ANNEX **81**

EBC
Energy in Building and
Communities Programme

International Energy Agency

Subtask A Report

IEA Annex 81 ‘Data-Driven Smart Buildings’

Energy in Buildings and Communities
Technology Collaboration Programme

February 2024

Technology Collaboration Programme
by iea

International Energy Agency

Subtask A Report

Energy in Buildings and Communities Technology Collaboration Programme

February 2024

Authors

Stephen White, CSIRO, Australia (stephen.d.white@csiro.au)
Chun Ping Gao, Building Construction Authority, Singapore
Gabe Fierro, Colorado School of Mines, USA
Pieter Pauwels, Eindhoven University of Technology, Netherlands
Rafael Gomez Garcia, RISE, Sweden

© Copyright CSIRO 2024

All property rights, including copyright, are vested in CSIRO, Operating Agent for EBC Annex 81, on behalf of the Contracting Parties of the International Energy Agency (IEA) Implementing Agreement for a Programme of Research and Development on Energy in Buildings and Communities (EBC). In particular, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of CSIRO.

Published by CSIRO, PO Box 330, Newcastle, NSW 2300, Australia

Disclaimer Notice: This publication has been compiled with reasonable skill and care. However, neither CSIRO, nor the Contracting Parties of the International Energy Agency's Implementing Agreement for a Programme of Research and Development on Energy in Buildings and Communities, nor their agents, make any representation as to the adequacy or accuracy of the information contained herein, or as to its suitability for any particular application, and accept no responsibility or liability arising out of the use of this publication. The information contained herein does not supersede the requirements given in any national codes, regulations or standards, and should not be regarded as a substitute for the need to obtain specific professional advice for any particular application. EBC is a Technology Collaboration Programme (TCP) of the IEA. Views, findings and publications of the EBC TCP do not necessarily represent the views or policies of the IEA Secretariat or of all its individual member countries.

ISBN ???

Participating countries in the EBC TCP: Australia, Austria, Belgium, Brazil, Canada, P.R. China, Czech Republic, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Republic of Korea, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States of America.

Additional copies of this report may be obtained from: EBC Executive Committee Support Services Unit (ESSU), C/o AECOM Ltd, The Colmore Building, Colmore Circus Queensway, Birmingham B4 6AT, United Kingdom
www.iea-ebc.org
essu@iea-ebc.org

Preface

The International Energy Agency

The International Energy Agency (IEA) was established in 1974 within the framework of the Organisation for Economic Co-operation and Development (OECD) to implement an international energy programme. A basic aim of the IEA is to foster international co-operation among the 30 IEA participating countries and to increase energy security through energy research, development and demonstration in the fields of technologies for energy efficiency and renewable energy sources.

The IEA Energy in Buildings and Communities Programme

The IEA co-ordinates international energy research and development (R&D) activities through a comprehensive portfolio of Technology Collaboration Programmes (TCPs). The mission of the IEA Energy in Buildings and Communities (IEA EBC) TCP is to support the acceleration of the transformation of the built environment towards more energy efficient and sustainable buildings and communities, by the development and dissemination of knowledge, technologies and processes and other solutions through international collaborative research and open innovation. (Until 2013, the IEA EBC Programme was known as the IEA Energy Conservation in Buildings and Community Systems Programme, ECBCS.)

The high priority research themes in the EBC Strategic Plan 2019-2024 are based on research drivers, national programmes within the EBC participating countries, the Future Buildings Forum (FBF) Think Tank Workshop held in Singapore in October 2017 and a Strategy Planning Workshop held at the EBC Executive Committee Meeting in November 2017. The research themes represent a collective input of the Executive Committee members and Operating Agents to exploit technological and other opportunities to save energy in the buildings sector, and to remove technical obstacles to market penetration of new energy technologies, systems and processes. Future EBC collaborative research and innovation work should have its focus on these themes.

At the Strategy Planning Workshop in 2017, some 40 research themes were developed. From those 40 themes, 10 themes of special high priority have been extracted, taking into consideration a score that was given to each theme at the workshop. The 10 high priority themes can be separated in two types namely 'Objectives' and 'Means'. These two groups are distinguished for a better understanding of the different themes.

Objectives - The strategic objectives of the EBC TCP are as follows:

- reinforcing the technical and economic basis for refurbishment of existing buildings, including financing, engagement of stakeholders and promotion of co-benefits;
- improvement of planning, construction and management processes to reduce the performance gap between design stage assessments and real-world operation;
- the creation of 'low tech', robust and affordable technologies;
- the further development of energy efficient cooling in hot and humid, or dry climates, avoiding mechanical cooling if possible;
- the creation of holistic solution sets for district level systems taking into account energy grids, overall performance, business models, engagement of stakeholders, and transport energy system implications.

Means - The strategic objectives of the EBC TCP will be achieved by the means listed below:

- the creation of tools for supporting design and construction through to operations and maintenance, including building energy standards and life cycle analysis (LCA);
- benefitting from 'living labs' to provide experience of and overcome barriers to adoption of energy efficiency measures;
- improving smart control of building services technical installations, including occupant and operator interfaces;
- addressing data issues in buildings, including non-intrusive and secure data collection;
- the development of building information modelling (BIM) as a game changer, from design and construction through to operations and maintenance.

The themes in both groups can be the subject for new Annexes, but what distinguishes them is that the 'objectives' themes are final goals or solutions (or part of) for an energy efficient built environment, while the 'means' themes are instruments or enablers to reach such a goal. These themes are explained in more detail in the EBC Strategic Plan 2019-2024.

The Executive Committee

Overall control of the IEA EBC Programme is maintained by an Executive Committee, which not only monitors existing projects, but also identifies new strategic areas in which collaborative efforts may be beneficial. As the Programme is based on a contract with the IEA, the projects are legally established as Annexes to the IEA EBC Implementing Agreement. At the present time, the following projects have been initiated by the IEA EBC Executive Committee, with completed projects identified by (*) and joint projects with the IEA Solar Heating and Cooling Technology Collaboration Programme by (☼):

Annex 1: Load Energy Determination of Buildings (*)
Annex 2: Ekistics and Advanced Community Energy Systems (*)
Annex 3: Energy Conservation in Residential Buildings (*)
Annex 4: Glasgow Commercial Building Monitoring (*)
Annex 5: Air Infiltration and Ventilation Centre
Annex 6: Energy Systems and Design of Communities (*)
Annex 7: Local Government Energy Planning (*)
Annex 8: Inhabitants Behaviour with Regard to Ventilation (*)
Annex 9: Minimum Ventilation Rates (*)
Annex 10: Building HVAC System Simulation (*)
Annex 11: Energy Auditing (*)
Annex 12: Windows and Fenestration (*)
Annex 13: Energy Management in Hospitals (*)
Annex 14: Condensation and Energy (*)
Annex 15: Energy Efficiency in Schools (*)
Annex 16: BEMS 1- User Interfaces and System Integration (*)
Annex 17: BEMS 2- Evaluation and Emulation Techniques (*)
Annex 18: Demand Controlled Ventilation Systems (*)
Annex 19: Low Slope Roof Systems (*)
Annex 20: Air Flow Patterns within Buildings (*)
Annex 21: Thermal Modelling (*)
Annex 22: Energy Efficient Communities (*)
Annex 23: Multi Zone Air Flow Modelling (COMIS) (*)
Annex 24: Heat, Air and Moisture Transfer in Envelopes (*)
Annex 25: Real time HVAC Simulation (*)
Annex 26: Energy Efficient Ventilation of Large Enclosures (*)
Annex 27: Evaluation and Demonstration of Domestic Ventilation Systems (*)
Annex 28: Low Energy Cooling Systems (*)
Annex 29: ☼ Daylight in Buildings (*)
Annex 30: Bringing Simulation to Application (*)
Annex 31: Energy-Related Environmental Impact of Buildings (*)
Annex 32: Integral Building Envelope Performance Assessment (*)
Annex 33: Advanced Local Energy Planning (*)
Annex 34: Computer-Aided Evaluation of HVAC System Performance (*)
Annex 35: Design of Energy Efficient Hybrid Ventilation (HYBVENT) (*)
Annex 36: Retrofitting of Educational Buildings (*)
Annex 37: Low Exergy Systems for Heating and Cooling of Buildings (LowEx) (*)
Annex 38: ☼ Solar Sustainable Housing (*)
Annex 39: High Performance Insulation Systems (*)
Annex 40: Building Commissioning to Improve Energy Performance (*)
Annex 41: Whole Building Heat, Air and Moisture Response (MOIST-ENG) (*)
Annex 42: The Simulation of Building-Integrated Fuel Cell and Other Cogeneration Systems (FC+COGEN-SIM) (*)
Annex 43: ☼ Testing and Validation of Building Energy Simulation Tools (*)
Annex 44: Integrating Environmentally Responsive Elements in Buildings (*)
Annex 45: Energy Efficient Electric Lighting for Buildings (*)
Annex 46: Holistic Assessment Tool-kit on Energy Efficient Retrofit Measures for Government Buildings (EnERGo) (*)
Annex 47: Cost-Effective Commissioning for Existing and Low Energy Buildings (*)
Annex 48: Heat Pumping and Reversible Air Conditioning (*)
Annex 49: Low Exergy Systems for High Performance Buildings and Communities (*)
Annex 50: Prefabricated Systems for Low Energy Renovation of Residential Buildings (*)
Annex 51: Energy Efficient Communities (*)
Annex 52: ☼ Towards Net Zero Energy Solar Buildings (*)
Annex 53: Total Energy Use in Buildings: Analysis and Evaluation Methods (*)
Annex 54: Integration of Micro-Generation and Related Energy Technologies in Buildings (*)
Annex 55: Reliability of Energy Efficient Building Retrofitting - Probability Assessment of Performance and Cost (RAP-RETRO) (*)
Annex 56: Cost Effective Energy and CO2 Emissions Optimization in Building Renovation (*)
Annex 57: Evaluation of Embodied Energy and CO2 Equivalent Emissions for Building Construction (*)

Annex 58: Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements (*)
Annex 59: High Temperature Cooling and Low Temperature Heating in Buildings (*)
Annex 60: New Generation Computational Tools for Building and Community Energy Systems (*)
Annex 61: Business and Technical Concepts for Deep Energy Retrofit of Public Buildings (*)
Annex 62: Ventilative Cooling (*)
Annex 63: Implementation of Energy Strategies in Communities (*)
Annex 64: LowEx Communities - Optimised Performance of Energy Supply Systems with Exergy Principles (*)
Annex 65: Long-Term Performance of Super-Insulating Materials in Building Components and Systems (*)
Annex 66: Definition and Simulation of Occupant Behavior in Buildings (*)
Annex 67: Energy Flexible Buildings (*)
Annex 68: Indoor Air Quality Design and Control in Low Energy Residential Buildings (*)
Annex 69: Strategy and Practice of Adaptive Thermal Comfort in Low Energy Buildings
Annex 70: Energy Epidemiology: Analysis of Real Building Energy Use at Scale
Annex 71: Building Energy Performance Assessment Based on In-situ Measurements
Annex 72: Assessing Life Cycle Related Environmental Impacts Caused by Buildings
Annex 73: Towards Net Zero Energy Resilient Public Communities
Annex 74: Competition and Living Lab Platform
Annex 75: Cost-effective Building Renovation at District Level Combining Energy Efficiency and Renewables
Annex 76: ☼ Deep Renovation of Historic Buildings Towards Lowest Possible Energy Demand and CO_2 Emissions
Annex 77: ☼ Integrated Solutions for Daylight and Electric Lighting
Annex 78: Supplementing Ventilation with Gas-phase Air Cleaning, Implementation and Energy Implications
Annex 79: Occupant-Centric Building Design and Operation
Annex 80: Resilient Cooling
Annex 81: Data-Driven Smart Buildings
Annex 82: Energy Flexible Buildings Towards Resilient Low Carbon Energy Systems
Annex 83: Positive Energy Districts
Annex 84: Demand Management of Buildings in Thermal Networks
Annex 85: Indirect Evaporative Cooling
Annex 86: Energy Efficient Indoor Air Quality Management in Residential Buildings

Working Group - Energy Efficiency in Educational Buildings (*)
Working Group - Indicators of Energy Efficiency in Cold Climate Buildings (*)
Working Group - Annex 36 Extension: The Energy Concept Adviser (*)
Working Group - HVAC Energy Calculation Methodologies for Non-residential Buildings (*)
Working Group - Cities and Communities
Working Group - Building Energy Codes

Summary

Subtask A of IEA EBC Annex 81 focuses on one of the most critical steps involved in implementing data-driven smart buildings services, *being establishment of the requisite IT infrastructure and data management services in a building*. This step is foundational for subsequent deployment of energy saving software applications. Barriers, associated with this step, have been identified as one of the most critical factors impeding uptake of data-driven smart building technologies and services.

Subtask A of IEA EBC Annex 81 aims to establish concepts, and provide overall guidance, on the various factors involved in non-residential buildings becoming ‘digital ready’. In this context, a ‘digital ready’ building is one where smart building services can be easily deployed, without significant further IT infrastructure upgrades. The focus is on mechanical and electrical services in buildings.

Section 2, of this report, discusses data governance and data quality issues. These are key factors that need to be considered when deploying software applications in smart buildings. On the one hand, data is considered more useful when it is Findable, Accessible, Interoperable and Reusable (FAIR). On the other hand, data governance requires consideration of commercial IP rights, ethics and privacy regulations.

Interoperability standards and stand-alone independent data platforms are powerful tools for enabling the FAIR data principles. To manage risks, permissions should be obtained to process relevant data streams. This can be achieved through appropriate data licenses, and/or identifying a relevant legal basis to use personal data.

Building technology providers should be required to use common interoperability standards and, in particular, to utilise a metadata schema to ensure that collected data has relevant context and meaning. Metadata schemas provide a standardised structure and vocabulary for contextualising data and describing the features of a building.

Section 3 discusses prevailing approaches to metadata for data-driven smart buildings, the value they provide to building stakeholders, and how metadata is a key enabling technology for future digitalisation. Metadata schemas provide stakeholders with greater flexibility and value creation opportunities. It does this by providing a means of storing information, required by data-driven software applications, in a way that does not depend on the choice of vendor or protocol, architecture and composition of building, or choice of data-driven consumers and processes.

Data Platforms are a key tool that can either enable or frustrate achievement of the FAIR data principles. They provide means for distributing data to where it can be used, and they provide important data management capability for enabling data-driven smart building services.

Section 4 provides an overview of relevant functions and features of data platforms in the context of an overall data infrastructure software/hardware stack, and its associated ‘layers’. It describes prevailing technologies that platforms have adopted, and features that make data accessible for users. It describes how data platforms help to realise other aspects of Subtask A, including data interoperability. Data platform implementation is discussed, with the aim of providing general guidance on practical considerations including goal setting, evaluation and selection, onboarding, compliance and training.

Policy makers can help to reduce barriers to adoption of energy productivity software services by (i) providing industry with information resources on digitalisation, (ii) supporting relevant data standards and (iii) certifying buildings as ‘digital ready’ (where they can demonstrate that they have achieved requisite digitalisation attributes). Policy makers could also help to reduce data sharing risks by providing some relevant pre-competitive data stewardship resources (particularly for low/mid-tier buildings), and by providing more explicit clarification on what energy related data is and isn’t private. Potential may exist for regulating some data sources to be collected and shared.

Table of contents

Figures

Tables

Abbreviations

Abbreviations	Meaning
AI	Artificial Intelligence
API	Application Programming Interface
App	Application
ANSI	American National Standards Institute
ASHRAE	American Society of Heating Refrigeration and Airconditioning Engineers
BACnet	Building Automation and Control Networks
BIM	Building Information Modelling
BMS	Building Management System
BOT	Building Topology Ontology
BSN	Building Services Network
CSV	Comma Separated Values
DBO	Google Digital Buildings
DERMS	Distributed Energy Resource Management System
EMIS	Energy Management Information System
EMS	Energy Management System
FAIR	Findable, Accessible, Interoperable and Reusable
FTPS	File Transfer Protocol Secure
GDPR	General Data Protection Regulation
HVAC	Heating Ventilating and Air Conditioning
IaaS	Infrastructure as a Service
ICN	Integrated Communications Network
ICT	Information and Communication Technology
IEA	International Energy Agency
IFC	Industry Foundation Classes
I/O	Input/output
IoT	Internet of Things
IT	Information Technology
LAN	Local Area Network
MAC	Media Access Control
ML	Machine Learning
MQTT	Message Queuing Telemetry Transport
PaaS	Platform as a Service
RDF	Resource Description Framework
REC	Real Estate Core
SAREFBLDG	Smart Applications REFerence Ontology
SSN/SOSA	Semantic Sensor Network Ontology

1. Introduction

IEA EBC Annex81 ‘Data-Driven Smart Buildings” is an initiative that aims to accelerate the adoption of digitalisation technology as a tool for deploying energy productivity analytics and control solutions.

The International Energy Agency (2017) explained the concept of digitalisation as:

> “*the increasing interaction and convergence between the digital and physical worlds” where “the digital world has three fundamental elements*:
>
> - ***Data**: digital information.*
> - ***Analytics**: the use of data to produce useful information and insights.*
> - ***Connectivity**: exchange of data between humans, devices and machines (including machine-to-machine), through digital communications networks.*
>
> *The trend toward greater digitalisation is enabled by advances in all three of these areas; (i) increasing volumes of data thanks to the declining costs of sensors and data storage, (ii) rapid progress in advanced analytics and computing capabilities, and (iii) greater connectivity with faster and cheaper data transmission*”.

Digitalisation offers new opportunities for saving energy in buildings. The International Energy Agency (2017) found that digitalisation could cut energy use by about 10%, by using real-time data to improve operational efficiency.

Digitalisation fundamentally takes a data-driven approach to management and control of energy consuming equipment in buildings. That is, data is ingested and processed in a way that determines logical correlation/causation based on observed outcomes (rather than by forcing correlation/causation through theoretical physics-based computational models).

This data-driven approach is potentially highly scalable, for industry, because:

- It reduces the need for skilled practitioners to devote time to understanding the features and operation of a building and avoids manual coding of rules-based computational models,
- It can utilise powerful analytical tools (eg machine learning algorithms) that have already been adopted and proven in other industries, and
- It utilises IT infrastructure and methods that support automated processing of data and digital communication between machines/devices, and with user friendly interfaces for humans.

Of course, the data-driven approach (and associated benefits of digitalisation) is somewhat predicated on access to relevant data of sufficient quality to drive these automated processes. Data collection and management must be cost-effective, trustworthy, and consistent with obligations to manage privacy and commercial rights.

The Energy Efficiency Hub ‘Digitalisation Working Group’ (2022), and various other roadmaps, identify access to data, interoperability, and privacy as three important barriers preventing uptake of digitalisation in buildings.

Industry consultation in Australia (Trianni et al, 2022) identified that the task of implementing digital energy performance strategies can be divided into two steps (1) establishing IT infrastructure and data management services and then (2) deploying data-driven energy productivity applications.

Error! Reference source not found. illustrates this two-step journey and highlights the barriers that are encountered, predominantly in the first step of establishing IT infrastructure and data management services. **Error! Reference source not found.** further highlights the consequent need to establish some concept of (and guidance for achieving) ‘digital ready’. The attributes and features of ‘digital ready’ would ideally provide a stand-alone target to guide the first step of the digitalisation journey for a building.

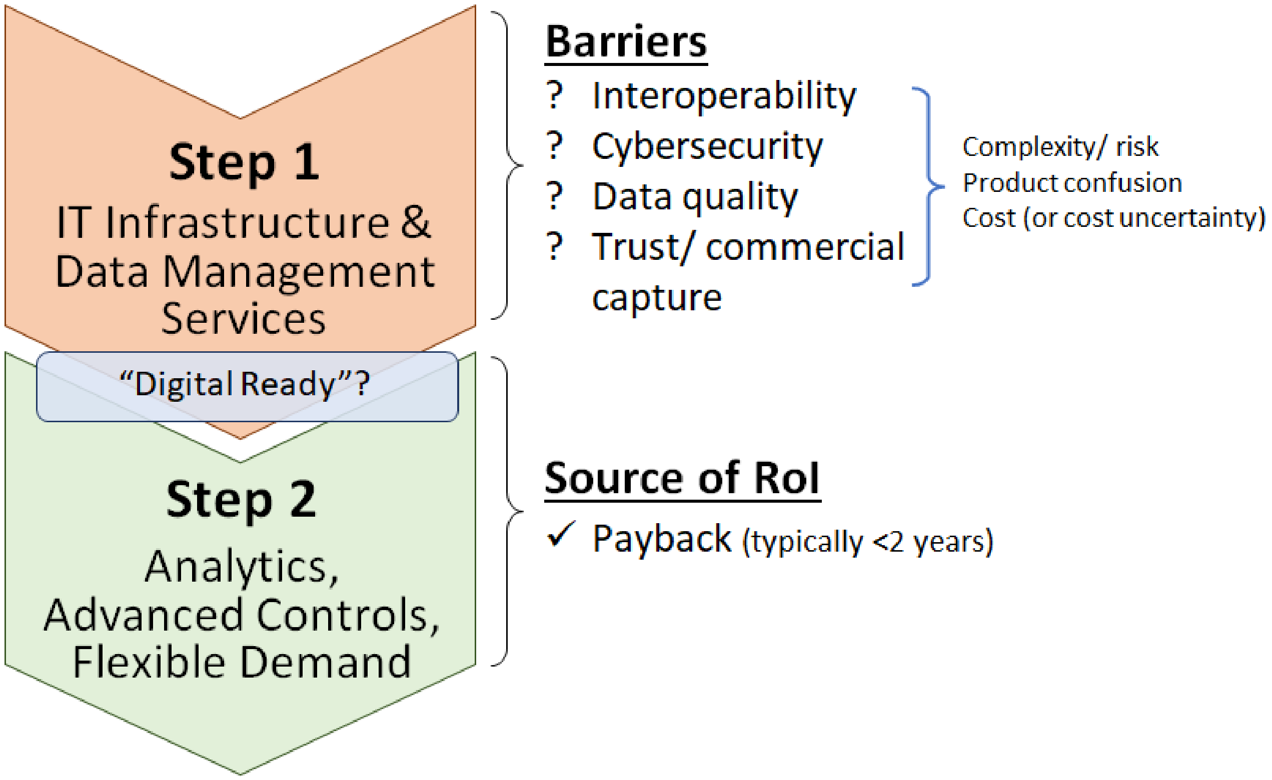


Figure 1.1 The two step journey for deployment of data-driven services (adapted from Trianni et al, 2022)

Subtask A of IEA EBC Annex 81 focuses on step 1 of the journey illustrated in **Error! Reference source not found.** – establishing the IT infrastructure and data management services that are needed to deploy data-driven software 'Applications'[1]. It provides guidance on many of the factors that should be considered for a building to achieve the proposed status of 'digital ready'.

This report summarises the work of Subtask A. The focus is on data infrastructure relating to electrical and mechanical services in buildings.

---

[1] 'Applications', in this context, are conceived as easy-to-configure and deploy software microservices, that are built on top of a common software infrastructure platform that manages input/output data. The software platform can be deployed on edge-computing devices or the cloud. This vision of Applications is somewhat analogous to the idea of 'Apps' that we use on personal mobile devices.

2. Making Data Shareable (data-sharing governance)

Subtask A prepared “A Data Sharing Guideline for Buildings and HVAC Systems” (White et al, 2023), with the aim of providing key concepts and language, relevant to data sharing. The focus was on data governance and data quality issues that may need to be addressed in energy productivity software applications, relating to electrical and mechanical-services in the non-residential buildings sector.

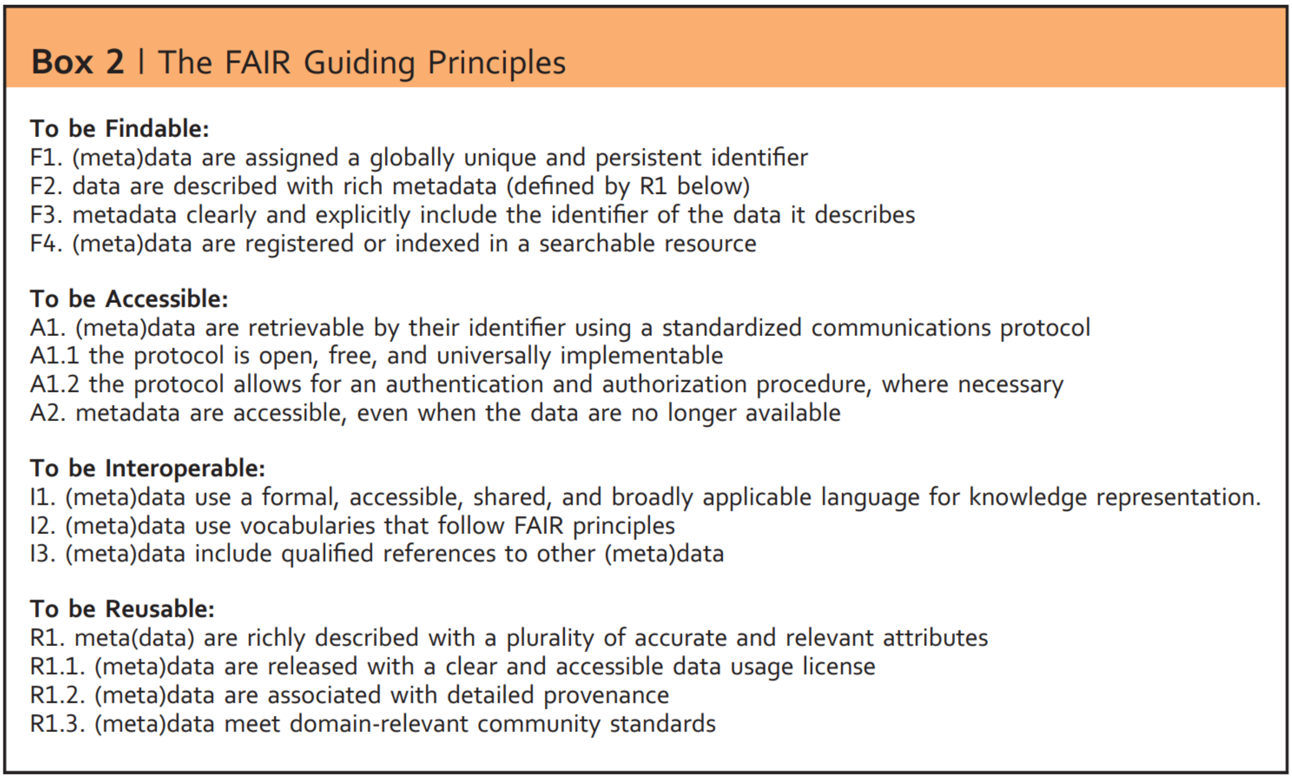
Some of the data governance issues that need to be considered, include management of commercial IP rights, ethics and privacy regulations. These issues create risks that need to be managed. The resulting risk-controls may constrain the type and quality of data that can be shared, and the situations where it is permissible to share data.

The ability to share data can be further constrained through the choices made when selecting a provider of data sharing platform services. These limitations may result from the provider’s technology and/or from the commercial interests and governance arrangements of the provider’s organisation.

These issues, and considerations for minimising their impact, are discussed in this section.

2.1 Interoperability and the FAIR Data Principles

The diversity and value of data-driven applications is generally enhanced when available data complies with the FAIR data principles. Putting it another way, data is more useful when it is Findable, Accessible, Interoperable and Re-useable (FAIR). These FAIR data principles are summarised in **Error! Reference source not found.Error! Reference source not found.**.



Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Figure 2.1 The FAIR guiding principles (Source: Wilkinson et al, 2016)

While various sensible constraints may make this difficult to achieve in all circumstances, the FAIR data principles should be the building-owner’s aspirational objective for best practice data collection and management.

2.1.1 Data Standards for Interoperability and Findability

Virtually all studies on barriers to digitalisation point out the significance of interoperability issues and the need for data standards to help overcome them.

Irrespective of commercial requirements, the technology used for collecting and managing data should be capable of achieving interoperability.

Device level (communications) interoperability barriers occur when the communication protocols used by one manufacturer’s devices are proprietary and therefore unable to talk with the devices from other manufacturers. This makes it difficult and expensive to integrate hardware components from different manufacturers into a coherent operating system for the building. This can also lead to vendor lock-in and high on-going service costs.

BACnet was introduced as an open communications protocol to address this issue. BACnet is both an international (ISO) and ANSI standard. It is maintained by ASHRAE. Unfortunately, implementation of BACnet is not always uniform, and interoperability issues can still exist. Care should also be taken with cyber-security as BACnet does not have native encryption.

Analytics level (informational/semantic) interoperability barriers occur when the data being collected from devices comes without any contextual information. Important contextual information includes the source and type of data, and the interrelationships between the data source and features/objects in the building. This contextual information is generally required to implement data analytics rules, data-driven algorithms and AI more generally (eg fault detection and diagnosis, model predictive control).

O’Reilly (2020) surveyed 1,900 people working in the field of Artificial Intelligence, to get their perspectives on the data quality issues they face. The survey highlighted critical issues relating to (i) poorly labelled or unlabelled data (ii) inconsistent data and (iii) disorganised data stores and lack of metadata.

Metadata schemas provide a standardised structure and vocabulary for contextualising data and describing the features of a building (see Section 3). Metadata can be used to understand the relationships between data and associated objects/entities (ie what belongs to what). They can also provide underpinning structure for software tools that manage storage (eg triple stores) and retrieve data (eg query languages) for data-driven applications.

Utilising metadata provides relevant information for automating searches for relevant data, somewhat analogous to a cataloguing system for libraries. In this way, meta data is core to data ‘findability’ and data ‘reusability’.

2.1.2 Data Platforms for Data Accessibility and Reuseability

Data can be exchanged locally, between devices on-premises, using various LAN technology options. However, accessibility for potential users is vastly improved by using cloud technology. The cloud enables a wider range of both on and off-premises data sources to be analysed together. And it enables information to be efficiently distributed to relevant people via remote PC and mobile devices.

Data platforms provide the cloud software layer for distributing data to where it is needed. The data platform is (for digitalisation of buildings) analogous, at least partly, to what a computer operating-system is for a personal computer; guiding computational workflows and exchanging data to/from storage. Importantly, the data platform consolidates data from disparate sources in one location and uses relevant data standards to provide a coherent, harmonised structure for the data.

In different contexts/applications the data platform could be called an IoT platform, an Energy Management Information System (EMIS) or Distributed Energy Resource Management System (DERMS).

Desirable technical features and attributes of a data platform are discussed in Section 4. Of importance in this Section, is that users will typically need to comply with the data capture and data management standards employed by the platform. This may impact on the degree of interoperability that can be achieved.

Furthermore, the commercial arrangement and institutional structure of the data platform provider may significantly impact on access to data and data-driven services. Ideally the data platform provides a means for collecting data once, only, and then reusing the data multiple times by sharing the data, via APIs, with an ecosystem of other applications and service providers (rather than being locked into single vendor solutions). These issues are discussed in Section 2.3.

2.2 Data Rights and Data Risk Management

Just because data can be collected, does not mean that it can be shared. Care should be taken to ensure that permission has been granted prior to using data.

2.2.1 Licence to use commercial data

Often this is obtained through a license from the data owner. However, the concept of the data owner is not always obvious.

It is common for people and businesses to refer to data as if it is something that can be owned. For example, individuals and businesses commonly refer to ‘my data’ or ‘our data’. However, in general, there are no property rights in individual points of data. Therefore, data cannot be owned.

Instead, deliberate collection and curation of data as a data-set can be viewed as a ‘creative work’ subject to copyright law. Various actors may have some claim to helping ‘make’ a data-set, with the resulting possibility of having some rights over the intellectual property. The maker will probably be the entity who made the commercial decision to collect the data and made the commercial investment in carrying out the collection and curation of the data-set.

Given that the building owner ultimately pays for all the services, and needs the ability to competitively source providers at regular intervals, it is generally assumed that data should belong to the building owner. This may not be the case, and (irrespective of legal rights) data has a practical tendency to find its way to the service provider and end up inaccessible to the building owner. Indeed, the uplift of data to access-controlled external IT systems can be used as part of a service provider’s business model for ensuring that the building owner retains their services.

2.2.2 Basis for using personal data and data controls

The General Data Protection Regulation (GDPR) regulates data protection and privacy in Europe. It has also been adopted, in full or part, in various other jurisdictions. The GDPR defines personal data as any information that relates to an identified or identifiable living individual.

Most energy productivity applications will use base-building data sources that relate to the operation of the whole building. As this data relates to the aggregate (rather than individual) needs of occupants, it is typically not considered to be personal data.

However, some applications could potentially utilise occupant data which is personal. For example, occupant movement data can be used as an input to drive allocation of energy consuming HVAC services. Occupants in the building may also wish to interact with building services using Apps on their mobile phone (eg to improve thermal comfort conditions, make meeting room bookings, or access other resources etc), which also has the potential to lead to the collection of personal data (eg email addresses, MAC addresses etc).

When using personal data, the GDPR requires that there be (i) a legal “basis” for processing the data, (ii) adherence to general data processing rules of transparency and fairness, and (iii) appropriate technical and organisational safeguards in place to ensure the security of the personal data. The GDPR takes a risk-based approach, where companies/organisations processing personal data are encouraged to implement protective measures corresponding to the level of risk of their data processing activities.

The six legal bases for processing personal data that are recognised by the GDPR (Article 6(1)) include:

1. The data subject has given consent
2. Performance of a contract, to which the data subject is party
3. Compliance with a legal obligation
4. Protecting the 'vital interests' of the data subject
5. Public interest or acting under official public authority
6. 'Legitimate interests'

The Five Safes Framework can be used to identify protective measures for safeguarding personal data. The data sharing principles of this framework are illustrated in Figure 2.2 below

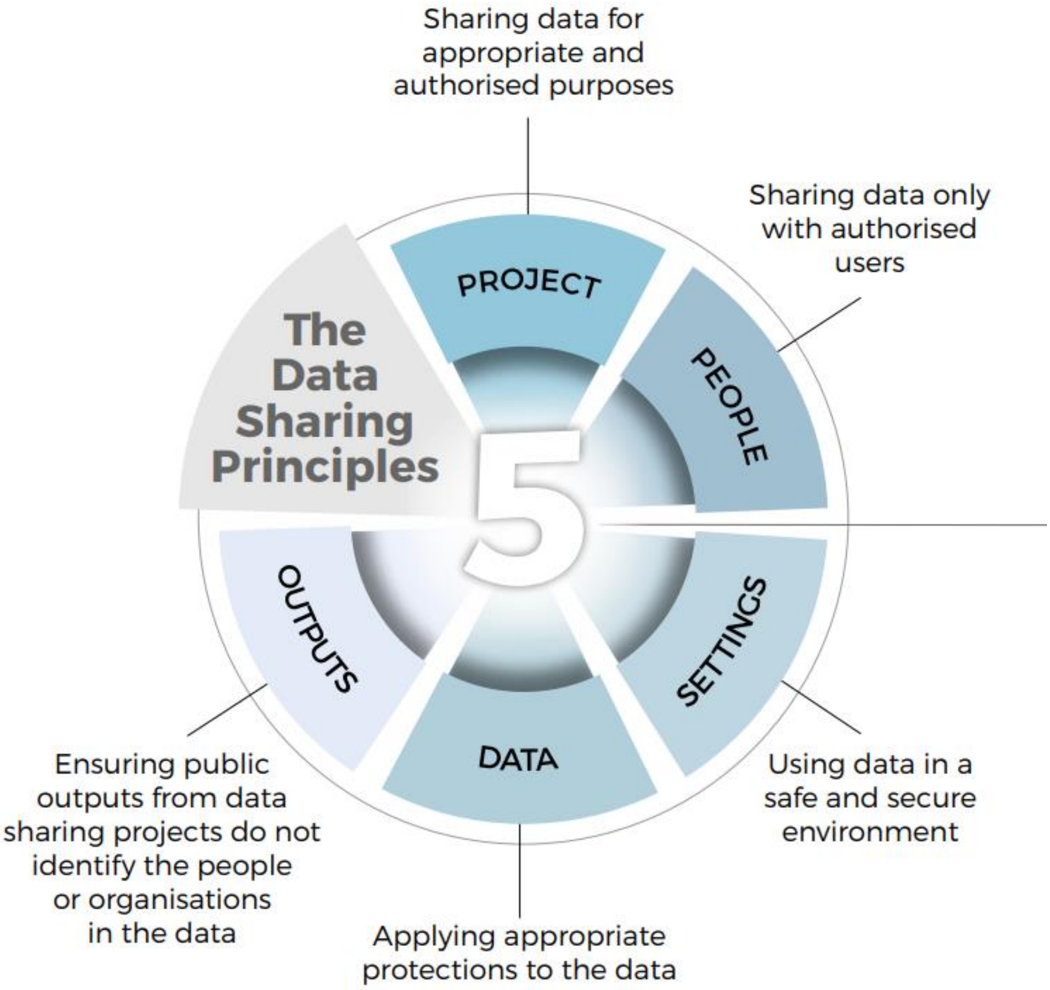


Figure 2.2 Principles for Sharing Data Safely (Source: Commonwealth of Australia, 2019)

Each of the Principles can be considered as a focus area, where the stringency of the required control mechanisms (risk management choices) can be adjusted to achieve an appropriate balance between openness and the level of sensitivity of the data being shared.

2.3 Data Platform Selection Strategy

2.3.1 Separating the data platform from data-driven applications

Data platforms provide an intermediating role between data-sets and data services, as illustrated in **Error! Reference source not found.**.

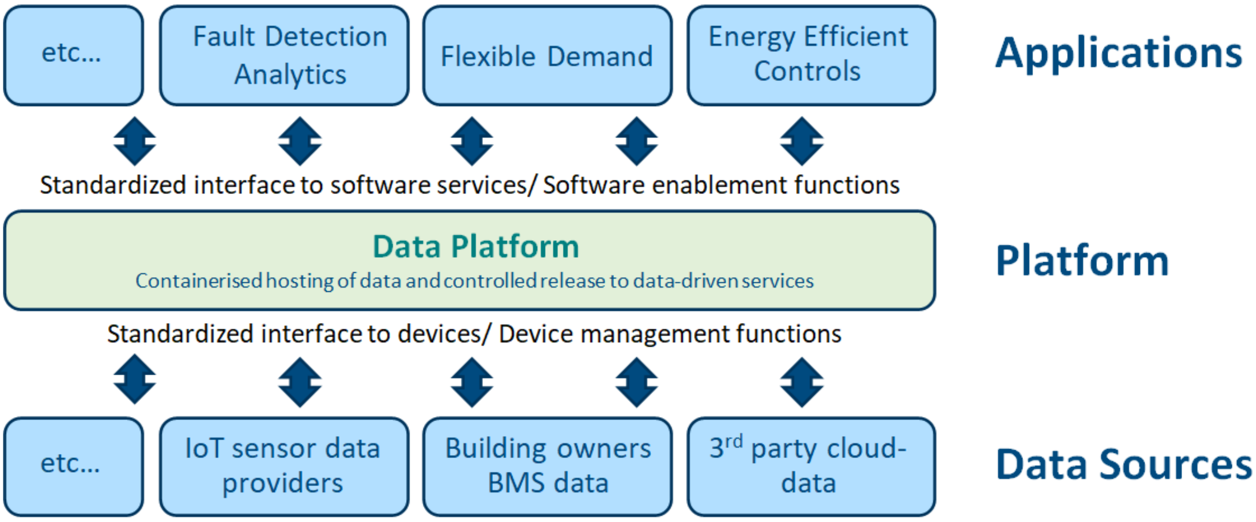


Figure 2.3 The role of a data platform to connect and enable software analytics services

The assembly of consistent yet individually curated datasets, on a platform, enables applications software to be cost effectively developed and deployed. Without this enabling aggregation, individual software application developers would need to engage with each individual data owner, potentially having to develop tailored applications based on the structure and semantics of each source of data.

Consequently, the platform plays a key role in brokering the relationship between data providers and software application developers who use the data to deliver value back to the providers. This brokerage entails:

- Setting consistent standards for data uploaded or deposited to the platform
- Qualifying/validating/authorising the applications that interact with the data on the platform

In this way, data platforms can (at best) facilitate desirable access to data but (at worst) can frustrate access to data and create a position of market power for platform owners.

Separating out the data platform role from individual application services, would be a useful step forward towards enabling the FAIR data management principles, supporting innovation, and driving value in the market. Unfortunately, the distinct value of the data platform role has not yet been widely recognized in the industry.

As a result, the data platform role is typically bundled with specific application services as a single vertically integrated market offering. This business model continues the industry's tendency to form siloes and to enshrine interoperability (data access) barriers. It can lead to multiple platforms being installed in a building (each for its own application), with reduced efficacy and with duplication of costs and data management risks.

2.3.2 Strategy for acquiring data platform services

Acquisition of a data platform is a significant strategic decision for the building owner. Key considerations for the building owner include

- Data Sovereignty: By collecting data and storing data on behalf of building owners, a data platform could potentially gain significant visibility of confidential or sensitive data, and/or have capability to use the data to provide services to others. Indeed, the business model of some IT platforms, in some other industry sectors, is to provide services for free on the condition that the collected data can be harvested to generate separate revenue streams for the platform owner.

- Vendor lock in: Once data is stored on an external provider's data platform, it may be difficult to recover that data if there is reason to change provider. This potential to lose historical data creates a disincentive to change provider, stifling competition and innovation. Similarly so-called 'network effects' have the potential to create natural monopolies for platform providers. The more a platform is adopted by users the more useful it becomes and the harder it is for others to compete.

Options for building owners include

1. Building owner establishes a private data warehouse for capturing, storing and managing data on the operation of their buildings. Data is stored on the building owner's servers and separately distributed to service providers by the building owner:
    - ➯ In this scenario, the cost of developing the data platform is high, but the building owner maintains full control and inhouse capability to enable more tailored/bespoke services.
2. Building owner utilizes a 3rd party data platform service to collect and manage their data, as a separate service to any analytics providers that the building owner may subsequently use:
    - ➯ In this scenario, the cost of developing the data platform is reduced compared with developing a bespoke data-warehouse. To the extent that the data platform is independent from analytics services, and provides the building owner with tools for accessing and self-managing their data, the data platform can provide access to a wide variety of 3rd party services.
3. Building owner obtains data platform services bundled up inside relevant software services.
    - ➯ In this scenario, the building owner does not explicitly pay for a data platform. Instead, the platform is installed (somewhat inadvertantly), creating limited access to data (loss of data sovereignty) and inability to reuse data for other services (vendor lock in).

While the property industry is generally behind the broader IT industry in considering many of these issues, there are various highly capable companies offering data platform services via a Platform-as-a-Service (PaaS) business model.

More broadly and philosophically (for policy makers), leading thinkers have identified data as being a key part of 21st century national infrastructure; akin to the roads, telecommunications and other infrastructures created in previous centuries. Considering data in a similar way, the Open Data Institute has proposed the need for "Data Institutions" as shared resources for stewarding data on behalf of users. A data institution for the property industry could address issues discussed in this Section, in a way that acknowledges the considerable digital skills gap present in the building-services industry.

2.4 Summary of Recommendations

This Section 2 focuses on data governance and data quality issues that need to be considered when deploying energy productivity software applications, relating to electrical and mechanical-services in the non-residential buildings sector.

Key conclusions and recommendations for building owners seeking to enable these services include:

- **Aim for your data to be Findable, Accessible, Interoperable and Reusable (FAIR)**. To do this:
    - Requirements should be placed on providers to use interoperability standards for both (i) device communications interoperability (eg BACnet) and (ii) mapping of data and data relationships (informational/semantic interoperability) (eg Brick Schema, Project Haystack).

  - A suitably independent cloud-based data platform should be deployed to make data easy to access for potential internal and external data users. The platform should utilise relevant interoperability standards and enable safeguard mechanisms for protection of data.

- **Permission needs to be obtained to use data**. This could require (i) obtaining a licence to use commercial data from the relevant dataset owner and/or (ii) identifying a relevant legal 'basis' to use personal data. Where personal data is used, control mechanisms should be put in place to safeguard the personal data, commensurate with the sensitivity of that data.
- Ideally, **procurement of data platform infrastructure should be separated from procurement of data-driven software applications**. This will help the building owner to achieve greater flexibility and data sovereignty, and avoid commercial vendor lock-in.

Additional recommendations for policy makers are:

- **Provide guidance and ideally certification for buildings that achieve 'digital ready capability'**. This guidance and certification of attributes and features relating to digital infrastructure can help to overcome awareness, confusion and skills gaps in the industry. It could potentially also provide a demonstrable mark of quality that could add to the asset value of buildings.
- **Consider supporting an independent 'data institution' to steward data resources on behalf of the property industry**. This would enable lower-tier buildings, who lack relevant digital skills, to participate in the digital economy (on a voluntary basis).

3. Making Data Meaningful (structuring data)

As digitalisation increases the availability, heterogeneity and scale of data in buildings, it is necessary to think about how that data will be identified, organized, and consumed by downstream data-driven applications. This can be accomplished through the use of metadata – “data about data” – which encodes salient properties of data, including their provenance (how the data was produced and managed) and their context (where the data comes from and how it relates to the building).

Metadata provides meaning to data by making these attributes explicitly available for consumption. This avoids the need for developers and their software applications to “memorise” the names of data sources and the details of buildings in order to make use of data. Instead, these data-consumers can query the metadata for a building to discover the desired data-sources automatically and dynamically. Metadata is a key facet of the FAIR data principles explored in the prior section.

Metadata’s essential role in enabling data-driven smart buildings has inspired investment and research by academic, commercial, and federal organisations. Subtask A prepared a “Survey of metadata schemas for data-driven smart buildings” (Fierro and Pauwels, 2023) with the aim of exploring the design space of metadata solutions for smart buildings, and calling out the challenges and benefits inherent to different approaches.

In this section, we discuss prevailing approaches to metadata for data-driven smart buildings, the value they provide to building stakeholders, and how metadata is a key enabling technology for future digitalisation.

3.1 Meta-data schemas and their value

Metadata schemas are organisational structures for metadata. Among other features, they define:

a) how data sources should be labelled or catalogued.
b) how the associations between data sources should be represented.
c) what attributes and properties can or must be attached to data sources.
d) how data sources relate to descriptions of the building and its assets.
e) the engineering units and enumeration definitions for the data itself.

Metadata schemas standardise what information should be captured, and in what format. From a technical perspective, metadata schemas provide value to stakeholders by providing a means of storing information useful for data-driven software applications in a way that does not depend on the choice of vendor or protocol, architecture and composition of building, or choice of data-driven consumers and processes. This technical reason has several useful outcomes for decision-makers, which can encourage adoption of metadata (in a way that enables digital building processes), including:

a) ability to (semi-)automate installation, configuration, and operation of building software applications, resulting in lower ‘soft costs’.
b) ability to automate and verify delivery of digital controls and analytics.
c) ability to provide increased transparency and visibility of building operations.

While metadata ***schemas*** provide the general framework for organising information about a given building, the schema is not, in itself, the information about any given building. The actual metadata about an individual building is contained in a metadata ***model***.

Models are digital representations of information about a specific building. They must be stored in a way that allows applications and other software processes to access and manipulate them. Models can store references to the equipment and other assets in the buildings, describe how these resources connect to each other and serve the building, and identify where the available digital input-output “points” are, e.g. in a building management system or other programmatic interface.

The difference between the schema and the model, in the example case of a sensor in a building, is as follows. The metadata ***schema*** will define the possible types of sensors and provide the options for what properties they may have (e.g. units of measure), as well as potential relationships of sensors to each other and parts of the buildings. Whereas, the metadata ***model*** contains the specific instances of sensors, for a given building. The sensors in the metadata model will be selected from one of the schema’s defined types and they will contain the schema’s defined properties relevant to those specific sensors. The metadata model does not contain the actual sensor readings.

Figure 3.1 further illustrates how these definitions combine, through a schematic representation of a building, its common subsystems (HVAC, lighting, electrical, and plumbing), and the digital interface to the assets within those subsystems. The metadata model is a logically separate entity, typically stored in a database (eg a graph database, or relational database), which contains references to the components of those subsystems and their respective data-sources. Software applications (Apps) execute queries against the metadata model to discover and retrieve important details and configurations necessary for execution.

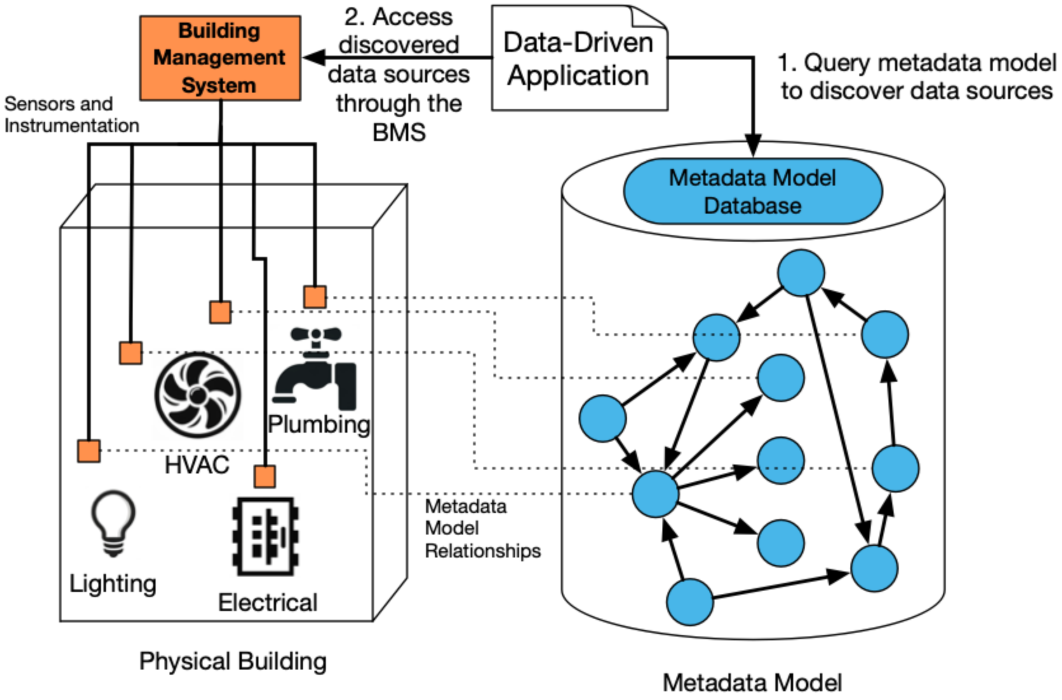


Figure 3.1 Representation of how a metadata model relates to and models the building, its subsystems, and its data sources

Ontologies are specific kinds of schemas which go beyond defining possible terms, attributes, and relationships by imposing additional rules and axioms that can ensure the consistent communication of *semantic* information. Put another way, schemas encode how metadata can be expressed in some data structure; ontologies additionally ensure that such metadata has human- and machine-interpretable meaning (i.e. it has *semantics*). Ontologies support an important operation called *validation*. A validation process takes a metadata model as an input and ensures that it obeys all of the rules, constraints, axioms and other requirements defined by the ontology.

Ontologies are commonly, but not necessarily, communicated as graphs using the Resource Description Framework (RDF) W3C standard. RDF ontologies can be expressed in one of several languages, including OWL (Web Ontology Language) and SHACL (SHApes Constraint Language). Descriptions of these language and their capabilities are beyond the scope of this document; we refer interested readers to the online documentation of those respective standards. For simplicity, we refer to both ontologies and schemas collectively as “metadata schemas” throughout this document.

3.2 Comparing meta-data schemas

Different approaches have been taken to the development of metadata schemas for data-driven smart buildings, exhibiting different preferences and the various needs of different use-cases. We apply our discussion of metadata schemas, and their respective value, to solutions for the operation and management stage of the building lifecycle. Schemas that are used solely or predominantly for the design and construction phase of the building are excluded.

When deciding to adopt a metadata schema, we recommend considering the following characteristics of the different schemas available:

1. **Structure of the models that will be created**: The structure of the model determines what questions can be answered through queries against the model. For example, while one can order and filter tabular data by the characteristics contained within, it is more difficult to inquire about relationships between the entities in a tabular model. Point naming schemes are simple to store, but they only support very basic string-based lookups. In contrast, graph-based models support richer queries that can relate multiple entities together in a manner more expressive than tabular models. The user must review the benefits, concerns, drawbacks, and advantages of available options relating to relational models, graph models and point labels. This can be done drawing from existing literature or experiences where appropriate. The choice of model should be informed by the questions that the data-consumers, the software applications, and other users of the model need to answer for their operation.
2. **Vocabulary organization and completeness and strictness/rigor**: In this context, “vocabulary” is the terminology used for describing different objects, relationships, and concepts relevant to a building. How are concepts organised and defined in the model? Are they generic or specific? Are concepts defined nominally (through labels) or structurally (through properties)? A detailed vocabulary helps to drive consistent representations but may reduce freedom to describe non-standard scenarios and may be onerous to implement.
3. **Alignment with other metadata schemas**: It may be useful to use more than one schema to cover different use cases and different aspects of the building. This can take advantage of the respective strengths of different schemas (rather than seeking to find a “universal schema”). Consequently, it is important to consider the ways in which metadata schemas may align with one another.

In addition to the technical characteristics of the schema, above, there are a range of implementation considerations that may impact on the viability of a given schema, including:

1. **Impact on smart building software architecture**: How is metadata stored and accessed by software processes to support data-driven buildings? For example, do these processes access a metadata model through a database service, or through another method? How does incorporating a specific metadata schema influence the development, deployment, and management of data-driven processes in the building?
2. **Required tooling / software support / expertise**: do metadata models need proprietary software? What features are required by supporting data platforms? What does this mean for its deployment for a data-driven building? Is the schema commercially supported? How can models created using the schema be used by the BMS, and relevant data platforms?
3. **Creation / bootstrapping /maintenance**: Engineering time is generally required to create a model for a building. It is important to consider how models get built and maintained. Can model development and management be done in an automated way, or does it require manual curation? Who owns the metadata model? When changes are made in the building, how does the model get updated and by whom?

The following metadata schemas were surveyed in the Annex81 report “Survey of metadata schemas for data-driven smart buildings” (Fierro and Pauwels, 2023): Project Haystack, Brick, RealEstateCore (REC), Building Topology Ontology (BOT), Smart Applications REFerence Ontology (SAREFBLDG), Semantic Sensor Network Ontology (SSN/SOSA), and Google Digital Buildings (DBO).

These schemas differ primarily in how data models are created and how they support data processing and data discovery in smart buildings (Table 3.1).

Table 3.1 Overview of metadata schemas and their model structure

Metadata Schema	Naming Convention	Tags	Relational	Graph	RDF Ontology
Haystack	No	Yes	No	Yes	No
Brick	No	Yes	No	Yes	Yes
RealEstateCore	No	No	No	Yes	Yes
BOT	No	No	No	Yes	Yes
SAREF4BLDG	No	No	No	Yes	Yes
SSN/SOSA	No	No	No	Yes	Yes
Google Digital Buildings	Yes	No	No	Yes	Yes[2]

Several schemas — Project Haystack, Brick, RealEstateCore and Google Digital Buildings — deal directly with the management and organisation of telemetry information in the building. Project Haystack and Google Digital Buildings explicitly define the format of the data and how it is accessed. Brick and RealEstateCore define more generic structures which can be incorporated into a variety of APIs and software platforms.

Other schemas — BOT and SAREF4BLDG — provide more contextual information about the building which can assist software applications to find relevant data. They typically focus more on asset management rather than telemetric data. As a result, these metadata schemas are much closer to the Architecture and Engineering Construction domain and the processing of BIM information. Conversely, SSN/SOSA provides all needed mechanisms to represent sensor data and actuator data on a large and detailed scale. It leaves the representation of actual building data to other ontologies like Brick, BOT, and SAREF.

Among the schemas, there are a variety of perspectives of what in the building should be modelled, and there are differences in the consistency and specificity of those perspectives. Brick and Project Haystack model many common building subsystems including HVAC, lighting, and electrical systems. Project Haystack’s tagging model affords a great deal of flexibility in describing these systems at the cost of consistency across Haystack models. In contrast, Brick prescribes more of the model structure in exchange for a consistent modelling and querying experience for the consumer of the model. Google Digital Buildings focuses primarily on collections of data coming out of the building, rather than the topology and composition of the building subsystems. RealEstateCore is similar to Brick, but focuses more on the property management aspects and includes a shallower hierarchy of equipment and data source types. Finally, BOT tends to focus much more on asset management and description of the building itself, with much less focus on HVAC systems or their telemetric data logs.

Generally, we can recognize a spectrum of schemas, from very flexible approaches (left in Figure 3.2) towards more rigid and formally defined approaches (right in Figure 3.2).

---

[2] The Google Digital Buildings metadata schema defines an OWL ontology export but it is not the intended mode of interaction, and does not support all features of the metadata schema

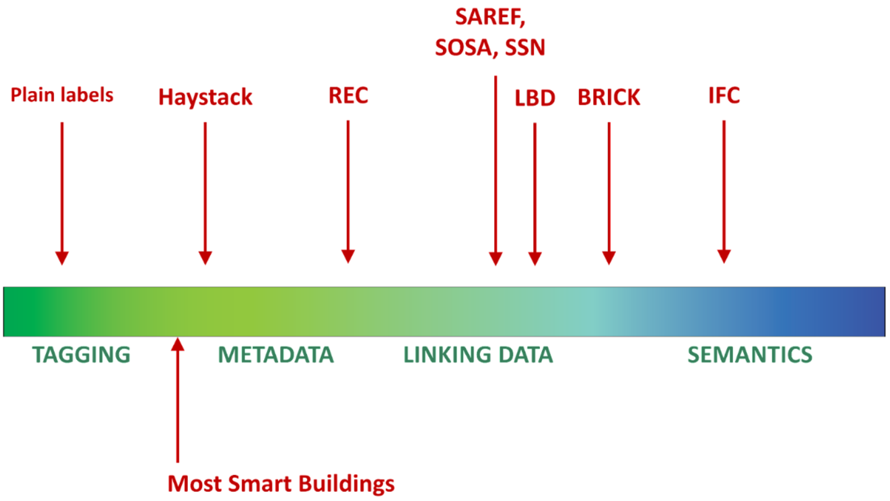


Figure 3.2 A spectrum of building data representation from more flexible and ad-hoc (leftmost) to more formal and semantically defined (rightmost); plus the estimated location on that scale for several existing metadata schemas

Despite the diversity of approaches and stakeholders for each metadata schema, there is a growing desire to ensure unity and alignment between the various groups. This can enable more than one schema to be used to take advantage of respective strengths in different use cases (rather than seeking to find a “universal schema”). We predict, hope, and recommend that future editions of metadata schemas will focus more on complementing each other through reductions in scope, rather than expanding the modelling scope to try to compete on all the different perspectives of data-driven buildings.

We also see RDF-based metadata schemas emerging as the dominant modelling approach. These demonstrate the highest degrees of interoperability and reusability compared to other proprietary models. New tools will emerge that raise the level of abstraction for interacting with RDF-based metadata schemas, ultimately democratising the use of rich metadata in data-driven smart buildings.

3.3 Applications of meta-data schemas and sufficiency

In this section, we discuss the broad categories of applications and use cases that can be enabled, (semi-)automated, or made cheaper by the integration of metadata schemas. This discussion provides a more technical perspective on the value proposition of adopting metadata. We begin with an explanation of the principle of *semantic sufficiency* (Fierro, et al 2022) which is a key part of making metadata-based solutions useful and practical.

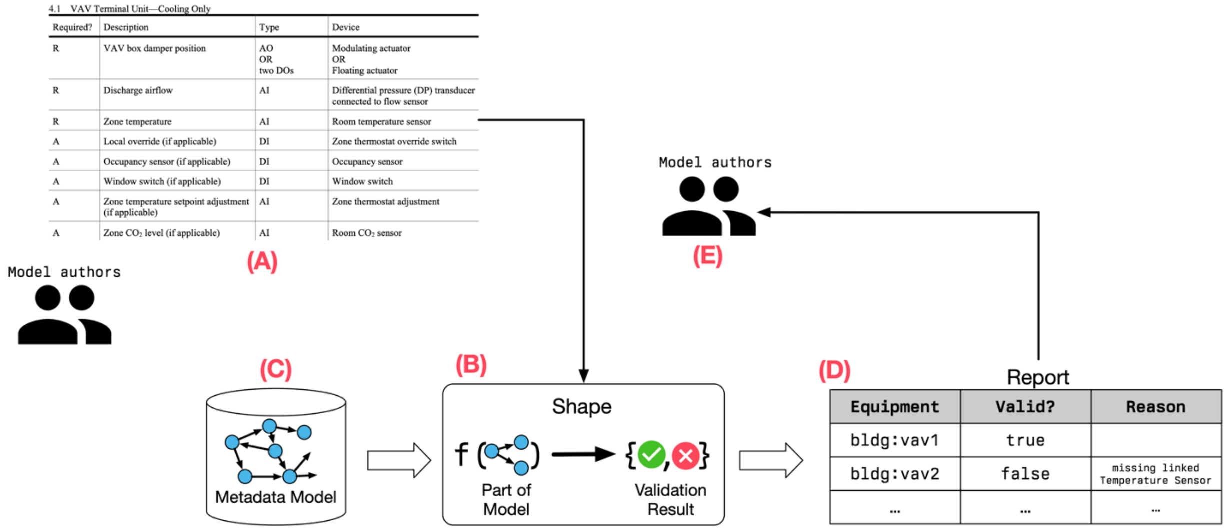
3.3.1 Semantic Sufficiency

One of the key challenges in adopting semantic metadata is how to create and validate the metadata model for a building (*validation* is the process by which a metadata model is checked against the rules, axioms, constraints, and other requirements defined by an ontology).

Because fully automated metadata model creation has yet to be developed, most deployments still require some manual effort to create the metadata model from existing sources of data (see Fierro and Pauwels, 2023 for more details). Without some guiding principle for what parts of the model to prioritize, model authors are left to either model everything (which can be cost-prohibitive) or guess at what metadata will be required by downstream analyses, controls, and other data consuming processes. The lack of such a principle means it is possible to create models which are valid with respect to the ontology, but do not actually contain enough metadata to support applications.

To address the lack of such a principle, Fierro et. al. (2022) introduced *semantic sufficiency*, a practical approach to creating and evaluating “complete” semantic metadata models. Semantic sufficiency pulls application-level requirements on metadata into the validation process. Specifically, the semantic sufficiency principle holds that a model is “complete” when it contains sufficient metadata to support a desired suite of applications. While straightforward at first glance, this principle exposes two important features enabled by semantic metadata. First, semantic metadata makes it possible to precisely express the configuration information required by a software application using the metadata schema. Second, semantic metadata makes it possible to *verify* that a given model of a building contains the correct (or “sufficient”) metadata to support a desired suite of applications.

Consider the workflow illustrated in Figure 3.3. Model authors are given a point list (A), usually from some standard or other control specification. The point list contains the informal names of the sensors, actuators, and other inputs and outputs required by some data-driven application. In Figure 3.3, the point list given is from the ASHRAE Guideline 36 document for high-performance sequences of operations for common HVAC systems. These point lists can be expressed as *shapes*, which are functions that validate part of a model (e.g., a particular piece of equipment) against the metadata requirements for a given application (B). Shapes allow metadata models (C) to be *validated* against a family of application specifications. The output of validation is a report (D) which tells the model authors (E) what metadata needs to be added or corrected to make the model *semantically sufficient*. The model is semantically sufficient if it contains all of the metadata required for relevant metadata-aware software applications to configure themselves and execute.



4.1 VAV Terminal Unit—Cooling Only

Required?	Description	Type	Device
R	VAV box damper position	AO OR two DOs	Modulating actuator OR Floating actuator
R	Discharge airflow	AI	Differential pressure (DP) transducer connected to flow sensor
R	Zone temperature	AI	Room temperature sensor
A	Local override (if applicable)	DI	Zone thermostat override switch
A	Occupancy sensor (if applicable)	DI	Occupancy sensor
A	Window switch (if applicable)	DI	Window switch
A	Zone temperature setpoint adjustment (if applicable)	AI	Zone thermostat adjustment
A	Zone CO_2 level (if applicable)	AI	Room CO_2 sensor

Report

Equipment	Valid?	Reason
bldg:vav1	true	
bldg:vav2	false	missing linked Temperature Sensor
...	...	...

Figure 3.3 Illustration of how semantic sufficiency guides the creation of metadata models

3.3.2 Metadata-Driven Cost Transparency

Currently, this semantic sufficiency information is almost non-existent, or at least not explicitly provided. However, in future we would expect data requirements lists to be provided by software application providers, so that building owners and their contractors can ascertain what information is missing for successful deployment, before purchasing a service.

This is an important step for the future of the data-driven software services industry.

Returning to the example points list in Figure 3.3: This table contains exactly which sensors, actuators, and other I/O points must be made available in a building for this control sequence to be run. This points-list allows stakeholders to determine the business case for adopting the application. That is, by reading the metadata requirements of an application, the building owner can determine the capital cost of buying the required sensors and equipment (if not already there), and they can determine the operating cost for making those data sources available in a building management system or other programming interface. These costs could then be compared with the expected financial benefits of adopting the software application.

For example, if a given fault detection rule claims to reduce energy consumption by 10%, a stakeholder can do the cost benefit analysis to determine if the financial savings of that energy reduction merit the investment in the equipment and data sources necessary to deliver that application. These types of use cases for semantic metadata are still emerging, but offer a compelling new direction for data-driven smart buildings.

4.Making Data Accessible (data-sharing platform deployment)

Section 2 highlights the role of Data Platforms as the means for distributing data to where it can be used, and its importance as a tool for enabling data-driven smart building services. Section 2.3 particularly focusses on governance aspects of data and data platforms, and how that can either enable or frustrate achievement of the FAIR data principles.

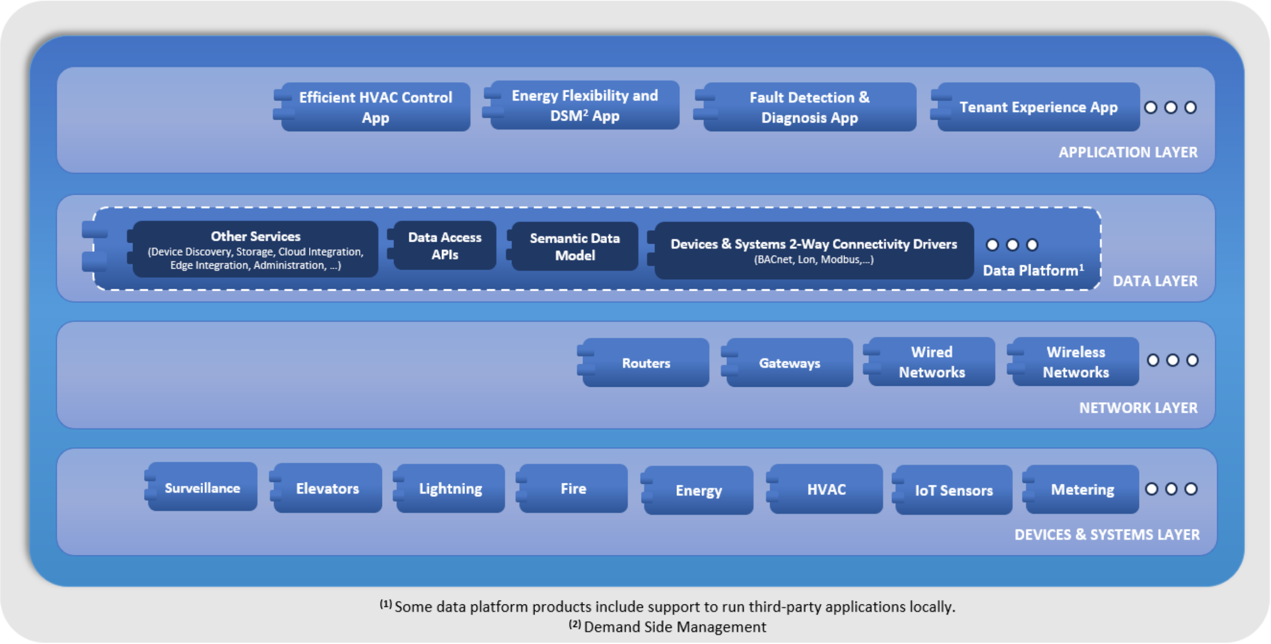
The building owner is encouraged to insist that common interoperability standards are used by its technology providers and, in particular, to utilise a meta-data schema (Section 3) to ensure that collected data has relevant context and meaning. These standards are all instantiated through the data platform.

This section focuses on the technology of data platforms, the platform features that make data accessible for users, and how the recommendations of Sections 2 and 3 are realised. It is assumed that data collection is in near real-time, to enable dynamic building performance optimisation.

More detailed information relating to this Section is available at Gomez-Garcia (2024).

4.1 Data Platforms in a Building's Digital Infrastructure Stack

The generalised digital infrastructure that would be expected for implementing data-driven smart-building solutions, is illustrated in Figure 4.1.



(1) Some data platform products include support to run third-party applications locally.
(2) Demand Side Management

Figure 4.1 The data layer in the digital infrastructure stack

In the '**device layer**' relevant data sources, from all the equipment and sensors in the building, communicate operational data to an on-premises controller (eg building management system (BMS), energy management system (EMS) etc) or data acquisition server, to enable staff to view, monitor and optimise building operations. This local communication is done via local wired or wireless protocols.

In the '**network layer**' data is collected on a server and/or relevant gateway/hub devices and transmitted to a central cloud data platform (the data layer). Communication can be via a range of wide area communications technologies including fibre, cable or wireless. This backhaul communications capability from the building may be dedicated to (i) the device (eg in IoT applications), (ii) to the broader suite of building services (dedicated building services network (BSN)), or (iii) even integrated with other communications services (such as internet, telephone, TV), through an integrated communications network (ICN). Much of the cyber security requirements, of smart buildings, are dealt with in the network layer.

Irrespective of the data source and network layer technology, the role of the '**data layer**' is to (i) consolidate the data in common formats, (ii) provide standardised interfaces to data sources and application software, (iii) provide structured data storage and (iv) make the data accessible to each of the deployed value-adding software services in the '**application layer**'. The data layer reduces the cost and complexity of deploying software 'Applications'. It may also provide greater choice of services, from third party Application providers.

The role of the data layer is delivered by the data platform; a common software foundation, providing data brokerage and data processing functions. Software 'Applications' request access to relevant data sources from the data platform, and the data platform then retrieves the data from cloud storage and potentially provides compute resources to host the processing of software analytics.

Ideally, as the industry matures, data platforms will become more like Platform-as-a-Service (PaaS) infrastructure (a domain specific version of common Infrastructure-as-a-Service (IaaS) cloud platforms such as AWS and Azure). The domain specificity can enshrine industry relevant nomenclature and data standards, so that it is easy to use for property industry practitioners. In future, this could enable data sources to be discovered and self-configured by Apps with minimal human intervention.

The focus of this Section 4 is on the data layer, of the smart building digital infrastructure stack, and the data platforms that perform this role.

4.2 Functionality and Features of Data Platforms

A desirable data platform will be able to perform a range of necessary technical functions and have a range of attributes that support the data governance and data management aspirations discussed in Sections 2 and 3. These key capabilities include:

A. Building Onboarding and Data-Source Connection: A building services contractor must be able to connect the data platform to, and ingest data from, the various on-premises data sources in a way that recognises context (ie meta-data information that can relate the data source to features of the building). These data platform functionalities include;

- *Building registration wizards*: These wizards are used for initial registration/ configuration/ partitioning of a new building on the data platform, and for describing the features of the building. The features of the building, entered into the system, can then be used as tags for ascribing meaning to data sources streaming into the platform. Ideally the building registration wizard will support the creation of a building model based on a recognised industry standard meta-data schema.
- *Device communication protocols*: The platform should be able to ingest data from a wide variety of devices using common communications protocols such as BACnet, Modbus, and MQTT. It should also be able to ingest data through APIs and perform scheduled FTPS ingestions of Comma Separated Values (CSV) files.
- *Data-source connection*: The platform should provide means for a building services contractor to find, authenticate and connect to data-sources.

B. Data Storage and Retrieval: The data platform should enable people and machines to find and retrieve data from storage in a way that facilitates data-processing and avoids data leakage. These data-platform functionalities include;

- *Data cleaning*: A range of tools should be available for processing incoming data to ensure data quality including (i) detection of anomalies and stale data and (ii) data interpolation to fill gaps and unify timestamps etc.
- *Time series database*: The platform will store time-stamped data so that trends in the data can be identified.
- *Meta-data store* (eg graph database): to process logical queries for finding data sources from the timeseries database on request from software applications. This store takes advantage of the meta-data schema to support software and data re-useability.
- *Data access permissioning system*: Data should be containerised in a way that provides data 'ownership' for the designated data-controller organisation. The data-controller organisation would appoint an administrator responsible for maintaining data sovereignty for the organisation. From there, only authorised people and software applications should be able to access data, through a permissioning system supervised by the administrator. The functionality for permitting users access to specific data (not just all or nothing), at different role-based levels of access (eg read or write) should be provided by the platform.
- *Data download*: Time series data should be able to be downloaded by the permitted user as a csv file, and be streamed to permitted third party software applications via APIs.

C. Data Utilisation and User Support: The data platform should provide basic support features (beyond just data-management/ data-access, as described in B above) that help users to deliver value adding data-driven software services

- *Basic visualisations and alerts*: The user should be able to view trendlines for timeseries data and thereby manually explore possible cause and effect relationships between different data streams. The user should be able to set alerts to reflect their preferred alert thresholds.
- *Compute resources and output signals*: Algorithms will ideally be able to (optionally) process data on-platform (as “Applications”), utilising Platform-as-a-Service (PaaS) compute resources. The results/outputs from these data-processing algorithms should be able to be automatically sent to remote devices, as informational alerts and/or supervisory control signals.
- *Application Marketplace and Third-Party Data*: These feature enables building managers to extend the data platform's built-in capabilities through the ability to browse available third-party data-sources and/or install third-party, data-driven, ready-to-use applications on the platform. A software development kit (SDK) would ideally be available for third party developers, so that independent software developers can be contracted to build customised applications.

In addition to the functionalities described above, the data platform should exhibit the following attributes:

Evolvability through open protocols and modular architecture – The ability to easily interface with third-party systems and devices helps to future-proof investment by maximising flexibility and choice for the building owner, and minimising the potential for vendor lock-in.

Abstraction of building features – The integration, onboarding process and database structure aims to provide a generalised digital representation of the building and its components, in a way that serves as the foundation for a structured understanding of component relationships and behaviours. This enables software to logically process data from diverse buildings without additional manual input or building specific software code.

Cybersecurity and data privacy – In concert with the device layer and network layer, the data-platform must provide robust protection against malicious attacks and attempts to gain unauthorised access to the platform and/or the integrated building systems.

Scalability and Performance – The data-platform should be able to maintain performance, irrespective of the amount of data and number of connected devices being processed. This further helps to future proof data-platform investment..

4.3 State-of-the-Art of Data Platforms

A review was conducted of available data-platforms in 2022. The aim of the review was to gain an indicative understanding of the extent to which the desirable features of data-platforms are readily available in the market. The review involved collecting responses from platform owners to a questionnaire/survey covering 11 thematic areas:

1. Governance
2. Data access and security
3. Data upload/building onboarding
4. Data capture
5. Data storage
6. On-platform programming
7. Data and application code recovery
8. Output signals and control
9. Applications marketplace
10. Screens and visualization
11. Platform development

The questions were both of a quantitative and qualitative nature. Questions included both (i) closed yes/no or multiple-choice answer questions, and (ii) open text-based questions, with no limits on the length of the answer. Responses were received from 12 data-platforms, covering both not-for-profit/government platforms (6) and commercial platforms (6).

Some findings from the survey include:

- Five out of the 12 platforms did not use any external cloud hosting services, for 1 it was optional, 2 used AWS and 3 used Azure.
- 9 out of the 12 platforms gave clients discretion over access to their data, and 2 could be fine-tuned for this possibility. All platforms claimed the ability for the client to make parts of the data (within a building) available to third parties. 10 out of the 12 platforms could provide different tiers of access based on role, both internally (within the data client organisation) and externally for authorised third parties.
- Platforms had connectors for the common open communications protocols (eg BACnet, Modbus), and various data cleaning functions. APIs were provided in 10 out of the 12 platforms, to assist with data export to external platforms.
- A structured schema (such as Brick), that enables data to be linked to physical spaces and systems within the building, was used by 7 of the 12 platforms. There was no consistency across the platforms regarding the programming languages used to query the database.
- An Applications marketplace (including means for 3rd party providers to list/advertise their software applications to data clients) was provided in 4 out of 12 platforms and it was under construction in 2 platforms. One platform had a payment gateway for charging ‘App’ usage fees. One platform hosted an open github community, but did not provide a marketplace.
- Six out of the 12 platforms were capable of dispatching high level interface outputs, for cloud-based supervisory control of building mechanical systems.

The survey highlights a growing technical maturity in the market with most data-platforms providing sophisticated data capture and data management capability. A significant point of difference relates to the availability of data-sharing capabilities that support third-party application developers. These points of difference include (but not limited to) availability of an application marketplace and use of common programming languages for querying the database. Presumably, this partly reflects differences in adopted commercial business models.

There also appear to be some different perspectives on the future role of data-platforms in smart buildings. A number of data platforms have opted not to utilise cloud hosting. And many of the platforms have opted not to target supervisory control services. This may reflect different perspectives on the viability of certain applications, when cyber security is considered.

4.4 Implementing Data-Driven Infrastructure

Intelligent data-driven buildings require a robust digital infrastructure which, at its core, includes a modern data platform. These data platforms simplify the collection of data from building components such as IoT sensors, BACS, and BMS. They also excel in the curation and integration of data; including cleaning, harmonization, aggregation, and description of the collected data. The resulting curated data is efficiently shared and consumed by both built-in and third-party software applications (including AI/ML algorithms) running on the platform. The data platform can then automate the processes for distributing the output of these applications, to notify building stakeholders and/or to adjust the behaviour of the underlying building systems and equipment.

Key implementation steps to achieve a data-driven infrastructure are outlined below.

Goals and Objectives – Define the overarching goals for the data-driven infrastructure (eg contributing to lower maintenance costs). Then, break down the goals into concrete and measurable objectives. For example, a goal of reducing maintenance costs can translate into a specific objective of a 10% reduction in expenses attributed to unexpected equipment failures (an objective which could be tackled by, for example, a third-party, smart, data-driven application purchased and deployed on the platform).

Smart Readiness Assessment – If the building is in the planning phase, acquaint yourself with the proposed infrastructure. For already constructed buildings, conduct a comprehensive inventory of the current infrastructure. In both scenarios, a detailed analysis of building equipment, devices, and systems is essential. This includes a thorough examination of network connectivity, data flow channels, and communication protocols. Additionally, analysing control logic, setpoints, and data collection points is vital to ensure alignment with established objectives. Then, conduct a detailed and comprehensive assessment of the building's smart readiness. Different methodologies, frameworks, and tools are available for this purpose, their suitability depending on the regional and contextual specifics. A number of rating systems exist (eg the European Smart Readiness Indicator (SRI)) that may help to unlock insights into a building's capacity to leverage the advantages of digital technologies.

Data Platform Selection – Research and select a modern data platform that aligns seamlessly with the data-driven infrastructure. Key considerations include ensuring that the platform has the necessary features and functionality to effectively achieve the established objectives. Additionally, the platform should be able to accommodate evolving needs and growth, ensuring it remains a sustainable and adaptable component of the infrastructure.

Onboarding – The process of connecting a building to a data platform for data acquisition purposes is normally known as onboarding. Aim for a robust onboarding process which starts with a thorough identification of primary data sources, i.e., other systems within the building’s data-driven infrastructure (e.g., a BMS, or a database) from where data can be retrieved. Continue with the creation of a digital model for the building being connected. Finally, conduct rigorous testing of the entire data acquisition process, including the correct representation of the collected data within the digital model.

Data Preparation and Sharing – Ensure that the collected data is optimized for use by the various applications deployed to the data platform. If necessary, implement automated data workflows within the data platform to pre-process the raw collected data without human intervention. Then, use the output of these workflows to populate the relevant structured data sources. Furthermore, carefully configure sharing of the prepared data with the target applications. Pay especial attention to access rights and adherence to data privacy policies, ensuring that data is shared securely and in compliance with established privacy regulations. This approach guarantees that the prepared data is efficiently and correctly utilized by the applications.

Feedback Loops – Configure feedback loops within the data platform to facilitate automated control and refinement of building operations. These feedback loops channel insights or results from the applications back into the integrated building equipment and systems. By utilizing feedback loops effectively, the data-driven infrastructure achieves a level of continuous improvement that enhances sustainability, cost-effectiveness, and overall functionality.

Regulatory Compliance – Ensure that the entire data-driven infrastructure complies with all pertinent regulations. Pay special attention to those concerning data privacy and protection. Compliance with the appropriate legal and regulatory frameworks is essential to guarantee that the data-driven infrastructure operates within the bounds of the law and maintains high standards of privacy and security.

User Training – Train pertinent building management personnel on how to use the data-driven infrastructure in general, and the data platform in particular. Empowering personnel with the necessary knowledge and skills ensures that they can effectively harness the capabilities of the infrastructure. Specifically, they learn how to navigate and utilize the data platform, enabling them to make data-informed decisions, troubleshoot issues, and optimize building operations.

5. Conclusions

Digitalisation, through data-driven approaches and advanced analytics, offers opportunities to save energy in buildings by improving operational efficiency. However, barriers such as access to data, interoperability, and privacy need to be addressed for the widespread adoption of digitalisation in buildings.

Subtask A of IEA EBC Annex 81 provides guidance on establishing the necessary IT infrastructure and data management services, to enable energy saving analytics and control services. It aims to establish a common understanding of what constitutes "digital readiness" in buildings.

Data is more valuable, and data-driven software applications are more deployable, when data collection complies with the FAIR data principles (where the FAIR acronym refers to data being Findable, Accessible, Interoperable, and Reusable). Implementing data standards and using independent data platforms can help to achieve these principles. Challenges such as interoperability, data rights and data risk management (including compliance with privacy laws such as the GDPR), are issues that need to be considered.

Metadata plays a crucial role in enabling data-driven smart buildings by providing a standardized way to organize and utilise data. Different metadata schemas have been developed, each with their own characteristics and means of creating value. There is a growing desire for unity and alignment between these schemas. RDF-based metadata schemas are emerging as the dominant approach for modelling metadata.

In future, we can expect energy-analytics and other software products to be transparent about their data needs. By understanding the ‘sufficiency’ of a building’s data collection capability (and the gap between this and the needs of the software product), building owners will be able determine the suitability of a given software product for their building and the likely cost-benefit of deploying these energy productivity services.

Data platforms are an important tool for distributing data and enabling data-driven smart building services. Careful consideration of a data platform’s functionality, attributes and commercial constraints on use, will help the building owner to realise effective control over their data (data sovereignty) and to streamline data management.

A modern data platform will simplify data collection, curation, and integration, allowing for efficient sharing and consumption by software applications. Key steps involved in implementing a data platform include defining goals and objectives, conducting a smart readiness assessment, selecting a suitable data platform, onboarding the building to the platform, preparing and sharing data, configuring feedback loops, ensuring regulatory compliance, and providing user training.

6. References

Commonwealth of Australia (2019), “Sharing Data Safely”, https://nla.gov.au/nla.obj-1856863320/view

Energy Efficiency Hub Digitalisation Working Group, (2022), “Roadmap on Digitalisation for Energy Efficiency in Buildings”, https://energyefficiencyhub.org/wp-content/uploads/2022/11/DWGRoadmap.pdf

Fierro G. and Pauwels P. (2022), “Survey of metadata schemas for datadriven smart buildings”, CSIRO, ISBN 978-1-4863-1896-4, https://annex81.iea-ebc.org/Data/publications/Survey%20of%20meta-data%20schemas%20(final)1.pdf

Fierro G., Saha A., Shapinsky T., Steen M., and Eslinger H. (2022), “Application-Driven Creation of Building Metadata Models with Semantic Sufficiency”, 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys’22), Boston, USA, https://doi.org/10.1145/3563357.3564083

Gomez-Garcia R. (2024), “A Guide on Data Platforms for Data-Driven Smart Buildings”, CSIRO, ISBN ??, https://annex81.iea-ebc.org/publications

International Energy Agency (2017), “Digitalisation & Energy”, https://iea.blob.core.windows.net/assets/b1e6600c-4e40-4d9c-809d-1d1724c763d5/DigitalisationandEnergy3.pdf

O’Reilly (2020), “The state of data quality in 2020”, https://www.oreilly.com/radar/the-state-of-data-quality-in-2020/

Trianni A., Bennett N., Hasan A.S.M., Katic M., Lindsay D., Cantley-Smith R., Wheatland F.T., White S., Dunstall S., Leak J., Pears A., Cheng C.-T. And Zeichner F. (2022), “RACE for Business - Opportunity Assessment: Industry 4.0 for Energy Productivity”, https://www.racefor2030.com.au/opportunity-assessment-reports/

White S.D., Marszal-Pomianowska A., Jin G.Y., Madsen H., Candanedo J., Harmelink M., Stensson S., Gori V. (2023), “A Data Sharing Guideline for Buildings and HVAC Systems”, CSIRO, ISBN 978-1-4863-1895-7, https://annex81.iea-ebc.org/Data/publications/IEA%20Annex%2081%20Activity%20A1%20-%20A%20Data%20Sharing%20Guideline%20for%20Buildings%20and%20HVAC%20Systems%20(final)2.pdf

Wilkinson M., Dumontier M., Aalbersberg I. et al. (2016), “The FAIR Guiding Principles for Scientific Data Management and Stewardship”, Scientific Data, 3:160018, https://www.nature.com/articles/sdata201618

ANNEX **81**

EBC

Energy in Building and
Communities Programme

www.iea-ebc.org

Technology Collaboration Programme
by iea