

HVAC Dataset Catalog

NexusForge AI Training Resources

26 Datasets | 13+ GB Total | 7 Equipment Types

AutomataNexus, LLC

Generated: December 25, 2025

Table of Contents

Dataset Overview	3
Quick Reference	3
FDD Datasets	5
Chiller Datasets	7
Boiler Datasets	8
VAV Datasets	9
Simulation Datasets	10
Reference Datasets	11
Controls Datasets	12
Lab Datasets	13
Heat Pump Datasets	14
Retrofit Datasets	15
Download Instructions	16
Figshare Datasets	16
Data.gov Datasets	16
OpenEI / LBNL Datasets	16
Kaggle Datasets	16

Dataset Overview

This catalog contains 26 curated HVAC datasets for training NexusForge AI models. These datasets cover fault detection and diagnostics (FDD), building simulation, equipment-specific data, and control systems. Total storage requirement is approximately 13+ GB.

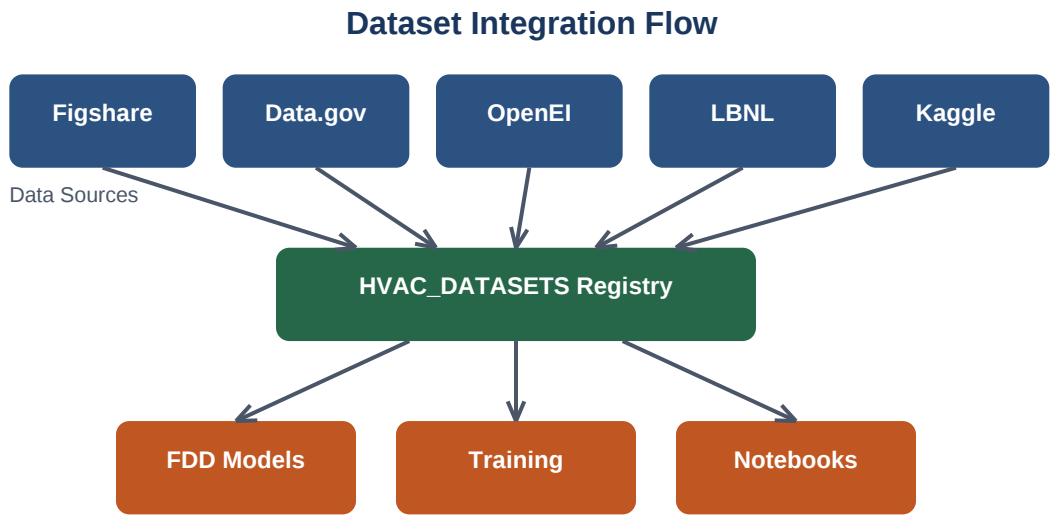


Figure 1: Dataset Integration Flow

26 Datasets | 13+ GB | 7 Equipment Types

Quick Reference

#	Dataset ID	Size	Format	Source
1	shifdr_michigan	500 MB	CSV	U-M Deep Blue
2	figshare_ahu_faults	100 MB	ZIP	Figshare
3	figshare_multibuilding_fdd	29 MB	CSV	Figshare
4	openei_910_fdd_evaluation	200 MB	CSV	OpenEI
5	lbnl_fdd_5763	500 MB	CSV	LBNL
6	lbnl_fdd	150 MB	CSV	LBNL
7	curated_fault	50 MB	CSV	Data.gov
8	fdd_evaluation	200 MB	CSV	Data.gov
9	figshare_chiller_operation	49 MB	CSV	Figshare
10	lbnl_chiller_plant_sim	8.14 GB	CSV	LBNL
11	figshare_lbnl_chiller_plant	1.14 GB	CSV	Figshare
12	lbnl_boiler_plant_sim	2.08 GB	CSV	LBNL
13	kaggle_industrial_boiler	11 MB	CSV	Kaggle
14	figshare_vav_fdd	32 MB	XLSX	Figshare

15	alphabuilding	500 MB	CSV	Data.gov
16	ref_warehouse	25 MB	IDF/CSV	Data.gov
17	ref_hospital	40 MB	IDF/CSV	Data.gov
18	ref_large_office	35 MB	IDF/CSV	Data.gov
19	ref_small_hotel	20 MB	IDF/CSV	Data.gov
20	ref_primary_school	25 MB	IDF/CSV	Data.gov
21	ref_supermarket	30 MB	IDF/CSV	Data.gov
22	thermostat_data	75 MB	CSV	Data.gov
23	hvac_hil	100 MB	CSV	Data.gov
24	gshp_model	45 MB	CSV	Data.gov
25	duct_sealing	15 MB	CSV	Data.gov
26	heat_pump_test	80 MB	CSV	Data.gov

FDD Datasets

SHIFDR: Michigan Buildings Dataset

ID: shifdr_michigan | Size: 500 MB | Format: CSV

Sub-metered HVAC data from 14 commercial buildings in Southeast Michigan. Contains device-level power, BAS data, temperatures, pressures, humidity, airflow, valve positions, water flow, and energy data.

URL: https://deepblue.lib.umich.edu/data/concern/data_sets/j6731449j

Labeled AHU Faults Dataset

ID: figshare_ahu_faults | Size: 100 MB | Format: ZIP

Eight labeled AHU datasets with injected faults and normal operations across various seasonal conditions. Ideal for transfer learning.

URL: <https://figshare.com/articles/dataset/29297999>

Multi-Building FDD (Auditorium, Hospital, Office)

ID: figshare_multibuilding_fdd | Size: 29 MB | Format: CSV

Real operational labeled data from auditorium, 24-hour hospital, and large office. Designed for transfer learning and IoT edge applications.

URL: <https://figshare.com/articles/dataset/27147678>

Multi-Lab FDD Evaluation (LBNL, PNNL, NREL, ORNL)

ID: openei_910_fdd_evaluation | Size: 200 MB | Format: CSV

Multi-lab FDD evaluation datasets including ASHRAE RP-1312. Contains AHU, RTU, and VAV simulation data.

URL: <https://data.openei.org/submissions/910>

LBNL Complete FDD (7 Systems)

ID: lbln_fdd_5763 | Size: 500 MB | Format: CSV

Comprehensive FDD for RTU, single/dual-duct AHU, VAV, fan coil, chiller plant, and boiler plant. Includes BRICK schema TTL files.

URL: <https://data.openei.org/submissions/5763>

LBNL FDD Datasets

ID: lbln_fdd | Size: 150 MB | Format: CSV

Labeled fault data for RTU, AHU, and chiller faults from LBNL.

URL: <https://data.openei.org/submissions/924>

NIST Curated Fault Data

ID: curated_fault | Size: 50 MB | Format: CSV

Curated fault scenarios for testing FDD algorithms with labeled faults from NIST.

URL: <https://catalog.data.gov/dataset/curated-test-fault-data-set>

NIST FDD Evaluation Dataset

ID: fdd_evaluation | Size: 200 MB | Format: CSV

Reference datasets for evaluating FDD algorithms in buildings from NIST.

URL: <https://catalog.data.gov/dataset/data-sets-for-evaluation-of-building-fault-detection-and-diagnostics-algorithms>

Chiller Datasets

Chiller Operation - ASHRAE RP-1043

ID: figshare_chiller_operation | Size: 49 MB | Format: CSV

Chiller fault diagnosis dataset from ASHRAE RP-1043. Records temperature, pressure, flow under fault conditions with labels.

URL: <https://figshare.com/articles/dataset/28435232>

LBNL Simulated Chiller Plant

ID: lbln_chiller_plant_sim | Size: 8.14 GB | Format: CSV

12-story building with 3 chillers. 7 faults at multiple severity levels. 1-year simulation at 1-minute intervals, 77 parameters.

URL: <https://faultdetection.lbl.gov/dataset/simulated-chiller-plant/>

LBNL Chiller Plant (Chen & Granderson)

ID: figshare_lbln_chiller_plant | Size: 1.14 GB | Format: CSV

World's largest labeled HVAC FDD time-series. Covers RTU, AHU, VAV, fan coil, chiller, and boiler systems.

URL: <https://figshare.com/articles/dataset/22338244>

Boiler Datasets

LBNL Simulated Boiler Plant

ID: lbnl_boiler_plant_sim | Size: 2.08 GB | Format: CSV

12-story building with 2 parallel boilers. 5 fault types, 1-year simulation at 1-minute intervals, 22 measurements.

URL: <https://faultdetection.lbl.gov/dataset/simulated-boiler-plant/>

Industrial Boiler Operations (Zhejiang)

ID: kaggle_industrial_boiler | Size: 11 MB | Format: CSV

High-frequency time-series from coal-fired boiler in China. 8.6% abnormal data. Pressure, temperature, flow, oxygen.

URL: <https://www.kaggle.com/datasets/nikitamanaenkov/time-series-of-industrial-boiler-operations>

VAV Datasets

VAV Terminal Units FDD

ID: figshare_vav_fdd | Size: 32 MB | Format: XLSX

Stuck dampers and biased airflow sensors in multi-zone commercial buildings. By Im, Jung, Yoon.

URL: <https://figshare.com/articles/dataset/27948192>

Simulation Datasets

AlphaBuilding Synthetic Dataset

ID: alphabuilding | Size: 500 MB | Format: CSV

Synthetic building operation data with HVAC, lighting, and occupancy. Good for control models.

URL: <https://catalog.data.gov/dataset/alphabuilding-synthetic-buildings-operation-dataset>

Reference Datasets

DOE Reference: Warehouse

ID: ref_warehouse | Size: 25 MB | Format: IDF/CSV

DOE reference building model for warehouse with HVAC simulation data.

URL: <https://catalog.data.gov/dataset/commercial-reference-building-warehouse>

DOE Reference: Hospital

ID: ref_hospital | Size: 40 MB | Format: IDF/CSV

DOE reference building model for hospital with complex HVAC systems.

URL: <https://catalog.data.gov/dataset/commercial-reference-building-hospital>

DOE Reference: Large Office

ID: ref_large_office | Size: 35 MB | Format: IDF/CSV

DOE reference building model for large office with VAV systems.

URL: <https://catalog.data.gov/dataset/commercial-reference-building-large-office>

DOE Reference: Small Hotel

ID: ref_small_hotel | Size: 20 MB | Format: IDF/CSV

DOE reference building model for small hotel with PTAC units.

URL: <https://catalog.data.gov/dataset/commercial-reference-building-small-hotel>

DOE Reference: Primary School

ID: ref_primary_school | Size: 25 MB | Format: IDF/CSV

DOE reference building model for primary school with occupancy schedules.

URL: <https://catalog.data.gov/dataset/commercial-reference-building-primary-school>

DOE Reference: Supermarket

ID: ref_supermarket | Size: 30 MB | Format: IDF/CSV

DOE reference building model for supermarket with refrigeration.

URL: <https://catalog.data.gov/dataset/commercial-reference-building-supermarket>

Controls Datasets

Multifamily Thermostat Data

ID: thermostat_data | Size: 75 MB | Format: CSV

Real thermostat data from multifamily buildings with setpoints and runtime.

URL: <https://catalog.data.gov/dataset/multifamily-programmable-thermostat-data>

Lab Datasets

Hardware-in-Loop HVAC Testing

ID: hvac_hil | Size: 100 MB | Format: CSV

Lab-tested HVAC and thermal storage performance data across US climates.

URL: <https://catalog.data.gov/dataset/hardware-in-the-loop-laboratory-performance-verification>

Heat Pump/AC HIL Testing

ID: heat_pump_test | Size: 80 MB | Format: CSV

Northeastern University lab testing data for heat pumps and AC units.

URL: <https://catalog.data.gov/dataset/benefit-hvac-hardware-in-the-loop-experimental-testing>

Heat Pump Datasets

Ground Source Heat Pump (TRNSYS)

ID: gshp_model | Size: 45 MB | Format: CSV

TRNSYS simulation data for ground-source heat pump systems.

URL: <https://catalog.data.gov/dataset/trnsys-building-energy-simulation-model>

Retrofit Datasets

Duct Sealing Evaluation

ID: duct_sealing | Size: 15 MB | Format: CSV

Before/after HVAC performance data from duct sealing retrofits.

URL: <https://catalog.data.gov/dataset/raleigh-housing-authority-duct-sealing-evaluation>

Download Instructions

Figshare Datasets

1. Navigate to the URL
2. Click 'Download all' or select specific files
3. No account required for public datasets
4. Check license (most are CC BY 4.0 or CC0)

Data.gov Datasets

1. Navigate to the URL
2. Click 'Download' or 'Access' button
3. Some redirect to original source (NREL, NIST)
4. No account required

OpenEI / LBNL Datasets

1. Navigate to the dataset page
2. Download individual CSV files or full ZIP
3. TTL files available for BRICK schema visualization
4. PDF documentation included

Kaggle Datasets

1. Navigate to the URL
2. Free Kaggle account required
3. Click 'Download' button
4. API: `kaggle datasets download -d <dataset-path>`