a clear institutional home. Below are the different options:

1. The first option is to establish a dedicated agency at the state level. This would tie the platform directly to regulatory authority, supporting compliance tracking and enforcement. However, it would require stable state funding and staffing commitments.
2. The second option is to contract a third-party, such as a university or other stakeholder to operate the platform. This model can leverage existing in-state technical expertise, offers staffing flexibility, and builds capacity within New Mexico's ecosystem. It would also make the platform less vulnerable to political or budget cycles. Regardless of the model, long-term sustainability will require stable financing.
   The State would retain ownership of all data, with the third-party contractor responsible for maintaining accuracy and integrity during the contract period. If the contract ends, two scenarios could apply: for public data only, all databases are transferred to the State; for public and private data, the third party needs to operate under an NDA, and upon termination, both public and private datasets are securely transferred to the State to ensure continuity and protection.
3. A third option is a hybrid approach that may also be the most effective: state agencies provide governance and oversight, while third-parties manage operations and technicalities. This structure combines the accountability and security of government with the technical skills of third-parties' stakeholders, ensuring the platform is durable, adaptable, and trusted over time.

6.2 Estimated Timelines and Minimum Cost

The minimum costs and timelines discussed here relate specifically to the development of the prototype of the DETECT data platform. This should be designed to integrate and build upon agency-level improvements. The DETECT platform will provide a centralized, scalable platform that unifies datasets and supports compliance with New Mexico's energy transition and climate goals.

To accurately estimate the minimum cost and timeline for building the data platform, it is necessary to define a set of assumptions. These include the anticipated volume of data to be processed, the frequency at which the platform will be updated, and the type and scale of infrastructure required. Establishing these parameters upfront ensures that both budget and delivery expectations are based on realistic projections.

6.2.1 Cloud Deployment Cost Breakdown - Estimates

Data platform assumptions:

- 2 to 6 GB/month, monthly refresh
- Software: Airbyte, dbt, Snowflake, Airflow, Metabase, CKAN
- All cloud-hosted

- Estimated fully loaded daily rate for key data roles:
  - Data Engineer (DE) / DevOps: $600/day
  - Data Analyst (DA): $500/day

Table 6: Assessment of minimal CAPEX and OPEX for DETECT prototype

Component	Description	CAPEX - Initial investment		OPEX - Recurring costs	
		Infrastructure costs	Labor Costs	Infrastructure costs	Labor Costs
Airbyte	Deployment, config, connectors	$0	$1,800 (3 days Data Engineer)	$800[1]	$600 (1 days/month, Data Engineering: monitoring & updates)
dbt	Initial models, tests, CI/CD	$0	$1,800 (3 days DE)	$0	
Snowflake	Computing costs	$0	$0	$3500–$4000[2]	
Airflow	Orchestration deployment	$0	$3,000 (5 days DE)	$800[1]	
Metabase	Data visualization / dashboards setup	$0	$2,000 (4 days Analyst)	$800[1]	$500 (1 day/month, Data Analyst: monitoring updates)
CKAN	Managed portal setup	$0	$3,200 (2 days DE + 4 days Analyst)	$800[1]	
Security & Backups	SSL, backups, monitoring	$0	$1,200 (2 days DevOps)	$300–$500	$600 (1 day/qtr, DevOps: monitoring & updates)

Notes:

- We propose hosting CKAN and Metabase on one AWS EC2 instance, and Airbyte, Airflow, and DBT on another EC2 instance. The annual OPEX infrastructure costs presented in the table are estimated based on the cost of two EC2 instances with appropriate capacity (4 vCPUs, 16 GB memory, and 100 GB EBS storage each).
- Snowflake annual OPEX is estimated based on the forecasted volume of data to be processed and the Snowflake compute rate, assuming approximately 60 hours of warehouse usage per month.
- Since this solution is entirely cloud-based, there is no upfront CAPEX, and costs are purely operational.

Total – Only for Platform Architecture Development:

- CAPEX:
    - Infrastructure/tools: **$0**
    - Personnel Costs: **$13,000**
    - Total estimated CAPEX: $13,000
- OPEX for 1st year:
    - Infrastructure/tools: **$7,000**
    - Personnel Costs: **$15,600**
    - Total OPEX/year: **$22,600**

The OPEX presented represents the cost of integrating a defined number of data sources into the DETECT platform and maintaining them over a one-year period. It does not include additional tools, maps, dashboard development nor analysis work. Because the platform is being designed to be modular, scalable, and robust, data integration can occur step by step rather than all at once. However, each time a new data source is added, there are associated costs for:

- Ingestion and integration (developing pipelines, harmonizing formats, applying QA/QC),
- Ongoing maintenance (updating connections, monitoring data quality, ensuring compliance), and
- Platform expansion (upgrading storage, processing capacity, UI, and visualization tools).

This means that the OPEX costs for integration and maintenance are minimal and will change widely as new sources or tools are developed and brought online. As the platform grows, the overall cost of development and upkeep increases proportionally with the number of datasets integrated and the staffing needs.

The modular approach ensures flexibility and allows the platform to evolve in phases, but it also means budgeting must account for incremental OPEX growth as the platform expands to cover additional energy, environmental, and compliance datasets and tools.

6.2.2 Breakdown for Pipeline Development Effort per Source

Table 7: Estimated timeline and tasks for new data source integration

Task	Estimated Data Engineer Effort
Connector setup & ingestion	2 days
Transformation with dbt (modeling, error handling,tests)	3 to +5 days (depending on the complexity)

Testing & deployment	1 day
Documentation	2 day
Initial dashboard adjustments	1 day

Total per source: ~9 days Data Engineer + Data Analyst

Note:

- A source is defined as a dataset from a specific website that is not already included in the architecture (e.g. EIA APIs, EPA webpages, NMOCD database…)
- This is a generic estimate provided for modelling purposes only. In practice, integrating a new data pipeline may require significantly more time, depending on its complexity and the systems involved.
- Each data pipeline also requires ongoing maintenance. For example, if the source system changes the data format, a data engineer will need to update the pipeline's code to accommodate this change. Without such updates, the pipeline may fail to operate correctly. Therefore, it is important to allocate dedicated time and resources for the maintenance of each pipeline to ensure continued reliability and performance over time.

For Phase 2 of the CETS project, we anticipate the integration of at least **20 data sources into the platform**. Based on our current assumptions, each integration is estimated to require around 9 days of dedicated work.

At this rate, the total effort amounts to roughly **180 days**. This estimate does not yet account for unforeseen complexities, such as incomplete documentation, unstable source systems, or unexpected format changes, which could increase the time required.

Development cost estimation = 180 days ($100,000 to $120,000)

Maintenance cost estimation = 20 days ($10,000 to $15,000)

Table 8: Minimal Cost Estimation Summary for Prototype Data Platform Development Only (for 1 year, 20 data sources)

	CAPEX - Initial investment	OPEX - Recurring costs
Infrastructure Costs	$13,000	$22,600
Development Costs	$100,000 to $120,000	$10,000 to $15,000

Total	$113,000 to $133,000	$32,600 to $37,600/year

6.2.3 Additional Staffing

In addition to the technical costs of integrating and maintaining datasets, there is a need for additional resources, staffing and coordination with state agencies to guide the development and sustain the platform.

For phase 2 development, a single Data Engineer with support from a Data Analyst cannot manage the entire project and needs of the DETECT data platform nor make it sustainable in time (given that it will grow in size and functionalities, and therefore increase maintenance needs). Like described in Chapter 1, DETECT will require a support team to ensure that ingestion pipelines, QA/QC processes, compliance dashboards, UI, and visualization tools operate effectively and are relevant. Additional energy and GHG experts should be engaged to interpret the data and develop storytelling, education, and outreach materials that make the platform interactive, and information accessible and relevant to diverse audiences.

Key roles that need to be added in future phases:

- Front-end developer;
- Support staff to manage routine data ingestion, formatting, and metadata documentation;
- Researchers and energy experts to guide development, interpret the data, develop sector-specific methodologies and validate development.

By incorporating these roles, the platform can move beyond raw data management and integrate the planned functionalities for Phase 2. These staffing costs represent an operational requirement, growing in proportion to the number of datasets integrated and additional tools development.

6.2.4 Timeline Estimation for Prototype Development

Figure 4: Time estimation for Prototype Development

The timeline for developing the **DETECT prototype** follows a clear, phased structure designed to deliver early functionality while allowing for iterative integration of additional data sources and features:

- **Project Start (Month 0):** Initiation of the project based on CETS phase 1 prototype and strategy, including team onboarding, scope definition, and finalization of data priorities and architecture design.
- **Infrastructure Setup (Month 1):** Establishment of the technical foundation: servers, databases, cloud environment, and data pipelines; ensuring the platform can securely host and process datasets. If possible, we could use the DETECT prototype from phase 1.
- **Operational Platform (Month 6):** A functional prototype becomes available with approximately **10 key data sources integrated** (e.g., EIA, EPA, NMOCD, NMED). This stage focuses on data ingestion, standardization, and the creation of initial dashboards and visual tools for user testing.
- **Fully Operational Platform (Month 11):** The system scales to include **around 20 integrated data sources**, improved analytics capabilities, and refined visualization tools, resulting in a stable, publicly accessible version of the platform.

This timeline represents a realistic step by step one-year path from concept to full prototype deployment, balancing technical development with data quality, interagency coordination, and user validation.

Conclusion

The development of the **DETECT** platform: *Data for the Energy Transition, Economy, Climate Change, and Technologies,* is an essential step toward establishing a modern, coordinated energy and climate data infrastructure for New Mexico. This report has outlined the technical, institutional, and strategic framework required to build the system that will integrate currently fragmented datasets into a unified, standardized, and interoperable platform to support energy planning, emissions tracking, energy literacy and policy development.

The phased design of the DETECT strategy and the phased design of the DETECT platform ensure adaptability to resource availability. In the short term, the platform can begin by integrating existing public datasets from federal and state sources to demonstrate early value through dashboards, visualizations, and metadata organization. In the mid-term, the system can expand through formalized data-sharing agreements with agencies and industry partners, incorporating additional datasets for more detailed analytics and near-real-time reporting. In the long term, with full staffing and legislative support, DETECT could evolve into a comprehensive, state-managed digital twin capable of predictive modeling, emissions verification, and long-term infrastructure planning.

Effective implementation will require close coordination with partner agencies and stakeholders, and a strong governance structure for data verification, validation, and updates. Workforce limitations at state agencies are a key consideration; therefore, a DETECT team should be built to lead technical integration and data management, reducing the workload on agency staff while ensuring that data collection and updates remain efficient and meet the current state needs. Establishing a data working group, standardized protocols, clear communication channels, and automated quality control processes will be critical to maintaining consistency and reliability over time.

Beyond a data platform, DETECT can play a broader role in energy outreach, education, and transparency. By making energy and climate data more accessible and interpretable, the platform can help inform policy discussions, strengthen stakeholder engagement, and support public understanding of the state's energy transition and its impacts. This integration of technical capability and communication value will ensure that DETECT not only serves as a data repository but also as a strategic asset for collaboration, accountability, and evidence-based decision-making.

Ultimately, DETECT positions New Mexico to become a national leader in data-driven energy governance. The foundation established in this report provides a clear roadmap for building a scalable, resilient, and transparent platform that supports both current needs and future energy and climate challenges.

References

Energy Information Administration (EIA)

- EIA Open Data (APIs and documentation): https://www.eia.gov/opendata/
- State Energy Data System (SEDS): https://www.eia.gov/state/seds/
- State Energy Profiles (SEP): https://www.eia.gov/state/
- Electricity: Hourly grid data (EIA-930): https://www.eia.gov/opendata/browser/electricity/rto
- Electric Power Plant Operations (EIA-923): https://www.eia.gov/electricity/data/eia923/
- Generator Data & Storage (EIA-860): https://www.eia.gov/electricity/data/eia860/

U.S. Environmental Protection Agency (EPA)

- Greenhouse Gas Reporting Program (GHGRP) data sets: https://www.epa.gov/ghgreporting/ghgrp-data-sets
- EPA GHGRP (40 CFR Part 98) overview: https://www.epa.gov/ghgreporting
- FLIGHT (Facility Level Information on GHGs Tool): https://ghgdata.epa.gov/
- Methane Emissions Data Explorer (MEDE): https://www.epa.gov/system/files/documents/2023-12/methane-emissions-data-explorer.pdf *(landing/example)*

State and Regional Resources

- Energy, Minerals and Natural Resources Department (EMNRD) – ECAM: https://www.emnrd.nm.gov/ecmd/
- Oil Conservation Division (NMOCD) – Geospatial Hub: https://ocd-hub-nm-emnrd.hub.arcgis.com/
- NMOCD Oil & Gas Map (ArcGIS app): https://nm-emnrd.maps.arcgis.com/apps/webappviewer/index.html?id=4d017f2306164de29fd2fb9f8f35ca75
- NMOCD Brine Wells Map (ArcGIS app): https://nm-emnrd.maps.arcgis.com/apps/webappviewer/index.html?id=5a610e64e6b1492d9f64449fccb6dcda
- NMOCD — division homepage & data links: https://www.emnrd.nm.gov/ocd/
- New Mexico State Land Office – Oil, Gas & Minerals Online Mapping: https://mapservice.nmstatelands.org/OilGasMinerals/
- New Mexico Environment Department (NMED): https://www.env.nm.gov/
- NMED OpenEnviroMap: https://gis.web.env.nm.gov/oem/?map=gonm
- New Mexico Public Regulation Commission (PRC): https://www.nm-prc.org/
- Interwest Energy Alliance — *Evaluating Transmission Opportunities in Integrated Resource Plans*: https://interwest.org/transmission-in-irp/
- New Mexico Water Data Act (2019) & Water Data Portal: https://newmexicowaterdata.org/
- New Mexico Geographic Information Officer (GIO) Report (2025): https://nmgic.com/wp-content/uploads/2025/05/01_NMGIC_GIOReport.pdf
- GO-TECH (NMT PRRC) petroleum data access: https://gotech.nmt.edu/
- NMOCD Statistics (summary files): https://www.emnrd.nm.gov/ocd/statistics/

- New Mexico Air Quality Control Act (overview gateway): https://www.env.nm.gov/air-quality/

Grid / Transmission (Markets, Queues, OASIS)

- CAISO Public Queue Report: https://www.caiso.com/library/public-queue-report
- CAISO Daily Energy Storage Reports: https://www.caiso.com/library/daily-energy-storage-reports
- PJM Interconnection Queue: https://www.pjm.com/planning/services-requests/interconnection-queues
- MISO Generator Interconnection Queue: https://www.misoenergy.org/planning/generator-interconnection/GI_Queue/
- ERCOT Interconnection information: https://www.ercot.com/services/comm/mkt_docs/gi
- OASIS overview (FERC Order 889 context): https://www.ferc.gov/industries-data/electric/power-sales-and-markets/oasis-open-access-same-time-information-system
- Tri-State (OATI OASIS entry point): https://www.oatioasis.com/TRLD/index.html
- El Paso Electric (EPE) transmission / operations info: https://www.epelectric.com/

Pipeline / Infrastructure Mapping

- PHMSA National Pipeline Mapping System (NPMS): https://www.phmsa.dot.gov/data-and-statistics/npms/national-pipeline-mapping-system-npms
- Homeland Infrastructure Foundation-Level Data (HIFLD) Open Data: https://hifld-geoplatform.opendata.arcgis.com/

Other States Tools Examples

- California Energy Commission — Energy Storage System Survey (portal/overview): https://www.energy.ca.gov/data-reports/energy-almanac/california-electricity-data/california-energy-storage-system-survey
- CAISO Integration / market-facing data (storage, operations): https://www.caiso.com/market/Pages/ReportsBulletins/DailyRenewablesWatch.aspx *(ops reports hub)*
- CPUC Hosting Capacity / ICA landing: https://www.cpuc.ca.gov/industries-and-topics/electrical-energy/infrastructure/hosting-capacity-analysis
- NYSERDA DER Integrated Data System (search): https://der.nyserda.ny.gov/search/
- NYSERDA Storage Data Maps: https://www.nyserda.ny.gov/All-Programs/Energy-Storage-Program/Storage-Data-Maps
- NREL ResStock & ComStock datasets: https://resstock.nrel.gov/datasets
- RTE-France interactive grid map (reference example): https://www.rte-france.com/

Appendices

Appendix 1 - Overview of EIA Data from the Federal Energy Administration Act

Forms by Major Category as of August 2025 - **some forms are suspended or discontinued as noted (EIA).**

Selection of EIA Forms, data collected, and reporting frequency under the FEAA. More forms are available here: (Link)

Form #	Name / Description	Data Covered	Reporting Frequency
EIA-782C	Petroleum Product Sales by Prime Suppliers	Monthly sales volumes of gasoline, diesel, heating oil, jet fuel, etc.	Monthly - Not available since XXX
EIA-800	Weekly Refinery and Fractionator Report	Crude oil inputs, production, stocks at refineries and fractionators	**Weekly** (every Monday)
EIA-810	Monthly Refinery Report	Inputs, outputs, fuel use, and production at refineries	Monthly
EIA-815	Monthly Bulk Terminal and Blender Report	Petroleum product stocks, receipts, and transfers	Monthly
EIA-816	Monthly Natural Gas Liquids Report	Plant-level NGL production and inventories	Monthly
EIA-877	Heating Oil and Propane Prices Survey	Retail prices of heating oil and propane by state	**Weekly** (during winter)
EIA-861	Annual Electric Power Industry Report	Electric utilities' retail sales, revenue, customer count	Annual
EIA-860	Annual Generator Report	Power plant attributes: nameplate capacity, fuel, retirements	Annual
EIA-923	Power Plant Operations Report	Fuel consumption, electricity generation, heat content, emissions	Monthly & Annual (mixed)
EIA-930	Hourly Electric Grid Operations	Balancing authority-level generation, demand, and interchange	**Hourly** (real-time)

EIA-176	Annual Natural and Supplemental Gas Supply and Disposition	Deliveries, receipts, storage, and sales by gas companies	Annual
EIA-757	Natural Gas Processing Plants Survey	Capacity, operations, and outages of gas processing plants	Annual & as needed

Alternative Fuels
- EIA-22M – Monthly Biodiesel Production Survey *(discontinued)*
- EIA-886 – Annual Survey of Alternative-Fueled Vehicle Suppliers & Users *(suspended)*

Coal
- EIA-20 – Emergency Weekly Coal Monitoring Survey (power producers)
- EIA-3 – Quarterly Survey of Industrial, Commercial & Institutional Coal Users
- EIA-6 – Emergency Coal Supply Survey
- EIA-7A – Annual Survey of Coal Production & Preparation
- EIA-8A – Annual Survey of Coal Stocks & Coal Exports

Electricity
- Hourly/Daily: EIA-930 (Balancing-Authority Operations)
- Monthly: EIA-826 (Utility Sales & Revenue) / EIA-860M (Generator Updates) / EIA-861M (Industry Report) / EIA-923 (Power-Plant Operations)
- Annual: EIA-860 (Generators) / EIA-861 (Industry) / EIA-861S (Short Form) / EIA-923 (Ops) / EIA-930A (Generator Inventory)
- Quarterly: EIA-111 (Electricity Imports & Exports)
- As Needed: OE-417 (Electric Emergency Incident & Disturbance)

Natural Gas
- Weekly: EIA-912 (Underground Storage)
- Monthly: EIA-191 (Underground Storage) / EIA-191L *(LNG storage – suspended)* / EIA-816 (NGLs) / EIA-857 (Purchases & Deliveries) / EIA-910 (Marketers) / EIA-914 (Oil & Gas Production) / FE-746R (Gas Imports/Exports)
- Annual: EIA-23L (Oil & Gas Reserves – county) / EIA-23S *(summary – suspended)* / EIA-64A (Origin of NGL Production) / EIA-176 (Gas Supply & Disposition)
- Stand-by: EIA-757 (Gas Processing Plant Survey)

Petroleum (incl. Biofuels & NGLs)
- Weekly: EIA-800 / 802 / 803 / 804 / 805 / 806 / 809 / 877 (Heating-Fuel Prices) / 878 (Gasoline Prices) / 888 (Diesel Prices)
- Monthly: EIA-14 / 22M *(biodiesel – discontinued)* / 182 / 782A/B/C *(all discontinued)* / 810 / 812 / 813 / 814 / 815 / 816 / 817 / 819 / 856
- Annual / Other: EIA-23L/S / 64A / 820 / 821 *(discontinued)* / 863 *(quadrennial – discontinued)*

Renewables (Non-hydro)
- EIA-63B – Photovoltaic Module Shipments (monthly/annual)

- EIA-63C – Densified Biomass Fuel Report (monthly/annual)

Uranium & Nuclear Fuel
- Quarterly: EIA-851Q (Domestic Uranium Production) / NWPA-830G (Spent-Fuel Contract Appendix G)
- Annual: EIA-851A (Uranium Production) / EIA-858 (Uranium Marketing)
- Every 5 yrs: GC-859 (Nuclear Fuel Data Survey)

Energy-Consuming End-Use Surveys
- EIA-457 (A-G) – Residential Energy Consumption Survey (RECS) *(quadrennial)*
- EIA-846 (A,B) – Manufacturing Energy Consumption Survey (MECS) *(quadrennial)*
- EIA-871A – Commercial Buildings Energy Consumption Survey (CBECS) *(quadrennial)*

Finance / Environment / Other
- EIA-28 – Financial Reporting System *(annual – suspended)*
- EIA-112 – Residential Utility Disconnections Survey *(annual)*
- EIA-1605 – Voluntary Greenhouse-Gas Reporting *(annual – suspended)*

Appendix 2 - Representative Table of Key FERC Data

MBR = Market-Based Rate
OMB = Office of Management and Budget

Dataset Name	Description	Industry	Link
Administrative Hydropower Actual	Actual (True-Up) assessment of hydropower-related charges.	Hydropower	Link
Administrative Hydropower Estimate	Estimated hydropower assessment data.	Hydropower	Link
Administrative Hydropower Fixed Charges (Actual)	Actual fixed charge assessment.	Hydropower	Link
Administrative Hydropower Fixed Charges (Estimate)	Estimated fixed charges for hydropower assessments.	Hydropower	Link
Electric Assessment Table	Summary of annual charges for electric utilities.	Electric	Link
Electric Debit/Credit Calculations	Annual charges adjustments for electric utilities.	Electric	Link
Gas Assessments	Annual charges levied on gas utilities.	Natural Gas	Link
Gas Debit/Credit Calculations	Gas utility charge adjustments.	Natural Gas	Link
Oil Assessment Table	Oil pipeline charges.	Oil	Link
Oil Debit/Credit Calculations	Debit and credit calculations for oil entities.	Oil	Link
Other Federal Agencies Cost Assessment	Cost allocation for non-FERC federal agencies.	Hydropower	Link
Information Collections – Approved	Metadata and listing of OMB-approved information collections.	Cross-sector	Link
Form 552 Master Table	Natural gas purchase/sale data summary.	Natural Gas	Link
Form 552 Page 3	Price index reporting and company list.	Natural Gas	Link
Entities to Entities	MBR affiliations (seller-entity relationships).	Electric	Link
Entities to Gen Assets	Generation assets tied to MBR sellers.	Electric	Link
Entities to PPAs	Long-term energy/capacity contracts under MBR.	Electric	Link
Entities to Vertical Assets	Assets used in vertical market power analysis.	Electric, Natural Gas	Link
Gen Assets	Generator asset information from MBR portal.	Electric	Link
Indicative Market Share Screens	Data for evaluating market share thresholds.	Electric	Link

Indicative Pivotal Supplier Screens	Data supporting pivotal supplier market analysis.	Electric	Link
MBR Authorizations	Market-Based Rate seller authorization records.	Electric	Link
MBR Category Status	Category 1 or 2 seller designation per market.	Electric	Link
MBR Mitigations	Mitigation plans imposed on MBR sellers.	Electric	Link
MBR Operating Reserves	Sellers authorized to provide ancillary services.	Electric	Link
MBR Self Limitations	Self-imposed restrictions on market-based rate authority.	Electric	Link

Appendix 3 – Metadata on Slide Deck

The data assessment slide deck and mapping identify available formats, sources, and metadata.

Available Data Categories
From the column "Data Category", the main domains include:

- Electricity
- Energy Supply
- Energy Supply & Demand

Subcategories (Data Granularity)
Under "Data Subcategory", datasets focus on:

- Electricity generation and capacity
- Plant-level generation and fuel consumption
- Generator characteristics (capacity, type)
- Utility-level electricity sales
- Energy production/consumption by state and sector

Source Organizations
Key agencies include:

- EIA (U.S. Energy Information Administration)
- EPA (Environmental Protection Agency)
- NMED (New Mexico Environment Department)
- NM OCD (Oil Conservation Division)
- BLM (Bureau of Land Management)

Public Access & Format

- Public Access: Most sources are publicly available.
- Formats: Common file types include CSV, Excel, and HTML.
- Access Links: Present for all EIA datasets.

Update Frequency

- Annual updates dominate (e.g., EIA 860, 861, SEDS).
- Some datasets update monthly (e.g., EIA 923, NM OCD).
- A few have variable or multi-frequency (e.g., profiles annually, tables monthly).

Metadata & Documentation

- Most datasets provide detailed documentation and metadata:
  - Tech notes
  - Methodology guides
  - Embedded file metadata
- Quality ranges from “Good” to “High”, often based on mandatory reporting.

Appendix 4 - EPA GHGRP Subparts (40 CFR Part 98)

General & Cross-Cutting

- **Subpart A – General Provisions:** Applicability, calculation methods, reporting requirements.
- **Subpart B – General Provisions for Monitoring Methods:** Additional requirements on monitoring.
- **Subpart C – General Stationary Fuel Combustion Sources:** CO_2, CH_4, and N_2O from boilers, turbines, engines, etc.

Oil, Gas, and Coal

- **Subpart D – Electricity Generation:** Emissions from electric generating units.
- **Subpart DD – Suppliers of Carbon Dioxide:** Reporting CO_2 supplied, imported, or exported.
- **Subpart PP – Suppliers of CO_2:** Specific requirements for CO_2 suppliers.
- **Subpart NN – Suppliers of Natural Gas and NGLs.**
- **Subpart W – Petroleum and Natural Gas Systems:** Covers production, processing, transmission, storage, LNG import/export, distribution.
- **Subpart Y – Petroleum Refineries.**
- **Subpart FF – Underground Coal Mines.**

Mineral & Metals

- **Subpart E – Adipic Acid Production.**
- **Subpart F – Aluminum Production.**
- **Subpart G – Ammonia Manufacturing.**
- **Subpart H – Cement Production.**
- **Subpart I – Electronics Manufacturing.**
- **Subpart J – Ethanol Production.**
- **Subpart N – Glass Production.**
- **Subpart O – HCFC-22 Production / HFC Destruction.**
- **Subpart P – Hydrogen Production.**
- **Subpart Q – Iron and Steel Production.**
- **Subpart R – Lead Production.**
- **Subpart S – Lime Manufacturing.**
- **Subpart T – Magnesium Production.**
- **Subpart U – Miscellaneous Uses of Carbonates.**
- **Subpart V – Nitric Acid Production.**
- **Subpart X – Petrochemical Production.**
- **Subpart Z – Phosphoric Acid Production.**
- **Subpart AA – Pulp and Paper Manufacturing.**
- **Subpart BB – Silicon Carbide Production.**
- **Subpart CC – Soda Ash Production.**
- **Subpart EE – Titanium Dioxide Production.**
- **Subpart GG – Zinc Production.**

Waste & Industrial Gases

- **Subpart HH – Municipal Solid Waste Landfills.**
- **Subpart II – Industrial Wastewater Treatment.**

- **Subpart JJ – Manure Management.**
- **Subpart KK – Suppliers of Industrial Greenhouse Gases.**
- **Subpart LL – Suppliers of Coal-based Liquids.**

Carbon Capture & Sequestration

- **Subpart RR – Geologic Sequestration of CO_2:** Requires MRV plan, monitoring, and reporting.
- **Subpart UU – Injection of CO_2:** Covers CO_2 injection for purposes other than geologic sequestration (e.g., EOR), without MRV plan.

Appendix 5 - Estimated Data Volume by EIA Form from ChatGPT

The estimated data volumes were calculated by combining the reporting frequency of each EIA form with approximate file sizes derived from typical datasets. For monthly and annual forms, yearly volumes were extrapolated from average file sizes per report, while weekly and hourly datasets were scaled based on the number of reporting periods per year. This approach provides an order-of-magnitude assessment of storage needs rather than precise totals.

This estimate reflects only EIA data; additional datasets will need to be incorporated over time, but this provides a necessary starting point.

Form #	Description	Frequency	Est. Annual Volume	10-Year Archive Est.
EIA-782C	Prime Supplier Sales	Monthly	~10 MB/year	~100 MB
EIA-800	Weekly Refinery Report	Weekly (~52)	~1 MB/week × 52 = 52 MB	~500 MB
EIA-810	Monthly Refinery Report	Monthly	~5 MB/year	~50 MB
EIA-815	Bulk Terminal/Blender	Monthly	~3 MB/year	~30 MB
EIA-816	NGL Production	Monthly	~2 MB/year	~20 MB
EIA-877	Heating Oil & Propane Prices	Weekly (seasonal)	~0.5 MB/week × 25 = ~12 MB	~100 MB
EIA-861	Electric Utility Sales	Annual	~10 MB/year	~100 MB
EIA-860	Generator Attributes	Annual	~5 MB/year	~50 MB
EIA-923	Power Plant Ops	Monthly/Annual	~20 MB/year	~200 MB
EIA-930	Hourly Grid Operations	Hourly	~40 MB/day × 365 = ~14.6 GB	**~150 GB**
EIA-176	Natural Gas Supply	Annual	~3 MB/year	~30 MB
EIA-757	Gas Processing Plants	Annual & periodic	~1 MB/year	~10 MB