**Table: Summary of open data related to building energy studies**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Database** | **Abbreviation** | **URL** | **Year** | **Region** | **Data Content (Key Words)** | **Category** | **Building Type** | **Time resolution** | **Full Detailed Introduction/ Description** | **Potential Usage** | **Location to access data** |
| 1 | Buildings Energy Data Book | NA | <https://data.openei.org/submissions/144> | 2011 (not updating) | U.S. by census region and state | Construction, building technologies, energy consumption, end use consumption, and building characteristics | Macro-economic data, Representative Building Data and Building Characteristics | Residential and Commercial Buildings | Annual | The Building Energy Data Book (2011) is a compendium of data from a variety of data sets and includes statistics on residential and commercial building energy consumption. Data tables contain statistics related to construction, building technologies, energy consumption, and building characteristics. The Building Technologies Office (BTO) within the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy developed this resource to provide a comprehensive set of buildings- and energy-related data.  Source: (<https://data.openei.org/submissions/144>) | Annual model calibration for UBEM / Future annual energy trend projection and prediction | Directory: (‘/Database\_1\_Building\_Energy/’) |
| 2 | Energy Information Administration | EIA | <https://www.eia.gov/>  <https://www.eia.gov/about/mission_overview.php> | 1973-Present, including projection to 2050 | U.S. by census region and state | Consumption (gas/electricity/energy), price, presentative building data. | Mix of Macro-economic data, Representative building data, and GIS data | Residential, Industrial and Commercial Buildings | Monthly+ Annual | The U.S. Energy Information Administration (EIA) is the statistical and analytical agency within the U.S. Department of Energy. EIA collects, analyzes, and disseminates independent and impartial energy information to promote sound policymaking, efficient markets, and public understanding of energy and its interaction with the economy and the environment. EIA is the nation's premier source of energy information, and, by law, its data, analyses, and forecasts are independent of approval by any other officer or employee of the U.S. government. EIA conducts a comprehensive data collection program that covers the full spectrum of energy sources, end uses, and energy flows. EIA also prepares informative energy analyses, monthly short-term forecasts of energy market trends, and long-term U.S. and international energy outlooks. EIA disseminates its data, analyses, and other products primarily through its website.  Source: (<https://www.eia.gov/about/mission_overview.php>) | Energy prediction for future years, compare the data between renewable energy sources of data. | Directory: (‘/Database\_2\_EIA/’) This also contains data from annual energy outlook, monthly energy review and annual electric power industry. |
| 3 | Energy Information Administration – Annual Energy outlook | EIA-AEO | <https://www.eia.gov/outlooks/aeo/>  <https://www.eia.gov/outlooks/aeo/narrative/index.php#Introduction> | 2010-2050 (projection) | U.S. by sector by census division | Different energy source by end use | Macro-economic data | Residential and commercial building | Monthly | The Annual Energy Outlook presents an assessment by the U.S. Energy Information Administration of the outlook for energy markets through 2050. EIA’s National Energy Modeling System (NEMS), was used to produce the Annual Energy Outlook (AEO), requires substantial updates to better model hydrogen, carbon capture, and other emerging technologies. w energy markets will develop. Graphs are retooled to emphasize the range of results, denoted by shaded areas, across the full suite of modeled cases. The Annual Energy Outlook 2023 (AEO2023) explores long-term energy trends in the United States. Since the release of the last AEO in early 2022, passage of the Inflation Reduction Act (IRA), Public Law 117-169, altered the policy landscape to develop the projections.  Source:(<https://www.eia.gov/outlooks/aeo/narrative/index.php#Introduction>) | Energy prediction | Directory: (‘/Database\_3\_EIA\_Annual\_Energy\_Outlook/’) |
| 4 | Energy Information Administration - Monthly Energy Review/ Total Energy/ Annual Energy | EIA-Energy | [https://www.eia.gov/totalenergy/data/monthly/index.php ­­­](https://www.eia.gov/totalenergy/data/monthly/index.php) | 1973-present | U.S. by sector by census division | Different energy source by endues | Macro-economic data | Residential and commercial building | Monthly | A publication of recent and historical energy statistics. This publication includes total energy production, consumption, stocks, and trade; energy prices; overviews of petroleum, natural gas, coal, electricity, nuclear energy, renewable energy, and carbon dioxide emissions; and data unit conversions values.  Source:(<https://www.eia.gov/outlooks/aeo/narrative/index.php#Introduction>) | Annual or monthly calibration or UBEM | Directory: (‘/Database\_4\_Total\_Energy\_Review/’) |
| 5 | Annual Electric Power Industry Report, Form EIA-861 (part of EIA) | EIA-861 | <https://www.eia.gov/electricity/data/eia861/> | 1990-present | U.S. by utility provider | Peak load, generation, electric purchases, sales, revenues, customer counts, energy efficiency, demand response, net metering programs, and distributed generation capacity | Macro-economic data | All buildings | Annual | Form EIA-861 and Form EIA-861S (short form) data files include information such as peak load, generation, electric purchases, sales, revenues, customer counts, energy efficiency, demand response, net metering programs, and distributed generation capacity. In 2012, this data chart was created Form EIA-861S to reduce respondent burden and to increase our processing efficiency; that year, about 1,100 utilities initially reported on this form instead of Form EIA-861. In 2020, the number of utilities increased to about 1,700 utilities. We reformatted the files for the years 1990–2011, but we didn’t change or update any data files. We reformatted the files to make them easier to understand and to match the format and titles of the current files.  Source: (<https://www.eia.gov/electricity/data/eia861/>) | Large-scale UBEM calibration | Directory: (‘/Database\_5\_EIA\_861/’) |
| 6 | EIA - Commercial Buildings Energy Consumption Survey | CBECS | <https://www.eia.gov/consumption/commercial/>  <https://www.eia.gov/todayinenergy/detail.php?id=21112> | 1992 – present | U.S. by census division | Building characteristics and energy consumptions by endues | Representative building data | Commercial building | Annual | The U.S. Energy Information Administration's Commercial Buildings Energy Consumption Survey (CBECS) is the only nationally representative source of statistical information on energy-related characteristics, consumption, and expenditures for the nation's 5.6 million commercial buildings totaling 87 billion square feet of floor space CBECS is the only independent, statistically representative source of national-level data on the characteristics and energy use of commercial buildings. EIA collects data for commercial buildings in two parts: collect building characteristics through an in-person or web survey. Respondents, such as building owners and managers, completed the survey for 6,436 buildings for the 2018 CBECS, representing 5.9 million buildings in the United States. Another way is collecting energy usage data from suppliers of electricity, natural gas, fuel oil, and district heat.  Source: (<https://www.eia.gov/todayinenergy/detail.php?id=21112>) | Build and calibrate sampled energy models | Directory: (‘/Database\_6\_CBECS/’) |
| 7 | EIA - Residential Energy Consumption Survey | RECS | <https://www.eia.gov/consumption/residential/>  <https://www.eia.gov/consumption/residential/reports/2015/methodology/> | 1993-2015 | U.S. | Building characteristics and energy consumptions by endues | Representative building data | Residential building | Annual | The Residential Energy Consumption Survey (RECS) is a periodic study conducted by the U.S. Energy Information Administration (EIA) that provides detailed information about energy usage in U.S. homes. RECS is a multi-year effort (Figure 1) consisting of a Household Survey phase, data collection from household energy suppliers, and end-use consumption and expenditures estimation. The Household Survey collects data on energy-related characteristics and usage patterns of a national representative sample of housing units. The Energy Supplier Survey (ESS) collects data on how much electricity, natural gas, propane/LPG, fuel oil, and kerosene were consumed in the sampled housing units during the reference year. It also collects data on actual dollar amounts spent on these energy sources.  Source: (<https://www.eia.gov/consumption/residential/reports/2015/methodology/>) | Build and calibrate sampled energy models. Benchmarking | Directory: (‘/Database\_7\_RECS/’) |
| 8 | Building Performance Database | BPD | <https://bpd.lbl.gov/>  [https://buildings.lbl.gov/cbs/bpd](https://buildings.lbl.gov/cbs/bpd%20%20%20%20%20%20%20%20%20%20%20%20%20)  <https://www.energy.gov/eere/buildings/building-performance-database-bpd> | NA | U.S. by climate zone/state/city/zip code | Building characteristics and energy consumptions by end use and energy source | Representative building data | Residential and commercial building | Annual | The Building Performance Database (BPD) is the nation's largest dataset of information about the energy-related characteristics of commercial and residential buildings. The BPD combines, cleanses and anonymizes data collected by federal, state and local governments, utilities, energy efficiency programs, building owners and private companies, and makes it available to the public.    Source: (<https://www.census.gov/programs-surveys/acs/about.html>) | Benchmarking and building sampled energy models | URL: <https://bpd.lbl.gov/explore> This is online data base, so we don’t download the data. Data Selection can be done from the website as per the requirement. |
| 9 | Microsoft building footprint | NA | <https://www.microsoft.com/en-us/maps/building-footprints>  [https://www.microsoft.com/en-us/maps/building-footprints#:~:text=Microsoft%20has%20made%20significant%20investments,identify%20map%20features%20at%20scale](https://www.microsoft.com/en-us/maps/building-footprints%23:~:text=Microsoft%20has%20made%20significant%20investments,identify%20map%20features%20at%20scale) | 2018 | United States, Canada, Uganda and Tanzania | Building locations and shapes | GIS data | All buildings | Time Data | Microsoft has made significant investments in deep learning, computer vision and AI that have been applied to mapping. Over the past few years, Bing Maps has generated high-quality building footprints leveraging AI and harnessing the power of computer vision to identify map features at scale. To achieve this, we rely on the Open Source CNTK Unified Toolkit which was developed by Microsoft. Using CNTK we apply our Deep Neural Networks and the ResNet34 with Refine Net up-sampling layers to detect building footprints from Bing imagery. Ensuring the best outputs, we remove noise and suspicious data, such as false positives, from the predictions. Then, apply a polygonization algorithm to detect building edges and angles to create a proper building footprint. With the goal to increase the coverage of building footprint data available as open data for OpenStreetMap and humanitarian efforts, we have released millions of building footprints as open data available to download free of charge.  Source: [(https://www.microsoft.com/en-us/maps/building-footprints)]((https://www.microsoft.com/en-us/maps/building-footprints)%20%20%20) | Large-scale coarse building modeling | URL <https://www.microsoft.com/en-us/maps/building-footprints> |
| 10 | The Building Data Genome Project I and II | BDG1 and BDG2 | <https://github.com/buds-lab/building-data-genome-project-2> | 2016-2017 | US and Europe | Timeseries building sensor data by building type and by meter types (electricity, heating and cooling water, steam, and irrigation meters) | Building Operating Data, AMI data | Commercial building and residential buildings by building types | Hourly | BDG2 is an open data set made up of 3,053 energy meters from 1,636 buildings. The time range of the times-series data is the two full years (2016 and 2017) and the frequency is hourly measurements of electricity, heating and cooling water, steam, and irrigation meters. A subset of the data was used in the Great Energy Predictor III (GEPIII) competition hosted by the ASHRAE organization in late 2019. (<https://github.com/buds-lab/building-data-genome-project-2>) The GEPIII sub-set includes hourly data from 2,380 meters from 1,449 buildings that were used in a machine learning competition for long-term prediction with an application to measurement and verification in the building energy analysis domain. This data set can be used to benchmark various statistical learning algorithms and other data science techniques. It can also be used simply as a teaching or learning tool to practice dealing with measured performance data from large numbers of non-residential buildings. The charts below illustrate the breakdown of the buildings according to primary use category and subcategory, industry and subindustry, time zone and meter type.  Source: ( <https://github.com/buds-lab/building-data-genome-project-2>) | Detailed energy modeling. Load shapes. | This data is hard to access. So, we will omit this data. |
| 11 | International Energy Agency | IEA | <https://www.iea.org/>  <https://www.iea.org/about/mission> | 1990-2022 | World-wide | Electricity consumption, CO2 emission, electricity generation by source | Macro-economic data | Commercial building and residential buildings | Annual | The IEA is at the heart of global dialogue on energy, providing authoritative analysis, data, policy recommendations, and real-world solutions to help countries provide secure and sustainable energy for all. IEA provide authoritative analysis, data, policy recommendations and solutions to ensure energy security and help the world transition to clean energy. The IEA collects, assesses, and disseminates energy statistics on supply and demand, compiled into energy balances in addition to several other key energy-related indicators, including energy prices, public RD&D and measures of energy efficiency, with other measures in development. This emphasis on sound data provides a unique platform for modelling work and tracking both short-term shifts and long-term trends in countries’ energy transitions, particularly for clean energy.  Source: (<https://www.iea.org/about/mission>) | Energy dataset of world | Directory: (‘/Database\_11\_IEA/’) |
| 12 | LandScan | NA | <https://landscan.ornl.gov/>  <https://www.ornl.gov/news/gis-landscan-goes-public> | 2000-present | U.S. | Population distribution database, capturing daytime and nighttime activity of the U.S. population at a resolution of roughly 90 meters | Macro-economic data, and GIS data | NA | Annual | Lands can, widely considered the gold standard of population and mapping data in the United States, captures daytime and nighttime activity of the U.S. population at a resolution of roughly 90 meters or about 300 feet. These data are collected from ORNL. These datasets are intended to aid in emergency preparedness, readiness, response, and recovery missions; risk assessments; site suitability studies; and a full range of other applications that depend on accurate population data.  Source: (<https://www.ornl.gov/news/gis-landscan-goes-public>) | Predict risk as per the mapping data. | Directory: (‘/Database\_11\_IEA/’) Files are in XML and TIF format which needs to be accessed through API to run. |
| 13 | FERC Form 714 | Form 714 | <https://www.ferc.gov/industries-data/electric/general-information/electric-industry-forms/form-no-714-annual-electric/data> | 2006 to 2020 | U.S. by utility company and by year | Utility company data | Macro-economic data | All buildings | Monthly or Annual | FERC represents for Federal Energy Regulation Committee. Form No. 714 – is Annual Electric Balancing Authority Area and Planning Area Report. FERC, pursuant to the Energy Policy Act of 2005, finalized on February 2, 2006, landmark new rules on the certification of an Electric Reliability Organization and the procedures for the establishment, approval, and enforcement of mandatory electric reliability standards. The Commission has inaugurated Docket No. RM04-2-000 in the wake of the August 14, 2003, regional blackout, and Congressional consideration of power-grid reliability legislation.  Source: (<https://www.ferc.gov/industries-data/electric/general-information/electric-industry-forms/form-no-714-annual-electric/data>) | Region-scale UBEM calibration | Directory: (‘/Database\_13\_FERC\_Form\_714/’) |
| 14 | American Community Survey | ACS | <https://data.census.gov/cedsci/table?d=ACS%201-Year%20Estimates%20Data%20Profiles&tid=ACSDP1Y2019.DP04&hidePreview=false>  <https://www.census.gov/programs-surveys/acs/about.html> | 2010 – present (2019 is the latest) | U.S. | HOUSING CHARACTERISTICS | Macro-economic data | Residential Building | Annual | The American Community Survey (ACS) helps local officials, community leaders, and businesses understand the changes taking place in their communities. It is the premier source for detailed population and housing information about our nation. In other words, The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about our nation and its people. Information from the survey generates data that help inform how trillions of dollars in federal funds are distributed each year. The American Community Survey (ACS) releases new data every year through a variety of data tables that you can access with different data tools.  Source: (<https://www.census.gov/programs-surveys/acs/about.html>) | UBEM building characteristics Sampling | Directory: (‘/Database\_14\_ACS/’) |
| 15 | AMERICAN TIME OF USE | ATUS | <https://www.census.gov/programs-surveys/atus.html#:~:text=The%20American%20Time%20Use%20Survey,childcare%2C%20volunteering%2C%20and%20socializing>. | 2012– present | U.S. | People characteristics | Macro-economic data | Residential Building | Annual | The American Time Use Survey (ATUS) provides nationally representative estimates of how, where, and with whom Americans spend their time, and is the only federal survey providing data on the full range of nonmarket activities, from childcare to volunteering.  Source: (<https://www.census.gov/programs-surveys/atus.html#:~:text=The%20American%20Time%20Use%20Survey,childcare%2C%20volunteering%2C%20and%20socializing>. ) | People energy usage as per gender, age, or other characteristics | Directory: (‘/Database\_15\_I/’) |
| 16 | Public Use Microdata Sample (PUMS) - Census Bureau (Part of ACS) | PUMS | <https://www.census.gov/programs-surveys/acs/microdata.html> | 2005 - 2019 | U.S. | A set of records from individual people or housing units | Representative building data | Residential Building | 1 year or 5 years | The Census Bureau’s American Community Survey (ACS) Public Use Microdata Sample (PUMS) files enable data users to create custom estimates and tables, free of charge, that are not available through ACS pretabulated data products. The ACS PUMS files are a set of records from individual people or housing units, with disclosure protection enabled so that individuals or housing units cannot be identified. The Census Bureau’s American Community Survey (ACS) Public Use Microdata Sample (PUMS) files enable data users to create custom estimates and tables, free of charge, that are not available through ACS pretabulated data products.  The ACS PUMS files are a set of records from individual people or housing units, with disclosure protection enabled so that individuals or housing units cannot be identified. The Census Bureau produces ACS 1-year and 5-year PUMS files. These files are accessible using the microdata access tool on [data.census.gov](https://data.census.gov/mdat/#/) and the Census Bureau's [FTP site](https://www2.census.gov/programs-surveys/acs/data/pums/).\* Only selected geographic areas are identified in the ACS PUMS, including nation, regions, divisions, states, and [Public Use Microdata Areas (PUMAs)](https://ask.census.gov/prweb/PRServletCustom?pyActivity=pyMobileSnapStart&ArticleID=KCP-2950). Of these, PUMAs are the most detailed geographic areas available. The Form 714 data for 2011-2020 in this database has been superseded by the XBRL Taxonomy found on the e-Forms portal.  Source: (<https://www.census.gov/programs-surveys/acs/microdata.html> ) | Residential Owners Characteristics | Directory: (‘/Database\_16\_PUMS/’) |
| 17 | American Housing Survey (Part of ACS) | AHS | <https://www.census.gov/programs-surveys/ahs.html> | 2011-2019 every two years | U.S. | The physical condition of homes and neighborhoods, the costs of financing and maintaining homes, and the characteristics of people who live in these homes. | Representative building data | Residential Building | 2 years | The American Housing Survey (AHS) is sponsored by the Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau. The survey has been the most comprehensive national housing survey in the United States since its inception in 1973, providing current information on the size, composition, and quality of the nation’s housing and measuring changes in our housing stock as it ages. The AHS is a longitudinal housing unit survey conducted biennially in odd-numbered years, with samples redrawn in 1985 and 2015. Since USA changes rapidly, policymakers in government and private organizations need current housing information to make decisions about programs that will affect people of all income levels, ages, and racial and ethnic groups.  Source: (<https://www.census.gov/programs-surveys/ahs.html>) | Dataset of people activities in house | URL <https://www.census.gov/programs-surveys/ahs/data/interactive/ahstablecreator.html?s_areas=00000&s_year=2021&s_tablename=TABLE1&s_bygroup1=1&s_bygroup2=1&s_filtergroup1=1&s_filtergroup2=1> This is online data base, so we don’t download the data. Data Selection can be done from the website as per the requirement. |
| 18 | National Residential Efficiency Measures Database | NREMD | <https://remdb.nrel.gov/>  <https://remdb.nrel.gov/about> | Released in 2013 | U.S. | Technology retrofit savings | A national unified database of residential building retrofit measures and associated costs. | Residential Building | Annual | The National Residential Efficiency Measures Database is a publicly available, centralized resource of residential building retrofit measures and costs for the U.S. building industry. With support from the U.S. Department of Energy, NREL developed this tool to help users determine the most cost-effective retrofit measures for improving energy efficiency of existing homes. NREL developed this database on behalf of the U.S. Department of Energy. The purpose of this project is to provide a national unified database of residential building retrofit measures and associated costs. These data are accessible to software programs that evaluate most cost-effective retrofit measures to improve the energy efficiency of residential buildings.  Source: (<https://remdb.nrel.gov/about>) | Retrofit savings | URL <https://remdb.nrel.gov/> This website gets access to data dictionaries. |
| 19 | Model America data and models of every U.S. building | NA | <https://www.osti.gov/dataexplorer/biblio/dataset/1774134>  <https://zenodo.org/records/6908189> | Released in 2021 | U.S. | Model of every building in the United States. | Building simulation data | Commercial and residential building | Minutes | The 5-year goal of the “Model America” concept was to generate a model of every building in the United States. This data repository delivers on that goal. Oak Ridge National Laboratory (ORNL) has developed the Automatic Building Energy Modeling (AutoBEM) software suite to process multiple types of data, extract building-specific descriptors, generate building energy models, and simulate them on High Performance Computing (HPC) resources. For more information, see AutoBEM-related publications (bit.ly/AutoBEM). There were 125,714,640 buildings detected in the United States and this dataset contains 122,930,327 (97.8%) buildings which resulted in a successful simulation. Future, annual updates have been proposed that may include additional buildings, data improvements, or other algorithmic enhancements.  Source: Source: (<https://zenodo.org/records/6908189>) | Dataset of Building in US. | URL <https://app.globus.org/file-manager?origin_id=62185ef6-7e40-44a6-913a-4b70dd111ef6&origin_path=%2FORNLNCCS%2F202104%2F10.13139_ORNLNCCS_1774134%2F> Model repositories which cannot be downloaded. |
| 20 | A synthetic building operation dataset | NA | <https://www.nature.com/articles/s41597-021-00989-6> | Released in 2021 | U.S. | A synthetic building operation dataset which includes HVAC, lighting, miscellaneous electric loads | Representative building data, Building simulation data | Office building | Minutes | A synthetic building operation dataset which includes HVAC, lighting, miscellaneous electric loads (MELs) system operating conditions, occupant counts, environmental meters, end-use and whole-building energy consumptions at 10-minute intervals. The data is created with 1395 annual simulations using the U.S. DOE detailed medium-sized reference office building, and 30 years’ historical weather data in three typical climates including Miami, San Francisco, and Chicago. Three energy efficiency levels of the building and systems are considered.  Source: (<https://www.nature.com/articles/s41597-021-00989-6>) | Building operation datasets | URL: <https://www.nature.com/articles/s41597-021-00989-6#Sec10> File size is 1.2 TB, which makes it hard to store in local computer. |
| 21 | Residential-Building-Stock-Assessment | RBSA | <https://neea.org/data/residential-building-stock-assessment>  <https://neea.org/get-involved/rbsa-information-for-participants> | 2016-2017 | U.S. | Representative sample of single-family, multi-family and manufactured homes gathered across the Northwest region | Represent usage of residence use of building applications | Residential | Annually | The RBSA is a comprehensive research study of energy use in single-family and multi-family Northwest residential buildings. The RBSA will provide the region with useful information for residences, utilities, and researchers to improve energy efficiency practices and identify opportunities for energy efficiency in the future. This study involves gathering information from on-site assessments at participating homes, apartments, and condos. These onsite assessments will take approximately 3-6 hours depending on the size of the residence. Evergreen Economics and its partner firms will collect information on energy-using equipment and building characteristics that affect energy consumption. They will also request authorization to collect electric and natural gas consumption data for all study participants from their electric and natural gas utilities. Participants will receive a $100 incentive for participation. There are also additional incentives available for residents who qualify for additional HVAC and ductwork testing. .  Source: (<https://www.census.gov/programs-surveys/acs/about.html>) | Residence application usage datasets/ | Directory: (‘/Database\_ 21\_RBSA/’) |