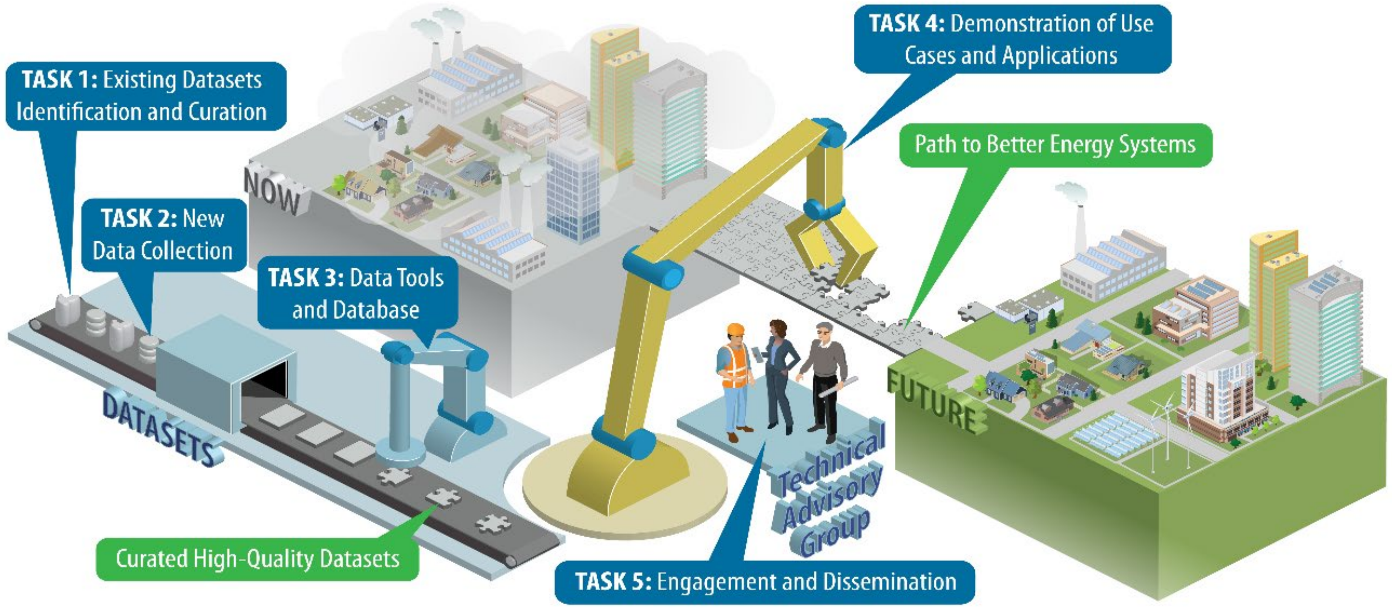
Benchmark Dataset Development and Applications



4-lab project team, Feb 2020, at NREL

LBNL
Tianzhen Hong
thong@lbl.gov

NREL
Lieko Earle
lieko.earle@nrel.gov

ORNL
Piljae Im
imp1@ornl.gov

PNNL
Vikas Chandan
vikas.chandan@pnnl.gov

Project Summary

Timeline:

Start date: 10/1/2019

Planned end date: 9/30/2022

Key Milestones

1. Data portal launched: 3/30/2020
2. 6 datasets curated and published: 3/31/2021
3. 2 case studies completed: 6/30/2021

Budget:

Total Project $ to Date:

- DOE: $1.98M
- Cost Share: N/A

Total Project $:

- DOE: $3.75M
- Cost Share: N/A

Key Partners:

LBNL	NREL
ORNL	PNNL

Project Outcome:

The focus of this project is to collect and curate a handful of high-resolution building datasets that have broad applicability to address highest-impact use cases. Main outcomes include:

- 12 high-quality and well-curated building datasets
- Enhanced data tools for metadata representation
- Demonstration of applications of the curated datasets
- A workflow to semi-automate the data curation process for scale up and industry adoption

Challenges

Problem Definition:

Reducing energy use in buildings and optimizing building operations require access to a diverse and integrated set of data. However, it is resource-intensive and often hard to find datasets with adequate data coverage, good data quality, and clear documentation to support research on energy efficiency and demand flexibility technologies for grid-interactive efficient buildings (GEBs).

Therefore, there is a strong need to:

- Collect, curate, and make publicly available high-resolution and high quality measured data from residential and commercial buildings that have broad applicability to a variety of high-impact use cases.
- Determine the level of data resolution required for most effectively optimizing building operations through advances in data analytics and control technologies.
- Provide a common, high-quality benchmark against which competing algorithms can be fairly

Impacts

- The curated datasets are made open access through a data portal to support data-driven research for grid-interactive efficient buildings (GEB) to save energy and improve operations and controls.
- The datasets can be employed for load forecasting and baselining, virtual sensing, building energy modeling, building performance benchmarking (at the whole building and system levels), and non-intrusive load monitoring.
- New and enhanced data tools to represent metadata supporting FAIR (Findable, Accessible, Interoperable and Reusable) principles.
- The project develops and demonstrates a semi-automatic workflow to curate datasets, which enables data contribution from other parties after the project is completed.
- A set of use cases demonstrating how users can utilize these datasets for their unique needs.

The BTO Sensors and Controls Sub-Program website acknowledges the role of curated datasets in meeting its objectives:

*"To complement the portfolio of competitive and interdisciplinary, early-stage R&D projects, testing frameworks and procedures, along with **curated datasets** are being developed by the sub-program to support the R&D community at-large through common baselines to evaluate performance improvements and encourage additional innovation in advanced analytics and control strategies"*

Key Activity 1 - Use Cases Identification & Prioritization

We developed a process to identify and prioritize use cases that would benefit from improved access to high-quality real-world building datasets:

- Identification and refinement of prospective use cases
- Development of a use case scoring method based on the potential benefits of each use case
- An iterative scoring process to prioritize the use cases

We collected use case suggestions from each of the laboratories in the project team. This yielded an initial list of 15 prospective use cases.

Example

UC#1
Design, size, or estimate the performance of demand management technologies (e.g., flexible loads, energy storage, supervisory controls) for different applications.

Map to Dataset Needs

1	Energy use data
2	**Indoor environmental data**
3	Outdoor environmental data
4	System and equipment operational data
5	Control setting and logic data
6	Occupant data
7	Design basis data
8	Building and system asset data
9	Utility and grid signal data
10	Onsite energy generation data
11	Cyber (IoT) device data
12	Dispatchable energy storage data

Thermal	air temperature, humidity, air velocity, barometric pressure, etc.
Visual	illuminances, glare, etc.
Air quality	CO2 concentration, PM2.5, VOC, etc.
Acoustic	noise level, sound level, etc.
Others	

Framework to describe the data needed to address any particular use case in terms of "bundled" categories.

Leverage the fact that data collection efforts typically contain several categories of data that are collected together.

Key Activity 1 - Use Cases Identification & Prioritization

List of 15 use cases (highlighted 5 high-priority use cases)

1 Design, size, or estimate the performance of demand management technologies (e.g., flexible loads, energy storage, supervisory controls) for different applications. Demand management technologies	**2** Validate methods for modeling how the electrical load of an aggregation of buildings is affected by building-sited energy efficiency or demand management technologies (e.g., HVAC controls, energy storage). Model validation for aggregated Bldgs	3 Evaluate tradeoffs between different levels of weather data collection (e.g., on-site measurement vs. data from miles away) for field studies of weather-sensitive technologies (e.g., space cooling, space heating, thermal energy storage, solar + storage). Sensitivity study on weather data	4 Validate methods for generating equipment schedule inputs for building modeling applications in which load-shaping outcomes are sensitive to the timing of occupant-driven loads (e.g., plug loads). Method of generating operational schedule for building energy models	**5** Use a combination of building time-series data (e.g., electrical load), measured weather data, and forecast data to evaluate the ability of predictive controls to improve the load-shaping capability of building-sited technologies. Development and validation of predictive controls
6 Develop and validate methods (statistical or machine learning) for using indirect observations or proxy data (e.g., WiFi data) to infer occupant count, which can be used in model predictive control Occupant count prediction	**7** Determine potential savings and grid services for smart- zoned residential HVAC systems with room-level zoning and control based on room-by-room occupancy. Smart-zoned residential HVAC control	8 Understand and quantify human-building interactions to improve building design and operation. (Example: the selection of smart thermostats and the design of operable windows.) Human building interaction	**9** Fault detection and diagnostics to identify faulty equipment and control, as well as evaluate their impact on energy. Fault detection and diagnostics	10 Evaluation of methodologies for development and calibration for different types of building energy models (e.g., blackbox, grey box, high fidelity models, etc.) to understand building operation, controls and performance Calibration of building energy models
use case #11 from original list was merged with use case #5	12 Optimization of occupant comfort, health, and safety (IAQ) and minimize cost of operations and/or energy consumption at the same time. Co-optimizing IEQ and energy	13 Ensure cyber-secure operation of energy systems in buildings. Identify, diagnose and isolate cyber-attacks on a building control system. Cyber security of building operations	14 Develop system-level (e.g., lighting, HVAC) key performance indicators for energy efficiency benchmarking and performance diagnostics. System level performance	15 Determine the real-world frequency of energy mismanagement by occupants and the energy savings/load flexibility potential of preventing mismanagement Identification and prevention of energy mismanagement

Key Activity 2 - New Data Collection Efforts

- **Objective**: To design experiments for data collection and develop sensor and instrumentation plan, based on the dataset needs and any existing instrumentation in the buildings
- **Four buildings (sites) were down selected**: From the **15 potential sites** that were identified to cover various **DNs (Data needs)** and **UCs (Use cases)** as well as different building types and climate zones

Flexible research platforms - ORNL

Building 59 - LBNL

Lab Homes - PNNL

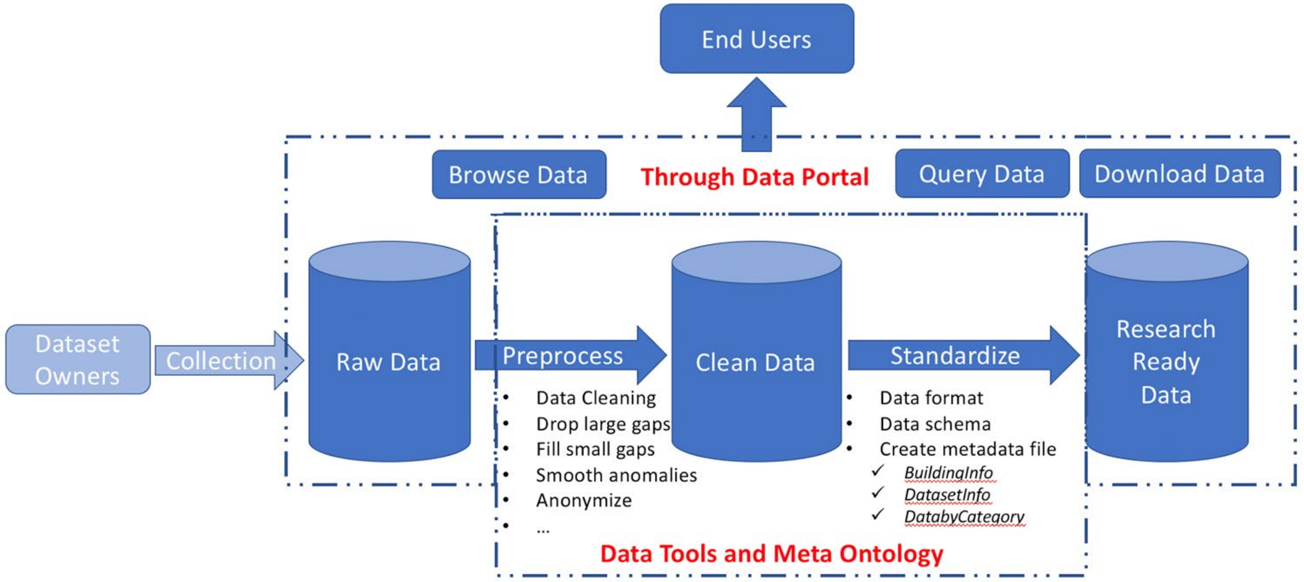
Bear Creek High School - NREL

Key Activity 2 - New Data Collection Efforts - Buildings

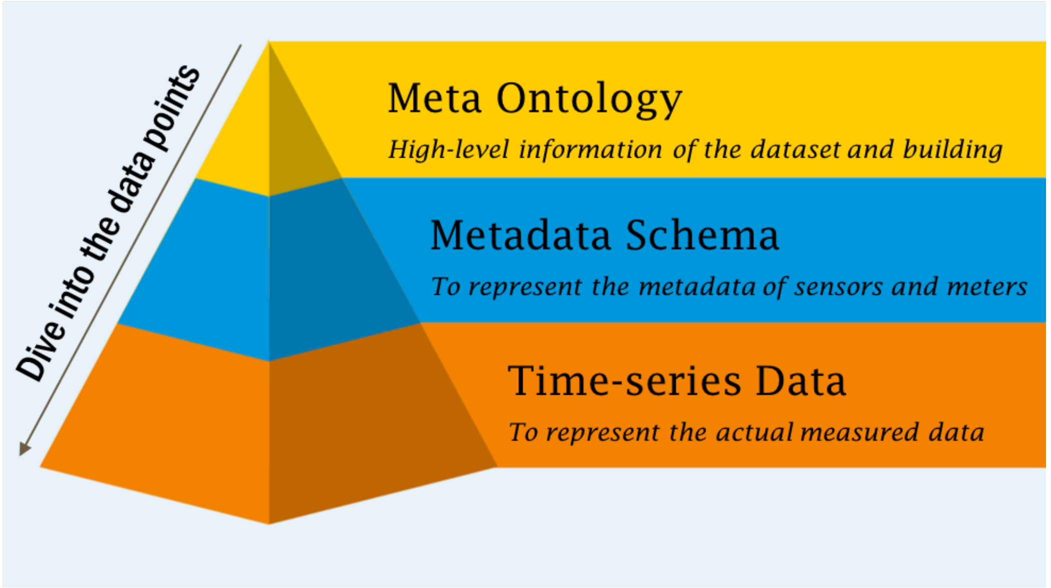
Building	Flexible research platforms (ORNL)	Building 59 (LBNL)	Lab Homes (PNNL)	Bear Creek High School (NREL)
Building type	Medium office (Commercial)	Medium office (Commercial)	Residential	Secondary school
Climate zone	4A	3C	5A	5B
Location	Oak Ridge, TN	Berkeley, CA	Richland, WA	Lakewood, CO
Data Needs	1,2,3,6,11,12,13	1,2,3,6,11,12,13	1, 2, 3, 4, 7, 12, 13	1,2,3,6,7,8,11,12,13
Use Cases	1, 5, 9	1,2,5,6,12,13	1, 2, 5	1,5,6,10,12,14,15
Expected test/data period	2021 summer to 2021 winter	2019 to 2021	2020 to 2021	2021 to 2022
Test methods	Normal operation; Thermostat Setback; Pre-cooling; Pre-heating		Normal operation;	

Key Activity 3 - Data Curation and Organization

- Curation of 6 datasets covering diverse building types, sizes, climate zones and use cases
- Three major steps of data curation (based on FAIR principles)
  - Data preprocessing/cleaning to reduce missing rates and outlier rates
  - Data standardization (metadata of dataset, sensor and meters)



Workflow diagram of the data curation process

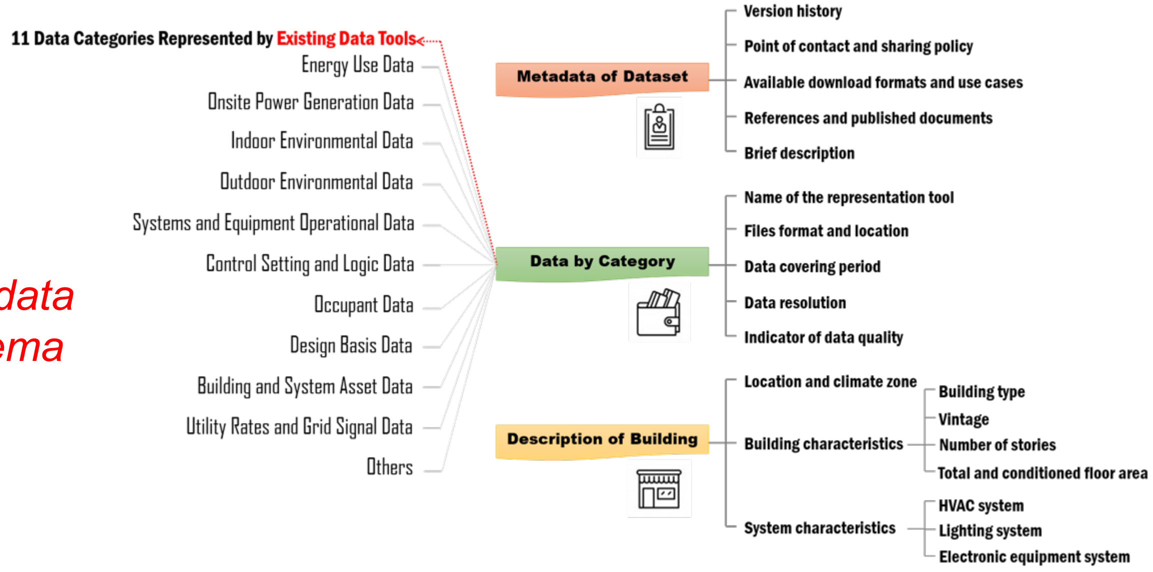


Three layers of data representation for the research-ready data

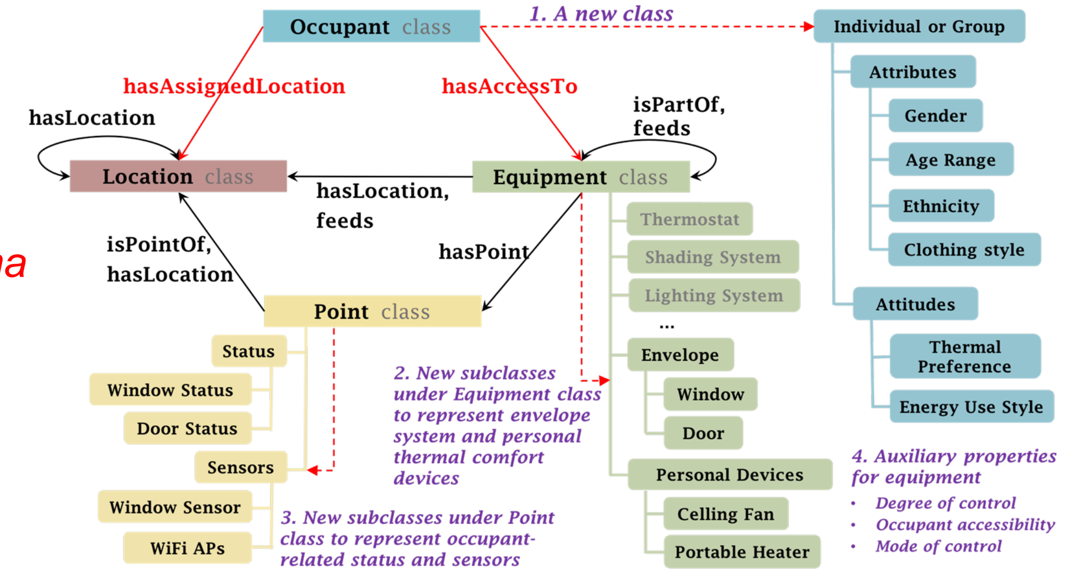
Key Activity 3 - Development and Enhancement of Data Tools

- A review of 24 existing data tools to identify their features and limitations
- A structured "readme" .json file to provide the high-level contextual information of the dataset, and to illustrate application perspectives
- Brick schema extension for representing occupant metadata

Metadata Schema



Brick schema extension



Key Activity 4 - Data Portal Development and Dataset Publishing

Search

Search Bar

FILTERS

State
- CA 1
- CO 1
- MD 1
- PA 1
- TN 1
- WA 1

Climate Zone
- 4A 3
- 5B 2
- 3C 1

Building Type
- Medium Office 2
- Campus 1
- Mixed-used Commercial 1
- Single Family Residential 1
- Small Office 1

HVAC System Type
- VAV with reheat coils 2
- Air Source Heat Pump with one dehumidifier 1
- Underfloor air distribution system with fan-powered terminal reheat coils for perimeter zones. Four roof-top units with VAV located on the roof. 1
- VRF with AHUs for dehumidification 1
- Various (hot/chilled water, ground source hear pumps, fan coil unites and other unitary equipment) 1

Fuel Type
- Natural gas 3
- Electricity 1

Category
- Indoor environmental 6
- Systems and equipment operational 6
- Energy use 5
- Outdoor environmental 5
- Occupant 3
- Control setting and logic 2
- Basis of design 1
- Building and system asset 1
- Onsite power generation 1

Floor Area (sq. ft)

4,165 1,100,000

Year Built

1925 2015

6 DATASETS (page 1 of 1)

Friends Center Building, Philadelphia, Pennsylvania

1 Files • 19 MB

The data from the Friends Center Building in Philadelphia, Pennsylvania, come from a one-year study of 24 U.S. office occupants. These office occupants recorded a set of possible drivers of personal comfort and behavior over time by completing online daily surveys, which provided information on individual thermal comfort, preference, and behavior information with datalogger readings of the occupan... show more

Lawrence Berkeley National Laboratory Building 59, Berkeley, California

2 Files • 434 MB

The Building 59 at the Lawrence Berkeley National Laboratory is a medium-sized office building with two office floors, one mechanical equipment floor, and one floor for NERSC data and computing center. The dataset only covers the two office floors. The building management system monitors and archives building-level electricity usage, HVAC and lighting system states (e.g., setpoint, temperature, fl... show more

National Renewable Energy Laboratory, South Table Mountain Campus, Golden, Colorado

4 Files • 8 MB

The National Renewable Energy Laboratory (NREL) South Table Mountain campus, located in Golden, Colorado, has multiyear datasets from buildings with approximately 1,100,000 sq ft of floor area. These 16 buildings include a large office, nine laboratories, and other education or public assembly facilities such as an education center, a warehouse, quick service restaurants, two site entrance securit... show more

Net-Zero Energy Residential Test Facility, Gaithersburg, Maryland

2 Files • 1 GB

The National Institute of Standards and Technology Engineering Lab operates the Net-Zero Energy Residential Test Facility in Gaithersburg Maryland. This unique laboratory is a net-zero energy home that produces at least as much energy as it consumes in one year. The facility has sensors and instrumentation on nearly every facet, from energy consumption to comfort.

Systems Engineering Building, Richland, Washington

2 Files • 2 GB

This dataset includes processed and raw data from the building management system for the Systems Engineering Building (SEB) located on the Pacific Northwest National Laboratory campus in Richland, Washington. SEB is a Leadership in Energy and Environmental Design (LEED) gold-certified facility. It is both a regular office building, and a resource for research in grid visibility, control, and resil... show more

Vanderbilt Alumni Hall, Nashville, Tennessee

1 Files • 22 MB

This dataset from the Vanderbilt Alumni Hall Building (VAH) on the University of Vanderbilt campus in Nashville, Tennessee, includes processed and raw data. A gold-certified Leadership in Energy and Environmental Design (LEED) building, VAH is a mixed-use commercial building that contains classrooms, meeting rooms, office rooms, conference rooms, an exercise room, and others. The dataset spans one... show more

Search facets customized for Buildings

Benchmark Datasets for Buildings

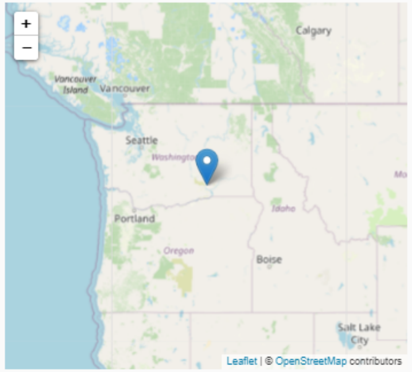
Systems Engineering Building, Richland, Washington

Summary

BROWSE DATA

Description

This dataset includes processed and raw data from the building management system for the Systems Engineering Building (SEB) located on the Pacific Northwest National Laboratory campus in Richland, Washington. SEB is a Leadership in Energy and Environmental Design (LEED) gold-certified facility. It is both a regular office building, and a resource for research in grid visibility, control, and resiliency—housing an Electricity Infrastructure Operations Center, a Buildings Operations Control Center, a Power Electronics Lab, and an Interoperability Lab. This SEB dataset spans November 2018 to May 2020 and includes time-series measurements corresponding to supply and return temperatures of air and water; air, hot water, and cold water flow rates; energy and power consumption; set points; and more in the form of CSV files. Also included are a metadata .json file, and a .ttl file to visualize the data according to BRICK schema.



50 Downloads

72 Files Downloaded

49.18 GB Downloaded

Building Characteristics

Location
Richland, WA 99354

Building Type
Small Office

IECC Climate Zone
5B

Total floor area (sq. ft)
24400

System Characteristics

HVAC System Type
VAV with reheat coils

Fuel Type
Natural gas

Data Details

Data Files Description

The raw .zip file includes a set of .CSV files containing time-series data, one file per measurement, as available from the building management system. The processed .zip file includes a metadata .json file, a .ttl file to visualize the data according to BRICK schema, and a .csv file containing all the time-series measurements after cleaning and curation.

Select SEB Files

Data Level	File Size
processed	111.91 MB
raw	1.44 GB

2 files selected (1.55 GB)

How would you like to download the data?

Cancel | Zip | Script

Dataset download options

Dataset information customized for Buildings

building management system. The processed .zip file includes a metadata .json file, a .ttl file to visualize the data according to BRICK schema, and a .csv file containing all the time-series measurements after cleaning and curation.

Data Begin
2018-11-30

Data End
2020-05-06

Categories

Systems and equipment operational · Energy use · Indoor environmental · Outdoor environmental

Data Formats
CSV

License
https://creativecommons.org/licenses/by/4.0/

Digital Object Identifier (DOI)

Processed Data: 10.25584/sebprocessed/1729908

Raw Data: 10.25584/data.2019-03.709/1509487

References

- Datahub
- Building description

Contacts

Name	Role
Vikas Chandan	Creator
Eddie Schuman	Curator
Jeremy Lerond	Curator

SEND MESSAGE

Data Files

Description	Data Access Method
SEB Data	Multi-Download

Key Activity 4 - Data Portal Development and Dataset Publishing

Link: https://bbd.labworks.org/

Registration required to access all features

Key Current Features

- Currently has **6 datasets** from across the US representing multiple climate zones, building types, HVAC system types and footprint (more to be added in FY22)
- Raw as well as processed data available
- Processed data includes
  - **Cleaned data** with gaps filled and outliers removed using ML techniques, with documentation of techniques
  - Data represented using a **custom-built metadata** schema based on extensive literature review
  - System and layout information with hierarchies represented using **BRICK** schema
- Features developed using scoring based prioritization matrix
- Easy navigation using **search facets** customized for the building domain
- **Usage metrics** are tracked

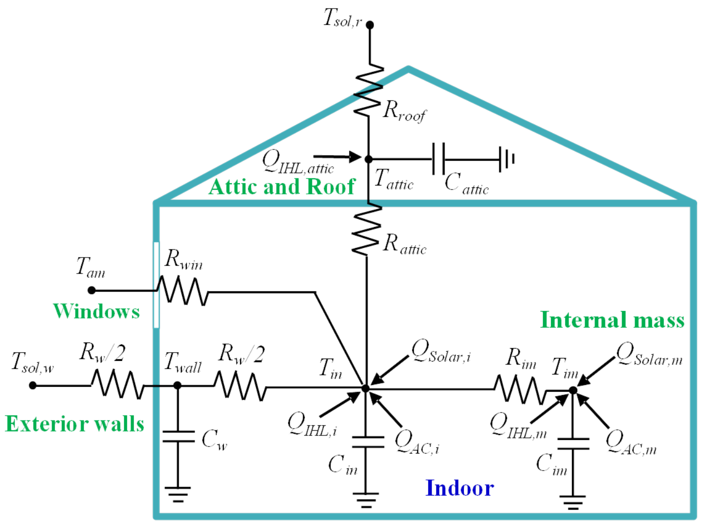
Platform

- Built on existing **mature data management capabilities at PNNL** that support atmospheric, wind, materials discovery, marine energy and other EERE domains
- Built on **FAIR** principles - findability, accessibility, interoperability, and reusability
- Content is driven by team members
- **Scalable**, easy to maintain and deploy
- The attributes maintained on a github repository by a few members closely aware of domain and dataset
- Future capabilities being developed
  - A data processing pipeline that **automatically** ingests data; cleans, curates, and standardizes data based on the metadata schema developed in the project; and provides data access via a "data lake" that can allow users to run custom applications (such as machine learning algorithms)
  - Support for **real-time data streaming** from a data source such as Building Management System, depending on the needs of the users
  - **Enhanced user insights** into time-series data, such as filtering by sensor types and time span without the need to download the entire dataset

Case Study 1 - Model Development and Calibration

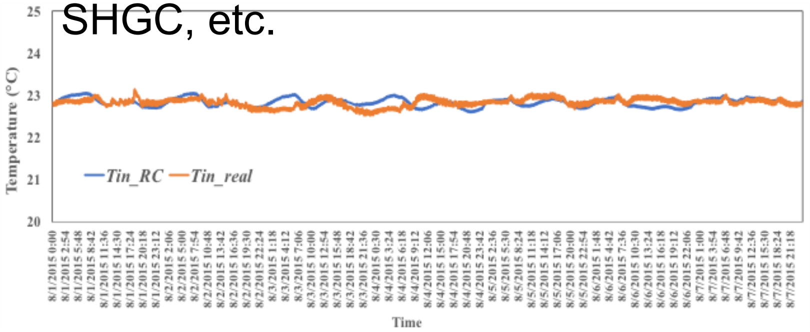
NIST Net Zero Energy Residential Test Facility (NZERTF)

Dataset: https://bbd.labworks.org/ds/bbd/netzeromd

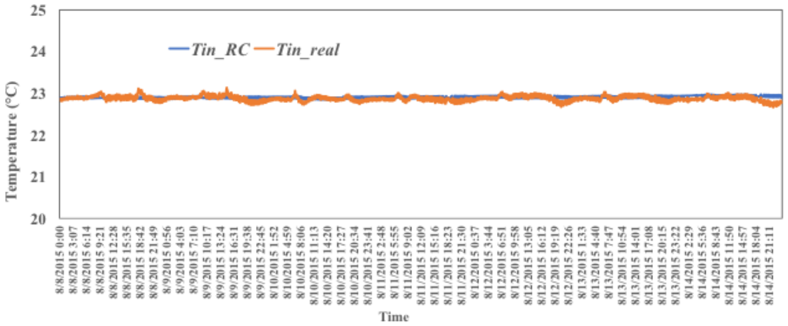


Schematics of the simplified building thermal network model

- **Development of a control-oriented building model using the data from NIST Net Zero Energy Residential Test Facility (NZERTF)**
  - Development of a gray box model (i.e., RC model), an ideal tool for optimal control design such as Model Predictive Control (MPC)
  - Prediction of average indoor air temperature (Tin) using the model
  - Model was developed based on available data and best guess for some building characteristics such as roof area, solar absorption rates of envelope and roof, SHGC, etc.



Training results of the RC model



Validation results of the RC model

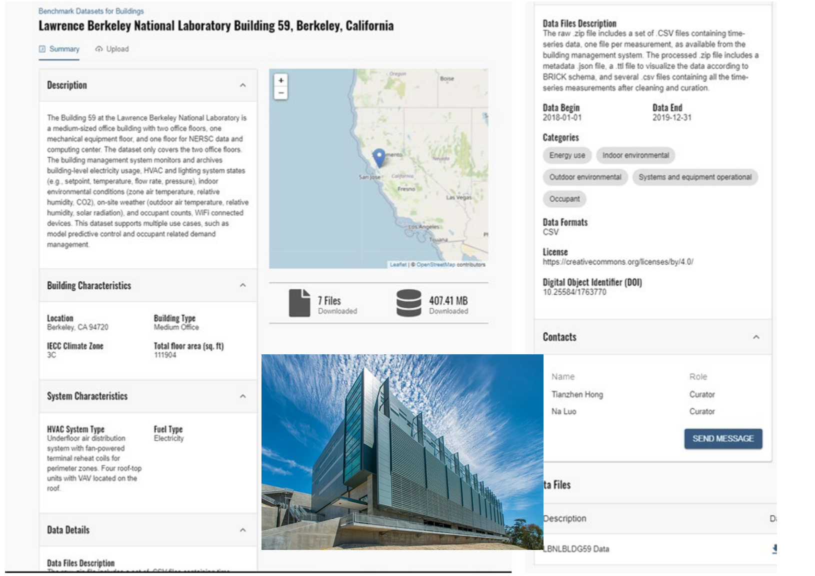
Model	Time	Training/validation	RMSE
Initial	Aug 1 to Aug 7, 2015	Training	0.31°C
	Aug 8 to Aug 14, 2015	Validation	0.95°C
Improved	Aug 1 to Aug 7, 2015	Training	0.13°C
	Aug 8 to Aug 14, 2015	Validation	0.12°C

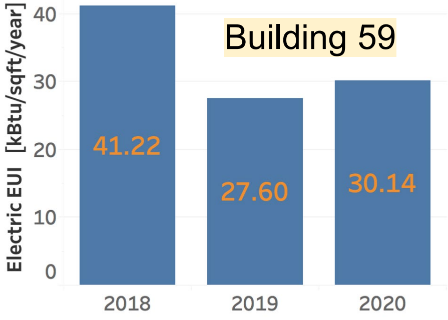
Model accuracy performance

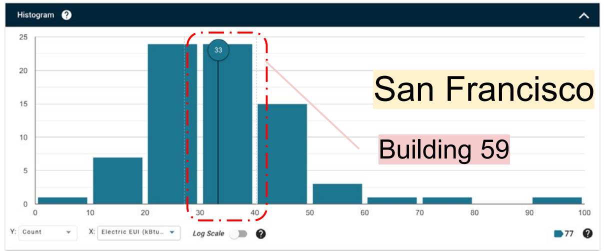
Case Study 2: Energy and Occupancy Analytics

- **Analysis on 3 research questions using the Building 59 dataset**
  - Building energy benchmarking and load shape analysis
  - Prediction of occupant count and its correlation with energy use
  - Impacts of COVID pandemic on occupancy and building performance

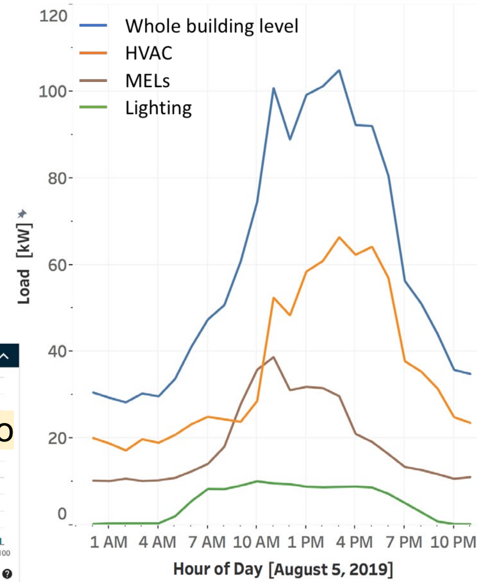
Dataset: https://bbd.labworks.org/ds/bbd/lbnlbldg59



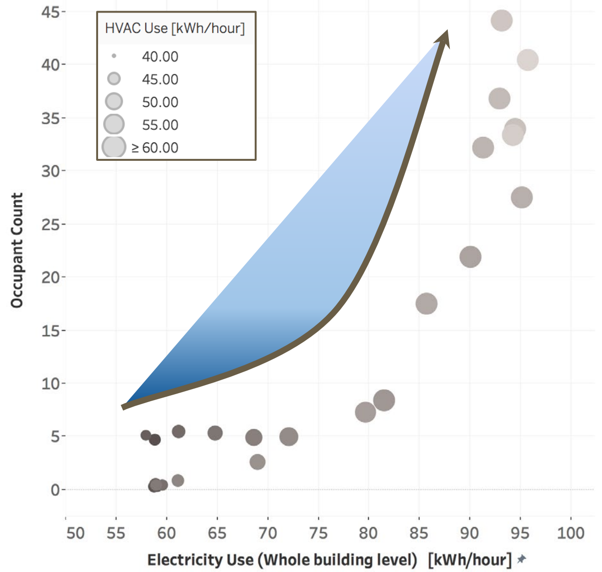




Load Shape - August 5, 2019



Electricity Use vs. Occupant Count (May-July, 2018)



Energy Use - end uses [kWh/day]

Change of HVAC operation (supply air flow rate) during the pandemic

Stakeholder Engagement and Outreach

Convened a Technical Advisory Group (TAG) of industry and academic experts to:

- solicit input on use cases, dataset needs, planned data collection efforts, data tools and repository
- pursue synergistic research opportunities

Industry	Academia

Clay Nesler JCI	**Krish Gomatom** EPRI	**Clayton Miller** National University of Singapore (NUS)
Zoltan Nagy UT Austin	**Rishee Jain** Stanford University	**Xiaofan Wu** Siemens
Joel Bender ASHRAE 223P/Cornell	**Sagar Rao** IBPSA-USA, AEI	**Omid Ardakanian** ACM SIGEnergy, University of Alberta
David Eric Schwartz Xerox Palo Alto Research Center (PARC)	**Goncalo Costa** Robert Bosch, USA	**Neera Jain** Purdue University
Hyojin Kim New Jersey Institute of Technology (NJIT)	**Michael Blasnik** Google (Nest)	**Paulo Tabares Velasco** Colorado School of Mines

Activity Highlights

June 2020 Concierge TAG: One-on-one meetings with TAG members for in-depth discussions
November 2020 Circulated responses to feedback, summarized actions taken
March 2021 All-TAG meeting (Virtual)
March 2021 BTO Sensors and Controls Lab Sharing Workshop
November 2021 A seminar at the ASHRAE BPAC conference

Remaining Project Work

Where are we?

- Use cases identified and mapped with data needs
- Data infrastructure identified and new sensors installed for new data collection
- 6 datasets curated and published on the project data portal
- Demonstrated two use cases using the curated datasets
- Conducted two rounds of TAG meetings
- Engaged communities on related efforts (e.g., IEA Annex 81, ASHRAE 223P)

Plan Ahead (FY22):

- Complete new data collection and publish the datasets
- Streamline and semi-automate the data curation workflow
- Demonstrate another two use cases using the new collected datasets
- Conduct the third round of TAG meetings
- Outreach: publications, seminars, community engagement
- Explore opportunities to continue the data portal as a resource serving the community

Thank You

LBNL
Tianzhen Hong
thong@lbl.gov

NREL
Lieko Earle
lieko.earle@nrel.gov

ORNL
Piljae Im
imp1@ornl.gov

PNNL
Vikas Chandan
vikas.chandan@pnnl.gov

REFERENCE SLIDES

Project Budget

Project Budget: $3.75M
Variances: N/A
Cost to Date: $1.98M
Additional Funding: N/A

Budget History					
FY 2020 (past)		FY 2021 (current)		FY 2022 (planned)	
DOE	Cost-share	DOE	Cost-share	DOE	Cost-share
$1.15M	N/A	$1.35M	N/A	$1.25M	N/A

Project Plan and Schedule

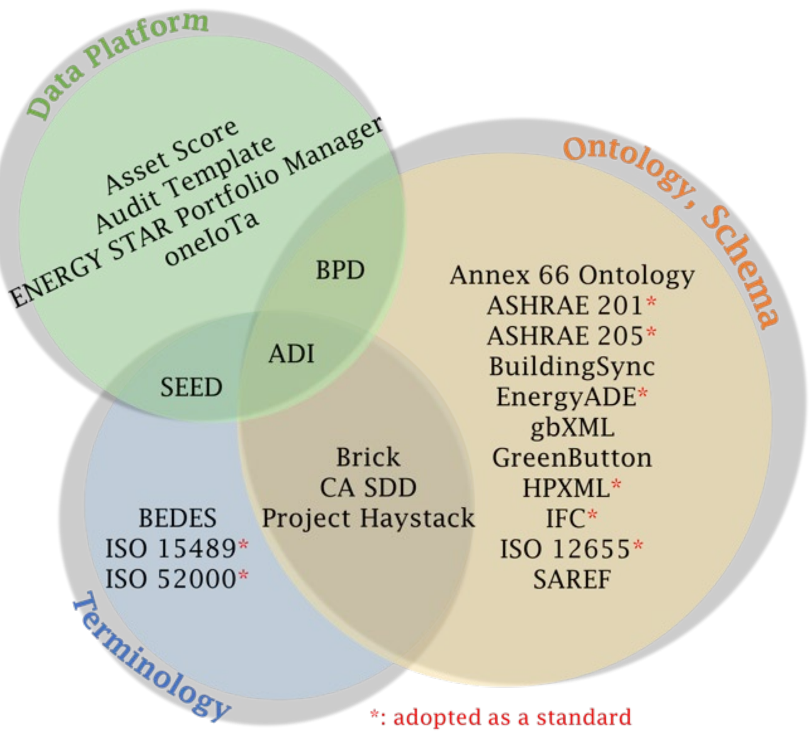
D = Accomplished AOP Deliverable/Milestone

G/N = Accomplished AOP Go/No-Go

Task		Deliverable/Milestone	Budget Year 1 (Fiscal Year 2020)												Budget Year 2 (Fiscal Year 2021)												Budget Year 3 (Fiscal Year 2022)											
			Q1			Q2			Q3			Q4			Q1			Q2			Q3			Q4			Q1			Q2			Q3			Q4		
			CY 2019			CY 2020												CY 2021												CY 2022								
			Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep
1	Existing data collection and curation (M1-21)	Schematic mapping data sources to use cases: Describe mapping of the data sources (existing and/or new datasets) to the identified use cases (M3)			D																																	
		Existing datasets curated: Collect and curate at least 4 existing datasets from the identified existing datasets. Curation scripts, curated datasets and dataset description documents reviewed with the project team and technical advisors. (M21)																		D			G/N															
2	New Data Collection (M1-33)	Report on data collection sites: Evaluate and finalize the facilities for the new data collection for the identified use cases (M6)						D																														
		Go/No-Go: Data collection plan, instrumentation design for selected buildings completed, and data representation requirements finalized (M12)												G/N																								
		SMART Milestone: Complete installations and start the collection (M24)																								D												
		Summary reports for all datasets collected in Task 2 (M33)																																	D			
3	Data tools and database (M1-30)	Propose enhancements to data representation tools and development of requirements of data warehouse tools: Built upon the reviews done in FY19, propose enhancements to data representation tools; develop requirements of data warehouse tools (M9)									D																											
		Metadata map for all the curated datasets (M24)																								D												
		SMART Milestone: Data portal with all datasets collected and curated in Tasks 1 & 2 uploaded (M30)																														D						
4	Demonstration of four case studies (M15-36)	SMART Milestone: Technical reports on four case studies using datasets collected in this project (two on M21, two on M36)																					D															D
5	Engagement and dissemination (M1-36)	Establish the Technical Advisory Group and organize three meetings									D						D			D																D		
		BPAC Seminar proposals and presentation slides																					D					D										

Development and Enhancement of Data Tools

- A review on 24 existing data tools to identify their features, applications and limitations



Categorizing of 24 existing data tools

		Tool Categories				Data Coverage										Data Relations				Flexibility and Extensibility				Stage of Application						Adoption	
ID	Tool Name	Terminology	Ontology	Schema	Data Platform	Energy Use Data	Onsite Power Generation Data	Indoor Environmental Data	Outdoor Environmental Data	Equipment Operational Data	System Control Setting/Logic	Occupant Data	Design Basis Data	Building and System Asset Data	Utility Rates and Grid Signal	Hierarchy	Typology	Association	Others	Allow various levels of details	Allow missing data/ Capability of adding new data	Extensible	Use standard terminology	Building Design	Operation	Audit	Code Compliance	Rating	Commissioning	Development status (U - under development P - pilot application A - widely adopted)	Applicable building type (C - commercial R - residential)
1	ADI	x		x	x	x		x	x	x	x				x	x			x	Yes	Yes	Yes	Yes		x					A	C/R
2	Annex 66 Ontology		x			x		x	x	x	x	x				x				Yes	Yes	Yes	Yes	x	x	x	x	x	x	P	C/R
3	ASHRAE 201			x		x		x	x	x	x				x		x			Yes	Yes	Yes	Yes	x	x	x	x	x	x	A	C/R
4	ASHRAE 205		x	x						x	x						x			Yes	Yes	Yes	Yes		x	x		x	x	A	C/R
5	Asset Score				x								x	x		x	x	x		Yes	No	Yes	No	x		x		x		A	C/R
6	Audit Template				x	x							x	x	x	x	x	x		Yes	No	Yes	Yes	x		x		x		A	C/R
7	BEDES	x				x				x			x	x					x	Yes	Yes	Yes	Yes	x	x	x		x		A	C/R
8	BPD		x		x	x				x			x	x				x	x	Yes	Yes	Yes	Yes		x	x		x		A	C/R
9	Brick	x	x	x		x		x	x	x	x					x		x		Yes	Yes	Yes	Yes		x				x	P	C/R
10	BuildingSync			x		x				x		x	x	x	x	x				Yes	Yes	Yes	Yes	x	x	x		x		A	C
11	CA SDD	x	x	x						x	x		x	x		x		x		No	Yes	No	Yes	x			x			A	C
12	Energy ADE			x		x				x		x	x	x		x	x	x		No	Yes	Yes	No	x	x					A	C/R
13	ENERGY STAR Portfolio Manager				x	x								x		x		x		No	Yes	Yes	No		x			x		P	C/R
14	gbXML			x		x		x		x			x	x		x	x			Yes	Yes	Yes	Yes	x		x				A	C/R
15	GreenButton			x		x	x													No	No	No	Yes		x					P	C/R
16	HPXML			x		x				x				x	x	x				Yes	No	No	Yes		x	x		x		A	R
17	IFC			x		x		x		x		x	x	x		x		x		Yes	Yes	Yes	Yes	x	x				x	A	C/R
18	ISO 12655-2013		x			x	x									x				No	Yes	No	Yes		x	x		x		A	C/R
19	ISO 15489-2016	x												x	x				x	No	Yes	Yes	No			x				A	C/R
20	ISO 52000-1:2017	x				x								x					x	No	Yes	Yes	No			x				A	C/R
21	oneIoTa				x	x		x	x	x	x				x	x				Yes	Yes	Yes	No		x					P	C/R
22	Project Haystack	x	x	x		x		x	x	x	x							x		Yes	Yes	Yes	Yes		x				x	P	C/R
23	SAREF		x			x				x	x				x					Yes	Yes	No	Yes		x					U	C/R
24	SEED	x			x	x				x			x	x			x			Yes	Yes	Yes	No	x	x	x		x		A	C/R

Key features of 24 existing data tools

Luo, N., Pritoni, M. and Hong, T., 2021. An overview of data tools for representing and managing building information and performance data. Renewable and Sustainable Energy Reviews, 147, p.111224. https://doi.org/10.1016/j.rser.2021.111224

Data Portal Development

Feature identification and prioritization

- 40 features were identified belonging to categories such as *collection of data, curating datasets, metadata, data discovery, security and download*
- Each feature was scored by multi-lab team on 2 parameters: *ease of implementation and value to users*
- A composite score was developed for each feature which was used to determine its priority: *Very High, High, Medium, Low, Trivial*
- FY20 Focus: *Very High, and High* priority features
- FY21 Focus (ongoing): *Medium* priority features

***Very High* priority features**
Explicit public license
Contact for dataset (distribution List)
Citation (if open) DOI
Climate zone in metadata
Data resolution
Access and Authentication
Secure web portal

***High priority* features**
Search through metadata
Measurements in metadata
Keyword based filtering of datasets to allow users find datasets with desired attributes
Adding google analytics to capture metrics such as page visits, most viewed datasets, and downloads
Processing data into standard form
Data cleaning, run manually
Downloadable data format

***Medium* priority features**
A link to the document that provides information about the structure/standards documentation
Clean/curated metadata
A link to the physical layout of the buildings for each dataset (if appropriate).
A link to the document describing how data was cleaned before sharing the data with users
Documentation of our curation/cleaning
Dataset usage example/recipe
Raw data downloadable
Subset data
Periodic batch collection/processing (cleaning and standardizing)
Make Data curation/cleaning scripts available to users
Transforming data resolutions (regridding)
Dataset time coverage information
Marked data quality for different time periods

Data Portal Development

Selecting Fields for the portal

- A second prioritization matrix was developed by multi-lab team to identify the appropriate level for each field: *project page, dataset page*, or *search queries*

1st level category	2nd level category	3rd level category	Project Page?	Dataset Page?	Search queries?
Building information	Geo information	City		x	
		State		x	x
		Zip Code			
		Climate Zone		x	x
	Building characteristics	Name	x	x	
		Type		x	x
		Year built		x	x
		Number of buildings			
		Number of floors			
		Total floor area			x
		Conditioned floor area			
		Number of occupants			
		More information (link to the building webpage)			
	System characteristics	Fuel type			x
		HVAC system type			x
		Lighting system type			
		Plugin system type			
		Water heating system type			
Dataset information	Creation	Data creator		x	
		Created date			
		Curated date			
		Curation approach			
		Version			
		Raw data availability		x	
	Sharing	Point of contact	x	x	
		License		x	
		Description		x	
	Application	Data formats		x	
		Use cases supported			x
		References (report, paper etc.)		x	
Data by category	11 data categories:	Data categories covered in this dataset		x	x
		Name of data tool used			
		Data formats			
		Data location			
		Data period			
		Data resolution			
		Data quality			
		Raw data availability			