Building Technologies & Urban Systems Division
Energy Technologies Area
Lawrence Berkeley National Laboratory

An Overview of Data Tools for Representing and Managing Building Information and Performance Data

Na Luo, Marco Pritoni, Tianzhen Hong

Building Technology and Urban Systems Division
Lawrence Berkeley National Laboratory

Energy Technologies Area
September 2021

https://doi.org/10.1016/j.rser.2021.111224

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the United States Department of Energy under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231.

An Overview of Data Tools for Representing and Managing Building Information and Performance Data

Na Luo, Marco Pritoni, Tianzhen Hong*

Building Technology and Urban Systems Division

Lawrence Berkeley National Laboratory

*Corresponding author: T. Hong, thong@lbl.gov

Abstract

Building information modeling (BIM) has been widely adopted for representing and exchanging building data across disciplines during building design and construction. However, BIM's use in the building operation phase is limited. With the increasing deployment of low-cost sensors and meters, as well as affordable digital storage and computing technologies, growing volumes of data have been collected from buildings, their energy services systems, and occupants. Such data are crucial to help decision makers understand what, how, and when energy is consumed in buildings—a critical step to improving building performance for energy efficiency, demand flexibility, and resilience. However, practical analyses and use of the collected data are very limited due to various reasons, including poor data quality, ad-hoc representation of data, and lack of data science skills. To unlock value from building data, there is a strong need for a toolchain to curate and represent building information and performance data in common standardized terminologies and schemas, to enable interoperability between tools and applications. This study selected and reviewed 24 data tools based on common use cases of data across the building life cycle, from design to construction, commissioning, operation, and retrofits. The selected data tools are grouped into three categories: (1) data dictionary or terminology, (2) data ontology and schemas, and (3) data platforms. The data are grouped into ten typologies covering most types of data collected in buildings. This study resulted in five main findings: (1) most data representation tools can represent their intended data typologies well, such as Green Button for smart meter data and Brick schema for metadata of sensors in buildings and HVAC systems, but none of the tools cover all ten types of data; (2) there is a need for data schemas to represent the basis of design data and metadata of occupant data; (3) standard terminologies such as those defined in BEDES are only adopted in a few data tools; (4) integrating data across various stages in the building life cycle remains a challenge; and (5) most data tools were developed and maintained by different parties for different purposes, their flexibility and interoperability can be improved to support broader use cases. Finally, recommendations for future research on building data tools are provided for the data and buildings community based on the FAIR principles to make data Findable, Accessible, Interoperable, and Reusable.

Highlights:

- Building information and performance data are grouped into 10 typologies
- 24 data tools are selected, categorized into 3 groups (terminology, ontology/schema, platform), and reviewed
- Most of these tools can be further enhanced to improve flexibility, standardization, and interoperability

- Ontologies or schema to represent metadata of occupants and basis of building design need to be developed
- Integrating building data across the building life cycle to support various use cases remains a challenge

Keywords: building information modeling, ontology, data schema, metadata, building performance data

Word Count: 11,188 (excluding references)

Nomenclature

ADI	ARM data Integrator
AHU	Air handling unit
API	Application Programming Interface
BAS	Building Automation System
BEDES	Building Energy Data Exchange Specification
BIM	Building information modeling
BPD	Building Performance Database
CA SDD	California Standards Data Dictionary
EBC	Energy in Buildings and Communities
EnergyADE	Energy Application Domain Extension
FAIR	Findable, Accessible, Interoperable, and Reusable
FDD	Fault Detection and Diagnostics
GEB	Grid-Interactive Efficient Buildings
HVAC	Heating, Ventilation and Air Conditioning
IEA	International Energy Agency
IFC	Industry Foundation Classes
IoT	Internet-of-Things
LBNL	Lawrence Berkeley National Laboratory
NIST	National Institute of Standards and Technology
NREL	National Renewable Energy Laboratory
ORNL	Oak Ridge National Laboratory
PNNL	Pacific Northwest National Laboratory
PV	Photovoltaics
SAREF	Smart Appliances REFerence
SEED	Standard Energy Efficiency Data Platform
VAV	Variable air Volume

VFD	Variable frequency drive
XML	eXtensible Markup Language

1. Introduction

Data are crucial to understanding and quantifying building performance, informing energy efficient design, optimizing operation and controls, and benchmarking and rating building energy efficiency [1]. With the increased adoption of building sensing and control technologies, a massive amount of data are being generated and employed for various applications [2]. Further, the digitalization of the design process through the use of building information models (BIM) has made available additional machine-readable data (e.g., geometry, materials, schematics of the building systems) that previously was stored in human-generated documents (e.g., drawings, manual annotations, spreadsheets, and text documents) [3]. As building data has increased in volume, lack of consistency in the representation, format, and meaning of these data has become evident [4,5]. These inconsistencies exist between phases of the building's life cycle; for instance, data about a building system generated in the design phase is not aligned with the building operations data [6]. Perhaps more surprisingly, they also exist within a single phase of the life cycle. For example, two brands of building automation systems (BAS) or even two installations of the same product in two buildings may use different naming conventions for their data points, often requiring manual mapping of the data by an expert, when they need to be used for another purpose [7,8]. More than 15 years ago, the National Institute of Standards and Technology (NIST) estimated that the United States building industry had lost $15.8 billion annually because of the lack of interoperability standards while storing the building data [9], and things have not significantly improved recently [4]. These data interoperability issues are not only caused by legacy systems; they also can be found in more modern Internet-of-Things (IoT) technologies. For instance, Smart Home hardware and software have recently gained popularity in residential buildings, but the adoption of several competing platforms and a lack of standardization among them has caused significant interoperability challenges [5].

While on the one hand, the building industry has been slow to come together to address these challenges, on the other hand, the research community has made significant progress in advancing building science and technology. For instance, many algorithms for analytics and controls have been proposed and demonstrated. Jia et al. reviewed the enabling technologies and applications for adopting IoT for the development of smart buildings [2], Bhattacharya et al. reviewed the metadata schemas for building datasets [10], and Kheiri et al. reviewed the optimization approaches employed in energy-efficient building design [11]. Despite these advances, lack of interoperability between data sources is still considered one of the main barriers to deploying these applications at scale [4,12]. To prototype, test, validate, and compare applications, the research community would benefit from having anonymized and publicly available building datasets. Examples of such datasets exist in other research areas, such as the MNIST [13] for digit recognition, the FERET for face recognition [14], as well as the Google ImageNet for storing and sharing images [15]. In the building domain, a few efforts have tried to create open datasets for specific applications. For instance, Fierro et al. recently developed an open testbed for portable building analytics called Mortar [16], Granderson et al. developed a dataset for benchmarking algorithms of fault detection and diagnostics (FDD) [17]. However, these datasets have limited coverage (i.e., they contain only limited types of data) and/or low data quality (e.g., missing and incorrect data, coarse sampling, and lack of descriptors about the data), limiting the ability to test applications developed by researchers.

To fill some critical gaps in existing buildings datasets, the Benchmark Dataset project aims to characterize potential use cases for buildings datasets, define an appropriate data infrastructure, inventory existing buildings datasets, and develop a new set of fine-grained and well-curated data from multiple buildings that can be used by researchers for various purposes. The project is funded by the United States Department of Energy (U.S. DOE) and performed by four U.S. DOE national laboratories: Lawrence Berkeley National Laboratory (LBNL), National Renewable Energy Laboratory (NREL), Oak Ridge National Laboratory (ORNL), and Pacific Northwest National Laboratory (PNNL).

Targeted use cases include load forecasting and baselining, virtual sensing, building energy modeling, building performance benchmarking, nonintrusive load monitoring, model predictive controls, and grid-interactive efficient buildings (GEB). The curation process for these datasets is to develop adequate metadata that describe the building and system characteristics (e.g., a standardized terminology and a model defining the relationship between objects) to clean errors and missing data, and to host the datasets on a data portal for public access. The first activity of this project consisted of a review of the open data tools used by researchers and practitioners to store and analyze building data. In this context, data tools include data dictionary, terminology, ontology, schema, and database management platforms.

Previous review papers have focused on three key areas: architectures of open data tools, metadata schemas, and use of building data for specific applications. Several articles proposed software architectures for centralized systems that host building automation systems and Internet of Things (IoT) data [2,18,19]. The features of interest in such platforms include data storage for time-series data and contextual metadata, a mechanism for data retrieval, and privacy and security features. Most of these papers present cutting-edge solutions that are not currently supported or have not been adopted by a broad user base. These papers fail to review more common open data tools that may have less innovative features but are supported by government and industry (e.g., Green Button [20] used for smart meter data). Review papers on metadata schemas [12,21,22] show the variety of schemas used for different applications (e.g., Industry Foundation Classes [23] or Project Haystack [24]), but they do not provide detailed information on the actual implementation of these schemas into tools. A group of review articles summarized how the building data are supporting different analysis and applications: Molina-Solana et al. reviewed how data science (algorithms and tools) has been applied to address the most difficult problems faced by practitioners in the field of energy management, especially in the building sector [25]; Volk et al. focused on the development and evolution of the BIM application for existing buildings, and identified several BIM tools to employ [26]; Coakley et al. conducted a review of using the measured data to support the development and calibration of the building energy model [27]; and Zhao et al. presented different methods to predict the building energy consumption using measured and synthetic datasets [28]. Other reviews categorize different methods and algorithms for solving particular research questions, and introduce the data tools needed afterward [29–31]. None of these papers comprehensively describes different tools and schemas necessary for curating and managing datasets that cover multiple use cases across various phases of the building life cycle.

Given these literature gaps, the present study sought to survey the capability and analyze limitations of a broad set of representative data tools, and to answer the following research questions: (1) What are the most popular data tools used to describe, store and exchange data for different phases of the building life cycle? (2) What type of information do they cover? (3) How can they be categorized? and (4) What are the limitations of these tools and opportunities to

improve them to enable big data analytics for buildings? This paper also discusses how the existing data tools support the FAIR principles to make building dataset Findable, Accessible, Interoperable, and Reusable [32].

2. Methodology

To answer these research questions, we surveyed 32 researchers at four U.S. DOE national laboratories to identify which of the data tools they have used are the most popular. Survey results were combined with information from the literature and input from a dozen members of a technical advisory group representing universities (U.S. and international), industry, and research institutes. Figure 1 illustrates the methodology we adopted to conduct the review and synthesize findings.

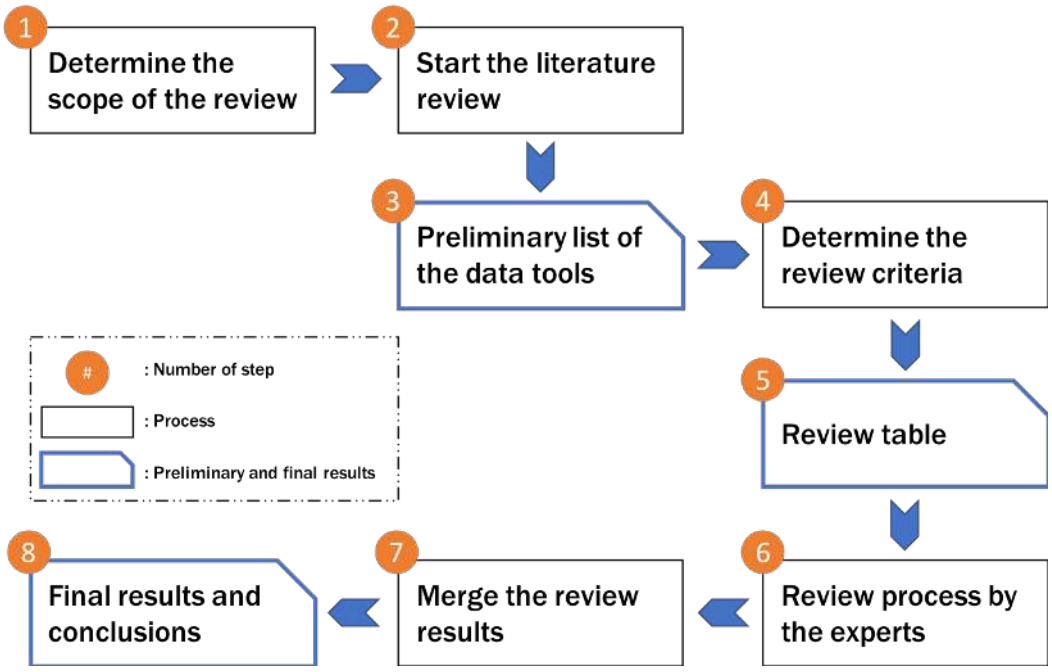

Figure 1. Overview of the review methodology, highlighting eight steps

A preliminary list of 24 tools was identified (Step 3 in Figure 1). These covered different phases of the building life cycle, from building design (e.g., BIM [26]) to building operation (e.g., Project Haystack, ISO standard 12655 [33,34]), and audit (e.g., BuildingSync, Asset Score [35,36]). The authors performed a preliminary analysis of these tools and categorized them into three major groups based on the hierarchical representation of building data, from the bottom up: (1) data terminology, (2) data ontologies and schemas, and (3) data platforms (databases and management tools), as shown in Figure 2. Data terminology tools (e.g., dictionaries) are the fundamental building block providing standardized terms for defining data ontologies and data schemas. Data ontologies and schemas utilize these forms of dictionaries to define ways of organizing related data items and denoting their relationships which allows linking and composing concepts together. They are used to represent metadata and data of various types. Some types of schemas include relational database schemas and ontologies. An ontology is a formal model that allows knowledge to be represented for a specific domain; while a schema describes the types of things that exist, the relationships between them, and the logical ways those things and relationships can be used together [21]. Data platforms are used for storing and managing large amounts of data organized in a uniform manner using a specific data

schema or ontology. Users can access the data by querying these platforms via an application programming interface (API) or by using a user interface. These platforms store both time-series data and their metadata, which are described by an explicit or implicit schema. Overall, these data platforms, schemas and dictionaries are used to facilitate the data curation process from raw data to research-ready data. The list of the tools is shown in Figure 3 in the Results section.

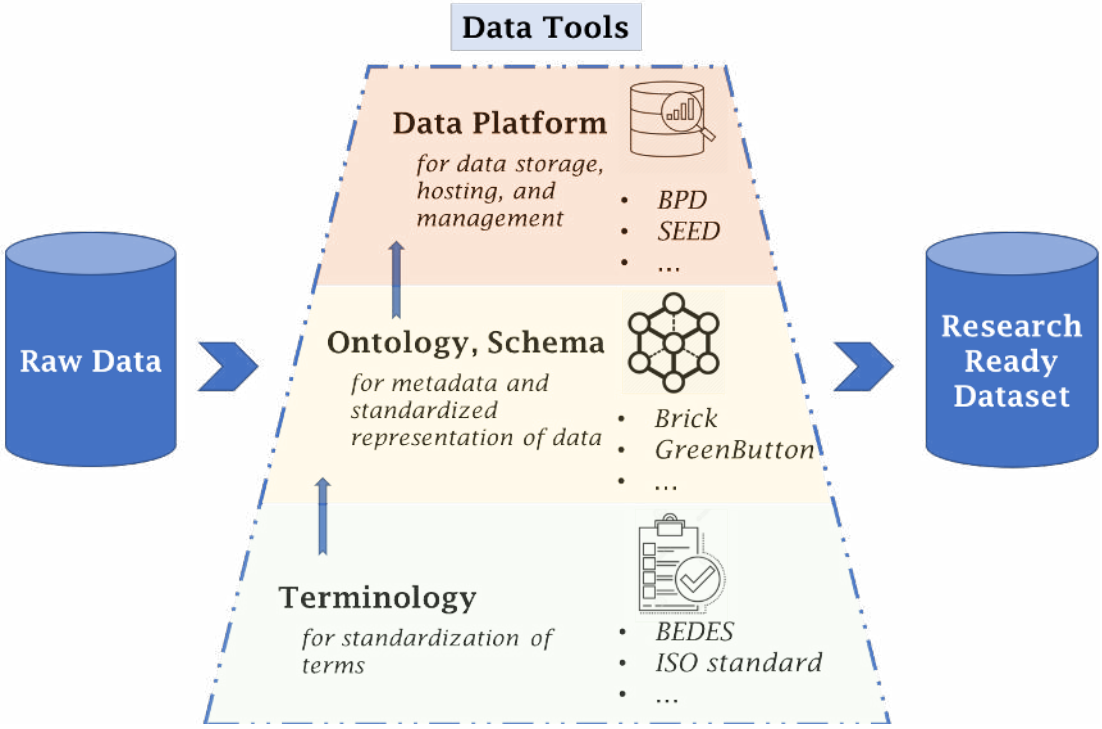

Figure 2. Data representation and management tools

In Step 4, we proposed seven feature aspects to organize the review of the selected data tools, as shown in Table 1.

Table 1. Seven feature aspects used to review the data tools

1. Tool Categories	2. Data Coverage	3. Data Relations	4. Flexibility and Extensibility	5. Stage of Application	6. Adoption Status	7. Pros and Cons
Terminology	Energy Use Data	Hierarchy		Design		
Ontology	Onsite Power Generation Data	Typology		Commissioning		
Schema	Indoor Environmental Data	Association		Operation		
Data Platform	Outdoor Environmental Data	Others		Rating		
	Equipment Operational Data			Audit		
	System Control Setting/Logic					
	Occupant Data					
	Design Basis data					
	Building and System Asset Data					
	Utility Rates and Grid Signal Data					

The 24 selected data tools were grouped into four categories, based on the use cases in which the data are involved, as well as their functionality during the data curation process:

1) Terminology: A collection of standardized definitions and terms.
2) Ontology: An ontology is a formal model that describes the types of things that exist, the relationships between them, and the logical ways those things and relationships can be used together [21].
3) Schema: A data schema is the skeleton structure that represents the metadata and hierarchy view of the entire dataset. It describes how the data points are organized and how the relationships among them are associated in the model [6].
4) Data platform: A data platform is a web-based interface to collect, store, host and manage datasets with specific uniform organizations, where users can upload, query and download the datasets, or create the models (e.g., web-based model builder).

Data coverage is one of the most important feature aspects. We categorized ten types of building data based on the categorization approaches in the previous studies [37] and typical use cases of building data. The ten categories cover most data types which are generated and collected during the building life cycle:

1) Energy use data: Data on energy use (electricity and fuel) in buildings. Resolution can be defined in various dimensions: (a) in spatial terms (e.g., micro-zones, rooms, floors, and whole buildings); (b) in systems or end uses (e.g., lighting, cooling, heating, and plug-in equipment); and (c) in temporal terms (e.g., sub-hourly, hourly, daily, monthly, and annual).
2) Onsite power generation data: Data on electricity generated from the onsite power equipment such as solar photovoltaics (PV), wind turbines, or combined heat and power systems.
3) Indoor environmental data: Information on indoor environmental quality, including thermal comfort related data (indoor air temperature, humidity, and air velocity), visual comfort related data (illuminance level and glare), indoor air quality related data (indoor air carbon dioxide (CO_2), fine particulate matter ($PM_{2.5}$), and volatile organic compounds (VOC) and acoustic data (indoor noise level).
4) Outdoor environmental data: Information regarding the relevant microclimatic/external conditions (e.g., outdoor air temperature, humidity, solar radiation, precipitation, CO_2, $PM_{2.5}$, and sound level).
5) Equipment operational data: Operation status and performance of equipment in buildings, including heating, ventilation and air conditioning (HVAC) equipment, central plant equipment (chillers, boilers, cooling towers), lighting, plug-in office equipment (e.g., computers and associated peripherals), and household equipment (e.g., appliances like washers, refrigerators, ovens, and dryers).
6) System control setting/logic: Setpoint of controllable variables (e.g., thermostat, humidistat, static pressure, indoor CO_2, indoor illuminance, frequency of variable frequency drives (VFD), on/off/stage), and control logics (e.g., reset control).
7) Occupant data: Occupancy of space (occupied status, number of occupants), and occupant interaction with building and energy systems (e.g., open/close windows, pull up/down blinds, dim or turn on/off lights, turn on/off plug-in equipment).
8) Design basis data: Basis for design, e.g., indoor temperature setpoint, ventilation rate, occupant schedule, internal loads and schedules, and design day weather data.

9) Building and system asset data: Data representing the physical characteristics of buildings and their energy systems (e.g., HVAC systems and configurations, envelope, lighting system, and plug-in equipment).
10) Utility rates and grid signal data: Tariff data representing utility rates and demand response signals from the electric grid.

While representing the building operational and control data, the metadata are usually structured semantically under a specific data relationship. In this review, we identified four types of data relations—namely, hierarchy, typology, association, and others—to capture the structure of different entries. A hierarchical model represents the data in a tree structure that links a number of disparate sensor data to one owner or parent. It allows one-to-one or one-to-many relationships between two different types of entities (e.g., terms such as *hasLocation* and *isPointOf* in Brick schema to identify the relationship between entities [38]). A typology structure refers to the relative positions of spatial features, which also involves a cause-and-effect relationship between each of two elements: the source element and target element. It indicates that the target element can be triggered only by the source element, thus showing that the target element will be executed only after the source element executes (e.g., entities of the energy audit data in the web-based Audit Template [36]). An association relationship refers to the class concept, which categorizes certain elements and defines the relationships between different categories (e.g., a combination of *tags* to describe each building *entity* in Project Haystack [24]).

We also evaluated the flexibility and extensibility of each data tool during the review. One criterion was whether a data tool could represent data at various levels of detail. The other criterion was the ability to allow missing data, as well as the addition of new data. We also indicate for each tool whether a standard terminology is applied. Next, data tools could be applied at various stages of the building life cycle, based on their functionality and use purpose. The application stages covered those from design to construction, commissioning, operation, audit, and retrofits. We also looked at the adoption status of a data tool as another essential feature. Some data tools are already widely adopted, while others are still in the academic research stage or under a pilot. More specifically, we also reviewed whether the application of a data tool was limited to specific building types (e.g., residential or commercial).

After defining the above set of criteria, we created a spreadsheet (Step 5 in Figure 1) and shared it with 32 researchers with expertise on data tools and applications from four U.S. national laboratories. The researchers helped review several data tools based on their experience and familiarity of those data tools (Step 6 in Figure 1). The results were compiled and synthesized by the authors (Step 7 in Figure 1).

3. Results

We first present the overview of the 24 selected data tools, then conduct an in-depth review of one or more representative tools from each tool category.

3.1 Overview of the 24 selected data tools

A list of 24 tools was identified based on the survey results. Figure 3 lists the basic information about the selected tools, including the name of the tool, the developer or maintainer of the tool, a brief description, and a reference link.

ID	Tool Name	Full Name	Key Maintainer	Brief Description	Reference Link
1	ADI	ARM Data Integrator	Pacific Northwest National Laboratory	A suite of tools, C libraries, structures, and interfaces developed to simplify the development of algorithms to analyze time-series data and decrease the costs associated with such development.	https://www.arm.gov
2	Annex 66 Ontology	IEA EBC Annex 66 - Definition and Simulation of Occupant Behavior in Buildings	Annex 66 community	An ontology for representation and incorporation of multiple layers of monitored building data in pertinent computational applications.	https://annex66.org
3	ASHRAE 201	ASHRAE Standard 201 - Facility smart grid information model	ASHRAE	A common basis for electrical-energy consumers to describe, manage, and communicate about electrical-energy consumptions and forecasts.	http://spc201.ashraepcs.org
4	ASHRAE 205	ASHRAE Standard 205 - Standard Representation of Performance Simulation Data for HVAC&R and Other Facility Equipment	ASHRAE/Argonne National Laboratory	A standard schema to facilitate automated sharing of equipment performance characteristics by defining data models and data serialization formats.	http://data.ashrae.org/standard205/
5	Asset Score	Building Energy Asset Scoring Tool	Pacific Northwest National Laboratory	A web-based software application that is designed to allow building owners, managers, and operators to assess the energy performance of their buildings.	https://buildingenergyscore.energy.gov
6	Audit Template	Asset Score Audit Template	Pacific Northwest National Laboratory	As a feature of Asset Score, it may be used to create a standard building energy audit report and submit to selected jurisdictions to comply with local ordinances.	https://buildingenergyscore.energy.gov
7	BEDES	Building Energy Data Exchange Specification	Lawrence Berkeley National Laboratory	A dictionary of terms, definitions, and field formats which was created to help facilitate the exchange of information on building characteristics and energy use.	https://bedes.lbl.gov/
8	BPD	Building Performance Database	Lawrence Berkeley National Laboratory	An anonymized database that contains energy use intensity data for hundreds of thousands of buildings in the United States	https://bpd.lbl.gov/
9	Brick	Brick Schema	Gabe Fierro Jason Koh	A unified semantic representation to capture the contextual information for building subsystems and their data points, especially the important relationships between data points that are explicitly or implicitly mentioned in a building's BMS.	https://brickschema.org
10	BuildingSync	BuildingSync	National Renewable Energy Laboratory	A XML Schema designed to easily exchange building related data. BuildingSync is based on BEDES terminology.	https://buildingsync.net
11	CA SDD	California Standards Data Dictionary	California Building Energy Code Compliance	A Standard Data Dictionary and file format used by CBECC-COM, a code compliance tool for California non-residential buildings	http://bees.archenergy.com/index.html
12	Energy ADE	CityGML - Energy Application Domain Extension	Special Interest Group 3D	An Application Domain Extension to CityGML which is an international standard of 3D city models.	http://www.citygmlwiki.org/index.php/CityGML_Energy_ADE
13	ENERGY STAR Portfolio Manager	ENERGY STAR Portfolio Manager	United States Environmental Protection Agency	A series of web services (APIs) to enable building owners or managers to document their facilities and their energy and water usage as well as their Energy Star ratings	https://portfoliomanager.energystar.gov/webservices/home/api
14	gbXML	Green Building XML	Carmel Software	An industry supported schema for sharing building information and enabling interoperability between disparate building design software tools and engineering analysis software tools	https://www.gbxml.org
15	GreenButton	Green Button Initiatives	United States Department of Energy	An industry-led effort that responds to a White House call-to-action to provide utility customers with easy and secure access to their energy usage information in a consumer-friendly and computer-friendly format.	http://www.greenbuttondata.org
16	HPXML	Home Performance eXtensible Markup Language	National Renewable Energy Laboratory	An open data schema that makes it easier and less expensive to collect and transfer home energy data among information trading partners.	https://www.hpxmlonline.com
17	IFC	Industry Foundation Classes	ISO/TC 59/SC 13	An open international standard for Building Information Model (BIM) data that are exchanged and shared among software applications used by the various participants in the construction or facility management industry sector.	https://www.iso.org/standard/70303.html
18	ISO 12655-2013	ISO 12655-2013 Energy performance of buildings — Presentation of measured energy use of buildings	ISO/TC 163	An ISO standard on representation energy use in buildings	https://www.iso.org/standard/51634.html
19	ISO 15489-2016	ISO 15489-2016 Information and documentation — Records management	ISO/TC 46/SC 11	An ISO standard to define the concepts and principles from which approaches to the creation, capture and management of records are developed.	https://www.iso.org/standard/62542.html
20	ISO 52000-1:2017	ISO 52000-1:2017 Energy performance of buildings	ISO/TC 163	An ISO standard to establishe a systematic, comprehensive and modular structure for assessing the energy performance of new and existing buildings (EPB) in a holistic way.	https://www.iso.org/standard/65601.html
21	oneloTa	Open Connectivity Foundation oneloTa	Open Connectivity Foundation	An open online tool created by the Open Connectivity Foundation (OCF) to encourage the design of interoperable device data models for the Internet of Things.	https://openconnectivity.org/developer/oneiota-data-model-tool/
22	Project Haystack	Project Haystack	Project Haystack Corporation	An open source initiative to streamline working with data from the Internet of Things, which is generated by the smart devices that permeate our homes, buildings, factories, and cities.	https://project-haystack.org
23	SAREF	Smart Appliances REFerence	Smart Appliances Project	A shared model of consensus that facilitates the matching of existing assets in the smart appliances domain.	https://sites.google.com/site/smartappliancesproject/ontologies/reference-ontology
24	SEED	Standard Energy Efficiency Data Platform	National Renewable Energy Laboratory	A database importing building characteristic data across various data sources, and mapping data from data sources to user specified fields, merges similar data, and pairs data between properties and taxlots.	https://seed-platform.org/

Figure 3. Basic information about the 24 selected data tools

Figure 4 shows the selected 24 data tools by the three categories. Sixteen tools fall into the category of data ontology or schema. Nine tools belong to the terminology category. Seven tools are data platforms. The Venn diagram in Figure 4 shows the overlap of some tools belonging to two or even three categories. For example, BPD is a database with its own data model, and it uses the standard terminology defined in the Building Energy Data Exchange Specification (BEDES). For each tool category, some of the selected tools have been widely adopted in the United States. For example, the Brick schema [39] is increasingly becoming used to represent the metadata of sensor and meter time-series data, the ISO standard 12655-2013 [34] is used to represent the energy use data for the whole building and end uses (e.g., lighting, HVAC, plug-in equipment), BEDES [40] is a dictionary of standardized terms for facilitating the exchange of information on building characteristics and energy use, and BPD [41] is a building performance database to benchmark and identify energy efficiency improvements.

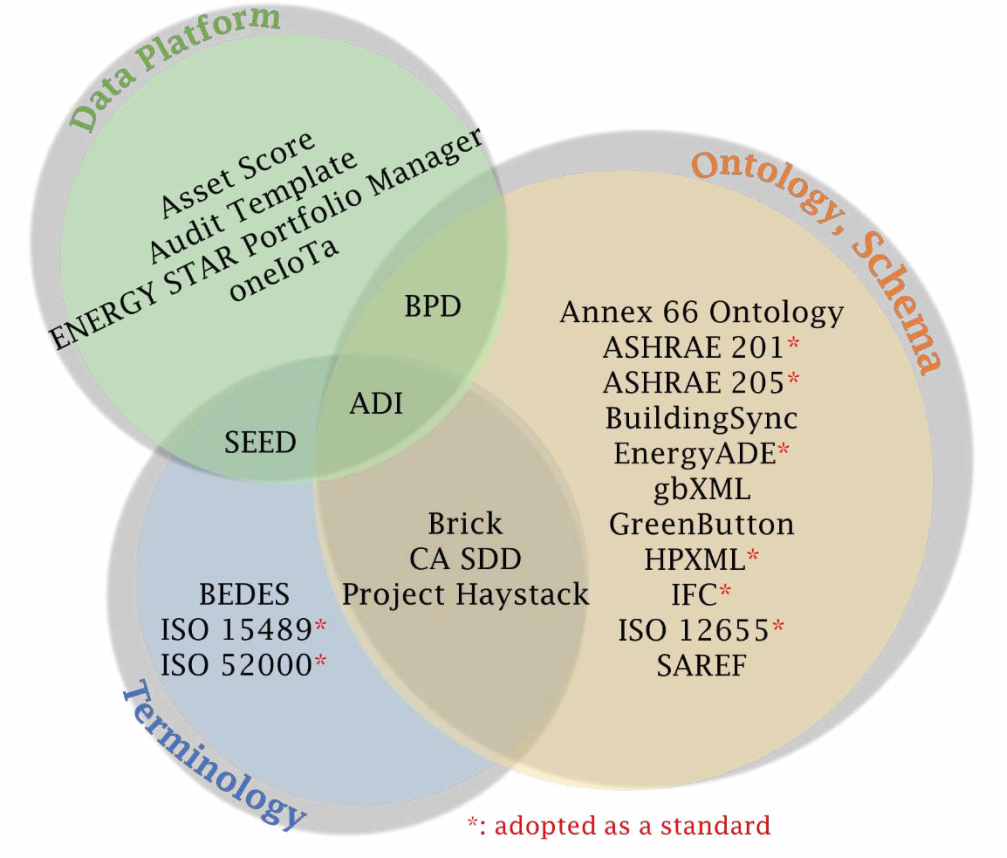

Figure 4. The selected 24 data tools by categories

The selected 24 data tools are summarized in Figure 5 using the proposed six feature aspects (except the pros and cons, which are described in the Discussion section). The left/top column shows the names of the 24 selected data tools in alphabetical order. The feature aspects listed are tool categories, data coverage, data relations, flexibility and extensibility, stage of application, and adoption status. Subcategories of each feature aspect are listed to guide the review of each data tool.

		Tool Categories				Data Coverage										Data Relations				Flexibility and Extensibility				Stage of Application					Adoption	
ID	Tool Name	Terminology	Ontology	Schema	Data Platform	Energy Use Data	Onsite Power Generation Data	Indoor Environmental Data	Outdoor Environmental Data	Equipment Operational Data	System Control Setting/Logic	Occupant Data	Design Basis Data	Building and System Asset Data	Utility Rates and Grid Signal	Hierarchy	Typology	Association	Others	Allow various levels of details	Allow missing data/ Capability of adding new data	Extensible	Use standard terminology	Building Design	Operation	Audit	Code Compliance	Rating	Commissioning	Development status (U - under development P - pilot application A - widely adopted)	Applicable building type (C - commercial R - residential)
1	ADI	x		x	x	x		x	x	x	x				x	x			x	Yes	Yes	Yes	Yes		x					A	C/R
2	Annex 66 Ontology		x			x		x	x	x	x	x				x				Yes	Yes	Yes	Yes	x	x	x	x			P	C/R
3	ASHRAE 201			x		x		x	x	x	x				x		x			Yes	Yes	Yes	Yes	x	x	x	x	x		A	C/R
4	ASHRAE 205		x	x						x	x						x			Yes	Yes	Yes	Yes		x	x	x	x		A	C/R
5	Asset Score				x								x	x		x	x	x		Yes	No	Yes	No	x		x	x	x		A	C/R
6	Audit Template				x	x						x	x	x	x	x	x	x		Yes	No	Yes	Yes	x		x	x	x		A	C/R
7	BEDES	x				x				x			x	x					x	Yes	Yes	Yes	Yes	x		x	x	x		A	C/R
8	BPD		x		x	x				x			x	x				x	x	Yes	Yes	Yes	Yes	x		x	x	x		A	C/R
9	Brick	x	x	x		x		x	x	x	x				x	x		x		Yes	Yes	Yes	Yes	x				x		P	C/R
10	BuildingSync			x		x			x	x		x	x	x	x	x				Yes	Yes	Yes	Yes	x	x	x		x		A	C
11	CA SDD	x	x	x						x	x		x	x		x		x		No	Yes	No	Yes	x			x			A	C
12	Energy ADE			x		x				x		x	x	x		x	x	x		No	Yes	Yes	No	x	x					A	C/R
13	ENERGY STAR Portfolio Manager				x	x								x		x		x		No	Yes	Yes	No		x			x		P	C/R
14	gbXML			x		x		x		x			x	x		x	x			Yes	Yes	Yes	Yes	x		x				A	C/R
15	GreenButton			x		x	x													No	No	No	No		x					P	C/R
16	HPXML			x		x				x						x	x			Yes	No	No	Yes			x				A	R
17	IFC			x		x		x		x		x				x	x			Yes	Yes	Yes	Yes	x	x	x			x	A	C/R
18	ISO 12655-2013		x			x	x			x						x				No	Yes	No	Yes		x	x		x		A	C/R
19	ISO 15489-2016	x											x	x	x			x	x	No	Yes	Yes	No			x				A	C/R
20	ISO 52000-1:2017	x				x								x					x	No	Yes	Yes	No				x			A	C/R
21	oneloTa				x	x		x	x	x	x				x					Yes	Yes	Yes	No	x						P	C/R
22	Project Haystack	x	x	x		x		x	x	x	x							x		Yes	Yes	Yes	Yes		x				x	P	C/R
23	SAREF		x			x				x	x			x						Yes	Yes	No	Yes		x					U	C/R
24	SEED	x			x	x										x			x	Yes	Yes	Yes	No	x		x		x		A	C/R

Figure 5. Summary of the selected 24 data tools

In terms of the data coverage, each data tool is capable of representing some types of building data. For example, the ISO Standard 12655-2013 [34] and Green Button [20] are used to represent the energy use data in buildings, the Brick schema [39] is used to represent the metadata for the systems and equipment operational data, and the ontology proposed by Mahdavi [37] is good at structuring the environmental data, the systems and equipment operational data, and the occupancy data. BIM tools such as IFC and gbXML [23] are designed to specifically represent a building's geometry data and its system and component data. However, coverage of other data such as occupant data and design basis data are still limited. Only a few data tools are capable of capturing certain types of data under those two categories. None of the selected data tools are capable of representing all the data categories collected from buildings. A combination of the data tools is needed when sharing and curating the building dataset.

As for the application perspective of these data tools, not surprisingly, none of the 24 data tools cover all data use cases across the entire building life cycle. For example, BIM tools like gbXML [42], IFC [3], and EnergyADE [43] are mostly used during the building design phase, but have limited data coverage for the building operational data. The Brick schema [39] is powerful at documenting the metadata for the sensor and meter time-series data during the building operation stage, but it is limited in covering the data for design. Audit Template [44] is designed specifically to collect, store, and report building and system characteristics for energy audit purposes. Because each data tool has its focus and limitations, it is important to choose the appropriate one from various candidates for different applications and use cases.

In addition, most of the 24 data tools (except BPD [41], SEED [45], and BuildingSync [35]) do not use a standard data dictionary, which can lead to issues of ambiguity and limited interoperability in data sharing and transformation. To address some of these limitations, new efforts such as ASHRAE Standard 223P [46] aim to standardize concepts for semantic information and the digital exchange of such information between machines.

3.2 Representative data tools from each tool category

3.2.1 Data terminology tools

The amount of available building energy data in digital form is increasing rapidly due to the increased availability of sensors and controls, the development of new applications for sensing and controls in buildings, and the digitization of previously paper-based building transactions. These changes will help stakeholders understand building performance variations from different perspectives and purposes. However, these generated datasets are hardly uniform in their quality, coverage, and level of description, since they are hosted in many decentralized databases with different formats. Data owners and users usually have to spend a significant amount of time on data integration, formatting, and cleaning during the data sharing and analysis process. Therefore, a common data format defined using the data dictionary could increase interoperability among different tools that share the same terms and definitions, and mitigate the risk of ambiguity while sharing and aggregating data. In this section, we select BEDES and ISO/TR 16344 as two representative data dictionaries, and further evaluate their ability to standardize the terms and definitions for building datasets.

BEDES

The Building Energy Data Exchange Specification (BEDES [47]) is a dictionary of terms and definitions designed to support the analysis of the measured energy performance (e.g., building characteristics data, energy use data, and efficiency measure data) for commercial and residential buildings. Currently, it is maintained by LBNL, and has been facilitating the exchange of information on building characteristics and energy use more consistently and at a lower cost [40]. BEDES utilizes data fields from Green Button, ENERGY STAR Portfolio Manager, and Home Performance XML. A few storage platforms (e.g., BPD) and schemas (e.g., BuildingSync) also leverage the standard energy data terminology defined in BEDES. Energy use data, systems and equipment operational data, design data, and building and system asset data are included in this dictionary.

The building performance data defined in the BEDES dictionary can be further applied in several potential use cases, such as energy efficiency investment decision making by building owners and managers, building performance tracking by public entities, as well as energy efficiency program implementation and evaluation by public entities and program administrators. One remaining gap for further improvement is establishing the mapping relationships between BEDES and other existing data formats for its broader applications.

ISO 52000-1

ISO 52000-1 [48] (formerly ISO/TR 16344) was developed by the European Committee for Standardization (CEN) Technical Committee CEN/TC 371, in collaboration with ISO Technical Committees TC 163 and TC 205. It establishes systematic and comprehensive terminologies for evaluating the energy performance of new and existing buildings in terms of primary energy or other energy-related metrics. The terms and definitions are defined in a structure [34] that includes: Building (characteristics of building), Indoor and Outdoor Conditions (temperature and solar irradiation), Technical Building Systems, Energy (e.g., energy use and building control), Energy Performance, and Energy Calculations. The current limitation of this dictionary lies in the lack of definitions of occupancy and service data, as well as sensor and meter data.

3.2.2 Data representation tools

Data representation tools are organized by what data they cover during the building's design, audit, and operation phases.

3.2.2.1 Design data

gbXML

The Green Building XML (gbXML [49]) is a widely adopted schema to facilitate the communications of different 3D building information models (BIM), currently maintained by an industry consortium. gbXML enables interoperability across disparate building design software tools by sharing the building information in a uniform language. It can be used for the design, operation, and maintenance of the building information models towards an energy efficiency building. Nowadays, gbXML has developed the capabilities of import and export in more than 50 engineering modeling tools used in the industry, which makes it the defacto industry standard schema within this area.

gbXML adopts a “bottom-up” approach which has fewer layers of complexity compared with other BIM schema such as IFC. The XML format provides a nonproprietary and robust file format for storing and exchanging data information between different vendors, devices, or platforms. In terms of application, gbXML is still limited to the energy simulation domain. A well-formatted gbXML file can drive the simulation of DOE-2, and export input files for further simulations in EnergyPlus and eQuest.

IFC

The Industry Foundation Classes (IFC [50]), adopted as an international standard, is an open data model for sharing and exchanging BIM data among various software tools such as Autodesk and other CAD-based tools. Unlike gbXML which is mostly applied to the building design stage and limited to the energy simulation domain, IFC was developed to represent the BIM data across the whole building life cycle, from building design, building construction, to building operation, building commission and other implementations) [51]. Even so, the interoperability of BIM data at different application stages is still limited due to the incomplete and ambiguously IFC attributes being used in reality. Most applications are still restricted to academic research.

Compared with the aforementioned gbXML, IFC adopts a “top-down” approach, which is more comprehensive and generic to represent an entire building structure. This “top-down” approach can trace all the semantic changes when one element in the schema is changed, and maintain an automatic semantic integrity with more complexity in program and implementation. This feature yields a more complex data schema and a larger data file size, making it a significant gap when implementing the IFC standard[23].

EnergyADE

The Energy Application Domain Extension (EnergyADE [52]) is an Application Domain Extension to CityGML which is an international standard of 3D city models. Energy ADE represents information on energy systems in buildings to provide input for building energy modeling. It’s designed to create a standard-based data model to allow: 1) energy simulation for single-building, based on sophisticated physical models for buildings and occupant behaviors, and 2) city-scale, bottom-up energy evaluations, specifically focusing on the building sectors. In terms of the application stages during the building life cycle, EnergyADE can be applied in both the design and analysis stage at either the building or urban scale, aiming at energy demand diagnostics and low-carbon energy strategies [53].

Overall, EnergyADE can model building performance at the district or city scales for design and operation purposes. It’s also part of the international standard of the 3D city model: CityGML. However, it cannot still represent some information (e.g., renewable energy systems such as PV), which is a gap they are trying to address.

3.2.2.2 Audit data

BuildingSync

BuildingSync [54] is an XML schema designed to standardize the collection and analysis of the energy audit data for commercial buildings. The schema was specifically built based on the subset of energy audit data terminologies defined in BEDES. BuildingSync provides a basis for comparing the audit analysis results across different software and facilitates the communications among audit tools. It can import the energy audit data collected from various building energy

analysis software with different formats and structures, and aggregate them into the BuildingSync format for further comparison and analysis [35]. The BuildingSync tool includes three elements: a data field dictionary which aligned with the terminologies defined in BEDES; an XML schema (.xsd) file documenting the relationships among each data field; as well as a list of energy conservation measures (ECMs) to enhance the building and system characteristics. In addition, BuildingSync's Use Case Selection tool helps users select the correct fields for the use case.

BuildingSync can standardize the collection of energy audit data across the building life cycle, and facilitate the aggregation of the audit analysis from different audit tools. Currently, it covers most of the essential data fields defined in ASHRAE's Procedures for Commercial Building Energy Audits to calculate the building's energy asset score. The new ASHRAE 211 standard on energy audits also recommends collecting data using the BuildingSync format [35]. However, the remaining challenge is that similar buildings might be represented in many different ways due to the large size of the schema.

3.2.2.3 Operation data

Building operation data includes data generated by the metering and control infrastructure. Since a single commercial building can have thousands of sensors recording readings every few minutes, operational data are typically a larger set than the design data. Until recently, there were no standard metadata schemas to describe the meaning of building operation data collected in the BAS and metering system [7]. This is partially due to the unique nature of each BAS setup (i.e., each building is different), but it is also due to the lack of standardization in the naming convention used by vendors and technicians [55]. Given the number of "points" per building (typically thousands) and the lack of standard naming conventions, significant time of an expert is needed to write new code (e.g., a new control sequence) or set up new tools (e.g., an FDD platform), limiting the scalability of these software applications [16]. The same issue applies to innovative applications; therefore, addressing the lack of standardization in metadata is a key to improving the potential market for these applications.

Meter Data: Green Button

Electromechanical electricity meters have been used for more than a century, but it was only with the introduction of smart meters [56,57] that the process of recording fine-grained energy consumption in buildings was digitized at scale. In the United States, smart meters have been deployed since the late 2000s as part of government efforts commonly known as grid modernization or "smart grid" initiatives [58]. Different smart meter technologies were deployed in different utility territories, resulting in a lack of consistency in how the data were collected and presented to utility customers. In 2011, the U.S. government issued a challenge to utilities to develop "Green Button [20]": a means of providing detailed customer energy-usage information available for download in a simple, common format [20]. According to the U.S. DOE [20] a total of 50 utilities and more than 60 million homes have access to their data via Green Button to date. Green Button technical specifications are based on the Energy Services Provider Interface data standard, which was released by the North American Energy Standards Board [59], and it is maintained by this organization. The standard consists of an XML format specification and a data exchange protocol for automatically transferring the data from a utility to a third party based on customer authorization [20]. This schema only supports smart meter data (i.e., building operation) and has no relationship with other schemas discussed in this section.

Building Automation Data

Lighting, HVAC, access control, and fire protection in large commercial buildings are typically controlled by a building automation system (BAS) [60]. In the early days of this industry (i.e., the 1980s), BAS vendors developed competing technologies based on proprietary protocols. This market strategy created systems that lacked interoperability and forced customers to be locked into one specific vendor ecosystem and line of products. To address this issue, in the late 1990s, ASHRAE developed a new standard communication protocol for automation systems called BACnet [61,62]. Two decades later, BACnet has been widely adopted by U.S. manufacturers, but the interoperability challenge is far from being resolved. BACnet imposes structure in the way information is communicated but does not provide rules to specify the meaning of the data. The "name" of an object can be anything, and it is left to the user to convey the purpose of a "point" within the system [7]. While naming conventions can be adopted, they are typically not enforced. Figure 6 shows 19 different names found to describe the same sensor type (i.e., discharge air pressure sensor) in just three buildings [55]. Given the number of "points" per building (typically thousands) and the lack of standard naming conventions, significant expert time is needed whenever control contractors must write or update control code and set up new tools (e.g., an FDD tool), which limits the scalability and flexibility of these software applications. The lack of consistency and clarity in the meaning of the data is commonly described as a lack of "semantic interoperability" [4,7].

Examples of Names for Discharge Air Pressure Sensors		
BAS Implementation 1	BAS Implementation 2	BAS Implementation 3
ACAD.AHU1.Supply Air Pressure	15 AHU 1 SA PRESS	30_ahu-001/dstpr
BJ1.AHU1_2.SSP	015-AHU-008.DA1-P	30_bl-023_024/statc_press
GIEDT.AHU.AHU1.SSP1	AHU00150.SA4-SP1	33_ahu-01/stat_press
GHA.AHU1.FAN SSP	70_BL184.DS-P	59-ahu_001/control_pressure
BRIG.SF1A.SUP STATIC	77 AHU 7 SA DUCT PRESSURE	59-ahu-004/da_stat_press
GBSF.AHU3.SPD	86 BL5-DP	
CHEM.AH2N.DUCT STATIC	90_BL4-5.DUCT-DP	

Figure 6. Example of lack of standardization in names of sensors in commercial buildings' BAS [55]

Project Haystack

An industry-driven initiative to address the problem of semantic interoperability is Project Haystack [63] [24]. Project Haystack[1] standardizes semantic data models for common building equipment such as meters, HVAC components, and lighting. Further, it provides specifications for API and serialization format for data exchange [7]. The technical specification and some of the tools shared within the community are open-source, while others are distributed commercially (e.g., Skyspark [64]). The Haystack data model goes beyond naming conventions and uses combinations of "tags" (i.e., key-value pairs) to describe building "entities" (e.g., a physical object in a building such as a pump). Tags are used to annotate data with categories (e.g., "site" identifies

[1] We are here describing Project Haystack 3, which is the public version of the data model at the time this paper was written. Several innovations have been introduced in version 4, but they have not yet been officially adopted.

a building), to specify values (e.g., the floor area of a building), or to identify the relationship between entities (e.g., a sensor point belongs to a piece of equipment). The model is machine-readable and can be queried via the specified API. One of the strengths of Project Haystack is its simplicity, as tags are easy to understand and to use by domain experts [7]. Another advantage is its flexibility, since people developing the models can extend them using customized tags for their particular application. However, these features come with significant drawbacks, since the lack of a formal structure (i.e., the combination of tags are not formally defined as concepts) prevents proper validation of conformance to a "standard" Haystack language [65]. As a result, two implementations of Haystack data models on different buildings may not be interoperable. Project Haystack is only used to represent operational and commissioning data and covers the categories of data indicated in Figure 5 (e.g., mostly BAS and meter data). The upcoming version 4.0 of Haystack is meant to address some of these issues by imposing a formal ontology on top of the existing tag-based model [7]. The project has been gaining traction in the last few years, and it is supported by a nonprofit organization sponsored by different commercial partners. Project Haystack has been coordinating its efforts with the team developing Brick (see below) and the ASHRAE committee working on semantic modeling [66].

Brick

Another emerging open-source metadata schema designed to describe building operation data is Brick [39]. Brick was developed by a consortium of universities to overcome some of the limitations of other schemas, such as Project Haystack and IFC. Brick provides "an extensible dictionary of terms and concepts, a set of relationships for linking and composing concepts together and a flexible data model [7] based on semantic web technologies [67]. Brick[2] covers the operation and commissioning phase of the life cycle and overlaps with Project Haystack in categories of information represented (Figure 5). Figure 7 illustrates an example of a Brick model. It depicts an air handling unit (AHU) supplying air to a variable air volume (VAV) box that conditions a thermal zone composed of two rooms. In one of these rooms a networked lighting system is also installed (luminaire). The Brick model is able to represent more clearly the relationship between these building components and to represent complex relationships between the lighting and the HVAC systems that were impossible to represent in Haystack 3.0. Compared to Haystack 3.0, Brick is more structured, formal, and expressive, but requires more specialized software tools [7]. Brick is an active development supported by academic institutions, federal agencies, and industry [68]. Reference implementations of Brick tools, as well as building data, are shared and open.

[2] This section describes Brick 1.1, which is the current version of the schema at the time this paper was written.

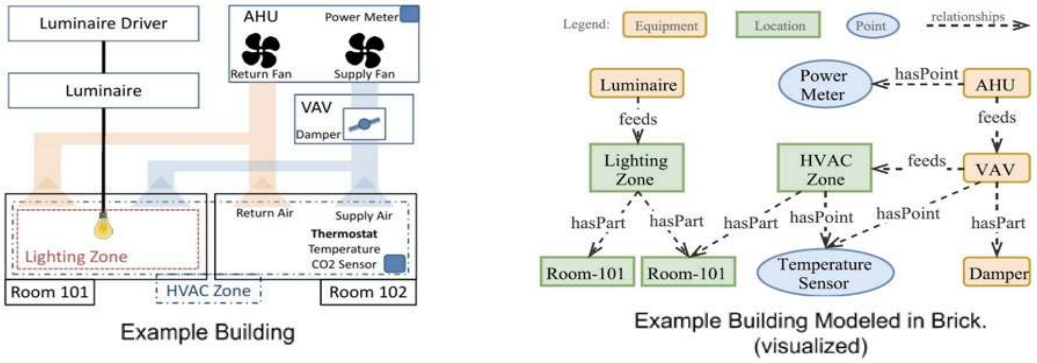

Figure 7. Example of a Brick model [65]

Annex 66 Ontology

As part of Annex 66's efforts [69], under the International Energy Agency's (IEA's) Energy in Buildings and Communities (EBC) Programme, an ontology was developed by Mahdavi to represent and incorporate multiple layers of data information obtainable from different categories of building monitoring systems, which can be used to support and streamline building data acquisition, storage, and processing in multiple computational applications [37]. The ontology is grounded on the identification of six basic data categories, namely, inhabitants, indoor and outdoor environmental conditions, control systems and devices, equipment, as well as energy flows.

Sensors, meters, and other data sources (e.g., simulated virtual sensors and human agents) in the aforementioned six categories generate streams of information (values of corresponding variables) subject to monitoring, storage, and processing. Given each data category and the respective subcategories, monitored variables are specified in terms of their values, associated sources, and possible actors. Currently, this ontology is highly theoretical, with pilot implementation for a few datasets [70].

3.2.3 Data platforms

Getting access to a large amount of building data is crucial for accelerating the evaluation of the building performance and energy efficiency. In this section, we review Audit Template, BPD, SEED, and OpenEI as four representative tools and further evaluate their ability to host and manage building datasets.

Audit Template

Audit Template is a web-based tool developed by PNNL, for collecting building energy audit information. It has built links with many other U.S. DOE's data tools, such as BuildingSync and SEED. Audit Template collected data can pass to the Asset Score tool for modeling and analyzing the building's energy asset score based on the local energy audit ordinances.

The Asset Score [44] of a building reflects the building's as-built physical characteristics and energy efficiency performance, independent of operations and occupancy. The asset score is calculated based on the building's envelope system, HVAC system, lighting system, service hot water system, and other major energy-used equipment such as plug load and elevator. It is

calculated by applying standard assumptions to an energy model to estimate the building's energy use intensity for the following four operational conditions: (1) occupant density, (2) building operating schedule, (3) plug-load density, and (4) indoor temperature set points and ventilation rates. The asset score report includes the building's current score, the recommendations for efficiency upgrades, and its expected score after the efficiency upgrades. It also includes an assessment of individual building energy systems, and a list of data used to score the building [36]. Asset Score can import data in the BuildingSync format.

BPD

The Building Performance Database (BPD [71]) is an anonymized database that contains energy use intensity data for all types of buildings. It has been widely adopted in the United States [41], and is maintained by LBNL. Currently, the BPD contains data from more than 1 million buildings [22], with a minimum data requirement of the basic building characteristics information, such as building type and location, as well as the energy use data. Some datasets also contain additional information such as the detailed characteristics of building systems and their operations. Users can filter specific building types based on their use cases, and compare performance trends among similar buildings by identifying and prioritizing cost-saving energy efficiency improvements. Based on the analysis, they can also assess the range of likely savings from these improvements during the operation, rating, and audit. In terms of time resolution, BPD can support high-resolution energy use data collected at daily, hourly, or 15-minute intervals.

BPD is typically used as an energy benchmarking tool. Its value lies in its ability to analyze custom-defined peer groups. Users can determine the peer groups by building location, building type, or even detailed building characteristics, and then compare the statistical relationship between variables within a peer group [72]. However, there are two gaps for further improvement: 1) insufficient or sparse data increases the uncertainty in defining peer groups and further calculating energy reductions, 2) BPD doesn't provide access to the raw data to protect the data privacy, which might sacrifice some data analysis during the applications.

SEED

The Standard Energy Efficiency Data (SEED [73]) Platform is an open-source, standardized data management tool supported by U.S. DOE. It was designed to import data from other data tools (e.g., Green Button, ENERGY STAR Portfolio Manager, and U.S. DOE's Building Energy Asset Score tools), and enable merging multiple sources of building energy data into one dataset which can be exported through other database platforms, such as BPD. It facilitates users to automate the process of data cleaning, formatting and validation. It also allows multiple parties to work on the same dataset while keeping track of activities. Besides, additional tools could be linked to the core function through an API call, to facilitate access to the building data from outside organizations. One remaining gap is that the data are stored in a flat database table, which makes importing structured data such as BuildingSync difficult.

OpenEI

The Open Energy Information (OpenEI [74]) is a website developed and maintained by NREL, to share and access energy data for multiple applications. There are currently more than 200 thousand raw and curated datasets on OpenEI, covering the topic from renewable energy to policy and regulations. Users on the OpenEI website can view, edit, contribute, and download data for free. OpenEI provides two approaches for sharing: 1) a semantic wiki for collaboratively

managed resources, using the MediaWiki and Semantic MediaWiki extension; and (2) a dataset-upload system for contributors to submit their recourses. The sharing information is made available via Linked Data (structured data which is interlinked with other data through semantic queries) standards whenever possible.

To be included in the OpenEI platform, datasets must be validated data with referenced sources. All submissions must be evaluated by data experts before acceptance. The user community also helps to expand the data and increase the accuracy afterward. The overall process of uploading the datasets to OpenEI ensures the data's quality and accuracy, which strengthens this platform's ability to help users make energy, market investment, and technology development decisions.

3.3 Limitations of existing data tools

Based on the review of the aforementioned data tools under each tool category, we summarized the unique features and critical limitations for each tool in Table 2.

Table 2. Summary of the selected data tools: unique features and limitations

Tool Categories		Tool Names	Unique Features and Limitations	General Limitations / Gaps
Data terminology		BEDES	A pure terminology for buildings and energy data. Originated from U.S., continues to evolve.	Need to supplement new terms for occupancy and occupant activities, indoor and outdoor environmental data, sensor and meter data, and other data for evaluating demand flexibility and
		ISO 52000-1	A set of standards on a systematic, comprehensive and modular structure for assessing the energy performance of buildings. Developed and widely used in European Union.	
Data ontology and schema	Tools for design data	gbXML	A BIM for individual buildings, mostly applied in energy simulations. Originated from U.S., widely adopted and easy to use. Representation of HVAC systems is limited.	None of the existing data tools cover all 10 data categories (Brick and Annex 66 Ontology provide best coverage). BuildingSync, BPD and SEED are using BEDES to standardize the terms, while others use their own terminology. Therefore, addressing the lack of standardization in metadata is key to improving the potential market for these applications. The capabilities of allowing various levels of details (flexibility) and adding new data attributes (extensibility) remain a challenge.
		IFC	A widely adopted International BIM standard. Very detailed and complex during the application.	
		EnergyADE	Energy domain application extension to CityGML for representing district or urban scale buildings, still in early development stage. Representation of energy systems especially HVAC systems is limited.	
	Tools for audit data	BuildingSync	An XML schema designed to capture energy audit data in line with ASHRAE Standard 211,. Originated from U.S., getting more adoption in US DOE data tool ecosystem. Similar buildings might be represented in many different ways due to the large size of the schema.	
	Tools for operational data	GreenButton	A simple XML schema representing time interval energy use data (i.e., smart meter data), used to exchange data from utility systems to third party providers or customers. Broadly adopted by many utilities. Only supports smart meter data and has no relationship with other schemas.	
		Project Haystack	A flexible tagging system to annonate data points and create metadata models, with good adoption in industry. It also provides an API and a query language. Lack of consistency between implementations and difficulty in validating models.	
		Brick Schema	A metadata schema (ontology) representing semantics of sensors and equipment in buildings. Builds on top of Haystack 3.0. More structured, formal, and expressive. Still under development, limited industry adoption, and requires more specialized software tools.	
		Annex 66 Ontology	A metadata schema developed by academics under international project IEA EBC Annex 66 project. Highly theoretical, limited use in industry.	
Database and management platforms		Audit Template	A web-based template to collect energy audit data, and save the data in the BuildingSync format.	Data privacy and security issues hinder sharing individual building data. These data platforms are still limited in the type of data they can store and exchange.
		BPD	A database and energy benchmarkig tool widely adopted in U.S.	
		SEED	An open-source platform to automate the cleaning, formatting, and integrating process of the building data.	
		OpenEI	A data portal for uploading and sharing datasets.	

A common data schema defined using a standardized data dictionary could increase interoperability among different tools that share the same terms and definitions, and mitigate the risk of ambiguity and transaction costs while sharing the data. While reviewing the two data dictionaries (BEDES and ISO 52000-1), we found that BEDES is designed to support the analysis of the measured energy performance (e.g., building characteristics data, energy use data, and efficiency measure data) for commercial and residential buildings. It utilizes data fields for Green Button, ENERGY STAR Portfolio Manager, and Home Performance XML. A few storage platforms (e.g., BPD) and schemas (e.g., BuildingSync) also leverage the standard energy data terminology defined in BEDES. However, since the goal of BEDES more specifically focuses on creating new data fields and formats (e.g., building equipment characteristics and occupancy data), it is crucial to establish the mapping relationships between BEDES and other existing data formats for its broader applications in future enhancements. Compared to BEDES, ISO 52000-1 is more targeted at assessing energy performance based on primary energy and other energy-related metrics. The two dictionary tools do not cover all the terms needed for the nine data categories (e.g., occupancy data and sensor and meter data). This is a gap that needs to be addressed.

A data platform stores a large amount of data in an organized structure and allows users to contribute, modify, query, analyze and export the datasets, which is crucial for accelerating the evaluation of building performance and energy efficiency. We reviewed three database management tools: BPD, SEED, and OpenEI. BPD, as the largest U.S. building energy database storing the energy-related characteristics of more than one million commercial and residential buildings, enables users to analyze custom-defined peer groups. It's not only considered an energy benchmarking tool, but also a retrofit suggestion tool. However, insufficient or sparse data increases the uncertainty in defining peer groups and further calculating energy reductions. Besides, users might not expose the raw data, except for the peer grouping criteria and the energy consumption data. The SEED platform is an open-source, standardized data management tool supported by U.S. DOE. It can directly import data from other data tools (e.g., Green Button, ENERGY STAR Portfolio Manager, and U.S. DOE's Building Energy Asset Score tools), and enable merging multiple sources of building energy data into one dataset, which can be exported through other database platforms, such as BPD. One remaining gap is that the data are stored in a flat database table, which makes importing structured data such as BuildingSync difficult. Another data portal, OpenEI, is a portal to share and access energy data, specifically for renewable energy and energy efficiency. Users are restricted to publishing their raw or derived data only when those datasets are validated with referenced sources.

For design data representation tools, we reviewed three BIM tools: gbXML, IFC and EnergyADE. Each of these tools has a specific application domain. gbXML is typically used in energy simulations of single buildings, EnergyADE is used for modeling building performance at the district or city scale (both for design and operation), while IFC covers application stages from building construction to the building commission domain. An additional difference between the three data schemas is the level of granularity. gbXML adopts a "bottom-up" approach which has fewer layers of complexity compared with IFC. Similarly, EnergyADE lacks the capability to represent building and system information at various levels of detail (a key feature of CityGML schema that EnergyADE builds upon), which is a gap to be addressed in a future revision of the standard. Conversely, IFC adopts a "top-down" relational structure, which yields a more complex data schema and a larger data file size. The "top-down" structure can trace all the semantic changes when one element in the schema is changed. It can maintain an automatic semantic integrity with more complexity in program and implementation, which makes it a major gap when

implementing the IFC standard. For the audit data representation tools, we reviewed the schema of BuildingSync and Audit Template. BuildingSync was developed based on the standardized energy data terminology defined in BEDES, which ensures consistency in naming and extensibility. However, the size of the schema is large and the schema is complex, which may cause different people to represent similar buildings in different ways. Audit Template was designed as a simplified web-based tool for specifically collecting building and system information during on-site audits. The data can be exported to a BuildingSync file.

Among the schemas analyzed Haystack and Brick overlap in scope. Haystack is more flexible and easier to use, but this flexibility may lead to a lack of consistency between implementations and difficulty in validating models. Brick, conversely, is a more structured, formal, and expressive ontology, but it requires more specialized software tools, and it is still under development. A more formal comparison of the two schemas is presented by Fierro et al. [75]. Another advantage of Haystack is the industrial adoption and large user community, while Brick is still at the early stages of development and has been embraced mainly by the academic community. Like Brick, the Annex 66 ontology is a metadata schema developed by academics under international project IEA EBC Annex 66 and had not seen adoption in commercial applications. Finally, Green Button is a simple schema that covers utility metering data it is narrowly focused on facilitating the exchange of data from utility systems to third-party providers or customers and cannot be easily extended to other applications.

4. Discussion

4.1 Implications

Buildings and IoT devices are producing a growing volume of data; however, analysis of the data and extracting insights to inform building energy efficiency or occupant comfort is limited. One of the challenges is the labor-intensive process of understanding the data and preparing data in a form for analytics, which has to repeat for every building dataset.

The reviewed 24 data tools can be used to provide standard terminology and metadata representation to curate building data to enable its use across the building life cycle. There are gaps revealed, including lack of data ontology or schema to represent metadata of occupant data, or design basis data, which can be addressed by enhancing the BRICK schema or BIM (e.g., gbXML or IFC). Currently, building design data or models (e.g., BIM) and building operation data (e.g., BRICK schema, Green Button) are represented with different data tools, which makes data reuse difficult without extra efforts to integrate these two types of data. As the building industry is moving to the future of digital twins, semantic data modeling of the physical and virtual buildings and their related data is crucial.

The reviewed 24 data tools contribute to making building data compliant with the FAIR principles. **Findable and Accessible**: BPD is a database platform hosting building performance data at the aggregated level for the public (individual building's data is not shared due to privacy concerns); BPD also allows data access via API calls. OpenEI is an open data portal enabling users to upload and share datasets with the public; SEED is an open-source database hosting platform that can help make datasets available to the public or specific customers. **Interoperable**: BEDES, BRICK schema, and Green Button schema can help standardize terminology and metadata representation for sensor data and smart meter data; BuildingSync, gbXML, IFC, and EnergyADE can help make the energy audit data and building information models easy to share between

users, applications, and workflows. **Reusable**: the metadata and data can be represented using these tools to enable their replication or integration into other datasets.

From the perspectives of the FAIR principles, the authors recommend the following to encourage curation and sharing of buildings datasets and to unlock the value of data for the building industry: (1) hosting datasets in data portals with public access (in various ways, such as direct download and API) and ease of data query and management, (2) providing rich metadata and representing metadata and data in standardized terminologies and schemas, and (3) maintaining the dataset for continuous quality improvements and addition of new data for a building during its life cycle.

Adoption of these data tools (e.g., BRICK schema) for building operation data is still limited. For example, only up to a few hundred buildings have BRICK models at the time of this writing. Various challenges need to be addressed to accelerate the adoption of these data tools including (1) value proposition: how standardized and semantic datasets can help streamline the analytics to unlock values of data for informing building life cycle to improve operations and reduce energy use and carbon emissions; (2) open-source or free codes/tools to facilitate the use of the data tools: currently it takes significant efforts to develop a BRICK model for a building, and there is no easy to use tool to help check the quality of the BRICK model; and (3) different types of building data may be collected by different stakeholders for different purposes across different stages of the building life cycle – there is no a single party responsible for the overall data collection design or use. Ideally, every building needs a dedicated data engineer responsible for all data-related efforts and can coordinate among various stakeholders, similar to the architect and MEP engineers of a building.

4.2 Related efforts on buildings data

There are synergistic U.S. and international efforts on data models and schema to address some of the aforementioned gaps to improve standardized representation and interoperability of buildings data.

ASHRAE is developing a new Standard 223P: “Semantic Data Model for Analytics and Automation Applications in Buildings” which aims to formally define knowledge concepts and a methodology to apply them to create interoperable, machine-readable semantic models for representing building system information for analytics, automation, and control [76]. This new standard strives to harmonize existing metadata schema (e.g., Brick and Haystack) to enable interoperability on semantic information across the building industry, particularly in building automation. The standard effort is jointly supported by the ASHRAE BACnet committee, Project Haystack, and the Brick initiative [66].

IBPSA-USA recently established a Building Data Exchange Committee, which aims to provide an inclusive forum to collaboratively support tool-agnostic data exchange through the development of consensus-based data models to inform building design and operations.

Annex 81: Data-driven Smart Buildings, an international collaborative project under the IEA’s Energy in Buildings and Communities Programme [77], aims to develop or integrate data models, dataset, best practices, and case studies to demonstrate digital solutions that can rapidly scale and provide energy efficiency knowledge that can be widely encapsulated and disseminated within highly accessible software applications.

4.3 Limitations

This data tools review study has limitations: (1) there are unavoidably some data tools we miss to cover, especially emerging data tools that may become available soon considering the rapid developments in data and analytics in the building sector; (2) Exercising these data tools using two or more building datasets with diverse data types and resolutions would be helpful to verify their usability and gain a deeper understanding of their limitations or gaps; (3) Other types of data and tools, e.g., from social media, IoT devices, and mobility, are not covered; and (4) tools that address issues of data privacy, data quality, and data security are not included in the review as they are big topics and existing literature cover them well.

5. Conclusions

To enable analytics of the increasing volume of data collected in buildings, there is a strong need for a toolchain to curate and represent building performance data in common standardized terms and common schemas to enable interoperability between tools and applications. This study selected and reviewed 24 data tools based on common use cases of data in the building life cycle. The selected data tools are categorized into (1) data dictionary or terminology, (2) data ontology and schemas, and (3) data platforms. The building data are grouped into ten typologies and mapped to their frequently used representation tools. Throughout the process, gaps and limitations of the existing data tools were identified.

The main findings of the review are (1) standard terminology such as BEDES should be adopted by all data tools to facilitate the communication and interoperability among multiple data sources; (2) none of the reviewed data tools can represent all ten typologies of building data; particularly, ontologies or schemas representing occupant data and basis of design should be enhanced or developed; (3) capabilities of allowing various levels of details (flexibility) and adding new data attributes (extensibility) remain a challenge for current ontologies and schemas; (4) data platforms are still limited to the type of data being able to store and exchange, as well as the data privacy and security issues which hinder the sharing of individual building data; and (5) integrating data across various stages (e.g., design and operation) in the building life cycle remains a challenge for solving problems such as the performance gaps. It is recommended that further development or maintenance of these data tools should engage diverse stakeholders (inside and outside the development party, both national and international), and ensure consistency and interoperability to support broader use cases.

As part of the U.S. DOE Benchmark Dataset project [78], a follow-up study will propose enhancements to the Brick schema to represent the metadata of occupant's monitoring data, and develop a data schema to describe the basis of design data.

Acknowledgments

This research was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the United States Department of Energy, under Contract No. DE-AC02-05CH11231. The authors benefited from technical discussions with project team members at NREL, PNNL, and ORNL. The authors thank the strong support of Erika Gupta and Harry Bergmann, technical managers of the project at the Building Technologies Office of the U.S. Department of Energy.

References

[1] National Institute of Building Sciences (NIBS), New Buildings Institute, National Environmental Balancing Bureau. Data Needs for Achieving High Performance Buildings: High-Performance Building Data Collection Initiative. Washington, DC: 2011.

[2] Jia M, Komeily A, Wang Y, Srinivasan RS. Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications. Autom Constr 2019;101:111–26. https://doi.org/10.1016/j.autcon.2019.01.023.

[3] Yang QZ, Zhang Y. Semantic interoperability in building design: Methods and tools. CAD Comput Aided Des 2006;38:1099–112. https://doi.org/10.1016/j.cad.2006.06.003.

[4] Hardin DB, Stephan EG, Wang W, Corbin CD, Widergren SE. Buildings Interoperability Landscape. 2015.

[5] Rossi L, Belli A, De Santis A, Diamantini C, Frontoni E, Gambi E, et al. Interoperability issues among smart home technological frameworks. MESA 2014 - 10th IEEE/ASME Int. Conf. Mechatron. Embed. Syst. Appl. Conf. Proc., Institute of Electrical and Electronics Engineers Inc.; 2014. https://doi.org/10.1109/MESA.2014.6935626.

[6] Fierro G, Prakash AK, Mosiman C, Pritoni M, Raftery P, Wetter M, et al. Shepherding Metadata Through the Building Lifecycle. Proc. 7th ACM Int. Conf. Syst. Energy-Efficient Build. Cities, Transp., 2020, p. 70–9.

[7] Bergmann H, Mosiman C, Saha A, Haile S, Livingood W, Bushby S, et al. Semantic Interoperability to Enable Smart, Grid-Interactive Efficient Buildings. ACEEE Summer Study Energy Effic. Build., 2020.

[8] Granderson J, Lin G. Building energy information systems: synthesis of costs, savings, and best-practice uses. Energy Effic 2016;9:1369–84. https://doi.org/10.1007/s12053-016-9428-9.

[9] Gallaher MP, O'Conor AC, Dettbarn JL, Gilday LT. Cost Analysis of Inadequate Interoperability in the U.S. Capital Facilities Industry. Natl Inst Stand Technol 2004:1–210. https://doi.org/10.6028/NIST.GCR.04-867.

[10] Bhattacharya A, Ploennigs J, Culler D. Short paper: Analyzing metadata schemas for buildings: The good, the bad, and the ugly. Proc. 2nd ACM Int. Conf. Embed. Syst. Energy-Efficient Built Environ., ACM; 2015, p. 33–4.

[11] Kheiri F. A review on optimization methods applied in energy-efficient building geometry and envelope design. Renew Sustain Energy Rev 2018;92:897–920. https://doi.org/10.1016/j.rser.2018.04.080.

[12] Benndorf GA, Wystrcil D, Réhault N. Energy performance optimization in buildings: A review on semantic interoperability, fault detection, and predictive control. Appl Phys Rev 2018;5:41501. https://doi.org/10.1063/1.5053110.

[13] LeCun Y, Cortes C, Burges CJC. The MNIST database of handwritten digits, 1998. http://yann.lecun.com/exdb/mnist/.

[14] Phillips PJ, Wechsler H, Huang J, Rauss PJ. The FERET database and evaluation procedure for face-recognition algorithms. Image Vis Comput 1998;16:295–306. https://doi.org/10.1016/s0262-8856(97)00070-x.

[15] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. 2009 IEEE Conf. Comput. Vis. pattern Recognit., Ieee; 2009, p. 248–55.

[16] Fierro G, Pritoni M, AbdelBaky M, Raftery P, Peffer T, Thomson G, et al. Mortar: An open testbed for portable building analytics. BuildSys 2018 - Proc. 5th Conf. Syst. Built Environ., New York, NY, USA: Association for Computing Machinery, Inc; 2018, p. 172–81. https://doi.org/10.1145/3276774.3276796.

[17] Granderson J, Lin G, Harding A, Im P, Chen Y. Building fault detection data to aid diagnostic algorithm creation and performance testing. Sci Data 2020;7:1–14.

[18] Luo XJ, Oyedele LO, Ajayi AO, Monyei CG, Akinade OO, Akanbi LA. Development of an

IoT-based big data platform for day-ahead prediction of building heating and cooling demands. Adv Eng Informatics 2019;41:100926. https://doi.org/10.1016/j.aei.2019.100926.
[19] Hossein Motlagh N, Mohammadrezaei M, Hunt J, Zakeri B. Internet of Things (IoT) and the Energy Sector. Energies 2020;13:494. https://doi.org/10.3390/en13020494.
[20] green button alliance. History of Green Button 2015.
[21] Pritoni M, Paine D, Fierro G, Mosiman C, Poplawski M, Saha A, et al. Metadata Schemas and Ontologies for Building Energy Applications: A Critical Review and Use Case Analysis. Energies 2021;14:2024.
[22] Bergmann H. DOE Building Energy Data Tools 2017.
[23] Dong B, Lam KP, Huang YC, Dobbs GM. A comparative study of the IFC and gbXML informational infrastructures for data exchange in computational design support environments. Proc Build Simul 2007 2007;1:1530–1537.
[24] “Project Haystack.” Project Haystack - Structure 2020.
[25] Molina-Solana M, Ros M, Ruiz MD, Gómez-Romero J, Martin-Bautista MJ. Data science for building energy management: A review. Renew Sustain Energy Rev 2017. https://doi.org/10.1016/j.rser.2016.11.132.
[26] Volk R, Stengel J, Schultmann F. Building Information Modeling (BIM) for existing buildings - Literature review and future needs. Autom Constr 2014;38:109–27. https://doi.org/10.1016/j.autcon.2013.10.023.
[27] Coakley D, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. Renew Sustain Energy Rev 2014;37:123–41. https://doi.org/10.1016/j.rser.2014.05.007.
[28] Zhao HX, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16:3586–92.
[29] Pelekis N, Theodoulidis B, Kopanakis I, Theodoridis Y. Literature review of spatio-temporal database models. Knowl Eng Rev 2004;19:235–74.
[30] Cerovsek T. A review and outlook for a ‘Building Information Model’(BIM): A multi-standpoint framework for technological development. Adv Eng Informatics 2011;25:224–44.
[31] Kjærgaard MB, Ardakanian O, Carlucci S, Dong B, Firth SK, Gao N, et al. Current practices and infrastructure for open data based research on occupant-centric design and operation of buildings. Build Environ 2020:106848.
[32] Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:1–9.
[33] Project Haystack. https://project-haystack.org/.
[34] Epb O. INTERNATIONAL STANDARD Overarching EPB assessment —. Iso 2017;2017.
[35] DeGraw J, Field-Macumber K, Long N, Goel S. BuildingSync® in Action : Example Implementations. 2018 ACEEE Summer Study Energy Effic Build 2018:1–12.
[36] Wang N, Gorrissen WJ. Commercial Building Energy Asset Score System: Program Overview and Technical Protocol (Version 1.0). Pacific Northwest National Lab.(PNNL), Richland, WA (United States); 2013.
[37] Mahdavi A, Taheri M. An ontology for building monitoring. J Build Perform Simul 2017;10:499–508.
[38] Balaji B, Bhattacharya A, Fierro G, Gao J, Gluck J, Hong D, et al. Brick: Metadata schema for portable smart building applications. Appl Energy 2018;226:1273–92.
[39] Balaji B, Bhattacharya A, Fierro G, Gao J, Gluck J, Hong D, et al. Brick: Towards a unified metadata schema for buildings. Proc. 3rd ACM Conf. Syst. Energy-Efficient Built Environ. BuildSys 2016, 2016. https://doi.org/10.1145/2993422.2993577.
[40] US Department of Energy. Building Energy Data Exchange Specification (BEDES). https://bedes.lbl.gov/bedes-online.
[41] US Department of Energy. Buildings Performance Database Overview 2014.

[42] Ham Y, Golparvar-Fard M. Mapping actual thermal properties to building elements in gbXML-based BIM for reliable building energy performance modeling. Autom Constr 2015;49:214–24.
[43] Nouvel R, Bahu J-M, Kaden R, Kaempf J, Cipriano P, Lauster M, et al. Development of the CityGML application domain extension energy for urban energy simulation. Build. Simul. 2015-14th Conf. Int. Build. Perform. Simul. Assoc., 2015, p. 559–64.
[44] Asset Score n.d. https://buildingenergyscore.energy.gov/.
[45] Alschuler E, Antonoff J, Brown R, Cheifetz M. Planting SEEDs: Implementation of a Common Platform for Building Performance Disclosure Program Data Management. 2014 ACEEE Summer Study Energy Effic Build 2014:4-25-4–35.
[46] American Society of Heating Refrigerating and Air Conditioning Engineers (ASHRAE). ASHRAE's BACnet Committee, Project Haystack and Brick Schema Collaborating to Provide Unified Data Semantic Modeling Solution 2018.
[47] BEDES n.d. https://bedes.lbl.gov/.
[48] ISO 52000-1 n.d. https://www.iso.org/standard/65601.html.
[49] gbXML n.d. http://www.gbxml.org/.
[50] IFC n.d. http://www.buildingsmart-tech.org/ifc/.
[51] Eastman C, Teicholz P, Sacks R, Liston K. BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors. John Wiley & Sons; 2011.
[52] EnergyADE n.d. http://www.citygmlwiki.org/index.php/CityGML_Energy_ADE.
[53] Benner J. CityGML Energy ADE V . 1 . 0 Specification 2018.
[54] BuildingSync n.d. https://buildingsync.net/.
[55] Pritoni M, Weyandt C, Carter D, Elliott J. Towards a Scalable Model for Smart Buildings. 2018 ACEEE Summer Study Energy Effic Build 2018:1–12.
[56] Cook B, Gazzano J, Gunay Z, Hiller L, Mahajan S, Taskan A, et al. The smart meter and a smarter consumer: quantifying the benefits of smart meter implementation in the United States. Chem Cent J 2012;6:1–16. https://doi.org/10.1186/1752-153x-6-s1-s5.
[57] Doris E, Peterson K. Government Program Briefing: Smart Metering - NREL. 2011.
[58] Owen J, Owen G. "Smart Meters: Commercial, Policy and Regulatory Drivers." 2006.
[59] NAESB. The NAESB Energy Services Provider Interface Model Business Practices Information Page 2015.
[60] American Society of Heating Refrigerating and Air Conditioning Engineers (ASHRAE). ASHRAE Guideline 13-2015, Specifying Building Automation Systems. ASHRAE Guidel 2015.
[61] Domingues P, Carreira P, Vieira R, Kastner W. Building automation systems: Concepts and technology review. Comput Stand Interfaces 2016;45:1–12. https://doi.org/10.1016/j.csi.2015.11.005.
[62] Samad T. Building Control and Automation Systems. Perspect. Control Eng. Technol. Appl. New Dir., IEEE; 2001. https://doi.org/10.1109/9780470545485.ch16.
[63] Project Haystack n.d. https://project-haystack.org/doc/Structure.
[64] Skyspark n.d. https://skyfoundry.com/product.
[65] Fierro GT. Design of an Effective Ontology and Query Processor Enabling Portable Building Applications. University of California, Berkeley, 2019.
[66] ASHRAE. ASHRAE's BACnet Committee, Project Haystack and Brick Schema Collaborating to Provide Unified Data Semantic Modeling Solution 2018.
[67] World Wide Web Consortium. Semantic Web - W3C. W3Org 2012.
[68] BRICK n.d. https://brickschema.org/community.
[69] Yan D, Hong T, Dong B, Mahdavi A, D'Oca S, Gaetani I, et al. IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. Energy Build 2017;156:258–70.
[70] Huebner GM, Mahdavi A. A structured open data collection on occupant behaviour in

buildings. Sci Data 2019;6:1–4.
[71] BPD n.d. https://bpd.lbl.gov/.
[72] Mathew PA, Dunn LN, Sohn MD, Mercado A, Custudio C, Walter T. Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. Appl Energy 2015;140:85–93. https://doi.org/10.1016/j.apenergy.2014.11.042.
[73] SEED n.d. http://seedinfo.lbl.gov.
[74] OpenEI n.d. https://openei.org/wiki/Main_Page.
[75] Fierro G, Koh J, Nagare S, Zang X, Agarwal Y, Gupta RK, et al. Formalizing Tag-Based Metadata With the Brick Ontology. Front Built Environ 2020;6:152.
[76] ASHRAE SPC 223P n.d. https://www.ashrae.org/technical-resources/standards-and-guidelines/titles-purposes-and-scopes.
[77] Annex 81 n.d. https://annex81.iea-ebc.org/.
[78] Benchmark Dataset n.d. https://bbd.labworks.org/.