DOCUMENTATION

Energy Data Documentation

PREPARED BY: Whitney Purdum

CONTENTS

1. Overview

1.1 Purpose

This dataset is a **curated subset of energy data derived from The Ohio State University's Energy Research Data Hub**, prepared specifically for use in this hackathon.

The purpose of this dataset is to provide participants with access to realistic, high-quality energy, building, and weather data to support **exploratory analysis, data processing, and applied analytics or AI experimentation** within a limited time frame.

The dataset includes:

- Smart meter data from a subset of campus meters, capturing delivered energy and demand for multiple utility types (including electricity, steam, heat, gas, and cooling-related services)
- Building-level metadata from Ohio State's Space Information and Management System (SIMS), providing physical and geographic context
- Historical weather data aligned temporally with energy measurements

This curated subset is intended to reflect the structure and complexity of real-world campus energy data while remaining accessible for learning and rapid prototyping. It is not intended to represent the full scope or completeness of the Energy Research Data Hub.

1.2 Data Use Guidelines

Because portions of the datasets provided for this challenge are internal to The Ohio State University, **all data access and use is limited to participation in this hackathon only**.

By participating in this challenge, teams agree that:

- The provided data may be used **solely for analysis, prototyping, and presentation within the scope and duration of the hackathon**.
- The data may **not** be copied, redistributed, published, or used for external research, commercial purposes, or projects outside of this event.
- Any demos, notebooks, repositories, or presentations shared publicly **must not include raw data or information that could reasonably be used to reconstruct the dataset**.

Participants are encouraged to showcase **methods, models, visualizations, and insights**, but not the underlying data itself. The intent of this guideline is to support learning and experimentation while respecting institutional data stewardship responsibilities (**https://it.osu.edu/data/institutional-data-policy**).

2. Available Data Sets

2.1 Dataset Overview

1. **Curated (flattened) smart meter data**
2. **SIMS building data**

3. **Historical weather data**

Together, these datasets provide a comprehensive view of building performance and utilization. When combined, they can reveal patterns that inform **maintenance, facility planning, staffing, energy management, and cost-reduction decisions**.

Data set	Source	Interval	Update frequency	Date range available
Smart meter - curated	Flattened smart meter data	1 hour	Daily	01/01/2025 – 12/31/2025
SIMS building (limited)	**SIMS Space API**	N/A	Once weekly	N/A
Weather	**https://open-meteo.com/en/docs/historical-forecast-api**	hourly	Daily	01/01/2025 – 12/31/2025

2.2 Data Set Documentation

SMART METER DATA

Source: Raw meter data

Format: CSV

Description: Each record includes the meter reading value along with the time window in which the reading occurred. The `simsCode` field can be used to join to a site's SIMS information, enabling researchers to link meter readings with building metadata.

Sample JSON return


```
{
    "meterId": 246014,
    "siteName": "East Regional Chilled Water Plant",
    "simsCode": "376",
    "utility": "ELECTRICITY",
    "readingTime": "2025-01-04T05:00:00",
    "readingValue": 151.05446750165862,
    "readingUnits": "kWh",
    "readingUnitsDisplay": "Kilowatt hour",
    "readingWindowStart": "2025-01-04T05:00:00",
    "readingWindowEnd": "2025-01-05T04:45:00",
    "expectedWindowReadings": 96,
    "totalWindowReadings": 96,
    "missingWindowReadings": 0,
    "filteredWindowReadings": 0,
    "readingWindowSum": 14142.396919778323,
    "readingWindowMin": 88.3088609341783,
```

```
  "readingWindowMinTime": "2025-01-05T03:30:00",
  "readingWindowMax": 157.55109096750593,
  "readingWindowMaxTime": "2025-01-05T01:45:00",
  "readingWindowStandardDeviation": 10.561087028350984,
  "readingWindowMean": 147.3166345810242
}
```


Data Dictionary

Key	Description	Data Type	Sample Value
meterId	Unique numeric identifier for the meter	integer	246073
siteName	Human-readable site name where the meter is located	string	"11th Ave, 33 W"
simsCode	Internal SIMS (Space Information and Management System) building code	integer	193
utility	Type of utility measured	string	"ELECTRICITY"
readingTimeEpoch	Timestamp of the reading in Unix epoch milliseconds	integer	1672574400000
readingValue	Recorded consumption or demand value	float	8.186482615928398
readingUnits	Standardized unit of measurement	string	"kWh"
readingUnitsDisplay	Display-friendly unit name	string	"Kilowatt hour"
readingWindowStartEpoch	Start time of the reading window	integer	1451624400000
readingWindowEndEpoch	End time of the reading window	integer	1451624400000
expectedWindowReadings	Number of readings expected for the window	integer	96
totalWindowReadings	Number of readings actually received	integer	0
missingWindowReadings	Number of readings missing from the window	integer	96
filteredWindowReadings	Number of readings excluded due to quality filters	integer	0
sumWindowReadings	Sum of all readings in the window	float	0.0

minWindowReading	Minimum recorded value in the window	float	112
maxWindowReading	Maximum recorded value in the window	float	12
windowStandardDeviation	Standard deviation of readings in the window	float	12

SIMS BUILDING

Source: SIMS Space API

Format: CSV

Description: Provides building-level metadata that is maintained by Facilities Information and Technology Services.

Sample JSON Return


```
{
        "buildingNumber": "311",
        "buildingName": "Mount Hall (0311)",
        "campusName": "Columbus",
        "address": "1050 Carmack Rd",
        "city": "Columbus",
        "state": "OH",
        "postalCode": "43210-1002",
        "county": "Franklin",
        "frameworkDistrict": "Western Lands",
        "geography": "Columbus Contiguous",
        "formalName": "Mount, John T. Hall",
        "alsoKnownAs": null,
        "schedulingAbbreviation": "MO",
        "grossArea": 75660.0,
        "floorsAboveGround": "2",
        "floorsBelowGround": "1",
        "constructionDate": "1974-07-01",
        "latitude": "40.00405648",
        "longitude": "-83.0367706"
    }
]
```

WEATHER

Source: https://open-meteo.com/en/docs/historical-forecast-api
Format: CSV

Description: Weather attributes by date

Data Dictionary

Key	Description	Data Type	Units/format	Example Value
date	Timestamp of observation in UTC	datetime	ISO 8601	2025-08-10 04:00:00+00:00
latitude	Latitude coordinate for location of weather observation (will always be the same – location of Ohio State University – main campus)	float	lat/long	40.08
longitude	Longitude coordinate for location of weather observation (will always be the same – location of Ohio State University – main campus)	float	lat/long	-83.06
temperature_2m	Air temperature at 2 meters above ground	float	°F	72.8798
dew_point_2m	Dew point temperature at 2 meters above ground	float	°F	64.871765
relative_humidity_2m	Relative humidity at 2 meters above ground	integer	%	76
precipitation	Total precipitation during the hour	float	mm or inches*	0
direct_radiation	Direct solar radiation	float	W/m^2	0
wind_speed_10m	Wind speed at 10 meters above ground	float	mph	4.297137
wind_speed_80m	Wind speed at 80 meters above ground	float	mph	12.164427
wind_direction_10m	Wind direction at 10 meters above ground	float	degrees	141.34016
wind_direction_80m	Wind direction at 80 meters above ground	float	degrees	147.77127
cloud_cover	Fraction of the sky covered by clouds	integer	%	0

apparent_temperature	Feels-like temperature based on air temp, humidity, and wind speed	float	°F	76.52037
shortwave_radiation		Big int		77
direct_radiation		double		36.0
diffuse_radiation		double		158
direct_normal_irradiance		double		16.9898

2.3 Utilities Overview

The smart meter dataset includes measurements for multiple **utility types**, each representing a different form of energy or thermal service delivered to campus buildings or systems. Utilities should generally be analyzed **separately**, unless teams explicitly convert or normalize values using appropriate engineering assumptions.

Not all buildings have all utilities, and the presence of a meter does not imply exclusive service to a single building or piece of equipment.

UTILITY DESCRIPTIONS

Utility	Description	Typical Interpretation
ELECTRICITY	Electrical energy consumption	Lighting, plug loads, motors, equipment, and building systems
ELECTRICAL_POWER	Instantaneous electrical demand	Real-time power draw (kW), often used for peak demand analysis
GAS	Natural gas consumption	Heating, hot water, cooking, or process loads
HEAT	Thermal energy delivered for heating	Hydronic or district heating systems
STEAM	Thermal energy delivered as steam	Space heating, hot water, or process uses
STEAMRATE	Steam flow rate	Instantaneous steam delivery rather than total energy
COOLING	Thermal energy delivered for cooling	Chilled water used for space cooling
COOLING_POWER	Instantaneous cooling demand	Real-time cooling load rather than total cooling energy
OIL28SEC	Fuel oil consumption	Legacy or backup heating systems

IMPORTANT INTERPRETATION NOTES

- **Utilities are not directly comparable** without appropriate unit conversion or normalization. For example, electricity (kWh) should not be summed with steam or cooling energy without justification.
- Some utilities represent **energy over time** (e.g., ELECTRICITY, STEAM, HEAT), while others represent **instantaneous demand or flow** (e.g., ELECTRICAL_POWER, STEAMRATE, COOLING_POWER).
- Meter data may reflect **delivered energy from centralized campus systems**, not on-site generation or individual equipment performance.
- Buildings may be served by **district energy systems**, and a single meter may represent aggregated service rather than a single end use.

Teams are encouraged to clearly state assumptions when comparing or combining utilities and to focus analyses on **consistent utility types** unless conversions are explicitly justified.

RECOMMENDED BEST PRACTICES (OPTIONAL GUIDANCE)

- Analyze **one utility at a time** when exploring trends or comparisons.
- Normalize energy-based utilities (e.g., ELECTRICITY, HEAT) by square footage or time where appropriate.
- Use power-based utilities (e.g., ELECTRICAL_POWER, COOLING_POWER) for **peak demand or variability analysis**, not total consumption.
- Clearly distinguish between **energy** and **power** in visualizations and interpretations

3. Schema joins

See next page.

Schema Join Map

Join Category	Table (db.table) A	Join key (A)	Join Type	Athena table (db.table) B	Join key (B)	Transform Required
Building metadata → Meter readings	Building_metadata	buildingNumber	direct	meter_data	simscode	None
Weather → Meter readings by time	weather_data	date	direct/partial	meter_data	readingtime	None if full date; SPLIT if partial date (e.g. day)