**Metadata file**

* Journal: Nature Scientific Data, under review, doi to be provided later.
* Title: LBNL Building 59: A three-year dataset supporting research on building energy, HVAC controls, and occupancy analytics.
* Author(s):
  + Na Luo, [nluo@lbl.gov](mailto:nluo@lbl.gov), Lawrence Berkeley National Laboratory
  + Zhe Wang, [zwang5@lbl.gov](mailto:zwang5@lbl.gov), Lawrence Berkeley National Laboratory
  + David Blum, [DHBlum@lbl.gov](mailto:DHBlum@lbl.gov), Lawrence Berkeley National Laboratory
  + Norman Bourassa, [njbourassa@lbl.gov](mailto:njbourassa@lbl.gov), Lawrence Berkeley National Laboratory
  + Jeffrey Broughton, [jbroughton@lbl.gov](mailto:jbroughton@lbl.gov), Lawrence Berkeley National Laboratory
  + Chris Weyandt, [cweyandt@lbl.gov](mailto:cweyandt@lbl.gov), Lawrence Berkeley National Laboratory
  + Mary Ann Piette, [mapiette@lbl.gov](mailto:mapiette@lbl.gov), Lawrence Berkeley National Laboratory
  + Tianzhen Hong, [thong@lbl.gov](mailto:thong@lbl.gov), Lawrence Berkeley National Laboratory
* Abstract: Short description of the dataset that would allow others to understand what the data is about.

This dataset was collected from an office building constructed in 2015 in Berkeley, California, which includes whole-building and end-use energy consumption, HVAC system operating conditions, indoor and outdoor environmental parameters, and occupant counts. The data was collected in three years from more than 300 sensors and meters for two office floors of the building. A three-step data curation strategy is applied to transform the raw data into the research-grade data: (1) cleaning the raw data to detect and adjust the outlier values and fill the data gaps; (2) creating the metadata model of the building systems and data points using the Brick schema; (3) describing the metadata of the dataset using a semantic JSON schema. This dataset can be used for various types of applications, including building energy benchmarking, load shape analysis, energy prediction, occupancy prediction and analytics, and HVAC controls to improve understanding and efficiency of building operations for reducing energy use, energy costs, and carbon emissions.

* Usage notes:
  + The “Bldg59\_clean data” folder contains 27 time-series data files in CSV format, covering energy use data, indoor and outdoor environmental data, HVAC control and operational data, as well as occupant-related data. The data has been processed using the data cleaning script (https://github.com/LBNL-ETA/Data-Cleaning) to fill the data gaps.
  + The TTL file, named as “Bldg59\_w\_occ Brick model.ttl”, is the Brick model of the dataset, which represents the metadata and hierarchical structure of the building, systems and sensors. We recommend users to explore the metadata of equipment and sensors in the Brick model by using the Brick TTL Viewer.
  + The metadata JSON file represents the high-level building information, data governance, contextual information for each data category, as well as application perspectives of the dataset.
* Methods: Any technical or methodological information that may help others to understand how the data were generated (i.e. equipment/tools/reagents used, or procedures followed)
  + We identified and modified the data gaps and outlier values to generate the clean version of the time-series raw data. Specifically, we used multiple algorithms (e.g., linear interpolation, K-nearest neighbors, Matrix Factorization) to fill the gaps by considering both the length of data gap and the sampling frequency of each data point. More detailed description can be found at: <https://github.com/LBNL-ETA/Data-Cleaning>.
* Funding Information: Name of the funding organization that supported creation of the resource, including applicable grant number(s).
  + The data development was supported by the model predictive control project, which was part of the U.S.-China joint Clean Energy Research Center on Building Energy Efficiency (CERC-BEE) 2016-2021. The data processing effort part of the Benchmark Datasets project. Both projects were supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the United States Department of Energy, under Contract No. DE-AC02-05CH11231.