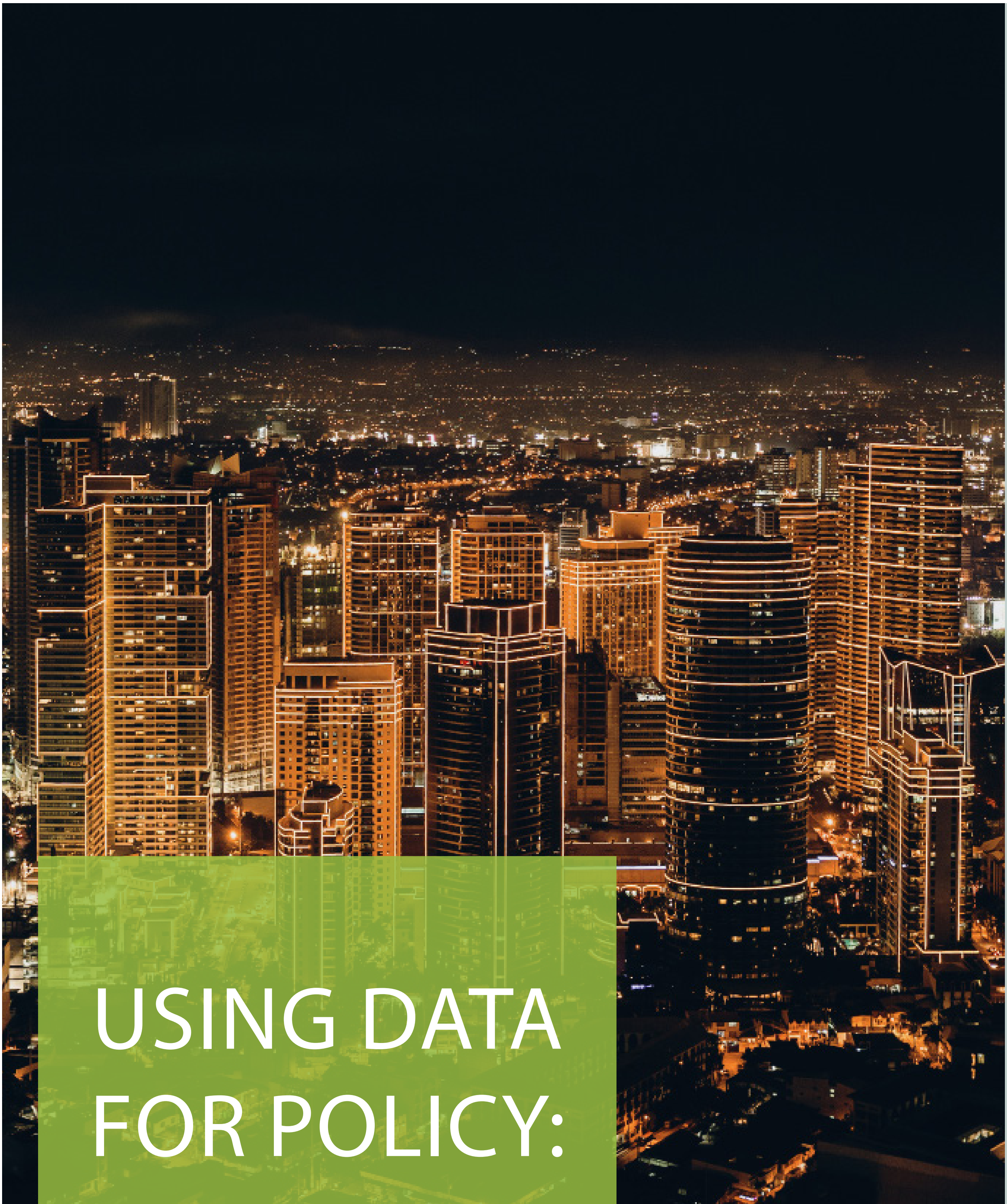
Version 1.0

March 2018



USING DATA FOR POLICY:

A MANUAL FOR C40 CITIES

C40 CITIES

BERKELEY LAB

C40 Cities

C40 Cities connects more than 90 of the world's greatest cities, representing 650+ million people and one quarter of the global economy. Created and led by cities, C40 is focused on tackling climate change and driving urban action that reduces greenhouse gas emissions and climate risks, while increasing the health, wellbeing and economic opportunities of urban citizens.

The C40/Arup Deadline 2020 report shows that increasing building energy efficiency is the most critical climate action for cities, as buildings represent around half of the scope 1 and 2 emissions in C40 cities on average. Cities should therefore prioritise the retrofitting of existing building stock, as well as establishing building energy codes and encouraging data reporting across new and existing estates. It is critical that most actions are deployed within the next couple of years, reaching 71% of total actions taken by 2020. Therefore, C40 is committed to helping cities create better buildings, by launching the C40 Building Energy 2020 Programme, in partnership with the Children's Investment Fund Foundation (CIFF) and ClimateWorks Foundation. In the framework of this programme, C40 commissioned Lawrence Berkeley National Laboratory (Berkeley Lab) to develop this manual on using data for policy. Berkeley Lab's Building Technologies and Urban Systems (BTUS) division performs analysis, research, and development leading to better energy technologies and reduction of adverse energy-related environmental impacts.

CIFF CHILDREN'S INVESTMENT FUND FOUNDATION

ClimateWorks FOUNDATION

CONTENT

1. INTRODUCTION

This manual provides guidance and best practices on how to use data for developing and implementing policy on building energy efficiency. The primary audience for this manual are the C40 cities in the Private Building Efficiency (PBE) network and the Municipal Building Efficiency (MBE) network. Most of the guidance is applicable to both PBE and MBE networks. Where guidance is specific to PBE or MBE, it is called out accordingly. The users of the manual include city policy makers, efficiency program administrators and data analysts as well as external consultants supporting them.

The manual is organized into five major activities:

1. Defining objectives and metrics
2. Data collection
3. Data cleansing
4. Data analysis
5. Communicating results

For each of these major sequential activities, we present two types of guidance:

- Process steps, i.e., basic guidance on how to conduct the activity.
- Best practices, including actual examples from various cities.

The manual aims to be concise rather than expansive. There already exist an array of in depth resources and examples of C40 cities using data for policy. Rather than duplicate their content here, the manual provides links to these resources as illustrative examples with more in-depth information.
Depending on your city's need, experience, and level of expertise, the manual may be used differently.

- Cities that have not yet begun to use data for policy, or are still new to it, should focus on the process guidance and follow each activity in sequence.
- Cities that have already begun to use data for policy may choose to focus on particular activities that they need more guidance on. For example, a city may already have a data collection system in place but may not have established data cleansing procedures.
- Cities that already have established procedures and experience with using data for policy may be interested in the best practices and can skip over the process guidance.

Appendices A and B includes quick assessment checklists for PBE and MBE cities respectively. Use these to assess your level of practice and expertise on using data for policy.

A note about scope: The scope of this manual is specifically limited to the use of data to develop and implement city-level policies related to building energy efficiency. Building energy data has many other uses not covered by this manual, e.g., using system and component level data to optimize building operation. Similarly, the manual is not intended to broadly cover all aspects related to policy-making — it is only focused on the data aspects of policy-making.

2. DEFINING OBJECTIVES AND METRICS

2.1. Process Steps

2.1.1. Define the primary objectives for collecting and using data

The first step in using data for policy is to define what you want to use data for, i.e., what questions are you trying to answer? Data analysis is only a means to an end. The scope, methods and level of effort for data analysis will vary based on the purpose of the analysis. Try to be as specific and precise as possible with the objectives and questions you are trying to answer with data. You may want to conduct a charrette to brainstorm analysis questions with a group of relevant stakeholders for your city.

Below are examples of data analysis objectives and questions. This is not an exhaustive list. Use it as a starting point to define the analysis objectives for your city.

Characterize and understand the energy use and GHG emissions of the building stock.

e.g., How much energy is used and GHG emitted by different types and sizes of buildings?

Some cities may not yet have basic data on the energy use and GHG emissions of the building stock that is needed to create a baseline. Others may have data on the overall use by the building sector, but may need to break that down into sub-sectors. All C40 cities are committed to compile a GHG inventory using the City Inventory Reporting and Information System (CIRIS) that is compliant with Global Protocol for Community-Scale Greenhouse Gas Emission Inventories (GPC). For more information about CIRIS and GPC, see this website[1].

Target specific buildings for retrofits.

e.g., Which buildings should be targeted in order to achieve the energy/GHG reduction goals?
Cities with efficiency programs may want to identify specific buildings for technical assistance, incentives, and retrofit actions.

Set energy / GHG reduction targets.

e.g., What should the % reduction target be for different building sectors? What should the target be for new construction vs. existing buildings?
Some cities may have established overall GHG reduction goals for their city. Such goals may need to be broken down into specific reduction targets for different types and ages of building. Cities may also be interested in setting goals for new construction as distinct from existing buildings.

Track and document energy/GHG reduction over time.

e.g., How much has energy and GHG reduced since 2010? Which sectors showed the greatest reduction? How much of the reduction is due to specific policies and programs?
Cities that have had policies in place for several years may be interested or have a requirement to document reductions from year to year.

Comply with the City Climate Planner

(CCP[2]) certificate program operated by the World Bank.

Comply with benchmarking and disclosure requirements.

A growing number of cities have passed ordinances requiring energy benchmarking and disclosure of energy performance (e.g. see this list for the US [3]). This requires cities to collect, analyze and display energy performance data. These data sets may also be used to support other policy objectives.

The scope of all ensuing activities flow from the objectives. Therefore, be sure to obtain explicit buy-in from all the key stakeholders and document the objectives.

Inform program development and implementation.

e.g., How does energy use relate to demographic metrics such as income, age, employment status? How does ownership type affect energy use? What is the relationship between energy intensity and economic activity?
Some cities that have implemented energy efficiency policies are interested in analyzing the relationship between energy metrics and demographics, economic activity, etc. This requires overlaying and combining energy data with other datasets such as census data, tax data, employment data, etc.

2.1.2. Determine scope, priorities and phasing

For each of the objectives, define the scope of the covered buildings. For example, many benchmarking ordinances in US cities only cover buildings that are greater than 50,000 sq.ft. Scope may be defined in terms of building type, size, electric demand, geographic region, age.

Prioritize the objectives. It may be desirable to pursue a host of objectives, but that may not be realistic given time and resource constraints. Each objective has its own set of data requirements. While there may be some overlap, there are some requirements that are unique to each objective and you may not have adequate resources to pursue all these objectives.

If the scope of the effort to too large to tackle at once, you may also consider phasing the effort. Phasing may be organized by building type, building size, or groups of buildings by zone. For example, New York City's law on commercial building energy audits covers all commercial buildings, but they only collect data on a subset of those buildings each year. For new efforts phasing is also an effective way to pilot an approach and then improve the approach before applying it more broadly. Figure 2.1 shows the phasing approach for Minneapolis' benchmarking requirements.

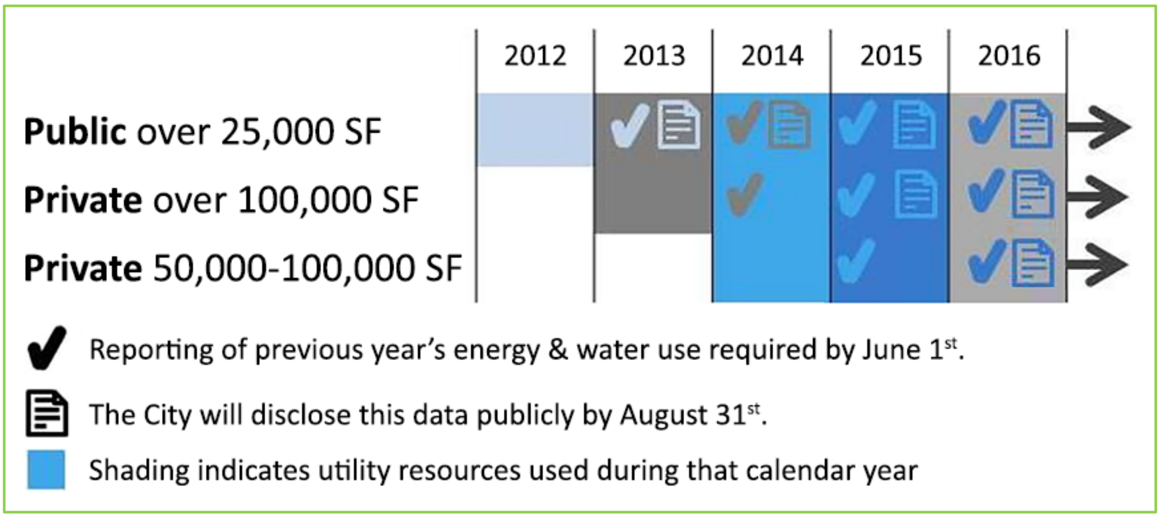


Figure 2.1: Phasing of building benchmarking requirements in Minneapolis.
Source: [2013 Energy Benchmarking Report[4]]

Quite recently, the Energy and Water Efficiency Ordinance issued by the City of Los Angeles required annual benchmarking reports to be submitted to the Los Angeles Department of Building and Safety in a phased schedule for private buildings, as shown in Table 2.1. Additionally, municipal buildings owned by the City of Los Angeles that are 7,500 square feet gross floor area or more must complete and submit an initial benchmarking report on or before December 1, 2017, and annually no later than June 1 thereafter.

Building Size (sq.ft.)	First Benchmark Reporting Date	Subsequent Benchmark Reporting Date
100,000+	December 1, 2017	June 1 thereafter
50,000+	June 1, 2018	June 1 thereafter
20,000+	June 1, 2019	June 1 thereafter

*Table 2.1. City of Los Angeles phased **deadlines**[5] for privately owned buildings.*

2.1.3. Identify key metrics

For each objective, identify key metrics and associated data needs. This may need to be done in an iterative manner based on the analysis approach (see Section 5). Table 2.1 shows examples of selected objectives and key metrics.

Objective	Key Metrics
Development of a GPC-compliant GHG inventory	Total city-wide electricity, natural gas, kerosene, diesel, biomass, furnace oil, LPG Broken out by sub-sectors: residential, commercial, Industries
Target specific buildings for retrofit	Individual building site energy use Individual building site energy use intensity (EUI)
Energy benchmarking/rating	Normalized individual building energy use intensity (EUI) Energy score or rating

Table 2.1: Examples of key metrics for selected objectives.
Note: this is only illustrative and not exhaustive.

Given that data collection is usually resource intensive and time consuming, we would reiterate that the metrics and data requirements should be carefully derived from the objectives and intended analysis, in order to ensure that all the required data are captured and also to avoid expending effort on unnecessary data. This is discussed in more detail in Section 3.

2.2. Best Practices

2.2.1. Ensure that policy statements and plans have measurable goals

Ensure that policy statements and plans have measurable goals that can be used to define clear objectives and implementation activities, and that these goals and targets are consistent with efforts to pursue limiting global temperature increase to 1.5 C above pre-industrial levels. Indeed, such action is needed to meet the collective ambition of the Deadline 2020: How Cities Will Get the Job Done[6] report, which lays out the GHG reduction targets and pathways cities must adopt in order to meet the 1.5 C goals of the Paris Climate Agreement. Strikingly, in order to meet such a goal, the report calls for 100% savings (net zero emissions) by 2050 for C40 Cities, against a 2015 baseline.

Copenhagen. In 2009, the Copenhagen Climate Plan[7] for CO2 neutrality in 2025, included an explicit goal for buildings: "The City of Copenhagen will achieve 10% of its total CO2 reduction by 2015 through construction and renovation projects."

New York City has developed several documents that state their measurable goals.

- 1.5 oC: Aligning New York City with the Paris Climate Agreement [8]
- New York City's Roadmap to 80 x 50 [9]
- One City Built To Last, Transforming New York City Buildings for a Low-Carbon Future [10]

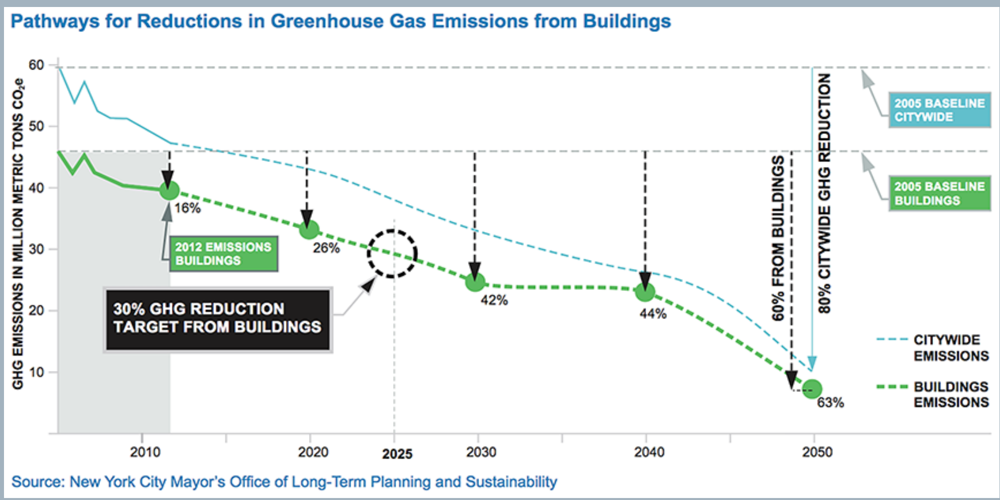


Figure 2.2 New York City GHG reduction target for buildings.
Source: One City Built To Last, Transforming New York City Buildings for a Low-Carbon Future[10]

The City of Seattle has the goal to be carbon neutral by 2050 and completed a detailed analysis to assess the reductions required in the building sector under various scenarios combining energy use reductions and fuel switching.

- Final Report: Building Energy Use Intensity Targets [11]

2.2.2. Convene technical experts

An advisory group of technical experts can help articulate the policy analysis objectives and define analysis use cases. The technical experts may be drawn from academia, real estate owners and operators, architects and engineers, energy consultants, and advocacy organizations. It is important to include a diversity of views and interests to ensure buy-in from a broad stakeholder group.

New York City convened a Buildings Technical Working Group [12] (TWG) that brought together dozens of leaders from a world-class real estate industry, architects, engineers, labor unions, academics, affordable housing experts, and environmental advocates to develop the right mix of policies and programs for new and existing buildings. This collaboration was, and continues to be, crucial to place the City on the pathway to an 80 percent reduction in greenhouse gas (GHG) emissions by 2050.

2.2.3. Conduct stakeholder engagement activities

Engage a diverse set of stakeholders to help develop the policies and determine the needs of potential program participants. This type of involvement in the early stage of policy formation can be a key success factor. Section 6.1 provides guidance on a stakeholder engagement process.

2.2.4. Create legislation to enable data collection and access

Often the greatest challenges in using data for policy are the legal and organizational barriers to accessing even the most basic data on building characteristics and energy use. Legislation can significantly reduce these challenges.

New York City has three laws that directly address data access:

- Benchmarking Local Law 84[13] requires buildings to submit energy benchmarking data through U.S. EPA Energy Star Portfolio Manager.
- Energy Audits and Retro-Commissioning Local Law 87[14] requires periodic energy audits and retro-commissioning measures. The audit data has to be submitted to the city.
- Sub-Metering Local Law 88[15] requires buildings to install electrical sub-meters for large non-residential tenant spaces and provide monthly energy statements

2.2.5. Develop interagency organization plan

If needed, develop a plan to organize and create consensus between different management divisions within a jurisdiction. This is especially critical if data analysis requires obtaining data from multiple sources. For MBE cities, it may be helpful to have a high-level coordinator who is empowered and authorized to direct data collection from multiple sources within the city administration.

3. DATA COLLECTION

The Sustainability Benchmarking Toolkit (Bosteels, et al., 2010), developed for sustainability benchmarking but applicable to all the policy objectives discussed in this document, lays out "Key data collection principles" in a very succinct way:

> "The success of the benchmarking process will be dependent upon collecting data which is accurate, consistent, replicable, verifiable, comparable and gathered over a sufficient time period to be able to discern trends. It is also important to ensure that data is collected over consistent time periods to enable the benchmarking process to take account of aspects such as seasonal variations in weather, which may influence the sustainability performance of a building. To ensure successful data collection, it is important that owners and occupiers engage and co-operate. Finally the data collection requirements should be realistic, achievable and practical."

3.1. Process Steps

3.1.1. Identify the data to collect

Based on the use cases and analysis objectives defined above, the next step is to identify the data that can be used to meet those requirements.

Generally, the use cases all involve obtaining the following types of building data (where a "building" can sometimes include multiple buildings, such as a campus).

General building data

The first steps is to compile a list of the individual buildings that fall into the categories defined in the use case definitions (e.g., Commercial buildings greater than a certain floor area).

- Municipal buildings

  A list of municipal 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.

- Privately-owned buildings

  Creating a list of privately-owned buildings can be much more complicated, and may involve merging data from multiple sources. There also may be data privacy and data access issues that complicate putting together this list.

Keeping in mind data privacy issues, the minimum data to collect would include the following:

- Building physical address
  The physical address of the building is the first piece of data to obtain, but different data sources may have different addresses for the same building. The minimum data should include the street number and street address and city. More detailed information could include GIS coordinates or similar physical location data.

- Contact information for the building owners and managers
  Contact information for the building is very useful, in order to follow up about building details. This contact could be the building owner, property manager, or building manager. If possible, gather contact information such as contact name, telephone number, email address, and possibly mailing address.

General Building Data: Basic	
Building ID	An ID of some sort to differentiate between individual buildings.
Building Name	Name of the building that may help in identification.
Building Address	Physical address of the facility including: ● Street number + Street name ● City ● State / Province
Contact Information	Contact person ● Name ● Telephone number ● Email address ● Role (building manager, owner, etc)

Building characteristics

It is necessary to compile at least a minimum amount of data of building characteristics that may be relevant to the energy consumption of the building that will be useful in the final analysis of the data. This can be a wide range of information, from basic to detailed.

- Basic information

Basic building characteristics can be a small number of fields, but should be relevant to identifying the building as well as possibly influencing the energy consumption.

Building Characteristic Data: Basic	
Total Gross Floor Area	The total floor area represented by the building. This may be associated with the energy consumption data (depending on how the building is metered for each energy type).
Year Built	The year the building was originally built, which can be used to generate assumptions about the building construction.
Occupancy Type	From a defined list, such as from the ENERGY STAR Portfolio Manager tool. This can be used to generate assumptions about how the building is used if more details are not available.
Number of Buildings	The number of buildings represented by the "building", if appropriate. In most cases it will be one, but there may be situations where there are multiple building associated with an energy utility meter.
Latest Renovation Year	The year the last major renovation was done on the building, if applicable. This could help determine what level of energy efficiency retrofit may have been applied to the building if more details are not available.

- Detailed information

Data from a building audit can include very detailed information about the building, including a breakdown of different use-types within a building, which may mean collecting different data depending on the use type. The ENERGY STAR Portfolio Manager (ESPM) documents listed in the box on page 19 give an example listing of property types and additional information to collect based on the property type. (ESPM uses the term property rather than building).

Here is an example of additional detailed data for a Bank Branch property type:

Building Characteristic Data: Detailed: Bank Branch
Weekly operating hours
Number of computers
Number of workers on the main shift
Percent that can be heated
Percent that can be cooled

Even more detailed information can be obtained from a building audit, where data is collected about the building construction, heating and cooling and cooling equipment, lighting, and so forth. Such detailed building characteristic data is useful when trying to determine the relationship between building characteristics and the energy performance of an individual building. It can also be used to correlate building characteristics with energy consumption levels on a broader set of buildings, if there is a statistically significant pool of data.

Building Characteristic Data: Detailed
Window to Wall ratio (WWR)
Type of glazing (for each orientation)
Wall insulation (Insulation value or Yes/No)
Type of cooling system and efficiency rating
Type of heating system and efficiency rating
Type, wattage and number of installed lighting fixtures
Heating and cooling setpoints and schedules
Building energy management systems (e.g., lighting sensors and controls)

Building energy data

It is necessary to have at least one year of energy consumption data for the buildings in the Building List, preferably with information about when the energy measurement period started and ended. It is important to get actual measured (not estimated or modeled) energy data.

This data can range from simple to detailed, such as simple annual energy consumption for an entire building (or a group of buildings such as a campus) to monthly energy consumption or even 15 minutes interval data for each meter in a building.

The energy consumption data should include the type of energy consumed (natural gas, electricity, diesel, etc) as well as the amount.

- ### Basic information

Basic building energy data should at a minimum include the energy use at the building site, which is what would be read from a meter or a utility bill. This is going to generally be divided up by the fuel type for the meter or the utility bill.

Building Energy Data: Basic	
Total Annual Site Energy Use by Fuel Type	Annual energy consumption by fuel type, such as electricity, natural gas, fuel oil, etc, from either meter data or utility bills.

- Detailed information

More time-resolved energy data will allow analysis of consumption trends and identification of potential efficiency opportunities. The time intervals are typically monthly, hourly, or 15 minutes.

Building Energy Data: Detailed	
Monthly Site Energy Use by Fuel Type (and/or meter)	A breakdown of site energy consumption by month by fuel type, such as electricity, natural gas, fuel oil, etc, from either meter data or utility bills.
Interval Site Energy Use by Fuel Type (and/or meter)	Site energy consumption, by fuel type, reported at an interval finer than monthly, such as hourly or every 15 minutes. This type of data can be used to see energy consumption trends and patterns that are not possibly at more coarse intervals.

Below are several examples of the types of data to collect:

- ENERGY STAR Portfolio Manager (US) [16]

ENERGY STAR Portfolio Manager (ESPM) is a web-based tool that can be used to track energy consumption for commercial buildings.

- Property Types, Definitions, and Use Details[17]

This file lists all the Property Types (examples range from Financial Office, to Museum to Worship Facility) as well as the detailed data that is useful to collect based on the property type.

India Commercial Buildings Data Framework: A Summary of Potential Use Cases[20], (Mathew, et al., May 2016)

This report contains example data to collect for different occupancy types, such as Hotel, Hospital, Educational establishment, Retail establishment, Office and Restaurant

- Data Collection Worksheet [18]

This web-based tool allows the user to select a country and a property type, and then the suggested data to be collected is displayed, and can then be saved out as a worksheet as either a MicrosoftTM Word or a PDF file.

- Full List of Portfolio Manager Reporting Metrics [19]

This is a list of data that is both entered into ESPM by the user as well as data that is calculated by the software. This is a very thorough list of property data.

Sustainability Benchmarking Toolkit [21], (Bosteels, et al., 2010)

This document contains a good summary of benchmarking best practices, with examples of fields to collect and how to move from simple to more complex data collection.

3.1.2. Identify data sources

There are many potential sources for data about buildings. As noted earlier, it should be easier to collect information about municipal buildings compared to privately-owned buildings.
Potential sources for building information include:

- Land / tax data

This data is a description of land parcels generally with a unique ID for each parcel, most likely obtained from a government entity. If it is tied to property tax collection it should have information about the property owner. However, it may or may not have detailed information about the building or buildings that are on the property. In addition, the property owner may or may not be the same as the building owner, although the property owner may be able to help in obtaining building and building owner information.

This data can be useful, but may not be enough information alone to identify the buildings for the use case.

- Research data

It may be possible to obtain data from universities and research institutions, although many times the data that they are working with has been made anonymous for privacy reasons and so may not be useful.

CoStar[22] : The largest private global real estate information service, with data on over 4.5 million commercial real estate properties. Growing focus on reporting energy and environmental performance, listings now include Energy Star and LEED related buildings in the U.S.

- Private consultants

Data may have been collected by private consultants that can be used to construct a building list. However, once again, there may be privacy issues for privately-owned buildings such that the consultants may not be able to release useful information (at least not without consent from the building owners).

- Government data

Particularly for municipal buildings, there are many government sources for building data including City, County/Province, State, National departments and colleagues.

- Real estate data

Real estate data is usually obtained from third-party data sources, based on data that has been collected with regard to potential or recent real estate sales transactions, and which is available for sale. It is generally based on building-specific information, and may have information about building owners and managers.

- Energy data from building owners, facility managers, and ESCOs

For benchmarking and analyzing building energy consumption patterns, at least one year of energy consumption data is needed. Annual energy consumption values, by energy type, should be available from a building owner or manager from their utility billing information. More detailed monthly data, that should also be available from utility billing data, can also be very useful for determining the energy consumption patterns of a building. For benchmarking and analyzing building energy consumption patterns, at least one year of energy consumption data is needed. Annual energy consumption values, by energy type, should be available from a building owner or manager from their utility billing information. More detailed monthly data, that should also be available from utility billing data, can also be very useful for determining the energy consumption patterns of a building.

For benchmarking and analyzing building energy consumption patterns, at least one year of energy consumption data is needed.

Keeping in mind data privacy issues, and the possible need for owner permissions, facility managers and ESCOs who are responsible for managing facility costs and energy consumption may be able to share collected or monitored data on buildings energy consumption, equipment energy performance, building activities and occupancy. (Iyer, et al., 2016)

Building energy data can be collected directly by the jurisdiction needing the data, or by using tools that have been developed for this purpose which allow building owners or managers to enter their own energy consumption data which can then be collected (via the software tool) by the jurisdiction.

Energy Star Portfolio Manager (United States and Canada):

For collecting building energy data (as well as general building characteristics). The tool is a web-based application.

NABERS[23] (Australia):

NABERS is a national rating system that measures the environmental performance of Australian buildings, tenancies and homes. Put simply, NABERS measures the energy efficiency, water usage, waste management and indoor environment quality of a building or tenancy and its impact on the environment. The tool is a web-based application.

Green Mark Scheme [24] (Singapore):

BCA Green Mark is a green building rating system to evaluate a building for its environmental impact and performance. It provides a comprehensive framework for assessing the overall environmental performance of new and existing buildings to promote sustainable design, construction and operations practices in buildings. The tools are Excel-based.

Building energy data can also be collected for specific buildings by installing automated building energy monitoring systems, which may include "smart meters" that typically provide 15-min interval data.

- Energy Key (Copenhagen):

For municipal buildings, the Copenhagen Properties and Procurement group is working with Copenhagen Utilities to establish a monitoring program for their entire municipal property portfolio by remotely reading heat, water and electricity meters on an hourly basis. This allows strategic planning for energy efficiency investments and evaluation of the success of the projects.

- BuildSmart DC Data[25] :

The District of Columbia (Washington DC) has collected data about their municipal buildings and made it available to the public. This is an example of a jurisdiction collecting data from utilities, such as electricity interval data from smart meters delivered daily by the local utility, as well as natural gas and water usage, and building profile data. The DC Department of General Services collects and manages this data, and then makes it available to the public (see more information under Section 6 Communication).

- Energy data from energy provider / utility data (automated)

Another option for gathering energy consumption data, rather than having the building owner or manager enter their energy consumption data by hand, is to set up an automated system for obtaining the data on a regular basis. This can be an ideal scenario, but can be difficult to set up. There are privacy issues which many times prevent anyone other than the owner of the data (such as the building owner) from obtaining the data directly from the energy provider. It is possible to set up programs with utilities which allow building owners to designate "third parties" (e.g., the jurisdiction needing the data) who can collect their energy data.

- Audit data

Detailed building audit data can provide useful information about the characteristics of a building that may influence the energy performance. Data typically collected in an audit will include floor area, wall area, building envelope construction types (including insulation levels), window types, mechanical system descriptions, etc. This type of data is only useful to collect if it can contribute to an analysis of the energy performance of a building, or be used to group buildings with similar characteristics. There are many calculation tools that can use audit data to calculate energy data to establish a baseline that can be used to evaluate the performance of energy retrofits. Audit data is also used across EU countries for issuing Energy Performance Certificates (EPC) according to the European Building Performance Directive.

EU Building Stock Inventory[26] :

The Directorate-General for Energy (DG Energy) is a department in the European Commission (EU) which focuses on developing and implementing the EU's energy policy. The EU Building Stock Observatory, a project of the DG Energy, monitors the energy performance of buildings across Europe, in order to assess improvements in the energy efficiency of buildings and the impact of this on the actual energy consumption of the buildings sector overall. The Observatory tracks energy efficiency levels in buildings in individual EU countries and the EU as a whole, different certification schemes and how they are implemented, financing available for renovating buildings, and energy poverty levels across the EU. The Observatory contains a database, a datamapper and factsheets:

EU Buildings Database[27] :

A website which allows access to a database of buildings in the EU, with filters by categories such as country, building stock characteristics, energy consumption, and certification, among others. The filtered data can be graphed, and downloaded as a MicrosoftTM Excel file.

Energy Performance Certificates (EPC) across the EU [29]

Published by the Buildings Performance Institute Europe (BPIE) in 2014 - a study evaluating the quality, availability and useability of EPC data and providing examples of good practices.

Building Energy Asset Score [30] (United States):

A web based national standardized tool, developed by the United States Department of Energy, for evaluating the physical and structural energy efficiency of commercial and multi-family residential buildings. It also has a template for collecting energy audit data.

EU Buildings Datamapper[28] :

The data mapper shows the information in the EU Buildings Database in graphic form, with filters for various characteristics such as floor area, age band, and so forth.

- Architectural or engineering drawings

Architectural or engineering drawings can potentially provide enough information and data to define basic, and possibly very detailed, building information. This method of acquiring data is potentially quite time consuming, but for situations where a detailed analysis of specific buildings is desired, this can be a good avenue for collecting that information.

3.1.3. Identify the form/format of the available data

The data being collected may be available in several different forms. If it is a file that has been exported from a database, likely it will be in one of the following forms.

Flat data vs hierarchical data

One key element of the data form is whether the data is "flat" or "hierarchical". Determining whether the data is flat or hierarchical is important because some tools will not have the ability to work with hierarchical data, and that data may need to be "flattened" before it can be used. Hierarchical data is the proper form for this type of data in a database, but can be difficult to manage outside of a database, such as in a spreadsheet because it is necessary to know how the different records (rows) are linked together, in this case by the Building ID.

- Flat data: each record in the data file represents one entity (such as a building), and all the fields associated with that entity are contained in that record. An example for a building with monthly data would be that all the monthly data is contained in the one building record, and there is a field for each month.

Building ID	Jan Elec Use	Feb Elec Use	Mar Elect Use	Apr Elec Use	May Elec Use
101	100	150	152	79	65

- Hierarchical data: data that has hierarchy means that there are potentially multiple records associated with one record. An example is that table would have records which represent buildings, and another table (set of data) that would have one record per month, and each building would be associated with one or more of the monthly data records. Hierarchical data can be represented in spreadsheets, but is generally represented either in a database or an XML or JSON file.

Building ID	Month	Elec Use
101	Jan	100
101	Feb	150
101	Mar	152
101	Apr	79
101	May	65

The relationship of the hierarchical data in a database is shown below. If the data management tool does not have the capability to deal with this type of hierarchical data, the records in the Building Energy Consumption table may have to be "flattened" so that all the monthly data is in one record (on one line) and then added to the associated record in the Building table.

If the data is only available in such as database-centric hierarchical form, it may be necessary to obtain help from a database expert to put the data into a form that is appropriate for how the data will be managed for the energy efficiency program.

The Building Energy Consumption table has a hierarchical, one to many relationship with the Building table.

Therefore, there may be many lines for one building in the Energy Consumption table.

The Energy Consumption data may need to be "flattened" ,depending on the data management tool capabilities.

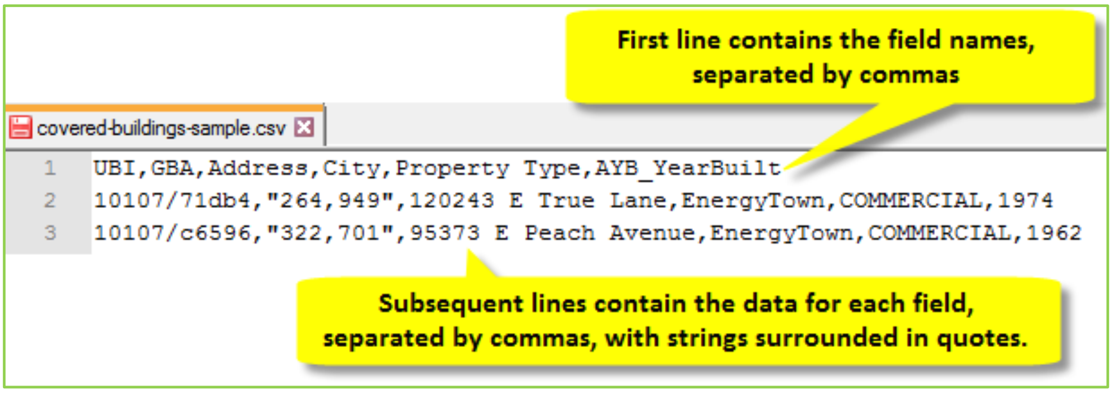
Building table

Building ID	Floor Area
101	55,000
102	45,203
103	33,403
104	12,399
105	3,209

Building ID	Month	Elec Use
101	Jan	100
101	Feb	150
101	Mar	152
101	Apr	79
101	May	65
102	Jan	235
102	Feb	322
102	Mar	40
102	Apr	22
102	May	10

Data formats

- CSV or other delimited format: A "comma separated values" (CSV) text file, which can be opened either with a text editor or most spreadsheet and database programs, will contain the data as a series of values separate by a comma delimiter, where each value represents a "field" or what would appear as a column in a spreadsheet. Generally, the first line should define what each of the fields is.



Here is what the CSV file above would look like if it was imported into a spreadsheet program.

	A	B	C	D	E	F
1	UBI	GBA	Address	City	Property Type	AYB_YearBuilt
2	10107/71db4	264,949	120243 E True Lane	EnergyTown	COMMERCIAL	1974
3	10107/c6596	322,701	95373 E Peach Avenue	EnergyTown	COMMERCIAL	1962
4						

Spreadsheet: Some data may be provided in a spreadsheet format. Depending on the type of spreadsheet the data is formatted for, there may be multiple "tabs" containing different types of data. In order to read this data, make sure that you have spreadsheet software that can open the spreadsheet format of the data.

Database: Getting data in a database format may be more unusual, because it is less flexible for opening and manipulating, but you may get data in this form. You will need to have database management software capable of opening the database, which will depend on the type of database that it is. Databases can be "relational", meaning that they contain data tables with hierarchical relationships.

- XML or JSON: Some data may be provided as one or more XML (Extensible Markup Language) or JSON (Javascript Object Notation) files. These are text files that are designed to be read by a computer (although they are quite human-readable also) that may contain a large quantity of data. They can also be used to represent hierarchical data with nested tags.

Generally, even though these formats are human readable, you will need a software tool to translate them into a usable form for data cleansing and analysis.

- XML format example: XML files have “tags” (words enclosed in brackets, e.g., <book>) that are used to describe a set of data. There is an opening and a closing tag to delineate the containment of the data

```
<books>
  <book>
    <title>The Cat in the Hat</title>
    <author>Dr. Seuss</author>
    <price>6.99</price>
  </book>
```

- JSON format example: The JSON format is less verbose than XML, and contains data in value: pair formats (e.g., “price: 6.99”). The data is grouped using brackets.

```
{
  "books" : [
    {
      "title" : "The Cat in the Hat",
      "author" : "Dr. Seuss",
      "price" : 6.99
    },
```

Energy data in XML format: Examples of data that may come in the form of either an XML or JSON file might be interval data from a smart meter.

Green Button: XML data format for energy data (United States)

- Green Button Alliance [31]
- Green Button Data [32]

Below is an example of Green Button interval data:

```
<interval>
  <duration>31622400</duration>
  <start>1357027200</start>
</interval>
<IntervalReading>
  <cost>709</cost>
  <timePeriod>
    <duration>3600</duration>
    <start>1357027200</start>
  </timePeriod>
  <value>147</value>
</IntervalReading>
```

Building and energy data in XML format: There are also several XML formats available for building characteristics.

Residential Buildings
Home Performance XML (HPXML) for Residential Buildings (United States)

- HPXML Online [33]
- HPXML[34] (National Renewable Energy Laboratory, Golden, Colorado, USA)

Commercial Buildings
- Building Sync[35] (United States)

3.1.4. Identify data management needs

Once you have identified the data that you want to collect and where to obtain the data, you need to consider how you will manage the data. The complexity of the system to manage the data will depend partly on how many buildings you are tracking, and how many different points of data you have for each building.

Data from different sources that needs to be merged

If you have a situation where you are collecting data from multiple sources, for example from parcel data and real estate data, that you need to merge together, you will need to determine how to manage that merge.

- Matching fields

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

- Address

A property address is probably the field that is most likely to be in most of the data sources that are being merged together. The accuracy of addresses can be quite poor, and it is possible to have both false positive and false negative matches based on building address. However, if it is the only field that is common between different data sets, that is the field that should be used. It will then be necessary to confirm that the addresses that were matched are actually the same building, which can be very time consuming.

- 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.

- Land parcel ID

Many jurisdictions assign unique IDs to their parcels for payment of taxes and other record keeping. Therefore, it might be relatively straightforward to have this ID in a property list generated from parcel data. In order to match on this ID, it needs to also be in other data, such as real estate or energy data. In the ENERGY STAR Portfolio Manager (ESPM) program, many jurisdictions have requested custom IDs for their particular parcel ID, so that when a property owner enters their building data into ESPM, there is a place to enter to parcel ID. Then that ID can be used to match the ESPM data to the property list.

Data storage and management

The amount and complexity of the data being managed (which is a combination of the number of buildings or records, and the number of characteristics or fields) will determine the complexity of the data management system. There are several possibilities for managing the data. Generally, it is good to start with a small set of data which can be managed by a simple data management system. However, as soon as the number of records increases, it may be necessary to either develop or adopt a tool designed specifically to manage data.

- Spreadsheet

Many times, the tool that jurisdictions start with for managing their data is a spreadsheet. This can be a good starting point, but depending on the complexity and amount of the data, you may not be able to effectively manage the data in a spreadsheet, particularly as you collect more data over time.

- Database

A database allows defining more complex relationships between data, as well as sophisticated data querying capabilities, but requires a knowledgeable database manager to set up and work with the data. If you have access to a person with those skills, setting up a database to house your data may be more efficient that trying to maintain a series of spreadsheets.

- Software tool for data management

Depending on the sophistication and complexity of the data you are managing, a software tool specifically designed to manage this type of data may be warranted. You can develop and in-house tool, or you can license or purchase a software tool. In either case it is strongly advisable to have a software tool consultant to help define software requirements, evaluate cost-benefit tradeoff for different features, and manage tool procurement, development and installation.

Identify how to take the data in raw form into the data management tool

In most cases, if the data is in a CSV or spreadsheet form, it should be straightforward to import it into a data management tool. However, if the data is in a more complex form, such as a database, XML or JSON form, it will be necessary to use tools that can read those files and put them in the form that your data management tool accepts.

Updating the Data over Time

The data management needs will include updating data when data quality checks have been resolved (and possibly tracking the fact that it was updated and when), as well as tracking final cleaned and verified data year over year. It may also be necessary to update the building characteristics as the occupancy changes or the building is upgraded with energy efficiency measures. (Bosteels, et al., 2010)

3.2. Best Practices

Data collection for building energy analysis is almost always resource intensive, time consuming and highly prone to data quality issues. Therefore, the scope and priorities for data collection should be carefully assessed and determined based on several key considerations. (Mathew, et al., 2016)[36]

3.2.1. Prioritize data collection requirements

Because data collection is expensive and time consuming, it is necessary to prioritize what data is actually needed for the desired use case analysis. It is good to determine the data that is actually required to generate the desired results, versus the data that might be "nice to have" but isn't required. It is important to consider the value-of-information tradeoff, e.g., what are the marginal costs and benefits of getting more data.

Start with the use case, not the data

Always use the specific analysis requirements of a use case to determine data needs and priorities. In other words, each data field should have an explicit reason for being included in a data collection effort — either as an input for a "key performance indicator" or a normalizing/clustering variable.

Assess the likelihood of poor data quality

Some fields may seem easy to collect but may be highly prone to poor data quality. For example, experience indicates that even a seemingly basic data field such as gross floor area can be significantly misreported. For certain building types, alternative measures of floor area may be more reliable. For example, net leasable area is likely to be more reliable because it is has a critical business purpose in leased buildings.

Consider the level of effort

The level of effort required to collect data varies significantly across data fields. Obtaining the number of guest rooms in a hotel is orders of magnitude easier than obtaining a detailed end use energy disaggregation. It may be worthwhile to assign a 1-5 score for level of effort required to collect the data for each field and use that as a consideration when prioritizing which fields to collect. For critical fields that are very difficult to collect, consider proxy fields that may require less effort. For example, use the nameplate efficiency of a chiller if the actual operational efficiency is not easily obtained.

Statistical sampling vs. 'opportunistic' data collection

Some use case analysis questions, e.g., obtaining a national or state-level estimate of sector-wide energy use, clearly require using formal statistical sampling methods. However, sampling may require collecting data from buildings for which data collection is especially difficult or even impossible. An alternative approach is to collect data 'opportunistically' i.e., pursue data collection from entities that are supportive and capable of providing data, e.g., large portfolio owners. In theory, such a dataset will not be a true statistical sample but may still be able to address most use case analysis questions with a reasonable level of rigor.

Breadth vs. depth of data collection

As with any data collection effort with a constrained budget there is a tradeoff between the number of buildings from which data is collected and the amount of data collected from each building. Use case priorities will determine this tradeoff. For example, an initial data collection effort may choose to focus on only a few geographic regions in order to afford more in-depth data for each building. For a program that is just starting out, it might be wise to collect detailed data for a small group of buildings, to see which variables are useful and which are not. That information can then be used to determine the most useful data for a larger group of buildings.

Remote vs. on-site data collection

In general, remote data collection (e.g., via telephone, web survey forms, email) requires less effort than on-site data collection. For the scope of data fields addressed with this set of use cases, it may be difficult to completely avoid site visits without seriously compromising data quality, especially for building system characteristics data fields. However, the time spent on-site could be minimized by collecting as much data as possible remotely.

Limit the number of touch-points for obtaining the data

No one person or documentation system will likely have all the data required for the use cases in any given building. However, as much as possible, the number of touch-points should be limited in order to ease data collection effort. For example, for large portfolio owners there may be a central repository that contains data across all buildings at least for certain data fields.

Minimize the burden on the data provider

Any tactics that help reduce the time spent by the data provider will help ease data collection. For example, if some data are located in certain documents (drawings, specifications, etc.), the data collector could offer to look up the data in those documents rather than requesting the data provider to do the same.

3.2.2. Assess tradeoff between tool complexity and difficulty of use

If the amount of data collected is relatively small (a few hundred records) without too much complexity (in terms of the number of fields collected), having a simple tool to collect and store the data may suffice. However, as the data sets become larger and more complex, it may be necessary to move to a more sophisticated data management tool. Such a tool may require more resources to learn, but may be worth it if it allows more data to be managed in an efficient manner.

3.2.3. Assess in-house vs third party tools

Depending on the sophistication and complexity of the data you are managing, a software tool specifically designed to manage this type of data may be warranted. You can develop an in-house tool, or you can license or purchase a software tool.

Develop an in-house data management tool

Developing your own internal data management tool is an option if you have an IT department (or other programming resource) and can afford to develop your own tool. Below are examples of cities that have developed their own data management tools.

New York City: New York city has Energy Audit and Retro-commissioning Data Collection Tools in MicrosoftTM Excel formats. The tool includes a list of Retro-Commissioning Measures (RCMs) which can be used to determine compliance with operating protocols and maintenance status , with a section to explain the correction if there was a deficiency and annual energy savings for each RCM.

Energy Audit Data Collecting Tool & Retro-Commissioning Data Collection Tool [37] (Excel)

Copenhagen: Energy Surveillance for Publicly Owned buildings

- Heat, water and electricity meters providing hourly data
- With utilities, establishing central energy surveillance on entire municipal property portfolio
- Allows strategic investment planning, provides management information
- Risk minimization — quickly discover problems (water leaks, etc.)
- Energispring (Energy Jump): sharing energy data between large private building owners, as well as renter of those buildings

Paris Supervision Project: The goal of this project is to develop a method to remotely manage the municipal building heating systems. Much of the infrastructure for this remote monitoring is in place, and the project should be completed by the end of 2018. Goals of the project include:

- Monitor 1500 municipal buildings on a data platform covering their energy performance
- Provide better insights of the performance of heating systems in these buildings with this new data infrastructure
- Study Consumers behaviors through cross data analysis
- Allow strategic energy efficiency plans and investments based on data in actual buildings

License / purchase software data management tool

There are many tools available that are specifically for tracking benchmarking or other building energy data. If you want to use such a software tool, we recommend that you have a person with specific expertise in databases (such as someone from an internal IT department) help evaluate the tools that are available.

Standard Energy Efficiency Data (SEED) Platform [38]: Benchmarking data management tool

3.2.4. Consider non-technical aspects of data sources

There may be many cases, particularly for privately-owned buildings, where the data sources will be difficult to obtain, for a variety of reasons listed below.

Data privacy

Because some of the data sources may contain personal or proprietary data, it is important to identify data privacy issues, which will likely determine what data you are able to collect.

Data access rights

In some situations, even if there are not data privacy issues, there may be issues with who owns the data and whether it can be given to another entity for analysis purposes. This can be true of energy provider data for private buildings. The owners or managers of the buildings may need to give "third-party" access to the data in order for a city to collect and analyze it.

Ease of data access

In some cases, obtaining the data is not necessarily easy. For example, detailed energy consumption interval data, probably available as an XML file, might not be easy to obtain for multiple buildings. In this situation, setting up an automated system to obtain the data, either with the energy provider or the building owner or manager, may be the best solution, but may also be complicated to implement.

3.2.5. Develop a unique building ID

Although it is not required, we highly recommend developing a unique building ID for all the buildings you will be tracking. There may already be such an ID in your data sources, but many times there is not, which requires that such an ID be developed. The United States Department of Energy is developing a "universal" unique building ID that may be useful for this purpose. More information will be provided as that project matures.

3.2.6. Develop robust methods for matching data from different sources

If you need to match data from different sources, for example tax records and real estate data, or your building list with energy provider / utility data, it is necessary to develop methods for matching those records together.

There needs to be at least one field that is common between the data sets being matched that can be used as the "matching" field, as discussed in Section 3.1. Possibilities include:

- Building address

Some form of physical address for a building is usually contained in most data sources, but the data may not be expressed in a similar way in each source. There may be abbreviations in one set of data and not in another, so that it is necessary to match "101 N Main St" to "101 North Main Street". Therefore, the addresses need to be "normalized" before they can be matched, which requires a level of sophistication that may be difficult to achieve without some sort of address interpretation algorithms. If you have access to a programmer, there are libraries available in various programming languages that contain standard ways to normalize addresses.

- Tax parcel ID

If the data sources have tax parcel IDs in them, that is a potential field for matching the data. For example, tax parcel information should contain a tax parcel ID, and many times real estate data will also contain the tax parcel ID. So it may be possible to match data based on the tax parcel ID. However this may not translate directly into a building, because many times there are multiple buildings on one parcel or a building spans multiple parcels.

- Building ID or other identifying ID

If the data sources being merged have another ID that is common between them, that can be used as a matching field.

4. DATA CLEANSING

Bad data is a common problem and can result from errors in collection, collation, transmission and transformation. It contaminates the data set and can lead to inaccurate analysis and erroneous conclusions and decisions. Consistently bad data can lead to a lack of confidence in the results, potentially undermining the credibility of the underlying program or policy.

Therefore it is necessary to have criteria to identify bad data and then decide how to handle it, which is generally called data cleansing.

4.1. Process Steps

4.1.1. Identify the potential errors in the data being collected

Examine the data to determine, to the extent possible, what types of errors it contains. This will not find all the possible error conditions, but is a place to start the process.

In many cases, at least some of the data is "self-reported" either by a building owner or manager who may not be an energy expert, and as a result, errors are inevitable. (Hart, 2018). If possible, identify the data that comes from such self reporting and give that data special attention during the data checking procedures. In comparison, data that is automatically generated and downloaded (as opposed to data that is entered by hand), such as energy consumption data from a utility database will potentially need less scrutiny.

4.1.2. Develop an explicit set of data cleansing rules and procedures

Determine what is an acceptable level of accuracy for a data record to be considered "complete". This will probably mean that there is no missing data in required fields and the data is reasonably accurate. (Hart, 2018). From that criteria, develop a document that contains the specifics of the data cleansing rules and procedures. It is best to start with a simple set of rules, and augment them over time based on errors found in the data as part of the data cleansing process. Look at examples of error checking rules from other jurisdictions to ensure that a robust set of error checking rules is developed. (see Best Practices below).

Categorizing data problems

When developing the data cleansing procedures, it is useful to develop categories for data issues. Possibilities might include:

- Required data that is missing

This is data that is required for the final data analysis but has not been included in the data being checked. Any record with this problem should be flagged as such, and procedures for obtaining the missing data should be defined (such as contacting the data provider).

- Data that exists but is suspect

This is data that exists in the data set, but after the data quality checking is found to be outside of the expected ranges, or might be default values. However, the data could be accurate, even if it is an outlier. This data should be flagged, and verified if possible.

What to do with bad data?

Included in the data cleansing procedures should be a clear definition of how to deal with bad data. Some options include:

- For missing and suspect data, attempt to contact the source of the data (such as a building owner) and request that they fix the data and resubmit it.

- For missing data, it may be possible to "fix" the data as part of the data cleansing process by interpolation or other statistical method. If the data is fixed in this way, that should be noted in the documentation about the quality of the data.

- The final option, particularly if required data is missing, is to delete those records from the final data set. Deciding to delete records that contain data that appear to be outliers is also an option, particularly if it was not possible to verify that the data was actually correct.

4.1.3. Determine how to implement the data cleansing rules

It is possible to define simple data cleansing rules that might be implemented "by hand" by filtering data or using simple functions in a spreadsheet or running simple queries in a database. However, having some sort of automated system for data quality and data cleansing of the data is crucial to obtaining consistent results as datasets grow larger over time.

Depending on the data available, comparing the same data between multiple data sources might be an option for including in the data cleansing rules. One data field that is usually difficult to obtain accurate information for is floor area. Comparing different data sources for floor area, such as comparing the values between floor area data from real estate, jurisdiction parcel and owner entered sources, and then evaluating the percentage difference between the sources, might lead to a better result that using data from just one source. For example, data that was more than 10-25% different between the sources could be flagged for further investigation for accuracy.

4.1.4. Communicate / publish the cleansing rules

It is good practice to make the data quality rules public in some fashion, usually by publishing a document on a website devoted to the project. This can include guidance on how the data can or cannot support various types of analysis and decision-making based on the quality of the data.

4.2. Best Practices

4.2.1. Allocate adequate resources to the data cleansing task

It is easy to underestimate the resources required to check and correct data, so it is worthwhile to identify the level of effort involved in the data cleansing task and allocate adequate resources to it. Work involved in this effort will undoubtedly include development of the data cleansing rules as well as their implementation. It is recommended that the personnel to do this task have previous experience with data and data cleansing. This might also mean developing an in-house automated data cleansing system, or using a third-party tool to do this.

It is also recommended, if possible, that resources are allocated for a "help center" which can interface with the building owners and managers to help identify and correct potential errors. The staffing for this could come from in-house employees or an outside consultant or contractor.

New York City Benchmarking Help Center[39]

This is an example of a help center set up by a jurisdiction to help building owners comply with an energy benchmarking requirement. It is open Monday through Friday from 9am - 5pm to answer call and emails on a case by case basis.

4.2.2. Review examples of existing data cleansing documentation

There are many jurisdictions engaged in energy efficiency programs that have developed and published their data cleansing rules and procedures. It is worth reviewing some of these documents in order to decide what data cleansing strategies are needed for your data. Several sources of such documentation are listed below.

Data Preparation Process for the Buildings Performance Database[40] . This process has been used to cleanse data for over 1 million US commercial and residential buildings, from over 50 different data sources.

Best Practices in Energy Data Collection and Tracking[41] . This document provides guidance on data collection and cleansing for US public sector agencies that are doing benchmarking.

- Link to Webinar[42]

City of New York LL84 Data Analysis & Quality Assessment [43]

The "Putting Data to Work" report (Beddingfield et al. 2018) provides guidance on data cleansing specifically for benchmarking data in US cities.

4.2.3. Start with simple rules, and expand based on needs

Before starting the data cleansing process, develop the minimum criteria for high quality data, based on your use cases. Determine what rules to apply, determine the tradeoff between accuracy and the level of effort to cleanse the data, and determine the cleansing options (e.g., is interpolation ok, and if so, how much is viable). Expand the data cleansing rules as needed based on analysis needs.

Initial set of data cleansing rules

The initial set of cleansing rules can determine records that should be deleted from the dataset or corrected (either by automated rules or resubmission with corrections from the original source). An example set of data cleansing rules is shown below. (Mathew, 2015).

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 ● 100,000 vs 100K vs 100000	Convert to common format
Inconsistent naming conventions	Convert to common terms
Missing data ● Buildings with no floor area reported ● Buildings with no energy use reported	Delete record or interpolate value
Obvious incorrect values / out-of-range checking 1. Buildings with EUI < 5 or > 1000 kBtu/sf 2. Floor area < 0 3. Hospital EUI < 10 kBtu/sf/yr	Delete record or interpolate value

Year over year data analysis

If the data collection and analysis is conducted going forward in time, comparing data between collection intervals can be an indication of problematic data. For example, if there is a large change in building consumption from one year to the next, the data from both years should be examined carefully. Establishing criteria for thresholds for acceptable percent increases or decreases in metrics will provide a standard method for checking the data year over year.

Statistically based methods for data cleaning

Developing statistically based algorithms is a valuable methodology for data cleansing, particularly when cleaning a large data set (such as New York City, which has over 14,000 buildings that need to comply each year with a benchmarking ordinance).

For example, when evaluating the accuracy of EUI values, the following method could be employed:

- Look at the distribution of source (or site) EUI based on building type
  - Extreme values relative to the overall distribution should be removed

The paper "DataIQ — A Machine Learning Approach to Anomaly Detection for Energy Performance Data Quality and Reliability" (Kontokosta, et al., 2016) discusses a data quality strategy based on statistical methods and machine learning.

4.2.4. Quantify the data quality and uncertainty

In order for policy makers to have confidence in the quality of the data being used to make decisions, it is important to quantify data quality and provide information about the uncertainty of the data. One approach is to develop a data quality scale and attach a quality score to each data item. For example, for energy data field-verified measured data would have the highest score while an estimate based on a simple model would have a low score.

4.2.5. Examine the data sources for systematic errors

Data that comes from a small number of sources, such as energy consultants, can have systematic errors. If such systematic errors are found, contact the sources and give feedback to improve their data collection process.

4.2.6. Validate data through control groups

If possible, establish controls groups that can independently validate a sample of the buildings within each building type. For example, jurisdiction staff or consultants may have in-depth knowledge of buildings that can be used to review the data. Staff or consultants who are known to have done a credible job of data reporting can be used to do field verification of a sample of buildings.

New York City: The Massachusetts Institute of Technology has performed studies to compare the energy efficiency policies of building in New York City, including comparisons to control groups. (Hsu, et al., 2016)[44] .

4.2.7. Automate the data quality checking and notification if possible

Establishing standardized metrics and criteria for data quality checking, and then automating those data quality checks will help ensure that the data has been consistently evaluated. If possible, automate these data checks, either through simple spreadsheet macros or filters, or by having someone write a tool for this type of quality checking.
In addition to automating the data quality checking, utilizing an automated notification system (such as sending emails) to inform building owners or managers of data quality problems, can save time and ensure that such notifications are delivered efficiently and consistently. Many "Customer Relationship Management" software systems have this type of automated functionality.

5. DATA ANALYSIS

4.1. Process Steps

5.1.1. Define desired output of the analysis

Start with the end in mind by defining the intended output of the analysis. There is a wide range of possible analysis outputs and the choice of which one to use will depend on the question being answered and the preferences of the audience. While it is not possible to catalog all of them here, we provide several examples. There are broadly three categories of analysis outputs:

- Stock-level aggregated data, e.g., total energy use or GHG by sector
- Data distributions, e.g., frequency distribution of energy ratings by building type
- Individual building data, e.g., Table showing weather normalized EUI for individual buildings

Within each of these broad categories, another key aspect is the time dimension. The output may be a single snapshot in time or may show variation over time. Categories 2 and 3 require individual building data, while Category 1 does not necessarily require that.

Aggregated data

Figure 5.1 shows an example of stock-level aggregated data. The top chart illustrates the total reduction for the city by sector and the bottom chart shows the breakdown of reductions within the building sector. Similarly, a chart could be constructed that shows the city's energy consumption or expected reductions split into different building types. When making decisions based on the behavior of a whole city, it is a good idea to start with this type of aggregate analysis. A key advantage of aggregated data is that it does not necessarily require individual building data. For example, it could be developed from aggregate data provided by the utility companies.

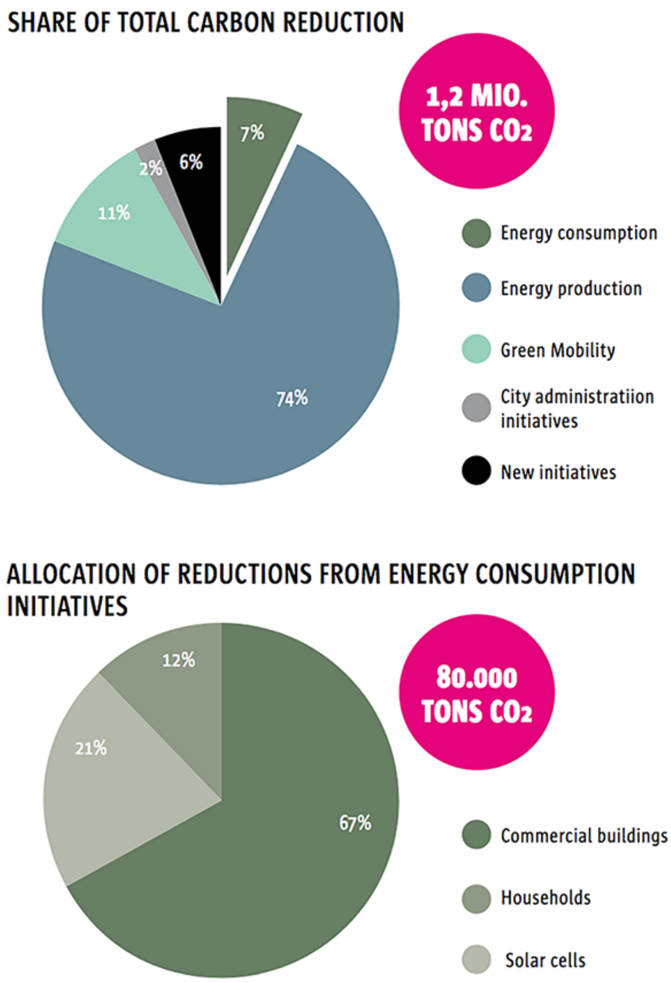


Figure 5.1. Aggregate analysis of carbon reduction in Copenhagen Climate Plan
Source: [CPH 2025 Climate Plan](#) [45]

Data distributions

Figure 5.2 is an example of data distribution analysis. The chart shows the number of buildings in Cambridge, Massachusetts, with Source EUI in several bins, and indicates the median Source EUI. This kind of chart illustrates what proportion of the building stock has high EUI, low EUI, etc., and makes it easy to see how many buildings would be affected if, for example, a city passed an ordinance requiring building with EUI above a certain threshold to take energy efficiency measures.

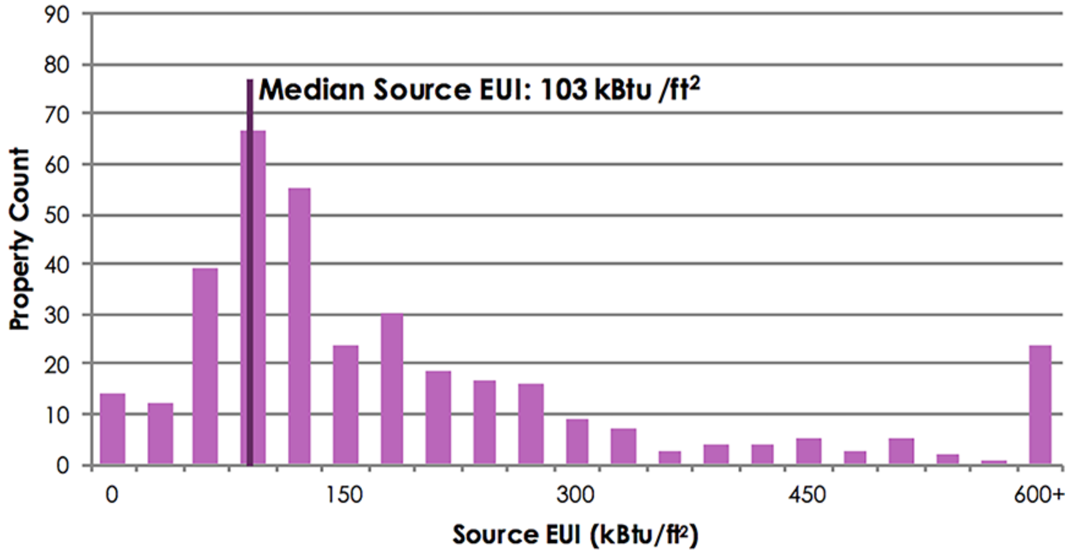


Figure 5.2. Median source EUI of buildings in Cambridge, MA
Source: 2015 Building Energy and Water Use Report [46]

Figure 5.3 is another example of analysis using data distributions. The plot shows the proportion of properties in Cambridge fitting into several categories, based on both building type and on ranges of energy rating. This kind of analysis can identify which types of buildings have the most buildings with low scores, allowing policies to target the building types that have the most room for improvement.

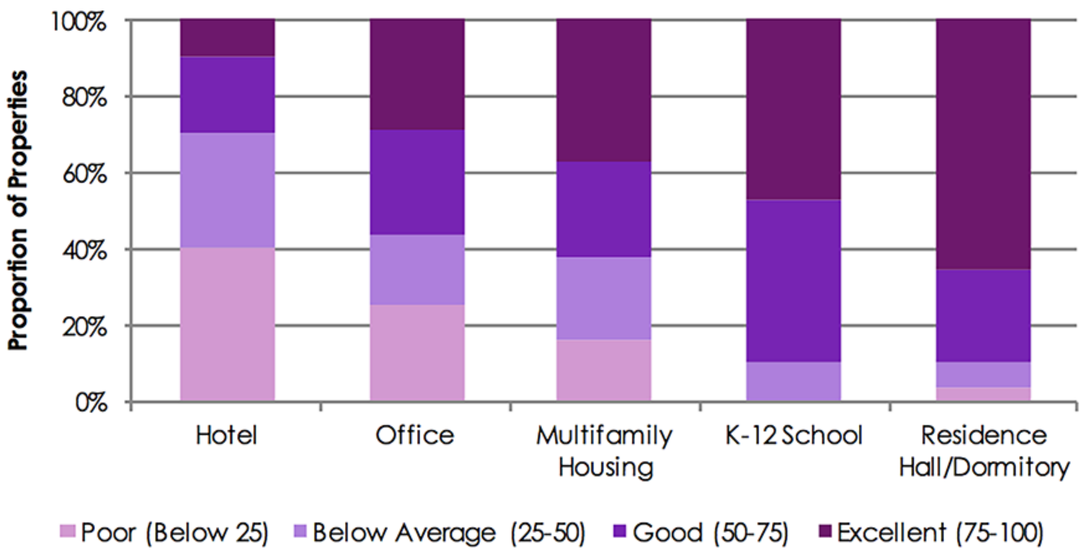


Figure 5.3. Proportion of buildings in Cambridge with different energy ratings
Source: 2015 Building Energy and Water Use Report [46]

Figure 5.4 is an example of both data distribution analysis and individual building analysis. The plot show EUI and year built for each building in Minneapolis. Each circle in the plot corresponds to an individual building, and can help to identify buildings that have interesting properties (for example, very new buildings that also have high EUI). In addition, the density of the circles at different locations in the plot can show ages of buildings that are more or less common (for example, very few buildings built between 1940 and 1950). Lastly, a scatterplot like this can help identify relationships between the two variables plotted. One may have suspected that newer buildings tend to have lower EUI, but this plot shows a very weak relationship between EUI and age.

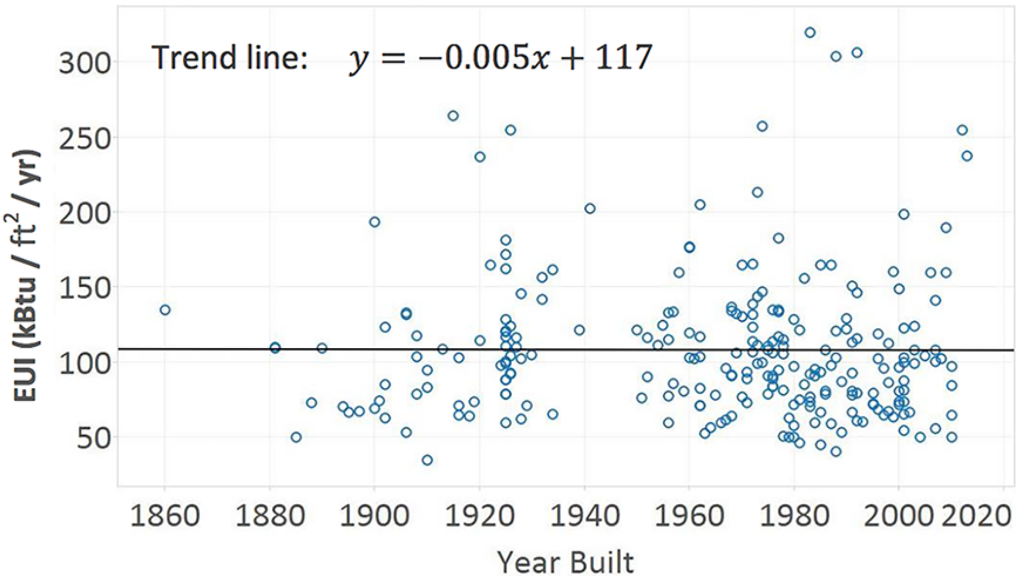


Figure 5.4. Relationship between EUI and year built for all Minneapolis buildings.
Source: 2013 Energy Benchmarking Report

Individual building data

Figure 5.5 is an example of individual building data analysis. The bar chart shows the cost of heating for several individual nursery schools in Warsaw, and uses color to indicate the type of heating (gas or electric) used in each school. The chart allows identification of the three schools with highest heating costs, and of the four schools with gas heating. These results can be used to target only the schools with high heating costs, without wasting effort on schools whose heating costs are already low.

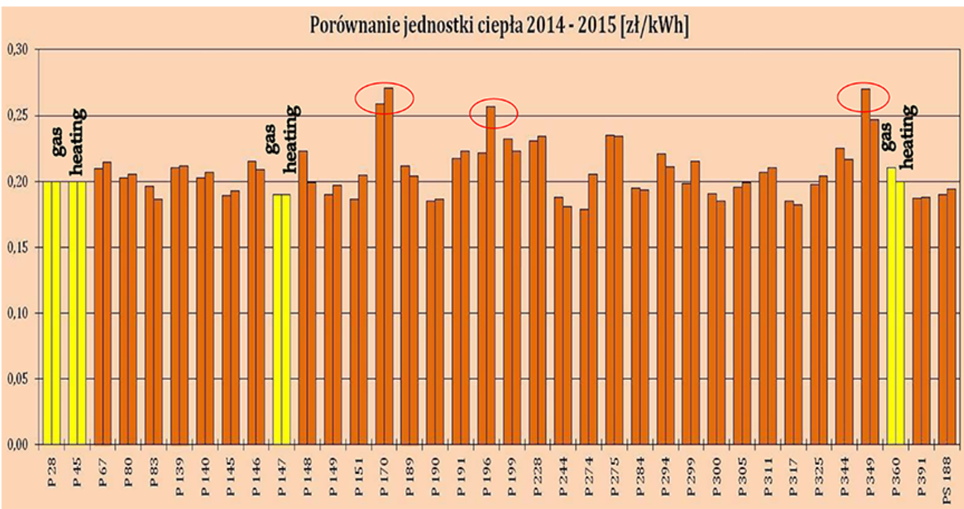


Figure 5.5. Heating unit cost in Warsaw nursery schools for 2014-2015
Source: Warsaw Municipal Energy Division 2015

Figure 5.6 is another example of individual building data. Each row corresponds to an individual building in Minneapolis, and several columns contain characteristics of that particular building. Examining the raw data like this is not always feasible (for example, when there are large numbers of buildings or of characteristics), but can provide insights not apparent when the data is viewed in aggregate. Data quality issues are often easy to identify when looking at raw data (for example, the building in the 4th row has an EUI of zero).

2016 Public Commercial Building Energy Benchmarking Data Published: 2017.09.12

Organization Name	Property Name	Address	Postal Code	ENERGY STAR Score	Primary Property Type	Floor Area (Buildings) (ft^2)	Floor Area (Parking) (ft^2)	Year Built	Total GHG Emissions (Metric Tons CO2e)	Site EUI ($kBtu/ft^2$)	Weather Normalized Site EUI ($kBtu/ft^2$)	Source EUI ($kBtu/ft^2$)	Weather Normalized Source EUI ($kBtu/ft^2$)	Water Use (kgal)
City of Minneapolis	10th and Hennepin Ramp	935 Hennepin Avenue	55403	0	Parking	552000	552000	1998	630	8	9	19	21	2770
City of Minneapolis	10th and LaSalle Ramp	915 LaSalle Avenue	55403	0	Parking	360000	360000	2001	433	9	10	20	22	175
City of Minneapolis	City Hall	350 South 5th Street	55415	95	Office	680000	0	1895	4565	57	59	115	116	12477
City of Minneapolis	City of Lakes	309 2nd Avenue South	55401	37	Office	47833	0	1958	793	149	161	288	300	595
City of Minneapolis	Currie Maintenance Facility	1200 Currie Ave North	55403	0	Other - Services	171200	0	1980	1319	77	86	136	143	3155
City of Minneapolis	Emergency Operations Training Center	25 37th Avenue NE	55421	0	Other	42581	0	2010	301	56	60	121	125	785

Figure 5.6. Individual public building data from Minneapolis
Source: Energy Benchmarking Results for Public and Large Commercial Buildings [47]

5.1.2. Select analysis methods and metrics

In general, data analysis methods for most policy objectives can be relatively simple. They mostly involve direct use of the data collected with minimal transformation. For example, generating a baseline of total site energy use simply involves summing up the total energy use for a given year. Similarly, data distributions can be easily generated within conventional spreadsheet tools. However, some types of analysis will require slightly more complex data transformation as described below.

Weather normalization

Building energy use can vary significantly when the weather is significantly hotter or cooler than usual. If energy use for a building with data measured in a hotter year is compared to energy use for another building with data measured in a cooler year, the effect of weather on energy use can be confused with the effect of some other difference between buildings. In order to avoid this confusion, if the data being analyzed was measured where weather can change significantly from year to year, we recommend comparing weather normalized energy use instead of measured energy use. Weather normalization requires data that are measured at least monthly, and is usually better with smaller intervals. Weather normalized energy use is an estimate of how much energy a building would have used under average weather conditions. For more information on how to do weather normalization, see the ENERGY STAR's technical reference[48] .

Normalization for building characteristics

Building energy use varies depending on several factors, and it is often useful to control or account for the some of the factors to gain the most insight. For example, larger buildings tend to consume more energy, but this is because they are larger, not because they are less efficient. An analysis on energy efficiency should take building size into account, rather than just comparing total energy use. In this case, the analyst can normalize energy use by size by defining energy use intensity (EUI) to be the ratio of energy use per unit of floor area and comparing EUI among buildings to judge efficiency. One could also normalize other variables to yield more meaningful analysis variables (for example, number of occupants per unit of floor area may be more useful than number of occupants alone). This normalization technique is typically only applicable when the relationship between the target variable and the normalizing variable is linear. For more complex relationships, other techniques are necessary.

When the relationship between two variables is complex and normalization is not applicable, it can be useful to simply conduct separate analyses for different ranges of the variables. Rather than trying to normalize energy use by the building's age, separate analyses could be carried out for old and new buildings. This approach can require more effort (especially when separating based on ranges of multiple variables), but is often necessary when the effect of the variables is unknown or not easily expressed.

Under certain conditions, more complex techniques can be used to control for the effects of some variables in order to isolate the impact of some key variables. One such technique is linear regression. Though typically used to find the influence of one variable on a second variable, regression can be used when several variables impact the outcome. For example, if we believe EUI is a function of both building age and of occupant density, we can construct a regression model that uses measured data to separate out the effects of the two variables and to quantify each of their impacts. Linear regression with multiple variables is often helpful in identifying variable that have minimal impact on the outcome and should therefore be excluded from further analysis. For an example of normalization for multiple variables using linear regression, see the technical reference for ENERGY STAR for offices[49].

GHG conversion factors

Depending on the goals of the analysis, it is important to consider which variables are focused on. Total energy use is a typical variable of interest from a cost perspective, but some analyses focus on environmental impacts instead. In this case, a more relevant metric may be the greenhouse gas (GHG) or carbon dioxide (CO2) emissions due to energy use, rather than energy use itself. When calculating emissions from energy use, it is important to separate energy use into different fuels (electricity, natural gas, etc.), and to apply the appropriate conversion factor for each fuel. C40 cities should use the conversion factors and approach in the CIRIS tool compliant with the Global Protocol for Community-Scale Greenhouse Gas Emission Inventories (GPC). For more information about CIRIS and GPC, see this website[50].

5.1.3. Select analysis tools and conduct analysis

Most simple analysis can be conducted in spreadsheet tools such as Excel. Spreadsheet tools are easy to use but are not as robust when it comes to data storage and management. You may also consider tools such as Tableau that are more specifically designed for data analysis and visualization. There are also third party analysis tools that have been developed specifically to help cities develop GHG inventories and reduction forecasts. Examples include the CURB scenario planning tool[51] developed by C40 and the World Bank, ESMAP's TRACE[52] tool, IBM Smarter Cities Assessment[53] , and Siemens City Performance Tool[54] .

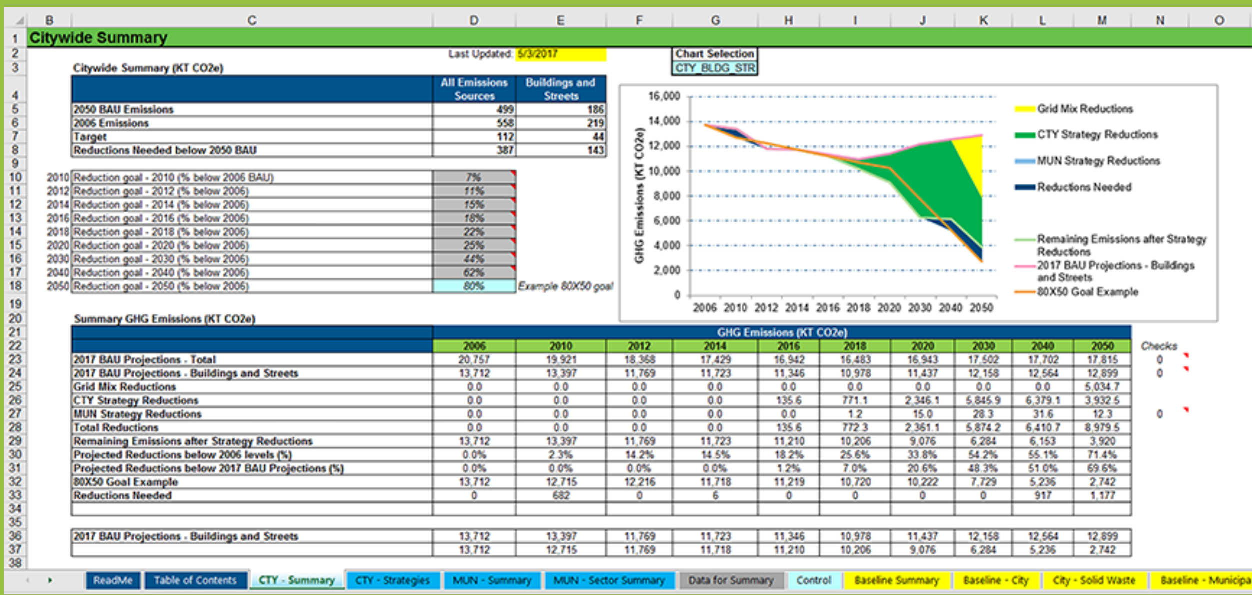


Figure 5.7. Philadelphia GHG baselining and forecasting tool.
Source: City of Philadelphia presentation at C40 PBE workshop, 2017

It is advisable to document each step explicitly to ease error checking as well as repetition by a different analyst, even for simple straightforward methods.

For more complex analysis methods, consider using a full-function programming language, such as R or Python. These software packages have many statistical analysis tools built-in, requiring less effort to implement complex algorithms and providing more confidence in the accuracy of the results. However, they may have a fairly steep learning curve, and modifying the analysis later can be difficult because it requires an analyst familiar with the programming language.

When conducting the analysis, it is important to develop explicit quality assurance/quality control (QAQC) procedures. Checks should be added to the analysis workflow to ensure results and logical, consistent, and reliable. This principle applies not only to the software itself, but how the software is developed. For example, if an analysis tool predicts the EUI resulting from an energy efficiency measure being applied, the software should check that the predicted EUI is within a reasonable range (e.g., greater than zero). In addition, the analyst responsible for generating the EUI predictions should have their work reviewed by other analysts or supervisors to ensure the techniques used are applicable and any assumptions being made are justifiable.

5.2. Best Practices

5.2.1 Explicitly link analysis results to overarching goals

Do not lose sight of the larger overarching goal. For C40 cities, this is the 1.5C pathway. Section 2 described a best practice of ensuring that policy statements have measurable goals. Likewise, analysis results should show how the results relate to these overarching measurable goals. Even when the analysis is broken out into targeted sub-sectors, it is helpful to show how it relates to the overall goal. For example, an analysis of energy reduction potential in non-residential buildings could show the reduction as a percentage of the overall building sector goal and the overall city goal. This helps put these results in perspective.

The city of Philadelphia did a detailed analysis of GHG reduction potential by sub-sector, and then compiled those results together to show how they each contribute to the city's overall building sector goal (see Figure 5.8). Interestingly, this example shows that measures that may be highly visible and popular, such as rooftop solar, do not necessarily have a high contribution to the overall goal.

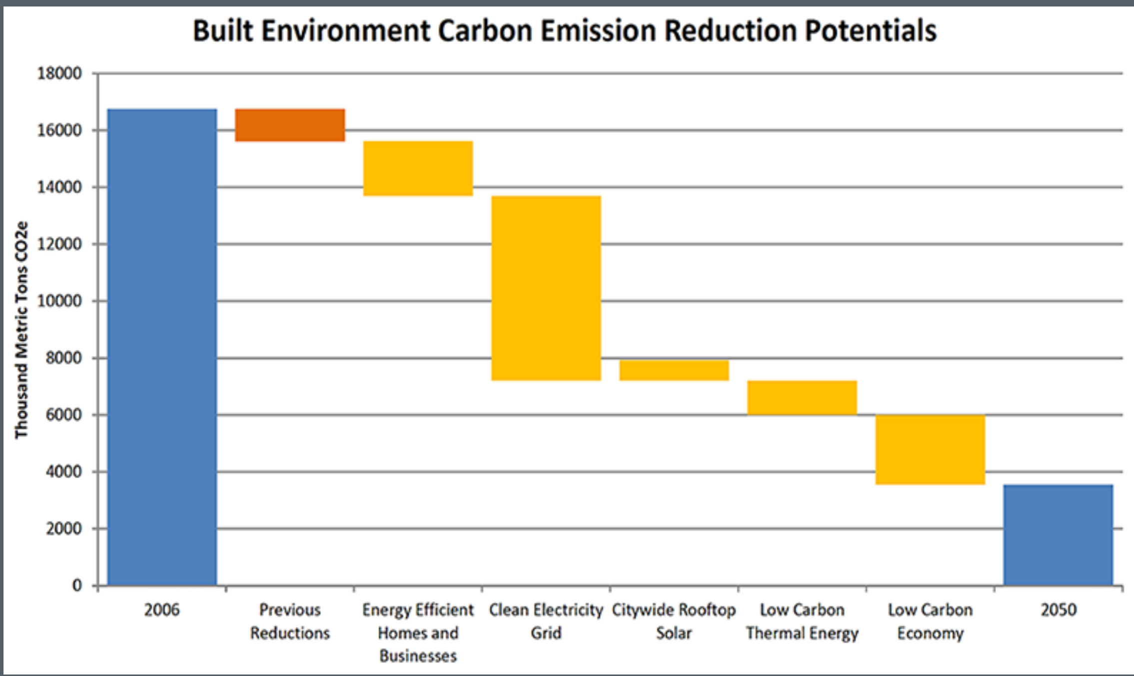


Figure 5.8. Philadelphia carbon reduction potential for the building sector.
Source: City of Philadelphia presentation at C40 PBE workshop, 2017

New York City used it's audit data to analyze energy consumption by end use which led them to a significant finding: 57% of building-based emissions city-wide come from on site combustion of fossil fuels for space heating and domestic hot water. Even with a 100% renewable grid, they cannot reach 80 X 50 goals without reducing on-site fossil fuel combustion. Therefore, they need to reduce thermal loads and consider alternate methods of heating buildings, such as electrification of these thermal loads.

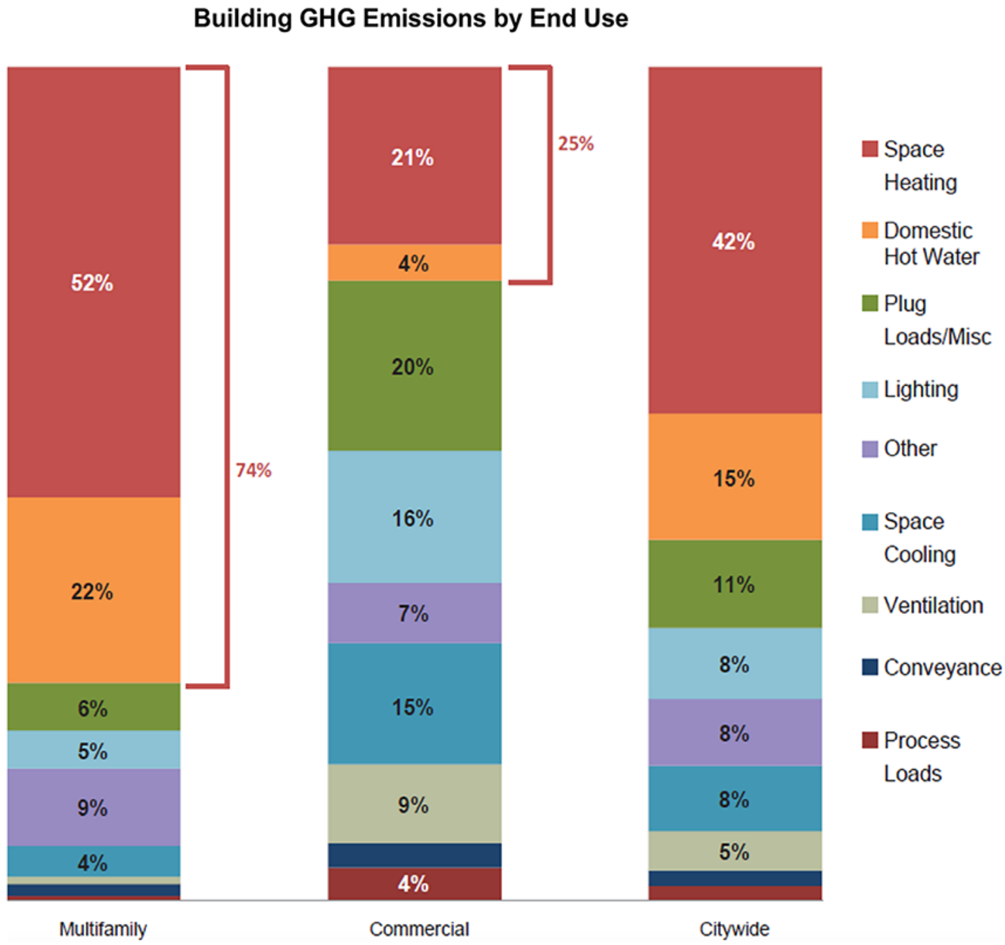


Figure 5.9. New York City analysis of end use energy showed that majority of building-based emissions city-wide come from on site combustion of fossil fuels for space heating and domestic hot water.
Source: New York City presentation at C40 PBE workshop, 2017

5.2.2. Be careful when interpreting results

An important step in any analysis is to document assumptions and limitations. Non-intuitive results can often be explained by violation of assumptions, or by assumptions that are unrealistic or not applicable. Results that seem reasonable when applied to specific circumstances may in fact be invalid due to limitations of the data or analysis methods. Documenting assumptions and limitations helps to formulate the analysis methodology, and to interpret the applicability of results.

A common mistake in data analysis is to infer behavior from limited data. For example, if one observes an increase in EUI to be associated with increased occupant density, but this is density may not be true in another climate (one would have to gather data from the other climate to find out). Be especially careful when trying to identify trends when very few data points are available because it is likely those few data points will exhibit relationships purely by chance.

It is also important to consider whether the data used to infer relationships is representative of the larger population the data was drawn from. Consider EUI data collected for several buildings throughout a city from an energy efficiency program. Participants enrolled in the program are more likely to care about their building energy use than building owners that are not enrolled. Trends identified from this EUI data might not be true for the whole city because the data collected was not from a representative group of buildings.

A final consideration when conducting an analysis is extrapolation. When learning a relationship between variables using measured data, keep in mind that the relationship is only valid under the conditions contained in the data. For example, if a model of EUI as a function of building age was fit to data only from building less than 10 years old, one should not assume EUI will depend on age in the same way for buildings more than 10 years old.

5.2.3. Carefully consider tradeoff between model sophistication and interpretation

It is important to weigh the benefits of more complex (and possibly more accurate) analysis methods against the costs of implementing the methods and interpreting their results. When considering more complex methods, consider whether it is well suited to the data quantity and quality, and consider the amount of effort and expertise required to make decisions using the results. More complex analyses can often be significantly more effort, while only providing marginally more useful results..

5.2.4. Develop uncertainty information

All data have errors and associated uncertainty. The sources of uncertainty range from measurement and transcription errors in the data, to uncertainty in models when doing data transformations. To the extent feasible, quantify the uncertainty and confidence level for all measured data and each model output. It is important to keep in mind the amount of uncertainty in analysis outputs when making decisions. The amount of confidence in the decision should reflect the confidence in the data and analysis on which the decision is based.

6. COMMUNICATING RESULTS

6.1. Process Steps

6.1.1. Identify stakeholder roles and interests

Stakeholders are any person or organization that can impact or be impacted by these policies, even if they are not directly involved with policy implementation. Building energy efficiency policies have a broad array of stakeholders — ranging from facilities staff to advocacy organizations. These stakeholders represent a diverse range in terms of their business and political interests, concerns, and knowledge of building energy efficiency. Therefore, it is necessary to tailor communications accordingly for different groups of stakeholders — one size does not fit all.

Table 6.1 is a partial listing of stakeholders to consider when communicating policy analysis, along with their key interests and concerns and applicability to PBE vs. MBE. Use this list as a template to develop a list specific to your city.

Stakeholder	Key Interests/Concerns	PBE	MBE
Building owners/managers	Are results meaningful and actionable information for individual buildings? Are peer comparisons fair ("apples to apples")? Concerns about publicizing individual building data, especially poor performers.	●	●
Building occupants	Are results understandable and transparent?	●	●
City political leadership	Do the results support or undermine policy objectives and political interests? How will the results be interpreted by key interest groups?	●	●
City energy program staff	Do the results provide actionable information for policy and program development? Robustness and credibility of the results.	●	●
Real estate industry organizations	Concerns about publicizing individual building data, especially poor performers. Will the results spur additional regulatory/compliance burden on the industry?	●	

Environmental organizations	Do the results properly account for all environmental impacts? Will the results impede or enhance efforts for more stringent policies.	●	○
Civic organizations	Are the results meaningful and understandable by the general public? Is the city meeting its obligations for public accountability and ensuring information is available to all concerned citizens?	○	
Utilities	Concerns about publicizing individual building data. Will the results spur additional energy efficiency requirements on utilities? Do the GHG calculations properly reflect the utility's fuel mix?	●	○
Energy Service providers	Do the results provide enough information on individual buildings to do targeting and analysis?	●	●
Academic/think-tank researchers	Is the underlying analysis rigorous and robust? Are the data available and adequate to conduct additional academic research?	●	●
General Public	How does this affect me? Will it impact city services?	●	○

● Highly applicable ○ Somewhat applicable

Table 6.1 Stakeholder interests and relevance to PBE and MBE.

Once the list has been developed, the next step is to broadly categorize each stakeholder in terms of their level of their involvement or influence in policy implementation as well as their interest in the details of the analysis. This will drive the communication strategy. The simple framework below can be used for categorizing the stakeholders. There are many resources that provide guidance on stakeholder engagement (see, for example, this website[55]). If your city's stakeholder relations are especially sensitive or contentious, you may also consider hiring professional facilitators and consultants.

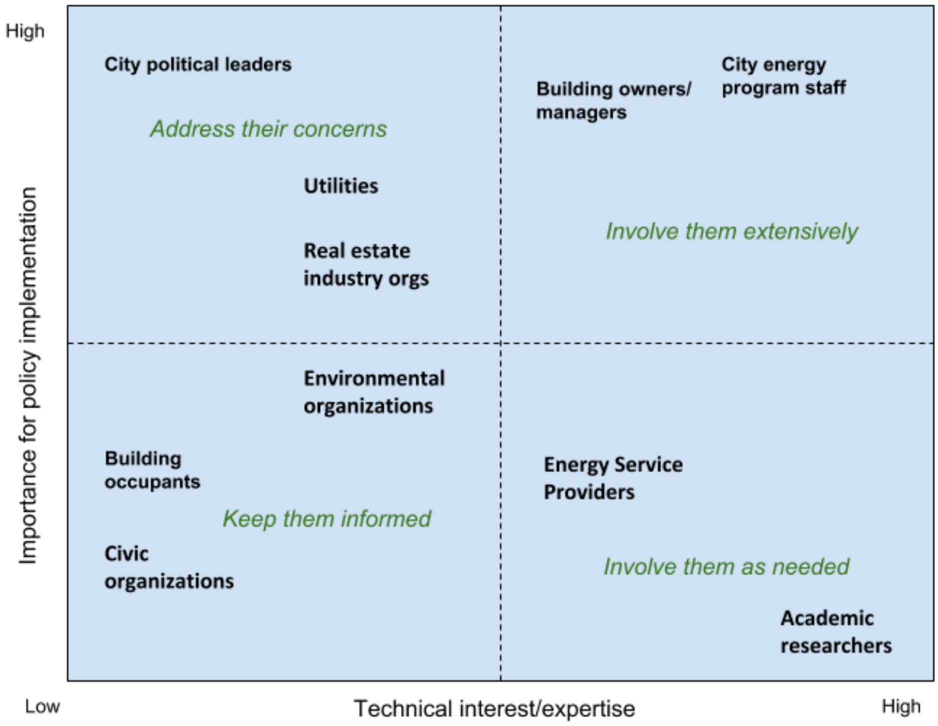


Figure 6.1. Framework for prioritizing stakeholder engagement and communications. This is only illustrative and should be tailored for each city. Adapted from Luc Galoppin CC BY 2.0

6.1.2. Develop communication strategy and resources

Once you have identified and characterized stakeholder interests, develop a strategy for outreach and related resources. Again, these may vary based on stakeholder influence and interest. Some stakeholders will want detailed technical reports, while others may only be interested in summary results. Also, the presentation style and language may vary based on the stakeholder, e.g., flyers for the general public would avoid using terms such as energy use intensity. Depending on the scope and scale of the data for policy effort, you should strongly consider engaging a consultant to help develop the communications strategy and resources.

Communication approaches should include both interactive channels (working groups, town hall meetings) and non-interactive channels (e.g., press releases, email blasts). Interactive channels may be especially important for stakeholders that are skeptical and need buy-in. Some stakeholders such as industry and advocacy organizations have their own meetings that may afford a direct opportunity to engage them. When the stakeholder group is critical but small, direct one-on-one discussions may be the most effective channel (e.g., for MBE, school administrators or city facilities department).

Figure 6.2 shows the applicability of various communication channels and resources for different stakeholders. This is only suggestive. You should tailor this to make it appropriate for your city's stakeholder environment.

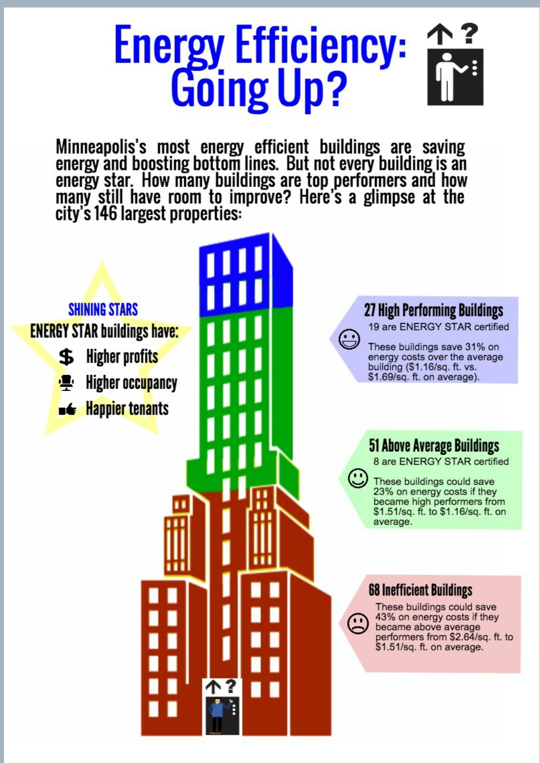
Stakeholder	Communication Channel								Communication Resource				
	Mailing Lists	Webinars	Working Groups	One-On-One Meetings	"Town Hall" Meetings	Press Conference	Presentations at Stakeholder Venues	Social Media	Websites	Technical Report	Fact Sheets	Videos	Press Releases / News Articles
Building Owners/Managers	●	●	●		○		○		●	○	●	○	
Building Occupants					○			●	○		○	●	
City political Leadership			○	●					○		●		
City Energy Program Staff	●		●	●					●	●	○		
Real Estate Industry Organizations	●	●	●		○	●	●		●	○	●		●
Environmental Organizations	●	●	○		○	●	●		●	○	●		●
Civic Organization	●	●	○		○	●		●	●		●	●	●
Utilities	●	●	●	●						●			
Energy Service Providers	●	●	●		○		●	○	●	●	●		
Academic/Think-Tank Researchers	●	●	●				●		●	●	○		

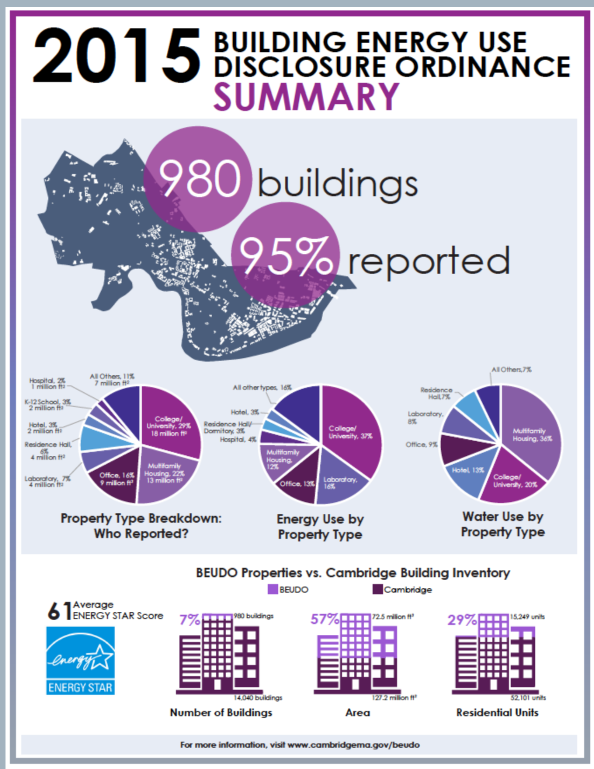
● Primary ○ Secondary

Figure 6.2. Applicability of communication channels and resources for different stakeholders

There are numerous examples of communications resources across the C40 cities. Below we highlight just a few as examples. We strongly encourage you to explore other examples as you develop resources tailored to your city. The "Putting Data to Work"[56] report (Beddingfield et al. 2018) provides guidance on communication and outreach strategies specifically for benchmarking data in US cities. Section 7.1 lists websites for various city energy initiatives, with examples of various types of resources.

Fact sheets: Below are two examples from Minneapolis (left), Cambridge, MA (right).





Technical reports: Almost every city that has analyzed data for policy has produced technical reports. Below are a few examples.

- City of Chicago Energy Benchmarking Report 2016 [57].
- Energy Benchmarking Report for New York City Municipal Buildings, Nov 2011[58].
- Singapore BCA Building Energy Benchmarking Report 2017[59] .

Short videos: These can be especially appropriate for the general public. Tokyo developed a short video to explain its Cap and Trade scheme [60].

6.2. Best Practices

6.2.1. Public disclosure of individual building data

Public disclosure of benchmarking and other data allow any city resident or stakeholder to assess the performance of specific buildings. It also supports an eco-system of energy efficiency services by allowing service providers to target specific buildings.

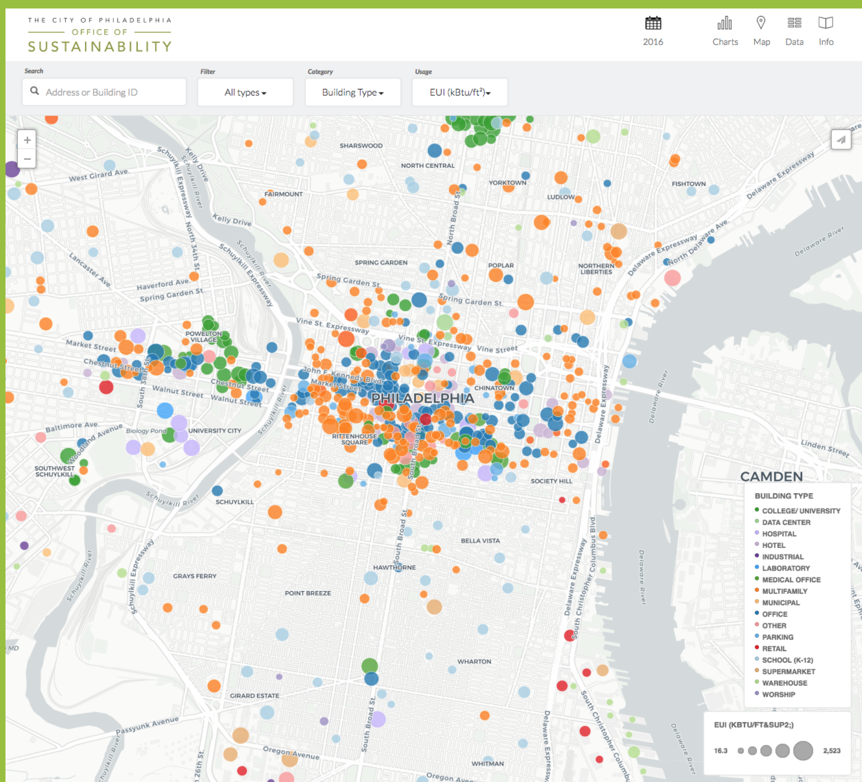
More than six US cities disclose individual building data publicly with downloadable spreadsheets. MBE cities may have even greater liberties with municipal buildings. For example, Washington DC discloses detailed 15-minute interval energy use data for over 350 municipal buildings and encourages various stakeholders to use these data to identify efficiency opportunities.

- Boston[61]
- New York City[62]
- San Francisco municipal buildings [63]
- Washington DC municipal buildings[64]

6.2.2. Data visualization through web based applications

Use maps to show the data analysis results to building owners, managers and the general public to increase interest in energy use and conservation. Their visual and interactive appeal can make them especially effective at engaging stakeholders to look at data.

- Boston[65]
- Philadelphia[66]
- Chicago[67]
- Minneapolis[68]
- New York City[69]
- Seattle[70]
- Washington DC[71]



6.2.2. Data visualization through web based applications

Many stakeholders may not have the time or wherewithal to explore and use the data. Targeted technical assistance can overcome this barrier. However, expert technical assistance can be resource intensive, and therefore it should be prioritized for high-impact stakeholders such as building owners and energy efficiency implementers.

- NYC Benchmarking Help Center
- New York City Retrofit Accelerator [72] provides one-stop free resource (team of efficiency advisors) for private building owners and operators to complete energy and water upgrades using building-specific data to target buildings.
- The city of Tokyo provides free energy audits to small and medium enterprises as a follow up to their Carbon Reduction Reports.

6.2.4. Encourage broad stakeholder engagement

There are innumerable ways in which to ensure broad stakeholder engagement. You will need to develop strategies that are suited to the culture and context of your city. Below we provide a few examples.

- City of Tokyo: As they developed their cap and trade scheme, the held stakeholder meetings open to the public that routinely drew over 200 participants. Tokyo stakeholder engagement on cap and trade scheme[73] .
- Copenhagen is exploring collaboration with public schools to use energy data for educational purposes.
- Sydney developed explicit tools and strategies for commercial tenant engagement, because tenants can be hard to reach with programs.

7. RELATED RESOURCES

7.1. City Programs

Chicago Environment and Sustainability
- https://www.cityofchicago.org/city/en/progs/env.html

City of Copenhagen
- https://stateofgreen.com/en/profiles/city-of-copenhagen

London Environment Energy Programs
- https://www.london.gov.uk/what-we-do/environment/energy

Los Angeles Better Buildings Challenge
- http://la-bbc.com/

Melbourne Sustainable Business
- http://www.melbourne.vic.gov.au/business/sustainable-business/Pages/sustainable-business.aspx

New York City Mayor's Office of Sustainability

- http://www.nyc.gov/html/gbee/html/plan/plan.shtml
- NYC LL84: Benchmarking
- NYC Retrofit Accelerator
- NYC LL87: Energy Audits and Retro-commissioning

Philadelphia Office of Sustainability
- https://beta.phila.gov/departments/office-of-sustainability/

San Francisco SF Environment, Buildings and Environment Green Building
- https://sfenvironment.org/buildings-environments/green-building

Seattle Office of Sustainability and Environment
- http://www.seattle.gov/environment

Singapore Building and Construction Authority Sustainable Built Environment
- https://www.bca.gov.sg/Sustain/sustain.html

Sydney Better Buildings Partnership
- http://www.cityofsydney.nsw.gov.au/business/business-support/greening-your-business/better-buildings-partnership

Tokyo Metropolitan Government Bureau of Environment
- https://www.kankyo.metro.tokyo.jp/en/climate/index.html

7.2. Third Party Tools

Green Building Information Gateway (GBIG)
Platform developed by U.S Green Buildings Council to identify green buildings activity globally. GBIG integrates hundreds of data sources, including some city building energy disclosure data and other publicly available rating data.

Global Real Estate Sustainability Benchmark (GRESB)
GRESB is a dynamic benchmarking used by institutional investors to assess sustainability performance of real estate portfolios around the globe.

International Sustainability Alliance (ISA)
Global network of leading corporate occupiers, property investors, developers and owners of commercial buildings. In 2013 latest benchmarking report from ISA covered 16.5 million m2.

Urban Land Institute (ULI) Greenprint Center for Building Performance
Offers guides, toolkits, global collection of property data (Greenprint Performance Report).

Putting Data to Work Toolkit
For U.S. jurisdictions with benchmarking and building performance policies, or those considering adopting them, this toolkit provides a guide for using city-collected data to identify efficiency opportunities.

8. REFERENCES

1. Beddingfield, Erin, Zachary Hart, Julie Hughes, Putting Data to Work. How Cities are Using Building Energy Data to Drive Efficiency. Report. Institute for Market Transformation. February 2018.

2. Bosteels, Tatiana, Niall Tipping, Christopher Botten, Matthew Tippet, Sustainability Benchmarking Toolkit for Commercial Buildings: Principles for best practice, Better Buildings Partnership, London, January 2010.

3. Hart, Zachary, Data Quality and Analysis Guide, Institute for Market Transformation. 2018. In Press.

4. Meng Ting, David Hsu, Albert Han, Measuring Energy Savings from Benchmarking Policies in New York City, Proceedings of the 2016 American Council for an Energy-Efficient Economy (ACEEE) Summer Study on Energy Efficiency in Buildings.

5. Iyer, Maithili, Satish Kumar, Sangeeta Mathew, Hannah Stratton, Paul Mathew, and Mohini Singh, Establishing a Commercial Buildings Energy Data Framework for India: A Comprehensive Look at Data Collection Approaches, Use Cases and Institutions. Energy Technologies Area, Lawrence Berkeley National Laboratory, Synurja, LLC, LBNL-1006427, October 2016.

6. Kontokosta, Constantine, Bartosz Bonczak, Marshall Duer-Balkind, Data-IQ — A Machine Learning Approach to Anomaly Detection for Energy Performance Data Quality and Reliability, Proceedings of the 2016 American Council for an Energy-Efficient Economy (ACEEE) Summer Study on Energy Efficiency in Buildings.

7. Mathew, Paul, Sangeeta Mathew, Satish Kumar, Mohini Singh, Hannah Stratton, Maithili Iyer, India Commercial Buildings Data Framework: A summary of Potential Use Cases, Energy Technologies Area, Lawrence Berkeley National Laboratory, LBNL-1005723, May 2016.

8. Mathew, Paul, Data Cleansing: Lessons from the BPD, Presentation at Better Buildings Summit 2015. Lawrence Berkeley National Laboratory.

9. Warsaw Municipal Energy Division. An analysis of the consumption and costs of electricity, heat, gas and water in the Mokotów District facilities in 2015. Report Provided to C40.

Appendix A: Using Data for Policy - PBE Quick Assessment

Defining objectives and metrics

	Y	N	Actions
Do you have a formal policy in place to reduce GHG?			N: Develop a policy statement
Does the policy have quantitative city-wide GHG reduction targets?			N: Define quantitative city-wide GHG reduction targets.
Does the policy have quantitative targets for buildings?			N: Develop quantitative targets for building sector or sub sectors.
Have you developed policy analysis objectives and metrics?			N: Define objectives and metrics using Section 2.1. Y: Review best practices in section 2.2

Data collection

	Y	N	Action Item
Have you developed a list of data requirements based on analysis objectives?			N: See section 3.1
Do you have a list of all target buildings with required information such as type and size?			N: See section 3.1 Y: Review best practices in section 3.2
Are you currently collecting building energy data for these buildings?			N: See section 3.1 Y: Review best practices in section 3.2

Data cleansing

	Y	N	Action Item
For data you collect, are you performing data cleansing on it?			N: See section 4.1 Y: Review best practices in section 4.2

Data analysis

	Y	N	Action Item
Have you defined what type of analysis outputs you want?			N: See section 5.1 Y: Review best practices in section 5.2
Have you developed an analysis methodology?			N: See section 5.1 Y: Review best practices in section 5.2

Communicating results

	Y	N	Action Item
Have you developed a stakeholder communication strategy?			N: See section 6.1 Y: Review best practices in section 6.2
Have you developed communication resources?			N: See section 6.1 Y: Review best practices in section 6.2

Appendix B: Using Data for Policy - MBE Quick Assessment

Defining objectives and metrics

	Y	N	Actions
Do you have a formal policy in place to reduce GHG for municipal buildings?			N: Develop a policy statement
Does the policy have quantitative targets for municipal buildings?			N: Develop quantitative targets for municipal buildings as a whole, ideally broken down by major typology (schools, government offices, fire stations, etc.)
Have you developed policy analysis objectives and metrics?			N: Define objectives and metrics using Section 2.1. Y: Review best practices in section 2.2

Data collection

	Y	N	Action Item
Have you developed a list of data requirements based on analysis objectives?			N: See section 3.1
Do you have a list of all target buildings with required information such as type and size?			N: See section 3.1 Y: Review best practices in section 3.2
Do you have access to and are you currently collecting building energy data for these buildings?			N: See section 3.1 Y: Review best practices in section 3.2

Data cleansing

	Y	N	Action Item
For data you collect, are you performing data cleansing on it?			N: See section 4.1 Y: Review best practices in section 4.2

Data analysis

	Y	N	Action Item
Have you defined what type of analysis outputs you want?			N: See section 5.1 Y: Review best practices in section 5.2
Have you developed an analysis methodology?			N: See section 5.1 Y: Review best practices in section 5.2

Communicating results

	Y	N	Action Item
Have you developed a strategy to engage the key city government stakeholders?			N: See section 6.1 Y: Review best practices in section 6.2
Have you developed communication resources?			N: See section 6.1 Y: Review best practices in section 6.2

Endnotes: Website URLs

1. http://www.c40.org/programmes/city-inventory-reporting-and-information-system-ciris

2. http://www.worldbank.org/en/topic/climatechange/brief/city-climate-planner-certificate-program

3. https://www.buildingrating.org/graphic/us-commercial-building-policy-comparison-matrix

4. http://www.minneapolismn.gov/www/groups/public/@health/documents/images/wcms1p-141979.pdf

5. http://www.betterbuildingsla.com/deadlines-and-penalties

6. http://www.c40.org/researches/deadline-2020

7. https://www.energycommunity.org/documents/copenhagen.pdf

8. http://www1.nyc.gov/site/sustainability/codes/1.5-climate-action-plan.page

9. https://www1.nyc.gov/site/sustainability/codes/80x50.page

10. http://www.nyc.gov/html/builttolast/pages/plan/plan.shtml

11. http://www.seattle.gov/Documents/Departments/Environment/ClimateChange/BldgEngy_Targets_2017-03-30_FINAL.pdf

12. http://www.nyc.gov/html/gbee/html/one-city/technical-working-group.shtml

13. http://www.nyc.gov/html/gbee/html/plan/ll84.shtml

14. http://www.nyc.gov/html/gbee/html/plan/ll87.shtml

15. http://www.nyc.gov/html/gbee/html/plan/ll88.shtml

16. https://www.energystar.gov/buildings/facility-owners-and-managers/existing-buildings/use-portfolio-manager

17. https://www.energystar.gov/buildings/tools-and-resources/list-portfolio-manager-property-types-definitions-and-use-details

18. https://portfoliomanager.energystar.gov/pm/dataCollectionWorksheet

19. https://www.energystar.gov/buildings/tools-and-resources/full-list-portfolio-manager-custom-reporting-metrics

20. https://eta.lbl.gov/sites/default/files/publications/lbnl-1005723.pdf

21. http://www.betterbuildingspartnership.co.uk/sites/default/files/media/attachment/bbp-sustainability-benchmarking-toolkit.pdf

22. http://www.costar.com/

23. https://nabers.gov.au/public/webpages/home.aspx

24. https://www.bca.gov.sg/GreenMark/green_mark_criteria.html

25. http://www.buildsmartdc.com/about/our-data

26. https://ec.europa.eu/energy/en/eubuildings

27. https://ec.europa.eu/energy/en/eu-buildings-database

28. https://ec.europa.eu/energy/en/eu-buildings-datamapper

29. http://bpie.eu/wp-content/uploads/2015/10/Energy-Performance-Certificates-EPC-across-the-EU.-A-mapping-of-national-approaches-2014.pdf

30. https://buildingenergyscore.energy.gov/

31. http://www.greenbuttonalliance.org/

32. http://www.greenbuttondata.org/

33. http://www.hpxmlonline.com/

34. https://hpxml.nrel.gov/

35. https://buildingsync.net/

36. https://eta.lbl.gov/publications/india-commercial-buildings-data

37. http://www.nyc.gov/html/gbee/html/plan/ll87_eer.shtml

38. https://energy.gov/eere/buildings/standard-energy-efficiency-data-platform

39. http://www.nyc.gov/html/gbee/html/plan/ll84_help_center.shtml

40. https://energy.gov/eere/buildings/downloads/data-preparation-process-buildings-performance-database

41. https://energy.gov/sites/prod/files/2016/09/f33/energy-data-management-part1.pdf

42. https://energy.gov/eere/slsc/downloads/energy-data-management-webinar-series-part-1-energy-data-collection

43. http://www.nyc.gov/html/gbee/downloads/pdf/120818_HSU_redacted.pdf

44. http://aceee.org/files/proceedings/2016/data/papers/9_988.pdf

45. https://stateofgreen.com/files/download/1901

46. https://www.cambridgema.gov/CDD/zoninganddevelopment/sustainablebldgs/~/media/B6C0D070B96D4A28A6E90C5DBD8C11A7.ashx

47. http://www.minneapolismn.gov/environment/energy/benchmarking

48. https://www.energystar.gov/sites/default/files/tools/Climate_and_Weather_August_2017_EN_508.pdf

49. https://www.energystar.gov/sites/default/files/tools/Office.pdf

50. http://www.c40.org/programmes/city-inventory-reporting-and-information-system-ciris

51. http://www.c40.org/programmes/climate-action-for-urban-sustainability-curb

52. http://www.esmap.org/node/235

53. https://www-935.ibm.com/services/us/gbs/bus/html/ibv-smarter-cities-assessment.html

54. https://www.siemens.com/global/en/home/company/topic-areas/intelligent-infrastructure/city-performance-tool.html

55. https://www.interaction-design.org/literature/article/stakeholder-maps-keep-the-important-people-happy

56. http://www.imt.org/resources/detail/putting-data-to-work-how-cities-are-using-building-energy-data-to-drive-eff

57. https://www.cityofchicago.org/content/dam/city/progs/env/EnergyBenchmark/2016_Chicago_Energy_Benchmarking_Report.pdf

58. http://www.nyc.gov/html/planyc/downloads/pdf/publications/Benchmarking-Report-11-23-11.pdf

59. https://www.bca.gov.sg/GreenMark/others/BCA_BEBR_Abridged_FA_2017.pdf

60. https://www.youtube.com/watch?v=ttPkGSeNu3o

61. https://data.boston.gov/dataset/building-energy-reporting-and-disclosure-ordinance

62. http://www.nyc.gov/html/gbee/html/plan/ll84_scores.shtml

63. https://data.sfgov.org/Energy-and-Environment/Energy-Benchmarking-San-Francisco-Municipal-Buildi/eg8h-pgn3/data

64. http://www.buildsmartdc.com/buildings

65. http://boston.maps.arcgis.com/apps/webappviewer/index.html?id=049576c7287f4ee09bcb0a062e43b55c

66. http://visualization.phillybuildingbenchmarking.com/#!/map

67. http://cityenergyproject.github.io/chicago/#chicago/2016?layer=energy_star_score&metrics[]=energy_star_score&sort=energy_star_score&order=desc&lat=41.91607416876307&lng=-87.75604248046875&zoom=10&building=659

68. http://cityoflakes.maps.arcgis.com/apps/View/index.html?appid=0f02e5a39c8c4ec49fa7b0c057b1c1a9

69. https://serv.cusp.nyu.edu/projects/evt/

70. http://www.seattle.gov/energybenchmarkingmap/#seattle/2015?layer=energy_star_score&metrics[]=energy_star_score&sort=energy_star_score&order=desc&lat=47.610213718421726&lng=-122.33053207397461&zoom=12

71. http://energybenchmarkingdc.org/#dc/2016?categories[0][field]=report_status&categories[0][values][]=In+Compliance&categories[0][other]=false&layer=energy_star_score&metrics[]=energy_star_score&sort=energy_star_score&order=desc&lat=38.889931&lng=-77.009003&zoom=12

72. https://retrofitaccelerator.cityofnewyork.us/

73. http://www.kankyo.metro.tokyo.jp/en/climate/cap_and_trade/index.files/kankyo4760.pdf

C40 Cities

3 Queen Victoria Street, City

London EC4N 4TQ

United Kingdom (UK)

Lawrence Berkeley National Laboratory

1 Cyclotron Rd

Berkeley, CA 94720

United States of America (USA)