U.S. DEPARTMENT OF
ENERGY

*Office of*
ENERGY EFFICIENCY & RENEWABLE ENERGY

*Quick Guide*

# Considerations for Creating a Covered Buildings List for a Building Performance Standard (BPS) or Benchmarking Program

## Introduction

Implementing a successful Building Performance Standard or benchmarking program requires accurate and comparable building data.

The first step is to develop a "Covered Buildings List" (CBL), which consists of all the buildings covered by the BPS or benchmarking regulations developed by the jurisdiction, for example, all commercial buildings greater than 50,000 square feet. The CBL should include key characteristics that will be used to determine the BPS targets for each building.

This document provides an overview of the process and key considerations for developing a CBL. It is intended for BPS or benchmarking policy administrators and implementers to help plan the process and level of effort needed to develop CBL. It does not provide detailed technical guidance on data processing.

## Generating a CBL

The data sources, process and level of effort for generating a CBL will vary based on ownership category.

- **Jurisdiction-owned buildings**
  A list of the jurisdiction's buildings may be relatively straightforward to compile, although it still may involve getting data in different formats from different departments within different government entities and pulling that data together into one unified list.
- **Other government-owned buildings**
  If buildings owned by other government entities are included in a jurisdiction's BPS or benchmarking ordinance, compiling the list of buildings may be relatively straightforward. Obtaining the data may involve data requests to those entities, requiring additional effort.
- **Privately-owned buildings**
  The primary data source for privately owned buildings are city tax and parcel records. It may consist of merging data from multiple sources (See "Data Sources" below) and may take significantly more effort than government-owned buildings. Additionally, there may be privacy and access issues that make it more difficult to compile a complete list.

Generating the CBL from these sources may require some data transformations. Two common aspects in particular are:

- **Condominiums** ("condos") could be defined as the collection of associated condos which would be considered one building, or the condos could be defined individually, depending on how the BPS or benchmarking guidelines are structured.
- **Multi-use buildings** can be defined based on each use type or the dominant use type, and the guidelines for determining that should be clearly defined in the BPS or benchmarking rule making. The ENERGY STAR Portfolio Manager Glossary has definitions for multi-use buildings with different use types that could be used as a guide.

## Primary Building Data Needed for a CBL

The minimum data to collect includes:

- **Building physical address:** Different data sources may have different addresses for the same building; reconciling the data to determine the correct address may take time. The minimum information should include the street number, street name, city and zip code.
- **Contact information for building owners and managers:** This information is needed in order to make sure the basic building information is correct, as well as communicating requirements for complying with BPS or benchmarking program.

It will also be useful to provide

- **Building ID:** A unique building identifier to differentiate between individual buildings because buildings may be listed under multiple or changing addresses. This may already exist depending on the data source, but if not, it will be necessary to generate it. One option is to generate a Unique Building ID (UBID) developed by the U.S. Department of Energy.
- **Building Name:** This information is not essential but may be useful to help with the building identification.

A "Covered Buildings List" (CBL) consists of all buildings covered by a jurisdiction's BPS or benchmarking program.

## Building Characteristics Needed for a Covered Buildings List

Basic building characteristics are needed for a covered building list, including:

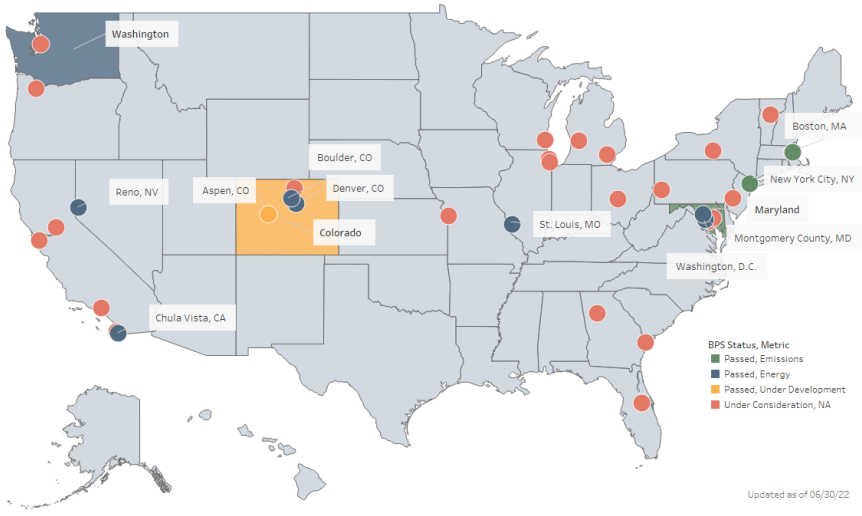
- **Gross Floor Area:** The total floor area represented by the building, excluding unconditioned spaces.
- **Year Built:** The year the building was originally built, which can be used to generate assumptions about the building construction.
- **Occupancy Type:** The primary occupancy of the building (e.g. office, retail, etc.), possibly from a defined list of types, such as from the ENERGY STAR Portfolio Manager, or other sources, such as the jurisdiction's occupancy codes or ASHRAE Standard 100. This can be used to generate assumptions about how the building is used if those data are not available from another source. Occupancy type is a key driver of the BPS targets.
- **Number of Buildings:** The number of buildings represented by the "building." In most cases, this value is probably 1, but in situations such as campuses or professional office parks, there may be multiple buildings on one energy utility meter so that the combined buildings are considered one building.
- **Last Renovation Year:** The year of the last major renovation of the building, if applicable and available. This might help determine the level of energy efficient retrofits that may have been applied to the building if more information is not available.

## Data Sources

There are many potential sources for building data, depending on whether municipal or privately-owned buildings are part of the CBL.

- **Land / Tax Parcel Data:** This data describes land parcels and typically has a unique ID assigned to each parcel, generated by the government entity that holds the data, such as a County Tax Assessor. It should have information about the parcel owner, but it may or may not have data about the buildings on that parcel. The parcel owner may not be the building owner, although the parcel owner may be able to provide information about the buildings and building owners. In many cases, this is the only data available and is where most jurisdictions start when trying to compile a CBL. Also, many ordinances may require BPS or benchmarking compliance to be tied back to the parcel owner, not the building owner. It may require quite a bit of work to identify specific buildings related to that parcel.

State and local building performance standards. Source: www.energycodes.gov/BPS




- **Government Data:** For municipal buildings, there may be multiple sources for building data including city, county, or state data.
- **Real Estate Data:** Real estate data is generally obtained from third-party sources based on potential or recent real estate sales transactions. There is usually a cost for this data, and there may be restrictions on how the data can be used, which might make it difficult to use for the CBL.

## Compiling the Data Sources

In many cases, it will be necessary to compile multiple data sources into one data set representing the CBL. The data may be in different forms, such as spreadsheets, text files, or databases, and it will be necessary to put them into a format that allows them to be merged together. The simplest form is "flat data," where each record (line) in the data file represents one entity, such as a building.

## Matching Data from Different Sources

In order to match data from different sources, there must be a field that is common between the data files. Examples of possible fields to match on are:

- **Address:** Addresses can be used to match records between data sources, but the accuracy of addresses can be quite poor and may result in false positive and false negative matches. However, it may be the only common field among the different data sources, and it can be time consuming to verify that the matches were made correctly.
- **Land / Tax Parcel ID:** This is a unique ID generated by the jurisdiction. Other building data may also contain this ID, and if so, it can be used to match the records in the data files.
- **Building ID:** If there is a common unique building ID in the data sources to be merged, this is ideal, although it is probably a rare circumstance where this will occur when generating the CBL for the first time.

Note that there can be different relationships between tax lots and buildings: single building on a single tax lot, multiple buildings on a single tax lot, one building across multiple tax lots, and multiple buildings across multiple tax lots.

## Data Quality

Data quality is a common problem resulting from data collection, transmission, and transformation errors. Erroneous data in the dataset can result in inaccurate analysis and misleading conclusions and decisions, which may result in a lack of confidence in the result and undermine the credibility of the BPS or benchmarking program.

Data cleaning criteria are critical for identifying erroneous data. Criteria should include:

- **Identify potential errors:** Although this is difficult, examining the data to determine its types of errors may be possible. For example, if the information is self reported it may be less reliable than if it is generated from a trusted source or automated process.
- **Categorize data problems:** It is useful to try to categorize the types of problems with the data. For example:
  - **Required data that is missing:** Records in this category should be flagged, and methods for acquiring the needed data should be developed, such as contacting the data provider.
  - **Data that exists but is suspect:** This might be data that is outside expected value ranges (such as a zip code that is outside of the jurisdiction zip codes or a floor area that appears too large or small for that building type), or maybe the default values were used rather than actual values. Data cleansing rules should be developed to flag these types of errors, and if possible, verify or correct the data.
- **How to handle erroneous data:** When developing the "data cleansing" rules, define how to handle erroneous data. For example:
  - **Contact the source of the data:** It might be possible to contact the building owner to verify and possibly fix data flagged as a problem.
  - **Delete the problem records:** If it is not possible to fix the data, or the records seem to be outliers, the best option is likely to delete those records from the dataset.

Examples of data cleansing rules.

| Data Issue | Cleansing Action |
|---|---|
| Building outside the defined geographic area (check for country, state, province, postal code) | Delete data outside range |
| Building outside the defined types for data collection | Delete data outside range |
| Duplicate entries | Delete duplicate records |
| Inconsistent units of measure | Convert to common units |
| Inconsistent formats (e.g. 100,000 vs 100K vs. 100000) | Convert to common format |
| Inconsistent naming conventions | Convert to common terms |
| Missing data, e.g.:<br>• Buildings with no floor area reported<br>• Buildings with no energy use reported | Follow up with owner to correct data |
| Obvious incorrect values/out-of-range checking, e.g.:<br>• Buildings with EUI < 5 or > 1000 kBtu/sf<br>• Floor area < 0<br>• Hospital EUI < 10 kBtu/sf/yr | Follow up with owner to correct data |

## Implementation Process and Effort

Collecting and compiling data is not trivial; do not underestimate the effort to compile and cleanse the data and allow enough staff time to complete these tasks. The work requires coordination with different departments or building owners, which can be quite time consuming. The data cleansing task can also be quite time consuming. It is important to have staff who can devote a substantial amount of time to this effort. DOE maintains several open-source data tools, and many private sector software applications exist to help entities compile and manage BPS and benchmarking programs.

## Resources

For general information on BPS and links to various resources, see: Building Performance Standards US Department of Energy, Energy Efficiency and Renewable Energy

For general information on benchmarking, see: Benchmarking and Building Performance Standards Policy Toolkit | US EPA

Github repository for generating geo-coded building footprints and UBIDs from street address data: https://github.com/SEED-platform/cbl-workflow ■

*Authors: Robin Mitchell and Paul Mathew, Lawrence Berkeley National Laboratory.*

U.S. DEPARTMENT OF ENERGY
Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

For more information, visit: energy.gov/eere/buildings

DOE/EE-2667 • November 2022