Developments in the Built Environment 20 (2024) 100530

Contents lists available at ScienceDirect

Developments in the Built Environment

journal homepage: www.sciencedirect.com/journal/developments-in-the-built-environment

Advances in building data management for building performance standards using the SEED platform

Nicholas Long [a,b,*], Katherine Fleming [a], Alex Swindler [a], Andrew Held [c], Robin Mitchell [d], Gregor P. Henze [b,a]

[a] National Renewable Energy Laboratory, Golden, CO, USA
[b] Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO, USA
[c] Department of Energy and Environment, Washington, DC, USA
[d] Lawrence Berkeley National Laboratory, Berkeley, CA, USA

ARTICLE INFO

Keywords:
Building data management
Building benchmarking
Building performance standards
Case study

ABSTRACT

Reducing energy consumption and greenhouse gas emissions in the built environment is a critical step in achieving emission goals to mitigate climate change impacts. Local, federal, and international jurisdictions are deploying several methods to reduce energy and emissions such as voluntary and mandatory benchmarking and building performance standards, requiring building owners to reach energy and emission targets.

Jurisdictions leveraging benchmarking and building performance standards require knowledge of the buildings covered; which is a large task due to staffing constraints, limited information on building characteristics and tax parcel data, and the need for advanced data management techniques to align datasets. This paper describes an open-source platform's recent advances to create consistent taxonomies, identify erroneous data, enable auditability, and track building performance. The paper concludes with two use cases on how the platform has been used by jurisdictions.

1. Introduction

Effective building data management is becoming increasingly crucial for pursuing sustainable urban development and climate change mitigation. Buildings are recognized as significant contributors to energy consumption and greenhouse gas emissions, and there is a growing need for organized data practices to ensure that energy and emission reduction goals are met. Governments, jurisdictions, and organizations worldwide are adopting strategies such as benchmarking and building performance standards (BPSs) to drive energy and emissions reductions in the built environment. The success of these strategies is not without complexity: challenges are encountered every step of the way, from data collection to alignment, cleaning, and finally, progress tracking. This paper delves into the complexities of managing building data for these purposes, exploring methods, challenges, and solutions. By discussing data organization processes such as cleansing, mapping, matching, merging, and linking, this paper aims to shed light on the intricacies of managing building data. The article also presents a case study on using an open-source platform for managing building data, providing insights into the platform's capabilities and potential for future research.

There are several pathways that jurisdictions have taken and are currently taking to reduce energy and greenhouse gas emissions in buildings. Since 2007, jurisdictions in the United States (US) started using benchmarking where building owners and managers must submit building characteristics and energy consumption to a jurisdiction (State and Local Energy Efficiency Action Network, 2012). In addition to the passage of the Energy Independence and Security Act of 2007 (EISA2007) (United States Congress, 2007) requiring 25% of federal building floor area to benchmark their energy use every four years. In 2015, the Energy Efficiency Improvement Act was passed by the US Congress to promote energy efficiency, encourage collaboration between stakeholders, and enhance data transparency in the building sector (United States Congress, 2015). In general, a benchmarking policy requires the building owner (or building manager) to submit high-level building characteristics (e.g., property name, property use type, floor areas) along with metered energy usage data to the authority having jurisdiction (AHJ). Often, these data are collected in United States Environmental Protection Agency (EPA)'s ENERGY STAR® Portfolio

* Corresponding author. National Renewable Energy Laboratory, Golden, CO, USA
E-mail address: nicholas.long@nrel.gov (N. Long).

https://doi.org/10.1016/j.dibe.2024.100530
Received 12 March 2024; Received in revised form 23 August 2024; Accepted 25 August 2024
Available online 5 September 2024

Manager® (ESPM). Early reports showed benchmarking savings of 2.4% annual energy use over time (US Environmental Protection Agency (EPA), 2012). Several jurisdictions passed auditing, re-tuning, retro-commissioning (RCx), or similar ordinances to encourage building energy performance upgrades. The city of Austin, Texas, was one of the first jurisdictions to pass an auditing requirement for multifamily buildings (City of Austin, 2008; Institute for Market Transformation, 2023a). BPSs are the latest type of ordinances that jurisdictions are putting into place, starting in 2018 with Washington, DC (District of Columbia, 2018).

While these policies are designed to help reduce the carbon footprint of buildings and promote energy efficiency, each jurisdiction must expand its data management practices to handle the large amounts of data required to track building performance over time. Implementing these policies is complex and requires a robust data management system able to handle the relationships between buildings and tax parcels; many jurisdictions have limited visibility on the details of the buildings on the tax parcel since they historically only track parcel-related information. This paper will further elucidate the complexities of managing building data for these purposes, exploring methods, challenges, and solutions. Lastly, the article will describe how an open-source solution called the Standard Energy Efficiency Data Platform™ (SEED) is being used and extended to help jurisdictions manage building benchmarking and BPSs programs.

2. Background

As benchmarking, RCx, and BPS ordinances are passed nationwide; there has been more scrutiny on the access to and quality of empirical whole-building characteristics, metered energy data, and water data. These data are typically accessible only to utilities, building owners, portfolio managers, bill aggregating companies, and energy service contractors. Furthermore, data inconsistencies across data owners are frequent. The progression of these ordinances has been from benchmarking to RCx (audits) and now to BPS ordinances. In many cases, the BPS requirements cover the earlier benchmarking steps and can include a pathway for compliance that requires an energy audit; thus, this section will focus mainly on the BPS as the most comprehensive policy. Many ordinances also contain a transparency requirement where some benchmarking and BPS data are publicly released. This section will provide an overview of the policies that require the collection of building data and the data management challenges that are associated with these policies.

2.1. Benchmarking and building performance standards

Benchmarking requirements are the predecessor to auditing requirements, RCx, and BPS ordinances. Benchmarking reporting requirements are nearly the same in each jurisdiction. Building owners or their energy providers must submit to the city's administering agency a 12-month history of all energy bills (electric, natural gas, district, fuel deliveries) and specific building details, such as gross square footage, year built, and operating hours (Palmer and Walls, 2017). In 2017, Palmer et al. (Palmer and Walls, 2017) cataloged that the building submission compliance rate from 8 jurisdictions was between 73% and 99%, and Institute for Market Transformation (IMT) in 2021 showed compliance of 14 jurisdictions was between 55% and 99% (Institute for Market Transformation, 2021). Many of the jurisdictions have exemptions and financial penalties for non-compliance (Bugnion et al., 2022). Focusing on San Francisco, which has been benchmarking since 2011, the compliance rate has been consistent over the last eight years, with lower compliance rates seen in smaller floor area buildings, see Fig. 1. The compliance rates of small buildings (as low as 10,000 ft) are challenging due to the lack of property management companies to handle the data collection and submission (San Francisco Environment, 2020).

BPSs are policies that regulate buildings' energy performance or

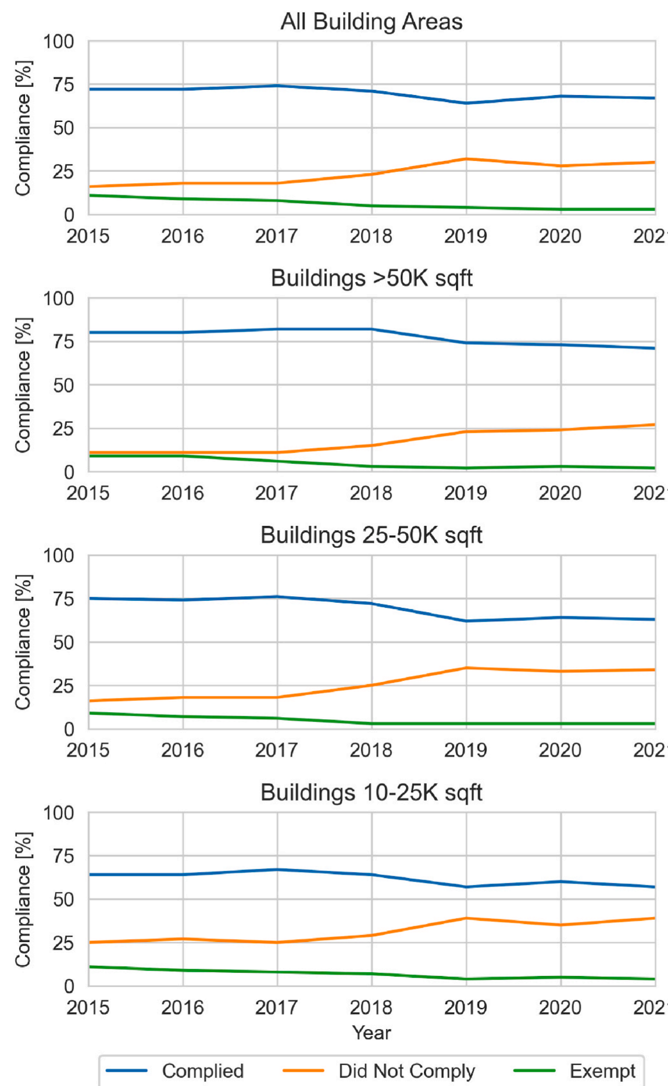

Fig. 1. San Francisco, CA compliance rates by building size and year (values might not sum to 100 due to rounding).

carbon emissions. These policies are designed to help reduce the carbon footprint of buildings and promote energy efficiency. BPS policies vary widely. They are more complicated than benchmarking programs since they require building owners to comply with specific pathways, such as meeting an energy or emission target, showing progress through energy audits, or choosing a financial compliance pathway. For example, the BPS policy in Washington, DC, gives building owners four pathways to comply; the program resets every six years, and the standards become more stringent (District of Columbia, 2018).

BPSs depend on tracking one or more building-level metrics over time and ensuring the metric is approaching a set target. As of early 2024, 53 jurisdictions (cities, counties, states, and the US federally owned and operated buildings) have enacted benchmarking ordinances. Of those, 15 have also enacted some form of a BPS ordinance (Institute for Market Transformation, 2023b; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023). There are two recently enacted state-level BPS ordinances that did not previously have benchmarking requirements (Maryland and Oregon), which brings the current total number of jurisdictions with BPS ordinances to 17.

The requirements of BPS vary by jurisdiction and can be broken into greenhouse gas (GHG)- and energy-based compliance. The common GHG metric is Scope 1 and Scope 2 emissions as calculated by ESPM in the form of yearly metric tonnes or kilograms of equivalent carbon dioxide per square foot of conditioned building area ($mtCO_{2e}/ft^2$ ·year or $kgCO_{2e}/ft^2$ ·year). The Federal BPS is currently only Scope 1 emissions (Council on Environmental Quality (CEQ and), 2022; The White House, 2021). The energy-based compliance pathways include annual weather-normalized site energy use intensity ($kBtu/ft^2$ ·year), annual weather-normalized source energy intensity ($kBtu/ft^2$ ·year), and/or an ENERGY STAR Score (1–100 with 1-poor, and 100-efficient) (US Environmental Protection Agency (EPA), 2021). Of the total jurisdictions with BPS ordinances, 5 include GHG-based compliance only, 9 include energy-based compliance only, 2 have pathways for either, and one is still undetermined (note that a jurisdiction can have an enacted BPS but may not have the rules written the same year).

In 2022, a National BPS Coalition of cities and states was launched with commitments to reduce energy and greenhouse gas emissions in their respective jurisdictions (Institute for Market Transformation, 2022). Cities are also leading the climate effort at a local level, with over 170 cities taking on commitments to reduce their greenhouse gas footprint (McCoy, 2019), many without local laws or policies enacted at the moment. At the US federal level, Executive Order 14057 (EO-14057) was passed requiring federal buildings to reduce Scope 1 emissions (Council on Environmental Quality (CEQ and), 2022). With the focus on buildings' contributions to climate change, it is expected that the growth of jurisdictions that enact BPSs will continue to increase; an additional 29 jurisdictions already have BPS ordinances under development or consideration.

After a BPS policy is enacted, the jurisdiction needs to define the specifics of their policy by determining the metrics, target values, building types, and data collection cycle periods. As shown, the building-specific metrics are typically GHG- or energy-based. Still, the target setting for each building can be involved and requires substantial effort and community engagement (ASHRAE, 2023). An important decision early on is determining the cycle period that will be used and the number of years between the target value and the compliance check. The compliance pathways and exceptions across BPS ordinances vary widely, with some commonalities around requiring approved audits or performance plans to be submitted to the jurisdiction if the target is unattainable within the cycle period (Institute for Market Transformation, 2023c). The BPS implementations look similar to the RCx and auditing policies where building owners have a cycle period to implement the upgrades. The majority of the BPS that allow an audit pathway are encouraging the use of American Society of Heating, Refrigerating, and Air-conditioning Engineers (ASHRAE) Standard 211–2018 (ASHRAE, 2018) and its definition of a level 2 audit for demonstrating compliance.

Table 1 shows the current list of BPS policies, the year they were enacted, the initial year of compliance, and the data sources. In many cases, the initial year of compliance has yet to be reached, with Boulder, CO, Chula Vista, CA, and Washington, DC being exceptions. The initial year of compliance is the first year the requirements apply to the building types and areas specified, not the first year of compliance evaluation (Nadel and Hinge, 2020).

The type of data collected by jurisdictions differs widely based on many factors such as jurisdictional conventions, access to tax lot/parcel data, Geographical Information System (GIS) department, and third-party evaluation of the jurisdiction (e.g., street maps, satellite imagery). The focus on data requirements for BPS is due to the complexities of tracking multi-year compliance across thousands of buildings for many types of data (e.g., monthly metered data, GHG emissions, building characteristics). In addition, the tracked data needs to include building owner contact information, communications, and support methods. Lastly, the time required to manage the program successfully is significant and costly. Engelman et al. surveyed three jurisdictions and showed that seven to nine people were needed to manage the entire program (Engelman et al., 2023). Moreover, Webb et al. determined that energy retrofits will be required in most buildings (over 50%) to reach BPS targets. Still, the overall BPS policy will produce considerable savings up to 45% of the energy or GHG goals (Webb and McConnell, 2023). This further emphasizes the need for adequately staffed programs with accurate, easily managed data at their fingertips.

Table 1
List of cities and states with enacted BPSs.

Location	Enacted/ Initial Year	Source
Aspen, CO	2022/TBD	(City of Aspen, 2022; DOE Building Energy Codes, 2023)
Boston, MA	2021/2025	(City of Boston, 2021; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Boulder, CO	2019/2021	(City of Boulder, 2024; DOE Building Energy Codes, 2023)
Cambridge, MA	2023/2025	(City of Cambridge, 2023; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Chula Vista, CA	2021/2023	(City of Chula Vista, 2021; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Colorado	2021/2026	(State of Colorado, 2023; State of Colorado, 2021; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Denver, CO	2021/2024	(City and County of Denver, 2021; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Federal BPS	2022/TBD	(Council on Environmental Quality (CEQ and), 2022; The White House, 2021; Institute for Market Transformation, 2023c)
Maryland	2022/2025	(State of Maryland, 2022; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Montgomery County, MD	2022/2028	(Montgomery County, 2024; Montgomery County, 2022; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
New York City, NY	2019/2024	(City of New York, 2019; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Oregon	2023/2028	(Oregon Legislative Assembly, 2023; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Reno, NV	2019/2026	(Reno Nevada Administrative Code, 2023; DOE Building Energy Codes, 2023)
Seattle, WA	2023/2027	(Durkan, 2021; City of Seattle, 2024; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
St. Louis, MO	2020/2025	(City of St, 2020; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Washington	2019/2026	(State of Washington, 2022; State of Washington, 2019; Washington State Department of Commerce, 2024; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)
Washington, DC	2018/2021	(District of Columbia, 2018; Institute for Market Transformation, 2023c; DOE Building Energy Codes, 2023)

2.1.1. Public disclosure data and transparency ordinances

Public disclosure data and transparency ordinances are a class of practices where the underlying data used for benchmarking, RCx, or BPS, are made public in an anonymous form. The type and granularity of the data vary by jurisdiction. They can include building location (a unique identifier, address), building characteristics (property use type, gross floor area, year built), energy use and water use (annual or monthly data aggregated or by meter), and other data, including BPS targets, compliance status, and/or property name. Nims et al. summarize that by promoting transparency, these programs encourage building owners and occupants to access and understand energy data. Transparent data can lead to informed decision-making, better resource allocation, and improved energy efficiency (Mims et al., 2017). Of the 54 jurisdictions with benchmarking ordinances, 47 have a transparency requirement with varying timelines for releasing subsets of the data. In some cases, BPS-related targets are also made available, allowing the public to see the progress throughout the building stock (Institute for Market Transformation, 2023b; Institute for Market Transformation, 2023c).

Researchers, program implementers, and third-party energy service companies have used public disclosure data to evaluate the effectiveness of the ordinances and to provide recommendations on energy-efficient building upgrades. For example, San Francisco leveraged city-collected data to determine building upgrade paths for existing buildings (Hooper et al., 2018); Chen et al. used public disclosure data to create city-wide urban models for energy use (Chen et al., 2019); and Yang and Papadopoulos et al. used public disclosure data to develop new target energy use intensity (EUI) models and elaborate on issues with ENERGY STAR Scores (Yang et al., 2018; Papadopoulos and Kontokosta, 2019). Roth et al. stated that "the feasibility of utilizing open data to construct robust data-driven urban energy benchmarking models" can be accomplished with "only a few key variables—Total Area, Type, Partial Areas, and Water Use" (Roth et al., 2020). Further, the study showed that data collection, modeling, and building ratings benefit a wide range of stakeholders including landlords, tenants, investors, energy service companies (ESCOs), policymakers, and many others.

Public disclosure formats vary by jurisdiction and include annual reports in PDF format, GIS-based visualizations, paginated tabular views, and downloadable spreadsheets. To date, there is no standardized data format for compiling, storing, or consistently sharing data. However, some conventions often stem from the use of ESPM as the originating data source. The ability to import data into a spreadsheet or database is essential for researchers and program implementers to conduct large-scale analyses. More information on the data management practices for public disclosure data can be found in the subsection 2.2.

2.2. Whole building data standards and data management platforms

The core of good data management is the use of standards and platforms, and the core of good analysis is the use of reliable, accurate, and consistent data. This section discusses commonly used whole-building data standards, common whole-building data management platforms, and the role of these tools in standardizing performance metrics. This section focuses on data collected for whole building energy tracking; it will not discuss the expansive field of semantic interoperability within the building controls and automation industry, asset tracking, or building design and construction, although there is an overlap in the data type required.

Within the field of ontology, the lowest level of formalization and expression is a glossary (Rebstock et al., 2008), typically aligned around the common use of terms and, more concretely, data dictionaries. The ability to agree upon terminology, definitions, and physical units directly affects how interoperable a data standard will be. Within the context of whole building data, the Building Energy Data Exchange Specification (BEDES) data dictionary provides a standard set of terms to facilitate consistent data exchange (LBNL, 2018; Pritoni et al., 2021). BEDES is supported by the US United States Department of Energy (DOE), and many organizations have adopted the terminology as canon. The terms are updated regularly and incorporate feedback from the building community. Many platforms in this section rely on BEDES terms.

Several expressive formats of whole building data provide more structure (formalization). BuildingSync is a eXtensible Markup Language (XML) schema designed for improving data exchange related to commercial building energy audits (Long et al., 2021a). BuildingSync is built upon the BEDES terms to ensure alignment with industry-standard definitions. It is also recognized as a recommended format within ASHRAE Standard 211–2018's appendix for commercial building energy audits (ASHRAE, 2018). High Performance Building XML (HPXML) is an analog format to BuildingSync but focused on residential data (Department of Energy, 2013). For utility-specific meter data, Green-Button is a popular XML format in which many utility companies allow their user to download their data (Green Button, 2020). Other whole-building formats exist, including (and not limited to): Green Building XML (gbXML), which is a common XML schema for building design and simulation data exchange (GbXML, 2024), and CityGML and GeoJSON which are practical formats for higher level building data and their connection to urban settings including 3D modeling and urban planning (Ali et al.).

In addition to established standards, various tools and platforms have emerged as assets in the realm of building data management, that is, managing portfolios of building characteristics and related information over time. At the scale of managing 10's to 100's buildings, spreadsheets (more specifically, Microsoft Excel) are the de facto solution and provide organizations with the flexibility to implement rapid solutions. The limitations of spreadsheets are commonly the lack of consistency, data quality, null cells, ability to codevelop or share, managing relational data, and overly complex implementations (Broman and Woo, 2018); further, a study showed that 44% of experienced spreadsheet users introduced errors into the spreadsheet (Brown and Gould, 1987). ESPM is a commonly-used tool for submitting and tracking large portfolios of buildings and meters (ESPM uses properties as the term for buildings) (US Environmental Protection Agency (EPA), 2022). ESPM is the recommended solution for benchmarking and BPS policies for building owners to submit data in the US (ASHRAE, 2023) and is also available in Canada. A comprehensive and open solution for managing building characteristics and performance data specifically for tracking BPSs progress has yet to emerge.

Additionally, capabilities of Urban Building Energy Modeling (UBEM) have advanced the ability to manage multiple simulated buildings. UBEM integrates city-wide data, building energy modeling, and urban scale analysis to stakeholders for assessing energy consumption patterns and optimizing energy efficiency at a city-wide level. However, the user focus of these UBEM tools is not often on jurisdictions needing to ensure building performance and compliance. When looking at building portfolios, multiple projects focus on integrating empirical data and model data to provide city-wide analysis (Hong et al., 2016; Kontar et al., 2020; Long et al., 2021b). The UBEM research area requires detailed data to conduct large-scale, urban-wide analysis of buildings. A significant effort is needed to synthesize data sets into coherent, consistent datasets that can be more easily sent to analysis programs (ultimately whole building energy modeling software). Chen et al. state that the data required for UBEM include "the GIS building footprint, building height, total number of stories, number of stories above ground, number of stories below ground, total floor area, heated floor area, number of dwellings, year of construction, year of refurbishment, use type (building type), heating system type, annual electricity use, annual natural gas use, annual site energy use, and annual source energy use" (Chen et al., 2019).

As jurisdictions strive to achieve sustainability goals and mitigate climate change, portfolio data management and its connections to detailed analyses (including UBEM) is a pivotal tool for informed decision-making and strategic urban development.

3. Requirements and evaluation terminology

The data collected for benchmarking and BPS purposes are, unsurprisingly, focused on building characteristics and generated data such as energy and water consumption, events occurring to the building (e.g., energy audits), and building-adjacent data (e.g., contacts, tracking unique identifiers). This section will discuss the challenges of developing data platforms for managing building data and how they relate to the best practice definitions enumerated below.

Based on section 2, several software development requirements are essential to implementing a solution for tracking benchmarking and BPSs policies. Due to the myriad of data sources present in jurisdictions and their individual formats (e.g., building data, tax parcel data, building footprints, ESPM data), it is essential for the workflow to allow multiple data imports (multiple here can mean both differently formatted data and repeated imports of updated data over time). Data ingestion occurs frequently in many jurisdictions; for example, Washington, DC, imports data from ESPM nightly. Furthermore, the fields of the data records being imported can change over time, and the units of those fields might differ (e.g., $mtCO_{2e}$ and tCO_{2e}).

The methodology to describe the advances in data management will use the following seven terms and definitions: data integrity, accuracy, consistency, shareability, extensibility, security, and transparency. The definitions are the following:

1. **Data integrity**: Data integrity refers to the data having only singular records, and changes to records are tracked over time. In the case of buildings, it denotes a single building of record within the data system.
2. **Accuracy**: Data accuracy refers to the degree to which data correctly represents the real-world object. Accurate data is generally free from errors, omissions, and inconsistencies.
3. **Consistency**: Consistent data ensures that data values remain uniform and coherent across different systems, unit systems, databases, and reporting cycles. Consistency ensures that data remains reliable and accurate.
4. **Shareability**: Shareability refers to the ease of access to the data by various people while preserving accuracy and consistency. For example, the ability to collaboratively edit within a system, methods to provide accessible knowledge transfer when staff changeover occurs, and potential access to public data through web feeds.
5. **Extensibility**: Extensibility is the ability of the platform to integrate into third-party applications or be extended through open-source software updates.
6. **Security**: Data security safeguards data against unauthorized access, alteration, or disclosure. In this context, rule-based access control (RBAC) is commonly required to ensure only authorized users can access the information needed. Secure data environments protect data integrity by preventing unauthorized modifications or tampering.
7. **Transparency**: Transparency refers to software and systems being inspectable to ensure accuracy with calculations and processes. While open-source software is often considered transparent, poorly developed open-source code is not always so. The data underneath the platform should also be completely accessible to the originating owner, i.e., data can be liberated from the platform.

The above terms have specific meanings in the context of benchmarking and BPS. This section will provide examples and describe the need for advances in data management specific to this type of data.

The core of the *data integrity* challenge is the ability to identify a building within a dataset uniquely. Note that the unit of record for benchmarking and BPS is the building or property. The building's address is most commonly considered first when performing the unique identification. Unfortunately, this leads to many integrity challenges: addresses are not always globally unique, often have alternate spellings or easy misspellings, and a single building can include multiple addresses, see Fig. 2.

Another critical component of *data integrity* is identifying which tax parcels have which buildings and vice versa since most jurisdictions are only aware of the existing tax parcels. Within the tax parcel, the jurisdiction's knowledge might be limited to the building type and the total gross floor area of all buildings. It is common in large cities to have configurations of buildings shown in Fig. 3, where the mapping between a building and a tax parcel is not straightforward. The tax parcel owner is not necessarily the building owner or manager, exacerbating the data collection challenge.

The final *data integrity* challenge is tracking all changes made to a

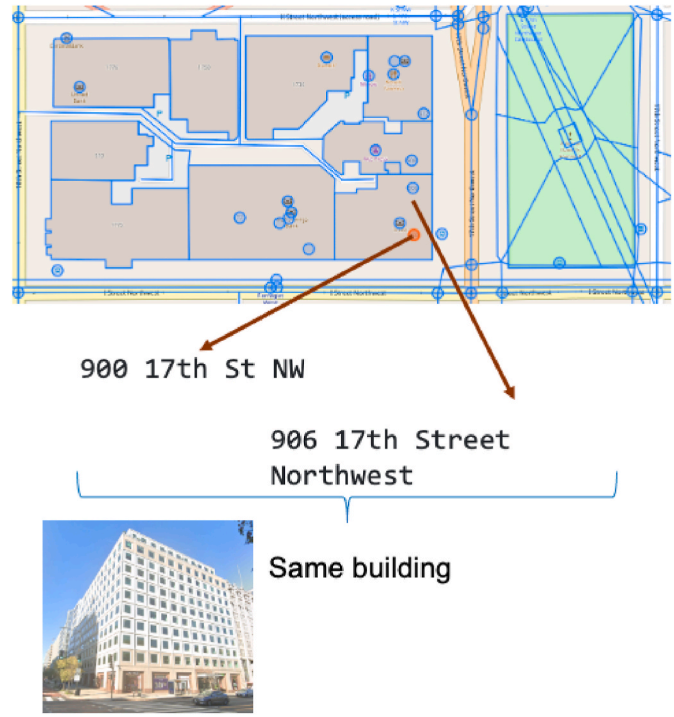

Fig. 2. There are many cases where buildings have multiple addresses, and often the convention of the names is not consistent.

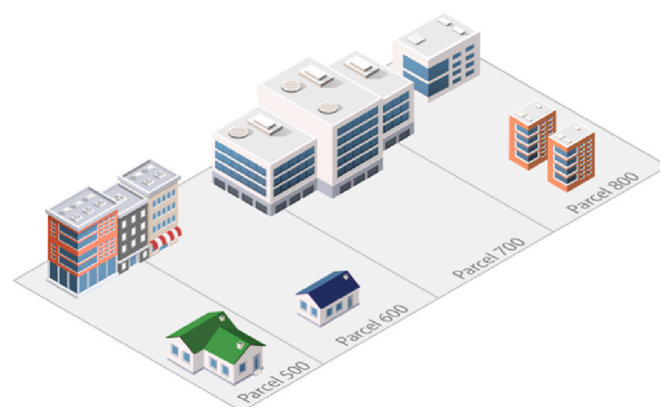

Fig. 3. Buildings and tax parcels are often not a simple one-to-one mapping, requiring software to handle the many-to-many relationships.

single building record over time. This level of auditability is needed to ensure accuracy and historical consistency. The ability to track which buildings are merged to form a new unique record, unmerged to create two distinct records, and updated allows future platform users to understand and verify the decision-making process and revert to more accurate data representations. *Data integrity* is similar to data reliability as they are both concerned about the data quality over time. The term integrity was used due to the annual scope of the data and the need for annual data integrity for reporting requirements. Data reliability is also important as it ensures data can be trusted over many compliance cycles.

Accuracy is another vital tenet to consider as improving data accuracy "could allow for a faster transfer of high-quality information from the jurisdictions to market actors" (Mims et al., 2017). Some of the values will inevitably be inaccurate when working with real data. The sources of inaccuracies or omissions can be user entry errors, faulty data collection devices (e.g., meters), incorrect records, and communication (human or computer) issues. For example, a user might type in the gross floor area of the building incorrectly or use incorrect units, or an energy meter may fail to report the monthly energy consumption for 2 of the 12 months of the year.

Building data sources vary greatly, including multiple spreadsheets, application programming interfaces (APIs), paper copies, and other formats. The data management structure must ensure data *consistency* with common naming conventions (data dictionary) and units. For example, gross floor area in specific systems is called square footage or conditioned floor area, and the physical units might vary, such as m^2 or ft^2. It is critical to understand the precise definition of these fields to ensure data *accuracy* and *consistency* and provide a common data dictionary and easily convertible unit systems.

Jurisdictions with more extensive building portfolios will have multiple people updating records at any given time. *Shareability* and database idempotence are critical to allow for a scalable and maintainable dataset of buildings. A central location where data is managed and updated will improve consistency and access. Further, processes and tools must be easily understood as teams working on the benchmarking and BPS compliance tracking move around and change.

Software solutions and data management tools within the building domain are aplenty, each with a bespoke use case. It is unreasonable to require a single solution for the entire domain, thus highlighting the need for *extensible* solutions. The data systems should enable integration with multiple existing tools and specifically allow for custom extensions through open-platform design or open-source community development for power users to enhance decision-making. Additionally, the number of fields collected for buildings can be numerous, and the database needs to handle the complexities of managing many fields.

Security is essential to protect building owners and personally identifiable information. Email, utility bill data, metered data, and energy costs are common in the database. *Security* needs to be balanced with accessibility to ensure that interconnected systems can readily access data without burdening the user (either human or another computer program).

Lastly, *transparency* is critical to build trust and ensure *accuracy*. Transparency helps stakeholders, including government agencies, developers, architects, and the public, to understand how decisions are made and to hold responsible parties accountable. It helps ensure that decisions are based on accurate and reliable information. Within the context of benchmarking and BPS, the ability to export all data related to an ordinance can encourage innovation and advanced visualizations and need introspection. Specific to BPS, many policies are being developed in parallel, each learning from previous policymakers. Having *transparency* enables future policymakers the ability to update benchmarking and BPS requirements (e.g., providing data to determine a better scoring methodology (Ding and Liu, 2020)).

4. Advances in data management for benchmarking and building performance standards

This section will discuss recent advancements for managing building data for benchmarking and BPS ordinances concerning the terminology of data integrity, accuracy, consistency, shareability, extensibility, security, and transparency. These advancements will highlight the challenges and how overcoming them will enable building owners and property managers to reach their climate goals on a building and portfolio basis. The ability to have consistent and robust data integrity readily allows users (jurisdictions and third parties) to develop actionable solutions.

This section will discuss the development of an open-source software tool and library to manage building characteristics, track multiple building metrics, and provide insights to a jurisdictional manager. The SEED project started in 2012 to meet the need for a new solution to manage the numerous benchmarking policies being enacted (Taylor et al., 2012; Alschuler et al., 2014; Long et al., 2020). SEED has evolved over the past years to help lead jurisdictions into the next phase of decarbonization through BPS (Bugnion et al., 2022).

Spreadsheets are ubiquitous in the engineering community and are often the preferred platform for quick prototyping and initial rollout. This is also true for benchmarking and BPS tracking; however, the data collected proliferates, and managing the data required to run a carbon reduction program for buildings becomes untenable. An open-source solution was deemed necessary to meet the needs of *transparency* and *data integrity* by providing solutions that will outlast the decades of tracking needed to bring the ordinances to fruition. Most jurisdictions have goals stretching out to 2045 and beyond, requiring data to be maintained and updated over many years and many different program administrators and through policy changes based on goals and advances in building technology.

SEED lowers the burden on jurisdictions implementing benchmarking and BPS programs by streamlining the process of collecting and managing data from diverse data sets and for large groups of buildings. The software identifies which facilities must comply with a jurisdiction's program, organizes and cleans the data, and interfaces with other programs to provide energy recommendations to decision-makers.

The following sections will discuss how SEED and the connected tools help address the concerns listed in section 3. The sections include data ingestion, mapping, matching and merging, inventory management, data quality, performance tracking, analysis pipelines, and program tracking.

4.1. Data ingestion

Initial data ingestion is arguably the most important portion of any database solution. A jurisdiction commonly creates the initial list of buildings covered by the enacted ordinance requirements, termed the covered buildings list (CBL). Many jurisdictions will create their building identifier when they generate their CBL. Determining a building identifier that can be used as the database's unique ID is critical. The building identifier can be used across multiple platforms, including SEED, Audit Template (a web application for collecting data for building energy audits) (Pacific Northwest National Laboratory, 2020), and tax parcel relationships. SEED and other tools support the use of Unique Building Identification (UBID). The UBID project aims to solve this challenge by generating a string-based identifier for each object on a map. The UBID can be assigned to a building or a tax lot. SEED has added functionality to handle the complex nature of UBIDs, including the ability to merge buildings or tax lots based on the Jaccard Index (Wang et al., 2019; Jaccard, 1901). It should be noted that addresses are not recommended to be used as the unique database identifier; however, SEED does apply address normalization on the field upon import to ensure that the address is consistent across all records.

There are also scenarios where a single building identifier is not sufficient. For example, some jurisdictions require building and portfolio manager identifiers. Combining the two becomes the "matching criteria" for records, and a positive match occurs when only both are equivalent.

The next critical consideration is establishing the length of the time cycles for the building programs being implemented (i.e., benchmarking or BPSs). A cycle is typically aligned to a reporting period, e.g., annual reporting. A set of cycles are often grouped in BPSs ordinances to track the entire compliance period (e.g., initial reporting year in 2019, evaluation in 2023, and compliance evaluation in 2024).

One of the prominent workflows of benchmarking and BPS policies is to import and manage data from many buildings continually. SEED provides a variety of methods to import data, including ESPM, most spreadsheets, BuildingSync, HPXML, GeoJSON, and direct API connections. Based on current benchmarking and BPS policies, the majority of the data import consists of importing first a CBL, a list of tax lots (normally used in large complex cities), then ESPM data either via a downloaded spreadsheet or a direct connection to ESPM. Many cases exist where no single data source provides all the information needed for the buildings and tax parcels, resulting in partial datasets. The partial datasets are common and require other features of data management to handle such as data quality checks to flag missing or partial data, field-by-field editing with logging to know who changed which field, and the ability to create derived columns based on multiple fields. The ability to import multiple partial data files to generate as complete picture of the entire jurisdiction's building stock is important.

The data import process can begin once solutions have been adopted around the sources of data, unique identifier(s), matching fields, and cycles. Fig. 4 shows the general process for data ingestion into SEED. The process undergoes multiple steps to ensure the data are consistent, clean, and have integrity. The imported data are initially stored in the SEED database entirely as a JSON data object (using Postgres' "JSONField" object). The imported data are considered ephemeral until the data are mapped and deduplicated. The following steps are conducted following the initial data import:

- *mapping*: After files are uploaded, file header columns need to be mapped to SEED columns. SEED columns can be canonical database columns (i.e., they are database fields), or the columns are "extra data" (i.e., they remain as objects within the "JSONField"). Each column can contain data types, physical units, display names, etc. SEED stores the mappings for all file imports, and the mappings can be stored in "mapping profiles" to be applied to subsequent file imports. Once the data are mapped, the entire object is hashed to create an easy-to-access identifier to check for duplicates. At this point, the data in the database are in an ephemeral state since an import often contains only a subset of the entire building record.
- *deduplication*: If two records are identical in the ephemeral storage (e.g., their hashes are the same), the last record is removed entirely from the import process. After this point, the remaining data is stored in the database as a new record, ready to be matched and merged with other records.
- *matching*: The matching and merging of two records is the process in which two building records are considered the same and will be joined together to provide the latest state. This provides a *consistent* data record that has *data integrity*. The matching process is based on the unique identifier(s) and matching fields. The matching process also matches and merges tax lots (if imported). The merging process handles conflicts by prioritizing the last-in record; however, the merging process is column-based, and a user can set a specific column not to merge the latest or to be "protected". After matching and merging, the data record is considered the principal record for the building or tax lot, and the record receives a new hash value that can be used in future matching processes (if needed).
- *pairing*: The pairing process creates the relations between the buildings and tax lots. The matching criteria for this process are based solely on a jurisdiction tax lot identifier. The import file can contain a semicolon-delimited list of tax lot identifiers associated with the building.
- *linking*: The linking process is reserved for linking records over time (or cycles). This functionality provides the cycle-over-cycle comparison capability. Data is easily compared over the years since the records have been imported with *consistent* mapping fields.

The import process is designed to be repeatable, and the data are stored in a way that allows for easy rollback; each merge within SEED is tracked and viewable. There is also a set of building-specific data stored as part of the main building, which includes relational data such as energy-saving scenarios, energy conservation measures, temporal sensor data, notes, analysis results, and other data.

4.2. Inventory management

SEED provides users with an interface reminiscent of spreadsheet programs, giving the user a familiar experience for data management and analysis but with many additional features specific to building and tax lot data. The features include sorting, filtering, and creating derived columns (columns calculated from other columns, providing *extensibility*), offering users familiar tools to manipulate and analyze data effectively. This supports the need for users to quickly filter and find potential issues for *accuracy*.

SEED provides flexible filtering capabilities for each column by providing customizable filters. Based on the column data type, the simple (e.g., > 50) or compound conditions (e.g., $< 50, < 100$) can be applied to one or many columns to refine the data views. This granular control enhances data exploration and facilitates the extraction of meaningful insights from large datasets.

In addition to filtering, users can apply unlimited labels to each record, facilitating granular data categorization and organization. It is common for jurisdictions to use these labels to flag buildings that do not meet compliance or are exempt. The labels can be easily recalled to show the label-filtered inventory view quickly.

Building on column filters and labels, SEED provides filter groups to improve data organization further. Filter groups combine labels, column filters, and sorting configurations to create tailored data views. Filter groups are used in the context of BPS to create a constrained data view for each of the performance tracks (discussed more in section 5). Due to many data fields, SEED allows column list profiles to be configured within the inventory, allowing users to display only the fields pertinent to their analysis. The column list profiles can be saved, updated, and applied.

Each building and tax lot has an accompanying detailed view with a comprehensive snapshot of each building and tax lot, displaying all

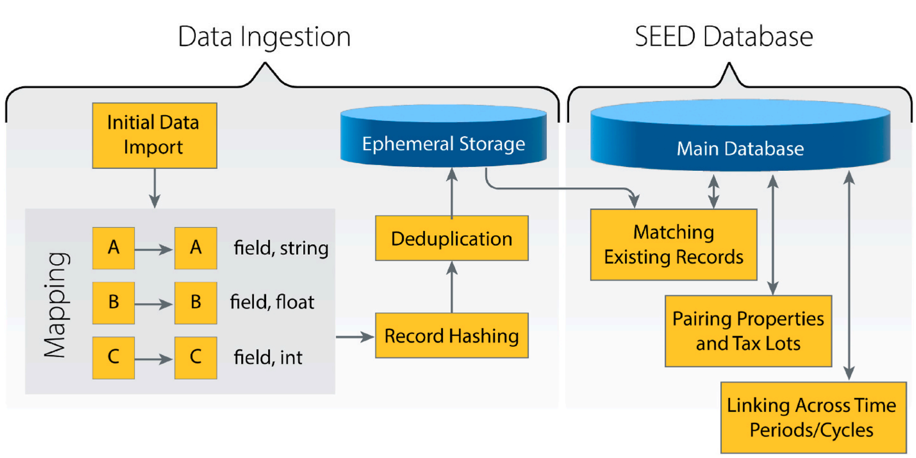

Fig. 4. Data ingestion process in SEED showing the steps of mapping, merging, matching, merging, and linking.

relevant imported files and the merged data (by column). This clear view offers additional functionalities such as note-taking, historical tracking, relationship mapping, and energy use analysis, facilitating thorough data exploration and evaluation within a single interface.

Combining these various inventory management features provides users with a comprehensive and flexible data exploration and analysis environment, enabling them to manage and analyze large datasets efficiently. The inventory management features in SEED are designed to support the *accuracy*, *consistency*, and *extensibility* of building and tax lot data.

4.3. Data quality checks

A fundamental tenet of SEED is the ability to ensure data quality and consistency, which is critical to maintaining the *accuracy* and *data integrity* of the data. Data issues include building outside the defined area, incorrect property types, duplicate entries, inconsistent units, format variations (e.g., 100K vs. 100,000 vs. 100000), various naming conventions, missing data, and out-of-range/incorrect data. The formatting of common issues and the unit types are addressed during the initial mapping process, but the other problems are addressed during the data quality checks. There are two stages of data quality checks: first, before the mapping is committed to the main database, and second, on demand from within the inventory management page.

Mitchell and Mathew documented a set of conventional data issues to be aware of when working with an initial set of building data from a jurisdiction tax assessor office, municipal data records, ESPM, or real estate data (Mitchell and Mathew, 2022). SEED has defaulted (but extensible) data quality checks to help detect and manage potential data issues. Table 2 shows the field-level checks that are defaulted in SEED. The field names are the canonical fields defined as part of the relational database. There are multiple checks, including *not null*, *range*, *must contain* and *must not contain* for strings, and *required*. If the data quality rule is checking a field that includes units, then the value is converted before it is compared to ensure validity. If the data quality rule is a *range* value, then if no minimum or maximum range is entered, it is unbounded on that end, respectively.

The on-demand data quality checks combine data quality rules with the ability to auto-label records that fail the data quality checks. Auto labeling can be used inversely to label records that pass the data quality checks. Auto labeling is a powerful tool to help users quickly identify records that need to be updated or those in good standing. The data quality checks can be run on any subset of records on an on-demand basis from within the inventory management view.

Advanced data quality issues can be created to ensure that values year-over-year (or across *cycles*) do not vary outside a predetermined threshold. Cross-cycle data quality checks are important for identifying potential issues with reporting inconsistencies. Unfortunately, many cases require a human to determine the root cause. Problems are often related to metering infrastructure, changes in building operations (e.g., warehouse converting to an office building), energy conservation measure implementations, and many others.

Although addresses are not recommended for use as matching criteria, SEED does provide access to third-party geocoding services to ascertain a building's latitude and longitude based on its address. Geocoding accuracy can vary (e.g., rooftop vs. street vs. locality) based on the geocoding service. However, visually verifying building data within a specified geographic area is an easy data quality check. For instance, Fig. 5 displays jurisdictions' imported public disclosure data on a map, facilitating rapid identification of issues like swapped latitude and longitude values.

The data quality checks within SEED provide jurisdictions with a powerful tool to ensure the *accuracy* and *consistency* of their building and tax lot data, thus resulting in all-around *data integrity*. The checks are designed to identify and manage potential data issues, facilitating the maintenance of high-quality data. The data quality checks are also designed to be *extensible*, allowing users to create custom checks to address specific data issues.

4.4. Building performance tracking

Several building-specific performance tracking features within SEED help jurisdictions track building performance over time. The features include monitoring energy use, water use, and carbon emissions. The tracking features provide users with a comprehensive view of building performance, enabling them to identify trends, track progress, and make informed decisions. The tracking features are designed to support the *accuracy*, *consistency*, and *data integrity* of building and tax lot data.

Energy and water data are most easily imported via a multi-tab spreadsheet containing links to the building identifier, meter metadata, and meter readings. The meters align closely to meter definitions in ESPM and include electricity, natural gas, fuel oils, steam, hot water, chilled water, and other less common onsite fuel uses such as wood, coal, and diesel. It is also possible to import meter data via an API connection to ESPM or through an uploaded GreenButton file provided by a utility.

Several BPS jurisdictions are receiving monthly energy consumption aligned to the months through ESPM. Intensity-weighted values are also provided, including annual EUI, annual water use intensity (WUI), and annual GHG intensity (GHGi). In some cases, the data within SEED are the raw utility bill data, which are rarely aligned to a calendar month; furthermore, the same building can have multiple meters with different billing cycles. SEED provides a few utility functions to "calendarize" meter data, which is the process of allocating meter readings to the months in which the energy was consumed (US Environmental Protection Agency (EPA), 2023). The calendarization process ensures the *accuracy* and *consistency* of the energy and water data.

The process of calendarization uses a weighted average approach. It involves dividing the total energy (or water) from the bill by the number of days in the billing cycle to get the per-day value, Equation (1). Next, the monthly allocation is calculated based on the per-day value and the number of days each month, Equation (2).

$$\overline{E}_{m,day} = \frac{E_{m,i}}{D_{m,i}} \tag{1}$$

where i is the billing cycle, m is the meter of interest, $\overline{E}_{m,day}$ is the daily average reading for the meter, $E_{m,i}$ is the meter reading from the bill, and $D_{m,i}$ is the number of days in the billing cycle.

$$E_m = \overline{E}_{m,day} \cdot D_{month} \tag{2}$$

where E_m is the monthly allocation, and D_{month} is the number of days in the month.

SEED also provides a few other utility functions to aggregate and clean meter data, including the following:

1. *extrapolate_meter_readings*: leverages the calendarization's average energy usage per unit of time and estimates the total use for each month before or after the period by extrapolating from the average.
2. *reject_outliers* Filters out readings whose z-score (a deviation from the mean) exceeds a specified threshold. This function helps remove outliers from aggregated meter readings; see Equation (3).
3. *interpolate_monthly_readings*: Interpolates missing months between the first and last readings in the provided list of meter readings. This function assumes that each reading represents a calendar month of data.

$$\text{mask} = \left[x_i \ \text{for} \ x_i \ \text{in} \ X \ \text{if} \ \left| \frac{x_i - \mu}{\sigma} \right| \leq \text{reject} \right] \tag{3}$$

where x_i represents each reading, i, in the list of meter readings, X, μ is the mean of the raw values, σ is the population standard deviation of the raw values, and *reject* is the threshold for rejecting outlier readings based

Table 2
Default data quality checks.

Field	Check Condition	Units	Min	Max	Severity
address_line_1	not_null				Error
conditioned_floor_area	range	ft^2	0	7,000,000	Error
conditioned_floor_area	range	ft^2	100		Warn
custom_id_1	not_null				Error
energy_score	range		0	100	Error
energy_score	range		10		Warn
generation_date	range		01/01/1889	12/31/2024	Error
gross_floor_area	range	ft^2	100	7,000,000	Error
jurisdiction_tax_lot_id	not_null				Error
occupied_floor_area	range	ft^2	100	7,000,000	Error
pm_property_id	not_null				Error
property_footprint	Invalid Footprint				Error
recent_sale_date	range		01/01/1889	12/31/2024	Error
release_date	range		01/01/1889	12/31/2024	Error
site_eui	range	$kBtu/ft^2 \cdot year$	0	1000	Error
site_eui	range	$kBtu/ft^2 \cdot year$	10		Warn
site_eui_weather_normalized	range	$kBtu/ft^2 \cdot year$	0	1000	Error
source_eui	range	$kBtu/ft^2 \cdot year$	0	1000	Error
source_eui	range	$kBtu/ft^2 \cdot year$	10		Warn
source_eui_weather_normalized	range	$kBtu/ft^2 \cdot year$	10	1000	Error
year_built	range		1700	2019	Error
year_ending	range		01/01/1889	12/31/2024	Error

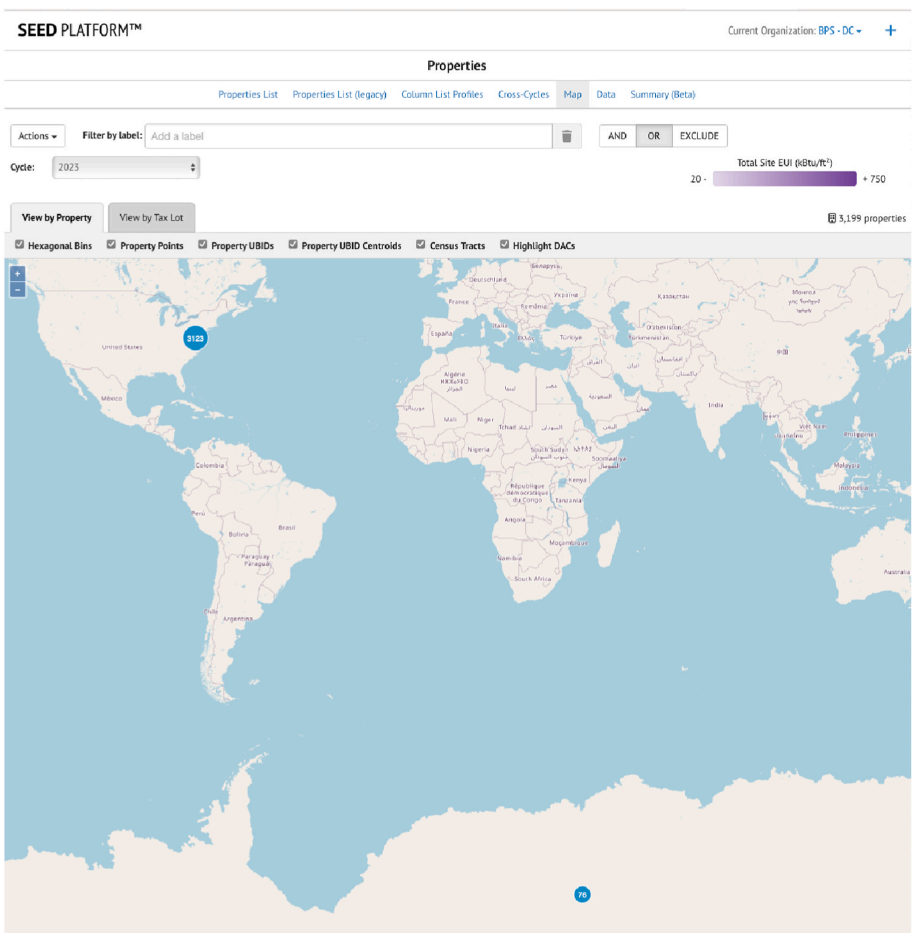

Fig. 5. The mapping feature allows a quick visual check on the location of all the buildings within a SEED organization.

on their z-scores, with one being the default. *reject* represents the standard deviations away from the mean.

SEED provides built-in analysis functionality to calculate and return the EUI and carbon emissions (GHG and GHGi); the process will be discussed further in subsection 4.5. The calculations leverage the calendarization and support methods defined above. The GHG calculations are based on Cambium's database (Gagnon et al., 2022) but should only be used if a jurisdiction does not provide locality-based emissions.

Lastly, a vital import capability within SEED is importing commercial building energy audit records that originate in the Audit Template tool. If a building has undergone an audit, the building data can live alongside the audit data. A building can also have multiple audits over time (or within a single year [e.g., level 1 audit and level 2 audit]). The data are imported into SEED through an API connection or directly from BuildingSync. The audit data contain scenarios and energy conservation measures (ECMs). A scenario is a set of ECMs with total performance and cost savings. An ECM is a single measure with implementation costs.

4.5. Analysis pipelines and asynchronous workflows

SEED enables *extensibility* through an analysis functionality that leverages asynchronous workflows. Multiple jurisdictions use this core feature to augment the building records with specific information. The process being asynchronous allows for longer-running analyses to be performed in parallel without blocking navigation within the website.

SEED's analysis pipeline serves as an abstract class for defining workflows for preparing, running, and post-processing analyses. In this context, an analysis is a self-contained software program requiring data from single or multiple buildings. The analysis pipeline is designed to be extensible, with various helper functions and programming hooks to enhance easily. These hooks provide flexibility and customization options for users to tailor analyses according to their requirements. The analysis can be "hard coded" in SEED, or a third-party web service in which SEED interacts. A SEED analysis runs entirely in SEED's background, and progress statuses are retrievable via the API and returned to the user interface (UI).

SEED is equipped with three embedded analyses, namely the CO_{2e}, EUI, and Energy Equity and Environmental Justice (EEEJ) analyses, which are executed entirely within SEED environment (albeit asynchronously). The results of these analyses are stored directly in the database alongside each building record, facilitating easy access for further analyses or integration with dashboards. Additionally, SEED supports two additional analyses conducted through third-party services: Building Efficiency Targeting Tool for Energy Retrofits (BETTER) and BSyncr. BSyncr serves as a data exchange layer utilizing an API and web service to execute the Normalized Metered Energy Consumption (NMEC) for R (the programming language) tool (KW-Labs and nmecr, 2024). The Normalized Metered Energy Consumption in R (NMECR) tool is instrumental in measuring and verifying energy conservation measures in commercial buildings, enhancing SEED's capabilities for comprehensive energy analysis and management. BETTER is a web application developed by Lawrence Berkeley National Laboratory (LBNL) that provides high-level ECM recommendations based on monthly metered energy data using change point models (Li et al., 2019.). BETTER has a well-defined API enabling easy integration into an analysis pipeline.

An analysis pipeline is extensible and requires developers to implement a few base class methods within the extended analysis. Fig. 6 shows the life cycle of an analysis and has the following key components:

1. *prepare_analysis*: Prepares the analysis by invoking the _prepare_analysis method of the concrete pipeline implementation. It locks the analysis object to ensure atomicity and updates its status accordingly.
2. *_prepare_analysis*: An **abstract method** that performs the necessary tasks for preparing an analysis, such as creating input files. This is required to be implemented within the extended analysis.
3. *start_analysis*: Initiates the analysis by invoking the _start_analysis method of the concrete pipeline implementation. Similar to prepare_analysis, it locks the analysis object and updates its status.
4. *_start_analysis*: An **abstract method** that starts the analysis, for example, by making HTTP requests to the analysis service or calling a "hard-coded" method within the application. This is required to be implemented within the extended analysis.

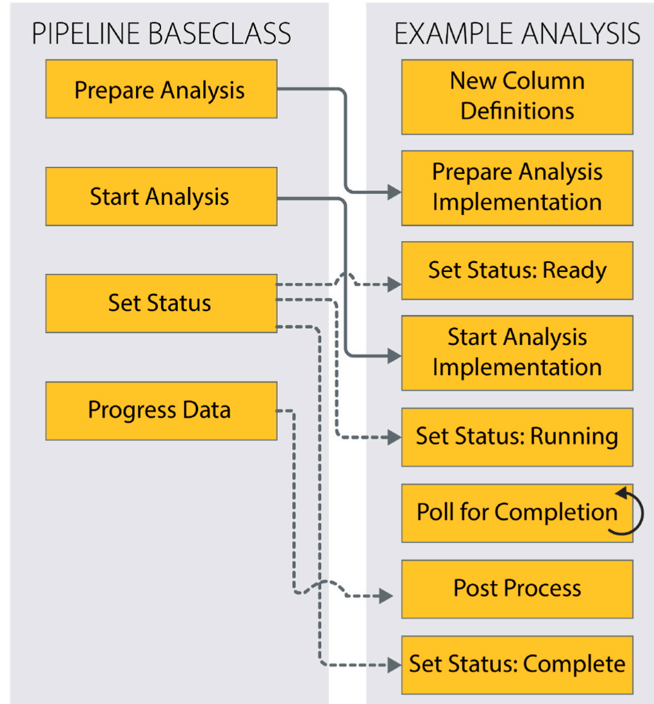

Fig. 6. Workflow of analysis pipelines.

5. *set_status*: Sets the status of the analysis to "READY", "RUNNING", or "COMPLETED" respectively. These methods ensure the analysis progresses through its life cycle properly.
6. *fail*: Marks the analysis as failed and logs an error message.
7. *stop*: Stops the analysis if it's not already in a terminal state.
8. *delete*: Deletes the analysis object.

Several helper methods are available for tracking analyses, including:

1. *_get_progress_data_key_prefix*: Generates a key prefix for progress data based on the analysis status.
2. *get_progress_data*: Retrieves progress data for the current analysis task.

The analysis pipeline is designed to provide *extensibility* and *shareability*. The analysis pipelines' open-source nature and platform-based development enable *transparency*. This feature, combined with inventory management, filter groups, data quality checks, and performance tracking, provides a comprehensive and flexible data exploration and analysis environment, enabling users to efficiently manage and analyze data sets for tracking, benchmarking, and building performance standards. The following sections will discuss two case studies on how SEED has been used to track building performance standards and how it is being extended by third parties to serve the needs of jurisdictions further. The ability to visualize the data throughout its life cycle is a critical feature of SEED. The results of an analysis pipeline can be visualized in the inventory management view, and more details of any particular analysis run can be viewed in the analysis itself.

In summary, there have been significant workflow and software advances for building- and tax lot-specific data sets. The data's uniqueness, sparseness, and error-prone structure have required the development of a new platform for tracking data. This is specifically important for jurisdictions with limited budgets for a custom solution but a large enough team with staff turnover where spreadsheets will fall short. This section focused on the importance of *data integrity*, *accuracy*, *consistency*, consistency, *shareability*, *extensibility*, and *transparency*. The topic of *security* will be broached during the second case study, section 6.

5. Case study: tracking building performance standards in Washington, DC

This case study focuses on an example implementation of a BPSs in the Washington, D.C. (District). The city has been a leader in the implementation of BPSs (note that the Washington, D.C. (District)'s BPS is termed Building Energy Performance Standard (BEPS)) due to its early adoption. Early on, the District used a collection of files, programs, and processes (e.g., multiple spreadsheets and Python glue code). As the complexities increased and more building data was collected, the District moved to a version of SEED in 2019. The District moved to a SEED-based platform called Building Energy Analysis Manager (BEAM) in 2021 to leverage additional requirements and will be discussed in section 6. This case study will provide an overview of importing and visualizing the District BEPS data within SEED. The data presented in this section are entirely public through Department of Energy and Environment (DOEE)'s public disclosure ordinance, thus the data can be used without concern for private or personal identifiable information.

5.1. Overview of Washington, DC's building performance standards

As discussed in section 2, the District was an early adopter of benchmarking laws in the US, enacting the Clean and Affordable Energy Act (CAEA) of 2008 (District of Columbia, 2008). This law mandates that owners of large private buildings benchmark their energy and water efficiency annually, reporting the results to the District's government for public disclosure. The earlier law covered buildings greater than 50,000 ft^2 gross floor area and the most recent law covers buildings greater than 25,000 ft^2 gross floor area. The District government also benchmarks and discloses the energy and water efficiency of its public buildings over 10,000 ft^2 gross floor area. These requirements aim to increase energy performance data availability, drive efficiency improvements, and reduce the city's greenhouse gas emissions, as buildings account for 74% of the District's emissions. Initially, managing the benchmarking requirement involved disparate files and processes across various software platforms, lacking a centralized system for access and management (Long et al., 2020). There was also no standardized Customer Relationship Management (CRM) system and minimal automation in processes or standard operating procedures. The passage of the *Clean Energy DC Omnibus Amendment Act of 2018* (District of Columbia, 2018) increased complexity by expanding the coverage of buildings under the CAEA and imposing specific performance targets on building owners.

5.2. Data ingestion

The District already leverages a SEED-based instance to track its buildings, this use case is a prototypical study on the ease of configuring and importing building-specific data into the SEED. The District publishes its annual benchmarking and BPS targets through its website.[1] This section provides an example of how the public portion of the data is loaded into SEED, quality checked, and visualized.

The District data was uploaded using the pySEED (Long et al., 2023) library. pySEED enables the data to be uploaded without user interaction of SEED's UI and provides a mechanism to reload the data multiple times since SEED's back end for identical records is idempotent. The need to upload data multiple times arose from generating a precise mapping profile, ensuring the correct fields from the source data are mapped to the proper fields in SEED. Within the SEED interface, there is a mechanism to create a mapping profile, and it reduces the chances of errors; however, the pySEED library has a simple comma-separated values (CSV) file format to define the mappings. The pySEED version of the mapping file requires more work to ensure accuracy. A sub-selection of the principal mapping fields from the District's data to SEED is shown in Table 3. The entire dataset has ~115 columns, whereas the dataset used by the District for their actual tracking includes over 500 columns. The table also shows which are canonical and which are extra data. The canonical fields in the database are mapped to the SEED column with underscores in the names (denoted below in *italics*). In contrast, the extra data columns are typically title-cased. The data mapping can support multiple unit types (if a field has units), and the conversion will automatically occur within SEED to make the data consistent.

In Table 3, *PS* is a property and *TL* is a tax lot. The table above indirectly highlights several inconsistencies in the data, including some having *EMISS* vs. *EMISSIONS*, *SOURCES* vs *SOUR*, incomplete spelling of *NORMALZED*, etc. Also, the names of the raw columns were updated to be readable columns for this table, as the originating data were all uppercase with no spaces.

When using pySEED or the SEED UI, the columns are automatically created for buildings and tax lots. The column descriptions are defaulted to the same name as the column, but users can change the description to be more descriptive. Once the mapping profiles were confirmed, the District's data set was broken into multiple files based on the reporting year. Each file (using pySEED) was uploaded into a newly created cycle set to the reporting year. The result was 12 cycles (years) of data starting in 2012 and ending in 2024 (no performance data for 2023 or 2024 yet). For the 2022 data, the number of records with a BPS target was roughly 4216 buildings and 1937 tax lots. After data ingestion, data quality checks are run to ensure the data is consistent and accurate, which is discussed in the following section.

Table 3
The District's data mapping into the database.

Raw Columns	Units	Table	SEED Columns
PID		PS	*pm property id*
SSL		TL	*jurisdiction tax lot id*
PROPERTY NAME		PS	*property name*
PM PARENT PROPERTY ID		PS	*pm parent property id*
REPORTING YEAR		PS	Year Ending
REPORTS TATUS		PS	Reporting Status
ADDRESS OF RECORD		PS	*address line 1*
OWNER OF RECORD		PS	*owner*
WARD		PS	Ward
CITY		PS	*city*
STATE		PS	*state*
POSTAL CODE		PS	*postal code*
YEAR BUILT		PS	*year built*
PRIMARY PROPERTY TYPE SELF SELECT		PS	*property type*
TAX RECORD FLOOR AREA	ft**2	PS	*gross floor area*
ENERGY STAR SCORE		PS	*energy score*
SITE EUI_KBTUFT	kBtu/ft**2/year	PS	*site eui*
WEATHER NORMALZED SITE EUI_KBTUFT	kBtu/ft**2/year	PS	*site eui weather normalized*
SOURCE EUI_KBTU_FT	kBtu/ft**2/year	PS	*source eui*
WEATHER NORMALZED SOUR EUI_KBTUFT	kBtu/ft**2/year	PS	*source eui weather normalized*
TOT GHG EMISSIONS_METRIC TONS CO2E	mtCOe/year	PS	*total ghg emissions*
TOT GHG EMISS INTENSITY_KGCO2E FT	kgCOe/ft**2/year	PS	*total ghg emissions intensity*
WATER USE_ALL WATER SOURCES_KGAL	kGal	PS	Water Use
NATURAL GAS USE_THERMS	therms	PS	Natural Gas Use
FUEL OIL AND DIESEL FUEL USE KBTU	kBtu	PS	Fuel Oil and Diesel
METERED AREAS_ENERGY		PS	Metered Areas (Energy)
METERED AREAS_WATER		PS	Metered Areas (Water)

[1] https://opendata.dc.gov/datasets/building-energy-benchmarking/.

5.3. Analysis of data tools' effectiveness for Detecting data quality issues

The first hurdle encountered was the complexity of importing ten years of building data and the realization that building identifiers were not consistent during the early phases of the project. The main building ID used in the District's data is the ESPM building identifier (or just PM Property ID). The number of buildings imported in each cycle shows that the last reporting year (2022) contained 3199 buildings, up from 3155 in 2021. The 4216 buildings now in SEED clearly show that not all buildings are being tracked, and some may have had a new PM Property ID generated over the last ten years. The number of buildings is expected to increase year-over-year as new buildings are built and policies are updated, adding new building compliance criteria. Regardless, the default data quality checks (presented earlier in Table 2) were run on the 2022 records. The results were exported and categorized, and the results are shown in Fig. 7.

The figure shows that "address line 1" was null (or blank) for over 800 records. This was an interesting result because once filtering out null address line 1 fields using SEED's column filtering, the total number of records in 2022 with non-null address line 1 fields was 3199. This result shows that the most recent data imported always contain addresses, and the total number of buildings being tracked is 3199. This research did not attempt to reconcile the historical data. It was also noted that of the 3199 buildings in 2022, 2701 used more than one fuel type. Per the data quality results, ~250 buildings had a null or out-of-range EUI, and another ~50 had a null or out-of-range ENERGY STAR Score. Lastly, with the default data quality checks, only a handful of buildings were marked as having issues with the gross floor area and/or year built.

Overall, the data quality checks are designed to ensure the *accuracy* and *consistency* of the building and tax lot data, and the results show that the data quality checks effectively identify potential data issues. Unfortunately, many of these issues require a human in the loop to determine the source of the error and correct it. SEED keeps track of all changes made to the building records to ensure auditability. With advances in artificial intelligence and machine learning libraries, it is possible that these human-in-the-loop checks could be mitigated with enough data and training.

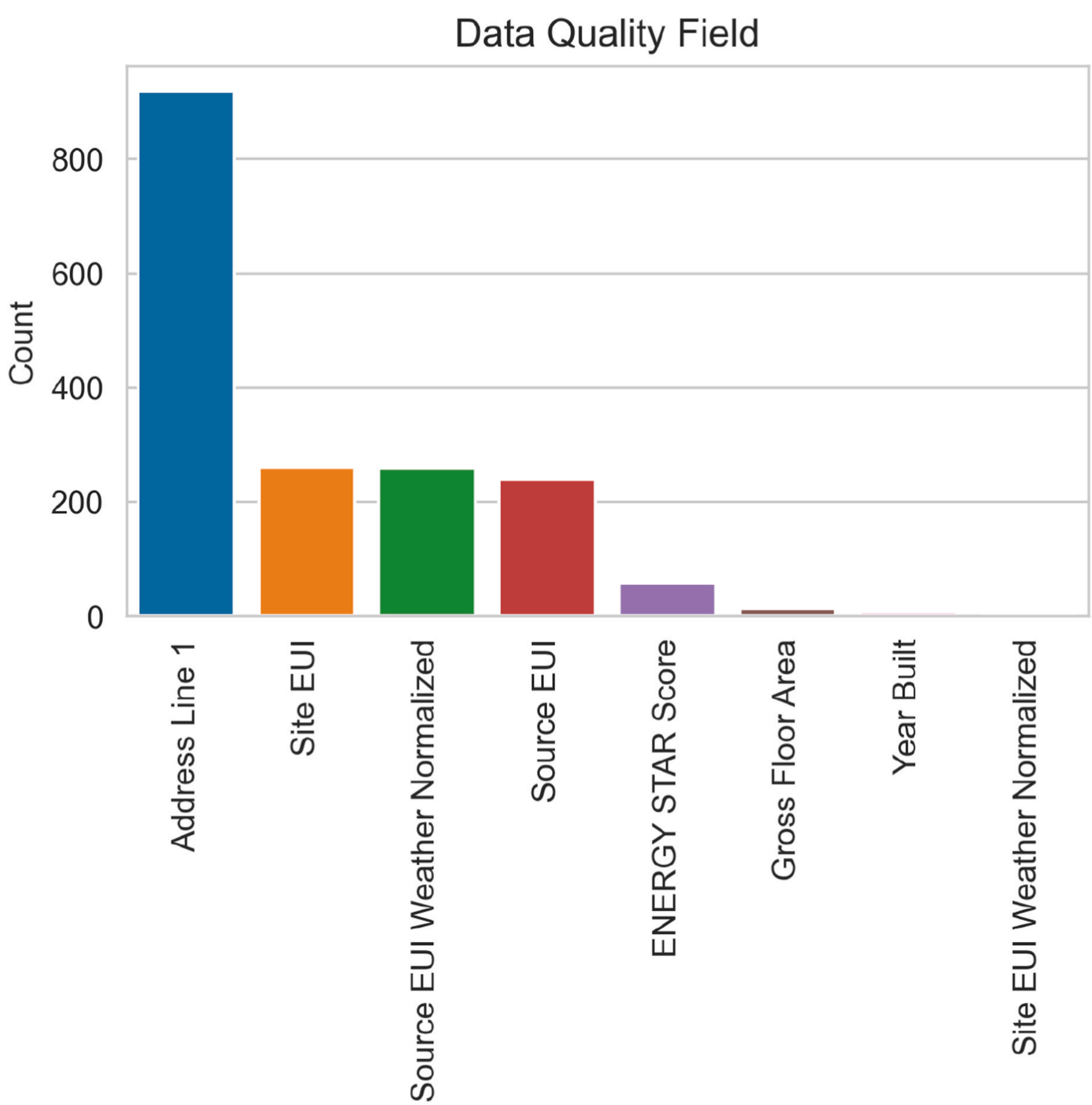

Fig. 7. Categorized data quality results for the District's 2022 compliance cycle.

5.4. Program configuration and tracking

For the District, buildings must be better than the local median ENERGY STAR Score or the equivalent metric of weather normalized source EUI. The buildings that do not meet the standard must complete a compliance pathway to come into compliance. This use case leverages only the buildings with the ENERGY STAR Score target set in the disclosed data. Of the 3199 buildings (in 2022), 1358 were assigned ENERGY STAR Scores.

Multiple programs can be tracked in SEED and are configured on an organization basis. The selected buildings that are tracked within a program are defined by a filter group. The filter group used for this use case filtered buildings with an ENERGY STAR Score compliance target, property type is non-null, target score is not null, and where the address was not null. Once the filter group is created, a program (e.g., energy target or GHG target) can be made, and the user selects the columns that denote the actual use and the targeted use (of the compliance metric). The user can also specify the compliance condition, that is, does the actual value need to be greater than the target (which is the case for ENERGY STAR Score), or is the value less than the target (used for EUI, WUI, and GHGi).

SEED provides visual representations of data through bar charts and interactive maps, facilitating intuitive data exploration and tracking of the programs. Fig. 9 shows the count of buildings per cycle of the ENERGY STAR Score buildings. The cyan colors are buildings in compliance, and red are those not compliant. The targets were configured per the District's public disclosure data.

Another valuable graph for jurisdiction administrators is the insights graph. This graph shows the target value and actual value for each building within the cycle year. Fig. 8 shows this as a rank-sorted graph. The graph shows how much each building must improve to achieve compliance. The graph is helpful for quickly identifying the buildings that are not in compliance and the improvement needed to reach the compliance goal. Each point is an individual building, and the details of the building can be viewed by clicking the point. The figure shows the target of the multifamily buildings is an ENERGY STAR Score of 66 (meaning that the building energy performance is better than 66% of the building's peers.

6. Case study: seed as a platform

When SEED was created, it was planned to be both a platform for jurisdictions to use directly and for third-party companies to leverage and build upon. To this end, SEED has been leveraged by a few third-party companies to support additional functionality and integration with other tools, such as adding a help desk, tracking milestones, and enabling CRM functionality. Curating an open-source project can be challenging and requires financial support and continuous updates by well-experienced developers. This section will provide examples of how SEED is being used and extended by third parties to meet the needs of jurisdictions, including two applications developed on top of SEED.

This case study will briefly discuss two applications built directly on top of SEED's code base, OPEN Technologies' GRID project and BEAM. GRID created a custom UI for their benchmarking and BPS, but the underlying data is stored and managed in SEED. This illustrates the flexibility of SEED as a platform due to SEED's well-developed API layer. SEED's UI is entirely driven through its own API, meaning that any functionality in the UI is available through the API. BEAM directly extended SEED's source code with regular syncing of the software repositories. BEAM enhances SEED's core functionality with the following:

- Alternative compliance and prescriptive compliance pathway tracking
- Tracking of milestones for buildings, e.g., due dates for BPS reporting
- Super-cycle integration to track building characteristics outside of a cycle or compliance period
- Customized pre-processing of jurisdictional data and importing into SEED via API
- Help desk linking
- Ticket tracking
- CRM integration

One of the features to elaborate on is the help desk and ticket tracking since a successful benchmarking and/or BPS program is only as good as the support it provides to building owners and managers. For example, in 2013, Seattle, WA received almost 10,000 emails and calls to support building wonders and managers (Mims et al., 2017). This burdens the jurisdiction support staff in addressing the ongoing

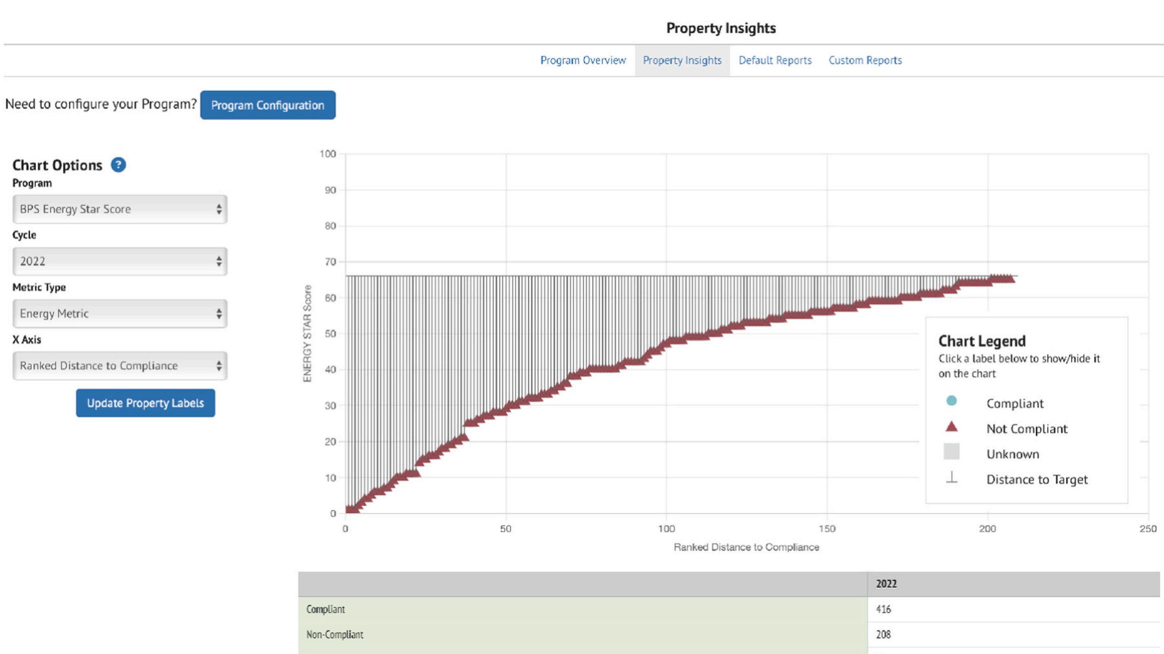

Fig. 8. Visualization showing the amount each building must improve to reach the compliance goal.

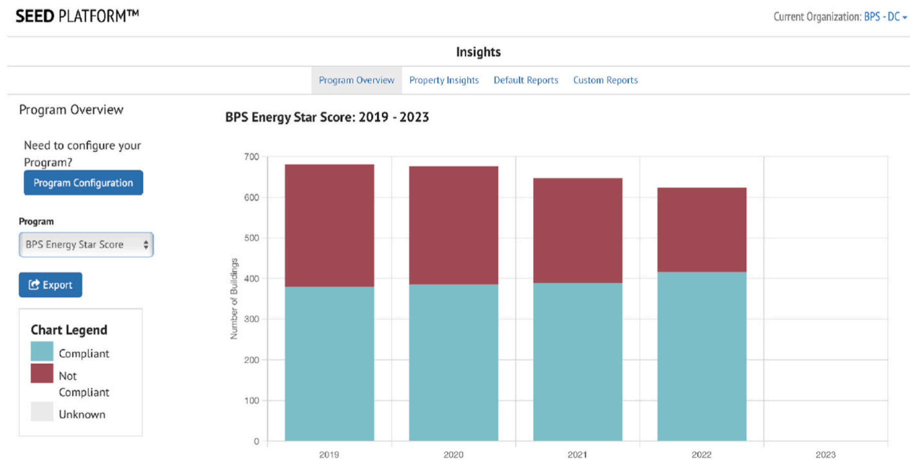

Fig. 9. Easy visualization tracking building compliance year over year.

questions and issues. Linking a help desk within SEED provides a comprehensive and consistent data management and support system. This is an excellent example of how SEED's core functionality can be expanded by third parties to provide jurisdictions with specific solutions.

SEED is developed as an open-source software application with a permissible license. The open-source approach enables the platform to be extended as needed. The structure of the platform shown in Fig. 10, which demonstrates how some users have built directly on the SEED stack while others have leveraged SEED as a developer platform. To date, SEED and SEED-based programs are a vital tool employed by 26 jurisdictions (cities, counties, or states) across the US and Canada.

Within SEED, adding connections to additional business tools is a common request. An example of this is SEED's interconnection with Salesforce. Salesforce is a popular CRM for generic business processes. This connection was added directly to SEED, which runs a syncing operation on a timer to ensure business systems are in sync. The connection is designed to be used as is but also provides an example to third parties on how SEED can connect to other CRMs or business systems.

Lastly, *security* is an important topic considering the amount and type of potentially stored data and the need to have easily accessible (and secure) authorization protocols. SEED supports basic authentication and OAuth. SEED is designed to be secure and is regularly updated with security patches. The platform is designed to be secure at the most fundamental layer, ensuring that all downstream platform users are secure.

7. Future directions and conclusion

Benchmarking and BPS are increasingly important policies for jurisdictions to manage energy consumption and reduce carbon emissions in buildings. As more jurisdictions enact BPS regulations, the demand for accurate and consistent data management becomes paramount. Data integrity, accuracy, consistency, shareability, extensibility, and transparency are core principles that underpin effective data management in this context. Ensuring the reliability and confidence of data is essential for auditability and provenance tracking. The cost to a jurisdiction of running a BPS program can be significant, and providing open-source and open-platform solutions (that can be extended at a reduced cost) can allow the jurisdiction to reallocate the cost savings to staff to help run a successful benchmarking or BPS program.

SEED's capabilities extend beyond basic data management tasks. It enables the import, cleansing, deduplication, and visualization of multiple years of building and tax lot data, as demonstrated in use cases within the District. These use cases highlight SEED's effectiveness in handling large volumes of data while providing visualization and tracking tools to monitor program performance and compliance over time. Furthermore, SEED serves as a platform for extending support to additional programs and integrating with other tools, such as BETTER, to provide a comprehensive and consistent preprocessing of data that even includes meter readings.

Looking towards the future, there is a clear direction for SEED's continued development and enhancement. This includes further system integration, enhanced tracking of building upgrades, and the evaluation

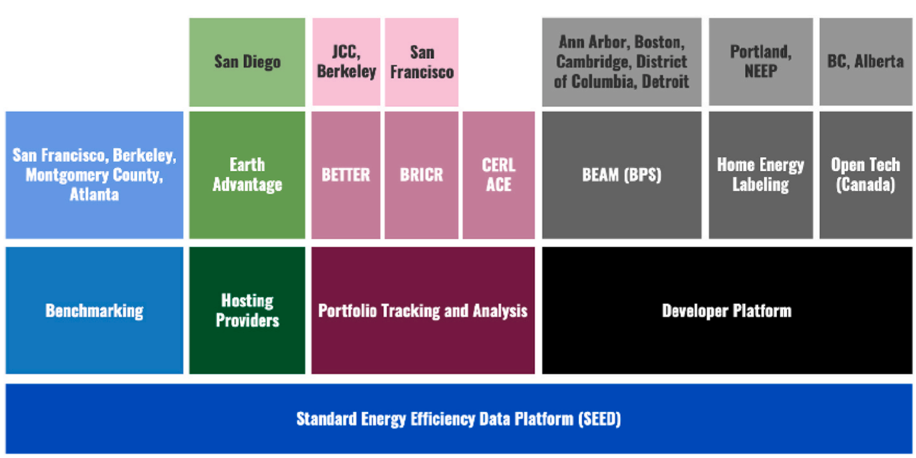

Fig. 10. SEED has been leveraged by various additional use cases to further standardize building data management for benchmarking and BPS.

of their effectiveness. Additionally, there is a growing demand for more advanced data quality checks (e.g., leveraging artificial intelligence to flag suspect data), increased visualization capabilities (e.g., exposing an API layer specific to third-party data analysis engines), and the development of roll-up features for organizations to gain deeper insights and actionable results from their data.

An essential aspect of SEED's future direction is its commitment to open-source principles and the establishment of standardized platforms for tracking building characteristics and meter data. By promoting open standards and interoperability, SEED facilitates data exchange with third-party tools through various formats such as GeoJSON, spreadsheets, and APIs through JSON. A user is able to select which buildings and which fields to export to compatible file formats, allowing for connections to third-party tools. This includes interoperability with building energy modeling tools and programming languages such as URBANopt, OpenStudio/EnergyPlus, and Modelica. This further enhances the overall efficiency and effectiveness of energy management initiatives by enabling scenario-based decisions to be evaluated. Having records of individual buildings with accurate and consistent data stored alongside multiple realizations of building energy modeling scenario results can help building owners and city planners make better decisions.

CRediT authorship contribution statement

Nicholas Long: Writing – original draft, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Katherine Fleming:** Software, Methodology. **Alex Swindler:** Visualization, Software, Conceptualization. **Andrew Held:** Writing – review & editing, Visualization, Data curation. **Robin Mitchell:** Methodology, Funding acquisition, Conceptualization. **Gregor P. Henze:** Writing – review & editing, Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

A link to the data and source code are provided in the manuscript.

Acknowledgements

The authors would like to thank the extended project team that helped develop the various mentioned software tools, including Isabel Langlois-Romero, Ross Perry, Hannah Eslinger, Alex Chapin, Devin Brady, Carolyn Szum, and Han Li. Also, the authors would like to thank Harry Bergmann, Sydney Applegate, and Billierae Engelman for their continued guidance and support for the project.

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U. S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided in part by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Building Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.

Appendix A. . Nomenclature

AHJ Authority Having Jurisdiction
API Application Programming Interface
ASHRAE American Society of Heating, Refrigerating, and Air-conditioning Engineers
BEAM Building Energy Analysis Manager
BEDES Building Energy Data Exchange Specification
BEPS Building Energy Performance Standard
BETTER Building Efficiency Targeting Tool for Energy Retrofits
BPS Building Performance Standard
CAEA Clean and Affordable Energy Act
CBL Covered Buildings List
CRM Customer Relationship Management
CSV Comma-Separated Values
District Washington, D.C.
DOE United States Department of Energy
DOEE Department of Energy and Environment
ECM Energy Conservation Measure
EEEJ Energy Equity and Environmental Justice
EISA2007 Energy Independence and Security Act of 2007
EO-14057 Executive Order 14057
EPA United States Environmental Protection Agency
ESCO Energy Service Company
ESPM ENERGY STAR® Portfolio Manager®
EUI Energy Use Intensity
gbXML Green Building XML
GHG Greenhouse Gas
GHGi GHG Intensity
GIS Geographical Information System
HPXML High Performance Building XML
IMT Institute for Market Transformation
LBNL Lawrence Berkeley National Laboratory
NMEC Normalized Metered Energy Consumption
NMECR Normalized Metered Energy Consumption in R
RBAC Rule-Based Access Control
RCx Retrocommissioning
SEED Standard Energy Efficiency Data Platform™
UBEM Urban Building Energy Modeling UBID Unique Building Identification
UI User Interface
US United States
WUI Water Use Intensity
XML eXtensible Markup Language

References

U. Ali, M. H. Shamsi, C. Hoare, E. Mangina, J. O'donnell, Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis, Energy Build.doi: 10.1016/j.enbuild.2021.111073. URL https://doi.org/10.1016/j.enbuild.2021.111073.

Alschuler, E., Antonoff, J., Brown, R., Cheifetz, M., 2014. Planting SEEDs: implementation of a common platform for building performance disclosure program data management. In: 2014 ACEEE Summer Study on Energy Efficiency in Buildings, 4–25 – 4–35. URL. www.buildings.energy.gov/BEDES.

ASHRAE, 2018. ANSI/ASHRAE/ACCA Standard 211-2018: Standard for Commercial Building Energy Audits.

ASHRAE, 2023. Building performance standards: a technical resource guide. Atlanta, GA: American Society of Heating, Refrigerating and Air Conditioning Engineers, 106pp, URL. www.ashrae.org/decarb.

Broman, K.W., Woo, K.H., 2018. Data organization in spreadsheets. Am. Statistician 72 (1), 2–10. https://doi.org/10.1080/00031305.2017.1375989.

Brown, P.S., Gould, J.D., 1987. An experimental study of people creating spreadsheets. ACM Trans. Inf. Syst. 5 (3), 258–272. https://doi.org/10.1145/27641.28058.

Bugnion, V., Long, N., Mitchell, R., Bergmann, H., Swindler, A., Beers, E., 2022. Building performance standards to drive market transformation. In: 2022 ACEEE Summer Study on Energy Efficiency in Buildings, p. 12. Pacific Grove, CA.

Chen, Y., Hong, T., Luo, X., Hooper, B., 2019. Development of city buildings dataset for urban building energy modeling. Energy Build. 183, 252–265. https://doi.org/10.1016/j.enbuild.2018.11.008.

City and county of Denver, bill 21-1310: energize denver (2021). URL https://denver.legistar.com/LegislationDetail.aspx?ID=5196421&GUID=641EBED8-31C9-4CA0-A8F9-946569B7C293&Options=ID\%7CText\%7CAttachments\%7C&Search=21-1310.

City of aspen, building IQ. URL. https://aspen.gov/1245/Building-IQ.

City of Austin, 2008. Austin city code: energy conservation. URL. https://austinenergy.com/energy-efficiency/ecad-ordinance.

City of Boston, 2021. Building emissions reduction and disclosure ordinance. URL. https://www.boston.gov/sites/default/files/file/2021/12/FinalAmendedDocket0775BERDO2_0.pdfreduction-and-disclosure.
City of Boulder, 2024. Building performance ordinance. URL. https://bouldercolorado.gov/services/building-performance-ordinance.
City of Cambridge, 2023. Building energy use disclosure ordinance. URL. https://cambridgema.iqm2.com/Citizens/Detail_LegiFile.aspx?Frame=&MeetingID=4267&MediaPosition=&ID=19814&CssClass=.
City of Chula Vista, 2021. Ordinance 3498: building energy savings. URL. https://buildingenergyscore.energy.gov/.
City of New York, 2019. Sustainable Buildings NYC (LL97).
City of Seattle, 2024. Building emissions performance standard. URL. https://www.seattle.gov/environment/climate-change/buildings-and-energy/building-emissions-performance-standard.
City of St. Louis, Ordinance 71132: Building Energy Performance Standards, 2020.
Council on Environmental Quality (CEQ), 2022. The federal building performance standard. URL. https://www.sustainability.gov/pdfs/federal-building-performance-standard.pdf.
Department of Energy, 2013. HPXML: A Standardized Home Performance Data Sharing System.
Ding, Y., Liu, X., 2020. A comparative analysis of data-driven methods in building energy benchmarking. Energy Build. 209, 109711 https://doi.org/10.1016/J.ENBUILD.2019.109711.
District of Columbia, 2008. Clean and affordable energy Act of 2008. URL. https://doee.dc.gov/sites/default/files/dc/sites/ddoe/publication/attachments/CAEA_of_2008_B17-0492.pdf.
District of Columbia, 2018. CleanEnergy DC Omnibus amendment Act of 2018. URL. https://code.dccouncil.us/dc/council/laws/22-257.html.
DOE Building Energy Codes, 2023. State and local building performance standards. URL. https://public.tableau.com/app/profile/doebecp/viz/BuildingPerformanceStandards/BuildingPerformanceStandards.
Durkan, J.A., 2021. Office of the mayor city of Seattle. URL. https://durkan.seattle.gov/wp-content/uploads/sites/9/2021/11/2021-09-Climate-EO_2.pdf.
Engelman, B., Kace, J., Mitchell, R., Jain, Y., 2023. Implementation and Administration of Building Performance Standards. Tech. Rep. June. Department of Energy, Washington, DC. DOE/EE-2791.
Gagnon, P., Cowiestoll, B., Schwarz, M., 2022. Cambium 2022 scenario descriptions and documentation. Tech. rep., National Renewable Energy Laboratory, 85pp, URL. www.nrel.gov/publications.
GbXML, 2024. Green building XML schema (gbXML). URL. https://www.gbxml.org/About_GreenBuildingXML_gbXML.
Green Button. URL. https://green-button.github.io/developers/.
Hong, T., Chen, Y., Lee, S.H., Piette, M.A., 2016. CityBES: a web-based platform to support city-scale building energy efficiency. In: 5th International Urban Computing Workshop, p. 9. San Francisco,CA.
Hooper, B., Hong, T., Macumber, D., Lee, S.H., Chen, Y., Long, N., Lee, E., Regina, I., Cruz, D., Piette, M.A., Berg, J., 2018. The BayREN integrated commercial retrofits (BRICR) project: an introduction and preliminary results overview of BRICR software tools and workflow. In: 2018 ACEEE Summer Study on Energy Efficiency in Buildings, pp. 1–12. Pacific Grove, CA. https://buildingsync.net/documents/Hooper-ACEEE-BRICR.pdf.
Institute for Market Transformation, 2021. Comparison of U.S. Commercial Building Energy Benchmarking and Transparency Policies, Tech. Rep. Institute of Market Transformation, Washington, DC.
Institute for Market Transformation, 2022. National BPS coalition. URL. https://nationalbpscoalition.org/#cities.
Institute for Market Transformation, 2023a. Comparison of U.S. Building audit, retuning, and retrocommissioning policies. Tech. Rep. September, Institute for Market Transformation, 5pp.
Institute for Market Transformation, 2023b. Comparison of U.S. Commercial Building Energy Benchmarking and Transparency Policies. Tech. Rep. December, Institute for Market Transformation.
Institute for Market Transformation, 2023c. Comparison of U.S. Building Performance Standards. Tech. Rep. January. Institute for Market Transformation.
Jaccard, P., 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull. Soc. Vaudoise Sci. Nat. 37, 547–579.
Kontar, R.E., Polly, B., Charan, T., Fleming, Katherine, Moore, N., Long, N., Goldwasser, D., 2020. URBANopt: an open-source software development kit for community and urban district energy modeling. In: 2020 Building Performance Analysis Conference and SimBuild, pp. 293–301. Chicago, IL. https://www.nrel.gov/docs/fy21osti/76781.pdf.
KW-Labs, nmecr. URL. https://kw-labs.github.io/nmecr/.
LBNL, 2018. BEDES online dictionary. URL. https://bedes.lbl.gov/bedes-online.
Long, N., Swindler, A., Bergmann, H., Reagan, A., Held, A., Longley, J., 2020. Standardizing city benchmarking and reporting: use cases in consolidating building data with SEED. In: 2020 ACEEE Summer Study on Energy Efficiency in Buildings, p. 12. Pacific Grove, CA. https://seed-platform.org/resources/Long-2020-Standardizing_City_Benchmarking.pdf.
Li, Han, Carolyn Szum, Sara Lisauskas, Ahmed Bekhit, Clay Nesler, Steven, C Snyder, 2019. Targeting Building Energy Efficiency Opportunities: An Open-Source Analytical and Benchmarking Tool." In *ASHRAE*. Transactions 125, 470–478. Atlanta, GA.
Long, N., Fleming, K., CaraDonna, C., Mosiman, C., 2021a. BuildingSync: a schema for commercial building energy audit data exchange. Developments in the Built Environment 7 (February), 100054. https://doi.org/10.1016/j.dibe.2021.100054.
Long, N., Gautier, A., Elarga, H., Allen, A., Summer, T., Klun, L., Moore, N., Wetter, M., 2021b. Modeling district heating and cooling systems with URBANopt, GeoJSON to modelica translator, and the modelica buildings library. In: Building Simulation 2021, p. 8. Bruges, Brussels. https://docs.urbanopt.net/doc_files/modeling_des_paper.pdf.
Long, N., Fleming, K., Swindler, A., Turas, F., Munday, P., 2023. pySEED: SEED API Client. URL. https://github.com/seed-platform/py-seed.
McCoy, C., 2019. U.S. City climate commitments: obstacles and opportunities in the building sector post-paris agreement. Md. J. Int. Law 34 (8), 249–300.
Mims, N., Schiller, S., Stuart, E., Schwartz, L., Kramer, C., Faesy, R., 2017. Evaluation of U.S. Building energy benchmarking and transparency programs: attributes, impacts, and best practices. Tech. rep., Lawrence Berkeley National Laboratory. Berkeley, CA, 131pp. https://escholarship.org/uc/item/3170j728.
Mitchell, R., Mathew, P., 2022. Considerations for Creating a Covered Buildings List for a Building Performance Standard or Benchmarking Program. Tech. rep., Lawrence Berkeley National Laboratory, Berkeley,CA.
Montgomery County, 2022. Building Energy Use Benchmarking and Performance Standards.
Montgomery County, 2024. Building energy performance standards. URL. https://www.montgomerycountymd.gov/green/energy/beps.html.
Nadel, S., Hinge, A., 2020. Mandatory building performance standards: a key policy for achieving climate goals. ACEEE White Paper. Washington, DC, 54pp. https://www.aceee.org/sites/default/files/pdfs/buildings_standards_6.22.2020_0.pdf.
Oregon Legislative Assembly, HB 3409, 2023.
Pacific Northwest National Laboratory, 2020. Audit template tool. URL. https://buildingenergyscore.energy.gov/.
Palmer, K., Walls, M., 2017. Using information to close the energy efficiency gap: a review of benchmarking and disclosure ordinances. Energy Efficiency 10 (3), 673–691. https://doi.org/10.1007/s12053-016-9480-5.
Papadopoulos, S., Kontokosta, C.E., 2019. Grading buildings on energy performance using city benchmarking data. Appl. Energy 233–234, 244–253. https://doi.org/10.1016/J.APENERGY.2018.10.053.
Pritoni, M., Paine, D., Fierro, G., Mosiman, C., Poplawski, M., Saha, A., Bender, J., Granderson, J., 2021. Metadata schemas and ontologies for building energy applications: a critical review and use case analysis. Energies 14 (7), 2024. https://doi.org/10.3390/en14072024.
Rebstock, M., Fengel, J., Paulheim, H., 2008. Ontologies-Based Business Integration. Springer.
Reno Nevada Administrative Code, 2023. Energy and water efficiency program. URL. https://library.municode.com/nv/reno/codes/administrative_code?nodeId=PT2READCO_TIT14BUCO_CH14.30ENWAEFPR.
Roth, J., Lim, B., Jain, R.K., Grueneich, D., 2020. Examining the feasibility of using open data to benchmark building energy usage in cities: a data science and policy perspective. Energy Pol. 139, 111327 https://doi.org/10.1016/J.ENPOL.2020.111327.
San Francisco Environment, 2020. Existing buildings ordinance performance report. URL. https://sfenvironment.org/energy/san-francisco-existing-buildings-performance-report. https://www.sfenvironment.org/existing-buildings-ordinance-performance-report.
State and Local Energy Efficiency Action Network, 2012. Benchmarking and disclosure: state and local policy design guide and sample policy language. Tech. Rep. May, Institute for Market Transformation, 20pp.
State of Colorado, 2021. HB 21-1286: Energy Efficiency in Buildings.
State of Colorado, 2023. Building benchmarking and performance standards. URL. https://www.sos.state.co.us/CCR/GenerateRulePdf.do?ruleVersionId=11123&fileName=5CCR1001-32.
State of Maryland, 2022. Climate solutions now Act (SB 528). URL. https://legiscan.com/HI/text/SB528/id/1472659.
State of Washington, 2019. HB 1257: energy efficiency. URL. https://www.commerce.wa.gov/wp-content/uploads/2019/06/HB1257.pdf.
State of Washington, 2022. Substitute senate bill 5722: clean buildings expansion. URL. https://lawfilesext.leg.wa.gov/biennium/2021-22/Pdf/Bills/SenatePassedLegislature/5722-S.PL.pdf?q=20220329121607.
Taylor, C., Mathew, P., Hernandez, G., Mercado, A., 2012. Standard energy efficiency data (SEED) platform: a tool to track and transact energy data. In: 2012 ACEEE Summer Study on Energy Efficiency in Buildings, pp. 289–297. Pacific Grove, CA.
The white House, executive order on catalyzing clean energy industries and jobs through federal sustainability. URL. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/12/08/executive-order-on-catalyzing-clean-energy-industries-and-jobs-through-federal-sustainability/.
United States Congress, 2007. Energy Independence and Security Act of 2007.
United States Congress, 2015. Energy Efficiency Improvement Act of 2015.
US Environmental Protection Agency (EPA), 2012. Benchmarking and Energy Savings, Tech. Rep. United States Environmental Protection Agency, Washington, DC. URL. www.energystar.gov/.
US Environmental Protection Agency (EPA), 2021. Technical reference: ENERGY STAR Score, tech. Rep. April, United States environmental protection agency. URL. energystar.gov/buildings/tools-and-resources/portfolio-manager-technical-reference-energy-star-score. www.energystar.gov/ScoreDetails.
US Environmental Protection Agency (EPA), 2022. Portfolio manager. URL. https://portfoliomanager.energystar.gov.
US Environmental Protection Agency (EPA), 2023. How does Portfolio Manager calendarize bills? URL. https://energystar.my.site.com/PortfolioManager/s/article/How-does-Portfolio-Manager-calendarize-bills.
Wang, N., Vlachokostas, A., Borkum, M., Bergmann, H., Zaleski, S., 2019. Unique Building Identifier: a natural key for building data matching and its energy

applications. Energy Build. 184, 230–241. https://doi.org/10.1016/j.enbuild.2018.11.052.
Washington State Department of Commerce, 2024. Clean buildings performance standards. URL. https://www.commerce.wa.gov/growing-the-economy/energy/buildings/clean-buildings-standards/.
Webb, A.L., McConnell, C., 2023. Evaluating the feasibility of achieving building performance standards targets. Energy Build. 288, 112989 https://doi.org/10.1016/J.ENBUILD.2023.112989.
Yang, Z., Roth, J., Jain, R.K., 2018. DUE-B: data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis. Energy Build. 163, 58–69. https://doi.org/10.1016/J.ENBUILD.2017.12.040.