Project acronym	B4B
Project full name	Brains for Building's Energy Systems
Grant No	MOOI32004
Project duration	4 year (Starting date May 1, 2021)

Deliverable D4.03

Study of data needs and requirements in smart buildings

Lasitha Chamari, Pieter Pauwels, Ekaterina Petrova (TU/e)

Elena Chochanova, Rizal Sebastian (TNO), Nico Mutsaers (Philips), Berno Veldhuis (RVB), Simon Hopkins (Simaxx), Niels de Jong (Cloud Energy Optimizer), Joep van der Velden (Kropman), Jan-Willem Dubbeldam (Kropman), Joep van der Weijden (TU Delft)

Work package	WP4
Result	8
Lead beneficiary	Eindhoven University of Technology (TU/e)
Due Date	30 April 2022
Deliverable Status	Final
File name	B4B WP4 D4.03_Data Needs and Requirements.docx
Reviewers	Olav Vijlbrief (TNO)

The project was carried out with a Top Sector Energy subsidy from the Ministry of Economic Affairs and Climate, carried out by the Netherlands Enterprise Agency. The specific subsidy for this project concerns MOOI subsidy round 2020

SUMMARY

Brains for Buildings Energy Systems (B4B) is a multi-year, multi-stakeholder project that aims at developing smart building methods to enhance the operations in buildings in terms of reducing energy consumption, increasing energy flexibility, increasing occupant comfort, and improving installation maintenance costs. This will be achieved by developing faster and more efficient Machine Learning and Artificial Intelligence models and algorithms. The project is geared to existing utility buildings such as commercial and institutional buildings.

There are five work packages in this project dedicated to above tasks. This report is an outcome from the Work Package 4 (WP 4) of the B4B project. WP4 works under "Data integration" theme. There are many heterogeneous data sources in buildings, e.g. sensor data, 3D model data, time schedules, occupant data, asset maintenance data, electrical system data, etc. Their data types, naming conventions and formats differ, making it difficult to use these data for developing models and algorithms. Data sources or systems communicate using different protocols and exchange data in different formats and standards. Therefore, these data are often siloed, they are not available via a single platform, meaning there is little to no interaction between them. WP4 aims to develop methods to integrate these data silos for developing Machine Learning and Artificial Intelligence models and algorithms while also guaranteeing privacy and ethics when collecting, storing, integrating, sharing, managing or utilizing data in smart buildings.

This deliverable "D.4.3 Data needs and requirements in smart buildings" reports the outcomes of T4.1.1: Study of data needs and requirements in a bottom-up and top-down manner. The study method consists of a literature study based on the smart building research field, a market study based on interviews carried out with real estate building owners and platform providers in the Netherlands, and meetings with other work packages in the B4B project. In this way, this report addresses the entire "smart building" spectrum to find out the use cases, their data needs and requirements, limitations and barriers when utilizing data, and finally, the idea for an ideal best-case scenario for data collection and integration.

Results of the study act as the foundation to develop a data collection and integration reference architecture for data integration in smart buildings in the upcoming deliverable (T4.2.1: Development of a reference system architecture and set of data flow procedures).

This report discusses various data needs and requirements in smart buildings in terms of five smart building use cases identified according to the literature and further elaborates on available systems in buildings, data types, data formats, and data storage techniques. It then presents the barriers and limitations to data collection. Finally, an idea of an ideal best-case scenario is presented, giving an overview of what needs to be achieved via a data collection and integration reference architecture.

In summary, developing smart building solutions depends on metered time series data and real time data from sensors, data from external services, occupant feedback, domain expert knowledge and contextual data/ metadata. Often, developing these solutions suffer from a) unavailability or partial availability of required data, b) low-quality data (accuracy, low resolution, timeliness, missing samples), c) unstructured data (naming conventions, metadata standards, data formats) and d) limitations due to accessibility, privacy and ownership.

Therefore, improving data availability, improving data quality, structuring the data and adhering to privacy and ethics guidelines can speed up the development of smart building solutions. Considering the requirements from the literature and market study, it also suggests that a reference architecture for data integration should be able to provide a) methods for data acquisition, storage, management and backup to improve data availability, b) provide standardisation to data originating from different systems using standard schemas and well-recognised data models, c) provide access to historical and real time data d) support bidirectional communication between the smart building applications and building controllers and e) provide scalable solutions that can accommodate the growing number of data sources and users.

SAMENVATTING

Brains for Buildings Energy Systems (B4B) is een meerjarig project met meerdere belanghebbenden dat tot doel heeft om slimme bouwmethoden te ontwikkelen om de activiteiten in gebouwen energie-efficiënter te maken, om flexibel energieverbruik mogelijk te maken, om comfort te verhogen van de gebruikers en om installatietechnisch onderhoud slimmer te maken. Dit wordt bereikt door snellere en efficiëntere modellen en algoritmen voor Machine Learning en Artificial Intelligence te ontwikkelen. Het project is afgestemd op bestaande utiliteitsbouw.

Vijf werkpakketten in dit project zijn gewijd aan bovenstaande taken. Dit rapport is een uitkomst van het Werkpakket 4 (WP 4) van het B4B-project, met als thema "Data Integratie". Er zijn veel heterogene gegevensbronnen in gebouwen, b.v. sensorgegevens, 3D-modelgegevens, tijdschema's, gegevens van bewoners, gegevens over systeemonderhoud, gegevens van elektrische systemen, enz. Hun gegevenstypen, naamgevingsconventies en formaten zijn heel verschillend van gebouw tot gebouw, waardoor het moeilijk is om deze gegevens te gebruiken voor het ontwikkelen van gestandaardiseerde rekenmodellen en algoritmen. Gegevensbronnen of systemen communiceren bovendien met verschillende protocollen en wisselen gegevens uit in verschillende formaten en standaarden. Daarom zijn deze gegevens vaak enkel beschikbaar in silo's. Ze zijn niet beschikbaar via één platform, waardoor er weinig tot geen interactie of koppeling of gezamenlijk gebruik is van meerderede dergelijke silo's. WP4 heeft tot doel methoden te ontwikkelen om deze gegevenssilo's te integreren voor het ontwikkelen van modellen en algoritmen voor machine learning en kunstmatige intelligentie, terwijl ook privacy en ethiek worden gegarandeerd bij het verzamelen, opslaan, integreren, delen, beheren of gebruiken van gegevens in slimme gebouwen.

Deze deliverable "D.4.3 Gegevensbehoeften en -vereisten in slimme gebouwen" rapporteert de resultaten van taak T4.1.1: Studie van gegevensbehoeften en -vereisten op een bottom-up en top-down manier. De onderzoeksmethode bestaat uit een literatuurstudie rond 'smart building', een marktstudie op basis van interviews met vastgoedeigenaren en platform providers in Nederland, en overleg met de overige B4B-werkpakketten. Het resultaat is een rapport dat het hele "smart building" spectrum behandelt om de use-cases, hun gegevensbehoeften en -vereisten, en beperkingen bij het gebruik van gegevens te achterhalen, incl. een gewenst best-case scenario voor gegevensverzameling en integratie. De resultaten van deze studie vormen de basis voor de ontwikkeling van een referentie-architectuur voor gegevensverzameling en gegevensintegratie in slimme gebouwen.

Dit rapport bespreekt verschillende databehoeften en -vereisten in slimme gebouwen, aan de hand van vijf smart building use cases die volgens de literatuur zijn geïdentificeerd. Het gaat verder in op beschikbare systemen in gebouwen, datatypes, dataformaten en dataopslagtechnieken. Vervolgens worden de barrières en beperkingen voor het verzamelen van gegevens weergegeven. Ten slotte wordt gewenst best-case scenario gepresenteerd, dat een overzicht geeft van wat moet worden bereikt via een referentie-architectuur voor gegevensverzameling en integratie.

Samenvattend is het ontwikkelen van slimme bouwoplossingen afhankelijk van gemeten tijdreeksgegevens en realtime gegevens van sensoren, gegevens van externe diensten, feedback van bewoners, kennis van domeinexperts en contextuele gegevens/metagegevens. Het ontwikkelen van deze oplossingen heeft vaak te lijden onder a) onbeschikbaarheid of gedeeltelijke beschikbaarheid van benodigde gegevens, b) gegevens van lage kwaliteit (nauwkeurigheid, lage resolutie, tijdigheid, ontbrekende voorbeelden), c) ongestructureerde gegevens (naamgevingsconventies, metagegevensstandaarden, gegevensformaten) en d) beperkingen vanwege toegankelijkheid, privacy en eigendom.

Daarom kan het verbeteren van de beschikbaarheid van gegevens, het verbeteren van de gegevenskwaliteit, het structureren van de gegevens en het naleven van privacy- en ethische richtlijnen de ontwikkeling van slimme bouwoplossingen versnellen. Het rapport concludeert dat een referentiearchitectuur voor data-integratie in staat zou moeten zijn om a) methoden voor data-acquisitie, opslag, beheer en back-up te bieden om de beschikbaarheid van data te verbeteren, (b) te zorgen voor standaardisatie van data met behulp van standaardschema's en goed erkende datamodellen, c) toegang te bieden tot historische en realtime gegevens, d) bidirectionele communicatie tussen de slimme gebouwapplicaties en gebouwcontrollers te ondersteunen en e) schaalbare oplossingen te bieden die geschikt zijn voor het groeiende aantal gegevensbronnen en gebruikers.

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

1.1 Background of the study

Commercial and residential buildings in Europe account for 40% of the total energy use and 36% of the carbon dioxide emissions, which mainly stem from the construction, usage, renovation, and demolition of buildings (European Commission, 2008). Furthermore, almost 50 % of Europe's total energy consumption is used for heating and cooling, 80 % of which is used in buildings. Therefore, there is great societal awareness and regulatory pressure to make buildings sustainable and energy-efficient enable the energy transition. The constantly increasing performance requirements are accompanied by rapid technological developments, advancements in the Information and Communication Technology (ICT) and Internet of Things (IoT) industries, as well as the richness and volume of data generated throughout the entire lifecycle of the built assets. Furthermore, the effects and benefits of cloud computing, data analytics and Artificial Intelligence (AI) are now also recognisable in the built environment. Leveraging ICT, IoT and AI enables the deployment of efficient methods for smart building control and optimised energy use. To achieve these, buildings are expected to interact with external systems such as electrical grids, renewable energy systems, and other buildings while responding to various demands and external conditions. In other words, buildings need to become smart and effectively utilise the deployed smart building operation strategies. A smart building must satisfy the occupant's increasing demands for a better indoor climate, energy efficiency, health, well-being, productivity, and safety.

All those use cases mentioned above rely on data from various systems in buildings. Typically, each building is different, and the data is distributed across multiple systems such as Building Management Systems (BMS), Asset Management Systems (AMS), IoT devices, smartphones, databases, etc. Available data, their naming conventions and data formats also differ significantly among different buildings and vendors. Depending on the use case, the data that is required also varies. Therefore, it is important to investigate what data is required for a given use case and the methods for acquiring and integrating those data to support the various and constantly growing data needs in smart buildings.

1.2 Research questions

There are various use cases that are related to the fulfilment of various objectives in buildings. These objectives can be improving energy performance, improving user satisfaction, detecting faults in building operation, performing building maintenance over time, etc. Data needs and requirements differ among those use cases, and it is important to highlight these differences in relation to their data needs and requirements. Defining these data needs and requirements is only possible when the use cases of the building are clearly understood and defined. One simple example of a use case is to determine how much energy a building uses per m^2. This energy KPI is a good indicator of how a building is performing energy-wise, and it requires certain data input to be calculated.

To provide useful information to support the given use case (e.g., derive the energy usage per m^2 as mentioned above), one needs to determine which data is needed. For instance, besides energy consumption, determining the KPI requires data about the type of building, its floor area, building usage, electrical equipment, and their wattage etc. Therefore, it is also crucial to **understand which data is needed** for the given use case.

Most use cases (e.g., anomaly detection in system operation, Model Predictive Control (MPC), etc.), need a holistic approach, meaning that the data should be collected from various distributed sources in a building. In other words, buildings are typically characterised by heterogeneous systems, data models, naming conventions, and data access control. Therefore, it is also important to investigate **which data is available** in which formats and which access control mechanisms are available to fulfil above needs.

Therefore, this report aims to answer following research questions:

1. What are the use cases related to fulfilling various objectives in buildings, in particular focused on data-driven smart buildings?
2. What are the data needs and requirements for those use cases?
3. What data is available and how to fulfil those needs and requirements ideally?

This leads to the overall structure and approach as shown in Fig. 1 for this report in which multiple use cases are listed, for each of them the available data is listed as well as the data that is needed, and finally those needs and available data are mapped to analyse how these use cases can be supported through the right type of data integration.

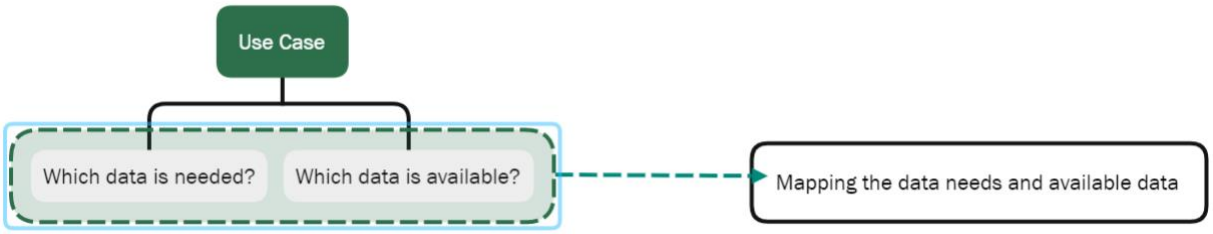

Figure 1 Mapping the data needs by use cases

1.3 Objective and scope of the study

The main objective of this deliverable is to investigate the data needs and requirements in smart buildings, thereby proposing future directions for data collection and integration for the upcoming tasks within the Brains4Buildings project.

Therefore, the main objectives of this report are to:

- identify the data needs and requirements in smart building applications according to the **state-of-the-art research** in the area.
- identify data needs and requirements in smart building applications according to the **state-of-practice** in commercial applications (market study on the real estate sector).
- provide a state-of-the-art review about the **smart building platforms**.
- provide a **market study** on smart building platforms based on input from platform providers.
- provide suggestions for a **data collection plan and a reference architecture** for smart buildings (ideal scenario).

2 METHODOLOGY

2.1 Research approach

The data needs and requirements in the smart building domain can be quite diverse. These needs and requirements vary according to their intended purpose. Therefore, this research takes a "use case"- based approach to understand and align those needs and requirements to specific use cases. This overall approach is schematically displayed in Fig. 2.

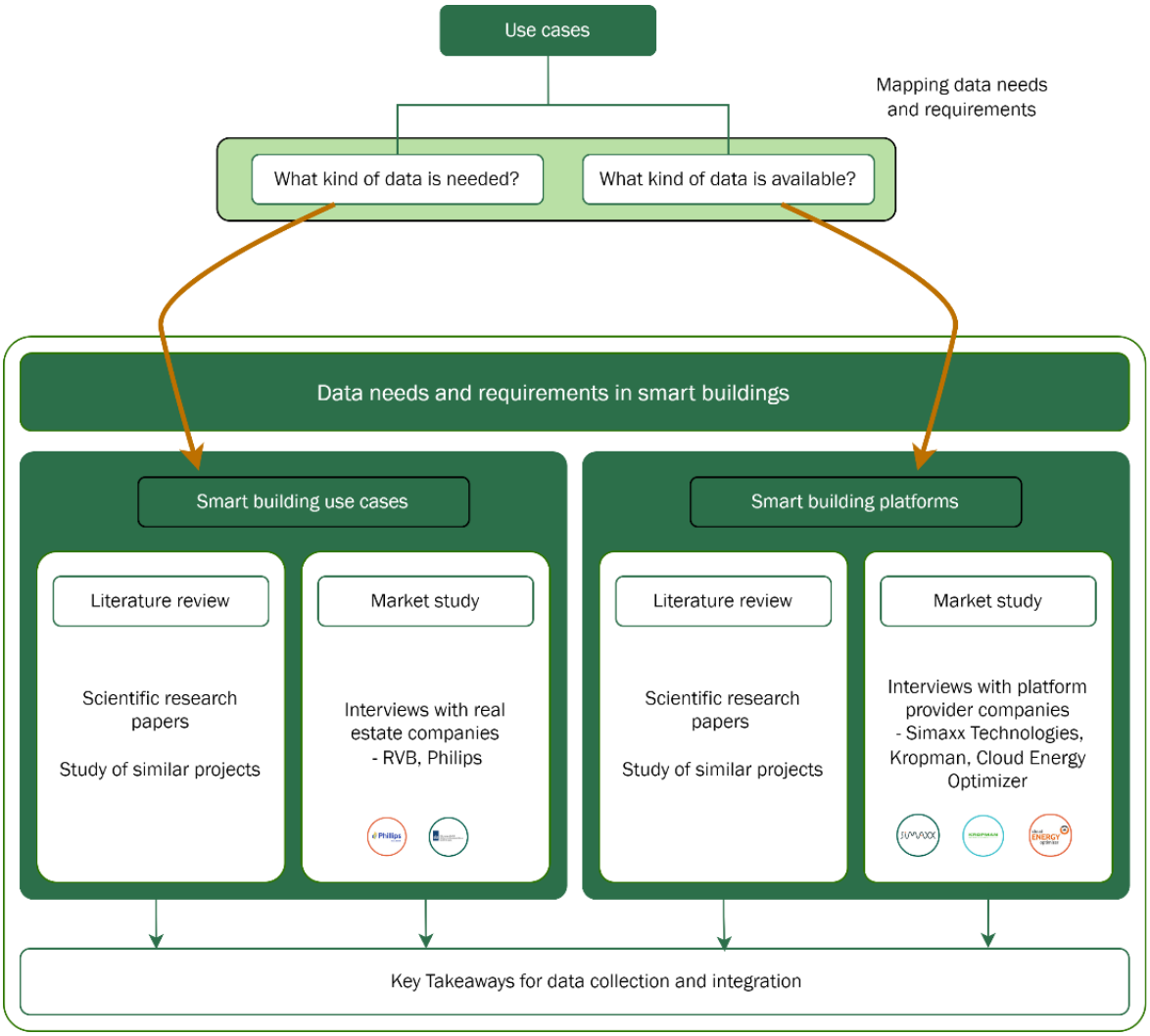

Figure 2 Research approach to map data needs and requirements leading to data collection and integration architecture

First, state-of-the-art research on smart buildings and smart building operation and maintenance is reviewed to identify different use cases, and their needs and requirements (leftmost Literature Review node in Fig. 2). In addition to the state-of-the-art research, we also conduct a market study by interviewing Dutch real-estate companies participating in the Brains4Buildings project (left Market Study node in Fig. 2). Real-estate companies need to comply with various regulations and operate buildings efficiently and cost-effectively. They provide an in-depth view of various data needs and requirements that will be analysed and structured to form a reliable overview of smart building use cases (Fig. 2, left).

Second, buildings rely on different software solutions and platforms to collect, analyse and visualise building data for a given use case (Fig 2, right). These platforms facilitate monitoring, reporting, diagnostics, and analytics combined in modules for building owners who seek useful information about their building. A state-of-the-art review of smart building platforms is performed to understand how different platforms provide different methods for data acquisition, collection, and integration. Also, in this case, a second market study presents the results of interviews conducted with platform provider companies participating in the

Brains4Buildings project (market study node towards the right in Fig. 2). The market study aims to demonstrate how the previously identified use cases are practically implemented in buildings. Available data, data formats, naming conventions, and integration methods are also identified and discussed. Finally, a bridge is made between use cases and platforms to identify key takeaways for data collection and integration.

2.2 Reading guide

The remainder of this report is structured as follows:

- Chapter 3 identifies the smart building applications (use cases) and their data needs identified from the scientific literature. Problems and limitations of the data collection are also discussed.
- Chapter 4 describes the current data needs identified according to the real estate sector. The future needs, challenges and limitations related to data collection are also identified.
- Chapter 5 presents a state-of-the-art review of smart building platforms and data sources in buildings.
- Chapter 6 defines the functions of smart building platforms.
- Chapter 7 discusses the existing gap between the required data and the available data, followed by initial suggestions on how to overcome the gap.

These five chapters align fully with the 5 lowermost nodes in the diagram in Fig. 2.

3 STATE-OF-THE-ART DATA NEEDS IN SMART BUILDINGS

3.1 State-of-the-art in smart building use cases

The following sections describe the main findings related to key use cases in smart buildings according to the scientific literature. Each of these use cases has different aims and goals, and the below section aims to find these different categories of use cases. Of course, the potential list of use cases is endless, hence we focus on the five most commonly found and identified use cases in the literature.

3.1.1 Use case 1 - Fault Detection and Diagnosis (FDD)

Faults in sensors, actuators, configuration errors in BMS etc., lead to significant energy losses in buildings. It is estimated that those faults account for 15 – 30% of energy costs (Taylor et al., 2005). Fault Detection and Diagnostics (FDD) has therefore gained a lot of attention in smart buildings. Roth et al. (2005) identified 13 faults in commercial buildings and their contribution to energy loss in buildings. These fault types are; keeping lighting and HVAC systems on when a space is unoccupied, duct leakage, dampers not working properly, airflow not balanced, insufficient evaporator airflow, software programming errors, improper controls hardware installation, improper controls setup/commissioning, control component failure or degradation, valves not closing properly, air cooled condenser fouling and improper refrigerant charge, waterside issues, and refrigeration circuits. The key idea of FDD is that faults can be detected in these active systems, they can be diagnosed, and then finally resolved. The prevalent purpose lately is to enable FDD in an automated manner using machine learning algorithms and AI libraries that detect unexpected outliers when expecting a certain pattern of data.

When it comes to potential causes for the found errors (i.e. diagnosis), plenty of very specific reasons can be named and found, which is a major challenge in FDD research. In a more general level, Lazarova-Molnar et al. (2016) distinguish **four common reasons for faults in smart buildings**; namely, **1) wrongly programmed BMS, i.e., wrongly programmed control logic, 2) wrongly configured or conflicting setpoints, 3) wrongly configured building equipment and 4) misplaced or wrongly wired sensors and actuators.**

There are several methods to detect and diagnose these faults, such as model-based, data-driven and hybrid methods (Lazarova-Molnar et al., 2016). The data-driven approach uses historical data to derive relationships and predictive models (black-box models), whereas model-based FDD is based on physical models (white-box models). A hybrid approach is a combination of the above two (grey-box models).

Typical data requirement for FDD consists of event logs that indicate faults or reconfigurations (Lazarova-Molnar et al., 2016), metered data obtained from sensors (Lazarova-Molnar et al., 2016; Schumann et al., 2011), occupant feedback received through real-time feedback or surveys (Lazarova-Molnar et al., 2016; Schumann et al., 2011), domain expert knowledge (most often in the form of "if-then" rules) (Schumann et al., 2011) and other implicit data, such as anonymised internet usage data, WiFi connectivity (Lazarova-Molnar & Mohamed, 2016).

An overview of these methods and ways of working is displayed in Fig. 3, where one can see that data collection happens either with real-time or historical data, after which black-box, white-box models are built that both are able to predict how the system in the building is meant to be functioning. As soon as real-time data does not align with these values in the system, an unexpected discrepancy is found, which is detected as a fault if it goes above a certain threshold. Using the correct FDD technique, the fault is then catalogued, diagnosed and resolved.

Figure 3 Data collection framework for FDD of smart buildings (Lazarova-Molnar et al., 2016)

Work Package 1 (WP1) of the Brains4Buildings project focuses on FDD methods for smart buildings. WP1 aims to develop a FDD method based on a Diagnostic Bayesian Network (DBN). Data required for developing such methods can be categorised under subjective data from the occupants and objective data from the BMS.

Subjective data from the occupants includes the occupant's interactions (door and window opening, light level adjustment, temperature adjustment) and the occupant's thermal comfort data.

Objective data from the BMS includes measurements of supply and return temperatures, humidity, air pressure, flow, gas, electricity meter data, CO_2 concentration, signals from actuators (opening/ closing), positioning of valves, etc.

Of course, the objective system data is of higher importance here, compared to subjective data from occupants, because the black-box, grey-box and white-box models specifically capture how the system is supposed to work. FDD analyses to what extent this is also the case. Subjective data from occupants may be additionally used to verify and further improve quality.

3.1.2 Use case 2 - Occupancy and occupant behaviour predictions

Occupant thermal comfort estimation, predicting occupancy, occupancy patterns and occupant behaviour are important aspects of the building operation. Predicting occupancy is important for demand-based HVAC control (Dong et al., 2018). Neglecting the true occupancy may result in overcooling, overheating and discomfort due to insufficient thermal and ventilation services (Wang et al., 2019).

Non-intrusive methods can be used to estimate occupancy and thereby provide more significant energy reductions than what would be the case when relying on traditional static schedules that do not take occupancy into account (Ardakanian et al., 2016). **Motion detectors, CO_2 level sensors, luminosity sensors,**

relative humidity sensors (Wagner et al., 2017)**, cameras and the users' WiFi connectivity** (Martani et al., 2012) are generally employed for detecting occupancy.

WP3 of the Brains4Buildings project focuses on occupant behaviour modelling. WP3 aims to develop occupancy models and comfort models based on dynamic neural network models, for example. Data required for developing such methods can be categorised under subjective data from the occupant and objective data from BMS/ other IoT devices.

Subjective data from occupants include occupant's interactions (door and window opening, light level adjustment, temperature adjustment) and occupant's comfort data. Occupant feedback can be collected via mobile apps installed on smart phones / smart watches. Objective data from the BMS or from other IoT devices includes room temperature, CO_2 concentration, relative humidity, outside temp, etc.

3.1.3 Use case 3 - Energy flexibility and Demand Side Management

Buildings nowadays rely on multiple energy sources such as on-site generation, electrical energy storage and thermal energy storage (H. Tang et al., 2021), rather than depending solely on grid power. This makes it possible to use multiple energy sources optimally to control the energy cost, as well as respond to the electricity grid needs. Future Building Energy Management Systems (BEMS) will require adapting to variations in the rate of energy production from renewable energy sources (Finck et al., 2019).

MPC and **EMPC** controllers are recognised to optimally control the energy usage in residential and commercial buildings compared to traditional controllers. MPC and EMPC are increasingly experimented with on heating systems to minimise the energy cost and maximise flexibility. In that relation, **historical weather data** is required to make weather forecasting models, which are used to predict the ambient temperature and solar radiation (Finck et al., 2019). **Temperature sensors** for measuring **room temperature, surface temperature, radiator temperature, heating supply and return temperature** are required. **Microcontrollers** like Arduino are also used for acquiring data from temperature sensors and flow sensors. Artificial Neural Networks (ANNs) are used for weather forecasting. Also, some studies (Wagner et al., 2017) use **surveys** to collect data about window opening and closing in a residential building context. The range of historical data required for weather predictions is around 10 years of hourly data. Prediction accuracy can be increased by using weather data with a shorter interval, but with an increased computational time.

Maximum demand (peak demand) charge constitutes a significant part of the energy cost of commercial buildings. Therefore, various mechanisms are utilised in commercial buildings to reduce this peak demand. Often, this is achieved by Demand Side Management (DSM) strategies like load shifting (Gellings, 1985), i.e. shifting load from on-peak to off-peak periods. DSM activities are categorised as energy efficiency, time of use, demand response (DR) and spinning reserve (Palensky & Dietrich, 2011). Predicting the energy consumption is important to determine the day-ahead baseline power consumption (Cox et al., 2020; Lymperopoulos & Jones, 2018).

MPC is mainly used for DSM, where coordination of building loads and distributed energy resources are required (Cox et al., 2020; Široký et al., 2011; R. Tang & Wang, 2019). Different prediction methods are used to predict energy consumption in different load groups with different characteristics (Cox et al., 2020). Data required for such predictions includes historical data from electricity energy meters, electricity consumed for lighting, HVAC, plug loads etc., which are available from a BMS or EMS. Also, energy meter readings from photovoltaics (PV) and other renewable energy systems, as well as weather data are required.

WP2 of the Brains4Buildings project focuses on energy flexibility. WP2 aims to develop models of HVAC and electrical systems in buildings, and energy prediction models. Data required for the energy flexibility use case mainly includes building models (data-based and physics-based), energy usage data, user behaviour data and data from external weather services. Historical BMS data plays a key role in this regard. Tools used for such tasks include Modelica, EnergyPlus, Matlab, Python etc.

3.1.4 Use case 4 - Efficient HVAC control

The fourth use case is directly related to efficient control of systems, in this case particularly HVAC systems. In general, control parameters are decided by engineers and architects during the design stage, often with limited knowledge of the actual building operation. A significant body of research shows that buildings do not operate as intended (Corry et al., 2015; Hu et al., 2016), leading to a sub-optimal operational performance. Therefore, a major part of the research aiming to improve building operation is dedicated to designing intelligent controllers that can "learn" from the collected operational data and adjust the control parameters according to the actual operation.

The idea behind efficient control of HVAC systems is to reduce energy consumption while maintaining the indoor climate at an optimum level. HVAC control in traditional buildings is done using simple control logic that responds to set points and schedules (Ma et al., 2012). The latter control logic does not take the varying occupancy, weather patterns, or availability of renewable energy into account. More advanced controllers with feedback and forecasts are also available. They use proprietary control sequences to optimise the system performance at a higher level. However, their implementation is not widespread (Ma et al., 2012). In contrast, complex control strategies like MPC (Mayne et al., 2000) have gained traction as a strategy for HVAC control, since it provides the opportunity to incorporate predictions of **weather, occupancy, renewable energy availability and energy price signals** (Ma et al., 2012).

3.1.5 Use case 5 – Facility management, monitoring and controlling building performance

The potential of Building Information Model (BIM) for enhancing Facility Management (FM) is also widely discussed in the smart building domain (C. Eastman et al., 2011; Edirisinghe & Woo, 2021). While this use case is much less related to systems and control operations, plenty of building owners have this 5th use case as a very important use case in relation to the management of their building. **Buildings need to be operated and maintained over time, including all the assets within these buildings**. Also in areas other than buildings, this is a very important use case. For example, in the infrastructure domain (tunnels, railways, roads, etc), infrastructure assets are critical to be maintained properly. In those cases, every road or tunnel or railway station forms an important asset to be managed over time. Each of these assets consists of elements and components (e.g. wall, beam, foundation) that together constitute the asset.

For smart buildings, the same technique applies and a building is typically decomposed then of smaller elements using a **Building Information Model (BIM model)**. Such a BIM model is a simplified 3D representation of the building and its elements, with a set of properties stored for each element separately, and with several semantically rich relationships between the individual elements of the BIM model (e.g. space x is bound by y walls).

While these 3D BIM models are commonly used in the design, engineering and construction phases (the typical AEC domain – Architecture, Engineering and Construction), they are much less often used in the operational phase of the building or asset life-cycle. Asset management in the operational phase instead is organised using asset management software and systems that are much more database-oriented. A typical 2D representation of the building (or a schematic) shows the context of the building and gives access to the properties and relations of a specific component. That information can then be used by the asset manager or facility manager to make decisions for that asset (operation and maintenance).

Data used for this use case typically contains the following: (1) **a 2D or 3D schematic representation**, using a 2D drawing or a 3D BIM model or similar, (2) **semantically meaningful data** that stores more specific qualitative information about certain parts of the building, e.g. building functions, element names, element IDs, material properties, etc., and (3) **usage data that is typically collected from sensors or time schedule databases** (often tabular data or relational databases).

3.2 Summary of data needs

The above **five use case categories require different data**, and also rely on different computational architectures and algorithms. For example, use case 5 is a very detailed use case where data consists mainly of semantically rich data and 2D or 3D schematics. On the other hand, use cases 3 and 4 are much closer to control and therefore typically preferably do not include any 2D or 3D schematic data. On the contrary, those use cases are much closer to operations and control, and require access to signals and systems and direct communication devices (e.g. WiFi access and routers). Finally, use cases 1 and 2 are entirely focused on monitored data, including data from diverse end users with data on their own personal devices. These two use cases therefore rely on yet again very different data.

Furthermore, the **algorithms used are also very different depending from use case to use case**. While use case 5 requires mainly dedicated software and rule-based and object-oriented algorithms for facility management as managed mostly by the end user, use cases 1 and 2 are much more data-driven and relying on data analytics algorithms (machine learning). Use cases 3 and 4 are more control-oriented and rely often an interplay between incoming measured or monitored data and a control or simulation model. These use cases 3 and 4 require then often a combination of data-driven and model-based approaches.

Key use cases for data-driven smart buildings are use cases 1, 3 and 4. In those cases, research typically considers the use of white-box, black-box and grey-box (or hybrid) models to develop smart building applications. Software packages like Energy Plus, TRNSYS, DOE-2, ESP-r, Modelica have a physics-based approach and are used in the literature (Mugnini et al., 2020). These are **white-box models**, with the assumption that the person building these models in this software also understands all the rules that she is implementing in these models. Alternatively, it is possible to use training data for a considerable period and develop **black-box models** using Machine Learning (ML) algorithms. Such black-box models are also able to predict future values in a system and so enable monitoring, FDD and system control. These models rely less on physical equations that govern systems, and so they are less work-intensive and less expensive to build (compared to white-box models that are built by humans). The downside of these black-box models is that one typically does not know the reasons behind certain predictions: its rules and reasons are opaque to the user and can then be considered less trustworthy. ANNs, Support Vector Machines (SVM) and statistical regression are among the most commonly used black-box models (Mugnini et al., 2020). Finally, a combination of the above two models leads to hybrid models, which are also used in building energy modelling applications. A hybrid model can be a combination of partly white and black box models. A grey-box model (for example a RC-model) is physics-based (however simplified) and parameters are trained based on historical data. These hybrid and grey-box models seem to be the most promising alternative for use cases 1, 3 and 4.

To summarise, the data needed for different smart building applications can be categorised as follows (see also Fig. 4):

- **Metered time-series data from sensors** - This data is in the form of time series and is valuable for machine learning applications, time series analysis and forecasting, real-time monitoring, etc.
- **Data from external services** - Data from weather services (e.g., temperature, relative humidity, solar radiation, precipitation, cloud cover, air pressure, wind speed), electricity grid (price signals).
- **Occupant feedback** - Data related to occupant comfort, presence and behaviour (interactions with the building) is required by many smart building applications. These are collected from BMS sensors, IoT devices like smartphones, smartwatches and surveys.
- **Domain expert knowledge** - Control logic, operating sequence, schedules, and rules provide knowledge on how the systems are operated in a building.
- **Contextual data/ metadata** - BIM models and drawings provide contextual and semantically rich information about the building, materials, spaces, and construction parameters.

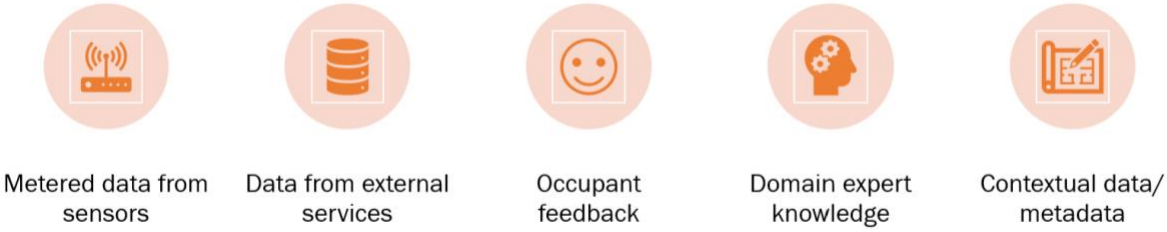

Figure 4 Summary of data needs

3.3 Problems and limitations in data collection for smart building applications

Collecting required data from buildings has been a challenge for developing smart building applications (Bolchini et al., 2017; Lazarova-Molnar & Mohamed, 2016, 2019; Wagner et al., 2017). It is nearly impossible to use the raw data as it is. It requires a major effort of pre-processing, depending on the quality of data and the application in which the data will be utilised. The decision of how to clean or not clean the datasets by handling missing and erroneous values and the choice of summarisation and aggregation strategies (Bolchini et al., 2017) has to be decided by the analyst. When the data gathering frequency does not match the application's sampling frequency, filtering or down-sampling is required. Often, a considerable amount of human input is required for cleaning the data before using it for any application. Some of the most commonly encountered problems can be categorised as follows:

- **Unavailability of required data from the buildings** – Many smart building applications rely on data (the five categories of data described above) and the most prominent data category is the metered sensor data

from the buildings. However, this data can be unavailable due to many reasons such as the absence of sensors, data not being recorded and/or stored, accessibility issues, etc.

- **Poor data quality** - Many smart building application methods are data-driven and the quality of the collected data significantly affects the accuracy of the results (Lazarova-Molnar & Mohamed, 2016, 2019; Ma et al., 2012). Missing samples, poor calibration and low resolution also contribute to the poor quality of the data.
- **Unstructured naming conventions** and tagging systems to represent syntax and semantics of data make it difficult to understand and interpret the data (Chen et al., 2021; Suzuki, 2015). Also, when these metadata standards are not properly documented, it makes users unclear about the provenance, quality, and purpose of data.
- **Limitations due to accessibility and privacy** - Accessibility and privacy-related issues (Lazarova-Molnar & Mohamed, 2016) often comes as a barrier to access the data, especially when the data involves sensitive and/or personal data. When data policies and licenses are undefined, it makes it difficult for a researcher to access the data.

These problems will always be present in existing data sets, and therefore it is of big importance that algorithms and systems that consume such data are made as robust as possible and can take into account the less than ideal data quality that is present in every system and building. In particular in the case of real-time data handling, important measures need to be taken at the data ingestion point to make sure that correct actions are made by the system or platform.

4 MARKET STUDY OF DATA NEEDS IN THE REAL ESTATE SECTOR

Data needs in the real estate sector stem from the sustainability goals and more demanding occupants. The traditional model of managing buildings independently, working in a highly manual way, and relying on highly experienced experts to run everything smoothly, is changing. The occupant demands more comfort and control, and the facility management is more data-driven. Therefore, the real estate sector is now facing various data needs and requirements driven by the increasing customer expectations and sustainability goals. Available sensors and various systems like BMS, EMS, and analytical platforms open new challenges and opportunities in the real estate sector when it comes to harnessing the full potential of their data. An example of how diverse the available data sources in the real estate sector can be is shown in a study that reveals that a UK government organisation managed real estate buildings with 70 siloed systems (Maslesa & Jensen, 2019).

To investigate the data needs and the requirements emerging from the real estate sector, we interviewed two real estate companies in the Netherlands that are inherent part of this Brains4Buildings project: Rijksvastgoedbedrijf (RVB) and Philips Real Estate. Since the real estate companies have a good relationship with a maintenance company to deliver the required information to them, the maintenance company Heijmans which provides services for the real estate sector in the Netherlands was also interviewed (also part of Brains4Buildings project).

4.1 Methodology

The investigation method followed semi-structured interviews with the participants. A questionnaire was prepared initially as a guideline for the interviews (see Appendix 1). Upon receiving the consent from the participants (See Appendix 3 for consent form), several interview sessions were conducted over a span of several months. As shown in Fig. 5, the questionnaire was organised in 5 themes, which a total of 33 questions. The themes for the questionnaire were use cases, data collection, data storage, data analysis and security. As such, this questionnaire aimed to collect data needs and requirements according to these 5 themes – as they were prepared in advanced by the B4B consortium and particular WP4 of this B4B project.

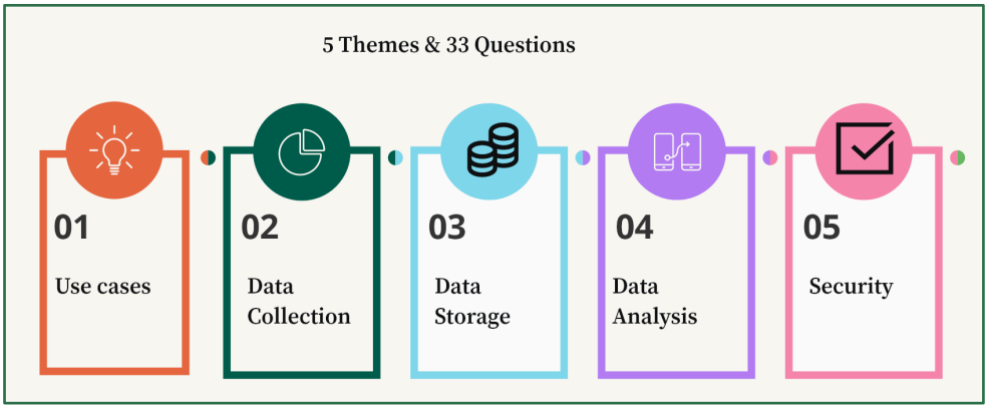

Figure 5 Categories of questions

After the interviews were done, a detailed analysis was performed according to the schema shown in Fig. 6. This first includes a procedure of transcribing, coding and categorizing. The results of that procedure are then analysed in a qualitative manner (method: qualitative analysis) and that leads to an understanding of the state of the art in practice, future data needs and requirements, and related barriers in the real estate sector. In the case of data needs, the survey has lead to the following needs, as will also be discussed later in this section in more detail:

1. Compliance to legislation
2. Occupant comfort and well-being
3. Asset performance
4. Energy performance

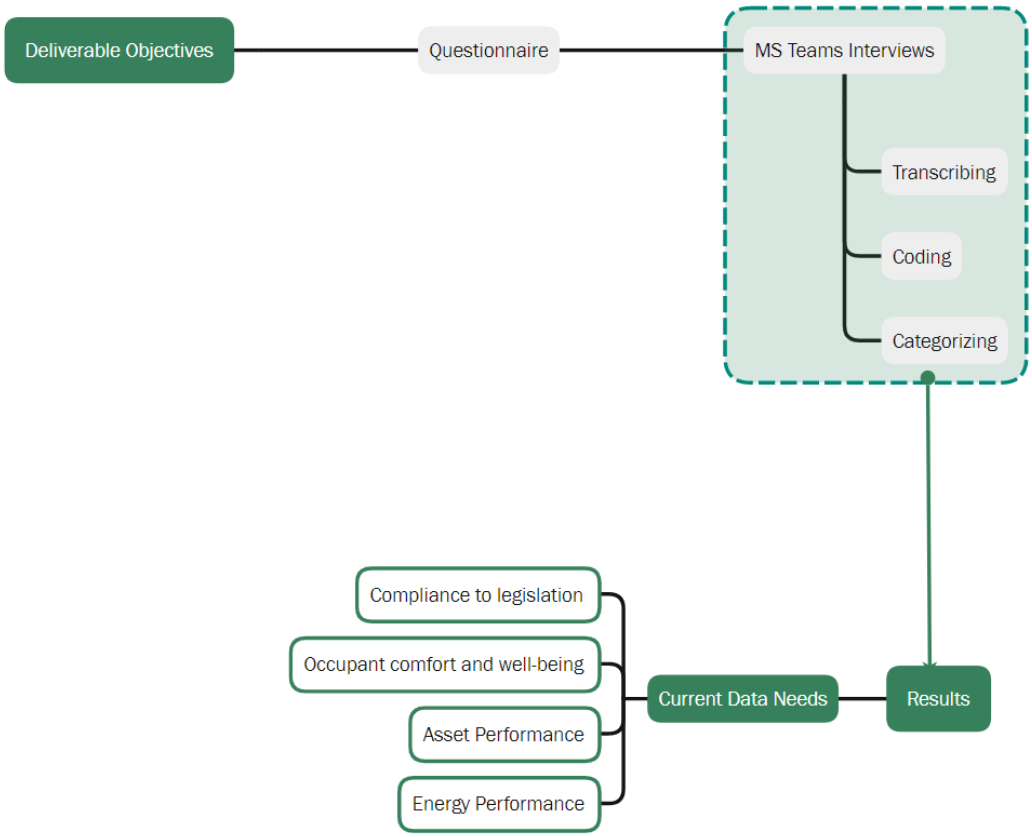

Figure 6 Qualitative research conducted on real estate data needs

This section summarises the needs arising from the real estate sector, from the point of view of the three parties that were being interviewed.

4.2 Current data needs in buildings

The complete list of needs can be consolidated and reduced to four categories. The real estate sector relies on data in order to comply with legislation, ensure that the occupants are comfortable and safe in the building, ensure its assets are performing well, and the building's energy performance is good.

4.2.1 Compliance to legislation

Buildings in the real estate sector need to comply with the Dutch government regulations such as energy efficiency (ISO 50001 energy management), maintenance (ISO 55000 asset management), health and safety ARBO regulations (TNO, 2017), sustainability regulation MJA3 (RVO, 2022), etc. Also, some buildings seek compliance with building standards such as BREEAM (*BREEAM - Sustainability Assessment Method*, 2015) and WELL (WELL Building Institute, 2020). Other organisation-specific regulations include RVB BOEI inspections for fire safety, maintenance, energy efficiency & sustainability (Central Government Real Estate Agency, 2018). For these purposes, buildings rely on data available from various systems such as BMS, EMS, and Asset Management Systems.

Please refer to *BRB-WP4-T4.1-AAA-Inventarisatie standaarden Brains for Buildings (B4B)* for a comprehensive list of standards related to buildings.

4.2.2 Occupant satisfaction and well being

A comfortable working environment needs adequate temperature, humidity, lighting, fresh air, CO_2 concentration and sound, and it enhances the well-being of people working on the site. As described above, maintaining comfortable working conditions is also required by legislation (adhering to table A at the ARBO regulations (TNO, 2017)). Especially with the COVID 19 situation, much more focus has been put on air quality

and ventilation. Occupants as well as building owners demand a well-performing ventilation system for health reasons. Hence, data about ventilation and air quality needs to be tracked, monitored, and assessed, including the control system. Furthermore, in relation to well-being, it is often the case that buildings monitor temperature, humidity, fresh air, and CO_2 levels; yet they do not measure lighting and sound levels adequately. Lighting and sound (acoustics) however, have a very important effect on the well-being and comfort of users in a space.

Although a wide area of research is dedicated to enhancing occupants' comfort and personalised environment, interview results show that this is rarely effectively implemented in practice. One of the respondents indicated that they rely mainly on a reactive approach, namely adjustments for temperature are made only in case a complaint comes in.

Further, according to the respondents, the existing configurations of buildings do not support personalised environments (due to the number of zones and available hardware). Yet, if improvements of personalised comfort and well-being is targeted, these zones and hardware are needed. In such cases the system also needs to be linked to the data and preferences coming from individual users, which is not straightforward technically and also not in terms of General Data Protection Regulation (GDPR) and privacy.

4.2.3 Asset performance

Building owners need to make sure that their assets are performing optimally. Various techniques make sure that the assets are up to standards.

- To understand which assets are most critical concerning the core process of the tenants, technical service provider Heijmans uses a technique called FMECA (failure mode, effects, and criticality analysis) to find out what the critical assets are and store that data in the maintenance management system. This technique is of use in particular for use cases 1 and 5 from the previous Section (Section 3).
- The status of the asset is determined according to NEN2767. This is carried out by an inspector using a set of defined measurements and a recording method. This can be used to detect possible defects, their extent, and their intensity. A score is offered from 1 to 6, one being very good and six being very bad. A condition score of 3 is sufficient for most real estate portfolio holders or users. Technical service provider Heijmans uses O-Prognose (O-Prognose, n.d.) for this process.
- Asset availability (downtime, malfunctions of the critical assets) is another asset performance index. Technical service provider Heijmans uses their maintenance management system where all assets are captured, and all maintenance information is made available.
- Another essential asset performance matrix is the costs incurred for corrective, preventive, and planned (replacement of the asset) maintenance. Technical service provider Heijmans uses its maintenance management system to create a Multi-Year Maintenance Plan.
- The location of the assets is based on BIM models, AutoCAD 2D drawings, and Excel sheets, as was also outlined for use case 5.

4.2.4 Energy performance

Building owners need data to know how well their buildings use energy. They need information about different KPIs that measure energy performance, such as energy consumption per m^2. They also compare the performance of buildings across the extensive portfolio of buildings. Energy consumption trends also provide insights into the growth or reduction of energy consumption compared to previous years.

Energy labelling is also an assurance of good energy performance. For example, all offices in Philips Real Estate should have energy label C from the beginning of January 2023. For new rental properties, Philips is looking for a BREEAM GOLD certificate. In that way, the owners can avoid evaluating individual systems' energy efficiency since the labelling is already a good indicator.

As mentioned earlier, energy efficiency in buildings is also imposed by government regulations. For example, Philips Real Estate is part of the MJA3 (meerjarenafspraken), an agreement with the Dutch government that needs to deliver a 2% energy reduction per year. Showing compliance with energy regulations such as *ISO 50001 Energy management* is an indicator that the building is energy efficient.

Buildings are making way towards harnessing renewable energy (mainly solar PV) and reducing fossil fuel usage. For example, Philips Real Estate is determined to realise a transition of 75% of all energy consumption by renewable energy sources by 2025 and 90% by 2030. Therefore, building owners are finding smart ways to utilise renewable energy, such as agreements with the national grid to lower electricity prices during peak

hours, battery storage, and solar for EV charging. This is therefore clearly related to Use Case 3 from Section 3 above.

For the above purposes, energy (electricity, water, gas) data is collected building-wise and sometimes equipment-wise.

4.3 Data collection

Building owners depend on data collected from many sources in buildings to achieve the above goals. Following sources and data are commonly used for the purposes mentioned above. However, the building owners do not usually influence the data models, naming conventions, metadata standards, and data storage techniques. Such responsibility is passed on to the platform provider who is providing the systems.

Table 1 Data sources and available data in buildings

Source	
Building Management System	Setpoints, measurements e.g., temperature, alarms, control algorithms for installations, malfunction signals, valve positions, drawings, solar power generation
Energy Management System	Energy flows (gas, electric, water)
Access control system	Personal data, date, time, level of access
Asset Management System	Downtime, maintenance costs
Transport installations (elevators)	Operating hours, energy consumption, communication (elevator cage), malfunction signals
BIM models, AutoCAD drawings	Building information, location, area, installation data

4.4 Future smart building needs for real estate owners

Although the use of data-based decisions is in its early stage, the interviews with respondents showed that building owners are highly enthusiastic about making data-driven decisions for optimising the building operations. Building owners' requirements for a future smart building are shown below, categorised according to the use cases identified in the literature.

Use case 1 – Fault detection and diagnostics

- Identify simultaneous heating and cooling automatically
- Identify faults easily (autonomous) with less involvement by engineers using AI
- Information-based decisions for maintenance, compared to scheduled maintenance.
- Minimise excess energy usage by using AI for fault detection
- Less maintenance by engineers (proactive monitoring, control, and automation)

Use case 2 – Occupancy and occupant behaviour modelling

- Increase occupant comfort level by managing HVAC, lighting, security, space booking holistically
- Provide flexible working spaces for occupants
- Optimal use of building space to reduce energy consumption

Use case 3 – Energy flexibility and Demand Side Management

- Regulating heating and cooling systems based not only on the set points or the number of occupants but also on actual occupancy, weather, and electricity prices.
- Smooth transitioning to aquifer thermal energy stores with electric heat pumps from gas-fired equipment
- Lower the electricity usage from the grid during peak hours by introducing batteries and agreements with the national grid
- Benefit from cheap energy when it is available by using batteries for storing solar energy.
- Understand where energy savings are possible using data.
- Using energy KPIs to benchmark energy usage

Use case 4 – Efficient HVAC control

- Different climate for different rooms, if possible, if it is consuming less energy
- Building's heating and cooling system needs to respond to the actual location of the equipment that generates heat
- Automated decisions to operate the building in the most sustainable way

Use case 5 – Facility management, monitoring and controlling building performance

- Integrate asset management systems with energy management systems for detailed analysis of energy consumption of assets
- Developing BIM models and digital twins to replicate the performance of the building

4.5 Problems and limitations in using data in real estate buildings

Based on the interviews, we recognised the following problems and limitations when utilising data to achieve the targeted goals.

4.5.1 Data collection and management is the responsibility of the service contractor, not the real estate owner

Usually, the data collection is a part of the service provided by a platform provider. Real estate owners do not interact directly with the data; they are rather interested in the information that they receive based on the data. Therefore, a sub-contractor is usually responsible for installing any additional sensors, data collection, storage, analysis, and delivering the information to real estate owners.

Due to the absence of a common information model, building owners are forced to accept the platform provider's information model. Due to the changing nature of contracts with platform providers, the information model that one contractor develops does not agree with one another. As a result, the owner suffers from a lack of understanding of their data.

Consequently, real estate owners (e.g. RVB) do not have a lot of information about the data model, set points, etc. It is considered the contractor's responsibility, and it appears to be a black box to them. According to RVB, real estate owners need to be able to manage their data, but awareness about the value of data and the position of data in business has to grow within the organisation before that.

4.5.2 Fragmented solutions due to large portfolio and many stakeholders

With a large portfolio of buildings, the real estate sector has too many stakeholders involved. The needs and requirements of these stakeholders differ; services and, therefore, data collected by each party are fragmented, making a less clear overall picture to the real estate owner about how to optimally use available data. They need to meet the user's needs while also being energy efficient. They need to know where the set points need to be on the portfolio level for better energy management. Also, users are more demanding of data; for example, with the COVID 19 situation, users demand more data about safety, ventilation, and hybrid work form. They seek answers for,

- Can we use available data to answer these new questions, or do we need new data?
- What data do we need to implement new applications such as a ‘work place checker’ and use spaces optimally?
- What data do we need to plan maintenance and cleaning for operational use?
- How can we provide overall solutions rather than several different solutions?
- How far can we go with available data, and what additional data do we need to collect?

In summary, these questions come down to identifying the target application, reusing available data and collecting any additional needed data.

4.5.3 A large amount of manual work involved in data analysis

The last identified problem is that a lot of manual work is needed to perform data analysis. This problem also relates to the (unsurmountable) "poor data quality" identified in the literature. The amount of manual work involved with data cleaning and analysis is identified as a major limitation and challenge. In RVB, data about energy usage in the buildings (gas, electricity) are analysed to indicate performance leakages and identify options for optimisation using R Suite twice every year, approximately for 100 buildings. Usually, manual data analysis is performed to plan preventive maintenance activities, annual operating plans and budget estimates, and it involves a lot of manual work.

Data validation is also a manual process. Data validation methods are used and indispensable, e.g. checking if the data from one day is different compared to the day or week before, and trying to find out why there is a difference. Automatic methods can be used to calculate values and make comparisons, but results from these methods often need to be validated manually. For instance, if energy use data are exactly two or five times higher than before, there can either be a fault, or any other explanation may be in place (e.g. values may have been set somewhere from a meter). Another limitation, according to RVB, is that there is no direct connection between RVB (data) storage and the analytics environment. Data analytics platform R Suite is hosted in the government's data centre.

5 STATE OF THE ART REVIEW OF EXISTING DATA SOURCES AND DATA PLATFORMS IN BUILDINGS

The built environment is a complex interconnected system, and the list of data sources in a building is potentially endless. We try to include the commonly available systems (non-exhaustive) as available in the literature and living labs selected for the Brains for Buildings project. Throughout a building's lifecycle, different forms of data are used for different purposes. In this report, we focus more on the operational phase of a building and the data that is important in this phase.

Main sources of data in the operational phase of a building include:

- Building Management systems (BMS)
- Energy Management systems (EMS)
- Smart meters and IoT systems
- Asset Management Systems
- Building Information Models (BIM)
- ICT systems and equipment
- External services such as weather
- Post occupancy surveys

However, the systems installed, and their functionalities differ from building to building. Available data in these sources are described under two categories, operational data and contextual data. Operational data can also be referred to as dynamic data. This data is usually available as time-series data continuously streaming from sensors. Contextual data refers to static data such as building information models, floor plans, semantic descriptions of sensors and devices etc. These two categories cannot be described distinctly since many systems described below consist of both categories.

5.1 Operational data

5.1.1 Building Management System

The Building Management System (BMS), also called Building Automation System (BAS), is the dominant source of information in buildings. It is used for monitoring and controlling mainly electrical systems like lighting and power, mechanical systems like HVAC, fire, elevators and securing systems like access control. In general, the main objective of a BMS is to fulfil the occupants' needs while maintaining energy consumption at a minimum. Operational data available from a BMS includes:

- Occupant comfort-related data - Air temperature, humidity, CO2, VOC, CO, air velocity, mean radiant temperature, light level, occupancy
- Energy consumption related data - water/gas/electricity consumption
- Machinery characteristics - refrigerant temperature, electrical machinery current, voltage, power factor, machinery vibration, return and supply air temperature, set points

Other than that, contextual data like floor plans and asset schedules are included in a BMS. Figure 7 shows the screenshot for a SCADA system as managed by Kropman InsiteView. This interface shows both the dynamic data (active state of individual components in the system) and the static data (system structure and properties and names of system components). In this case, the static data is represented using a typical HVAC system schematic. Other BMS systems can use other representations, such as 2D floor plans or 3D model views in the case of facility management use cases (see Section 3).

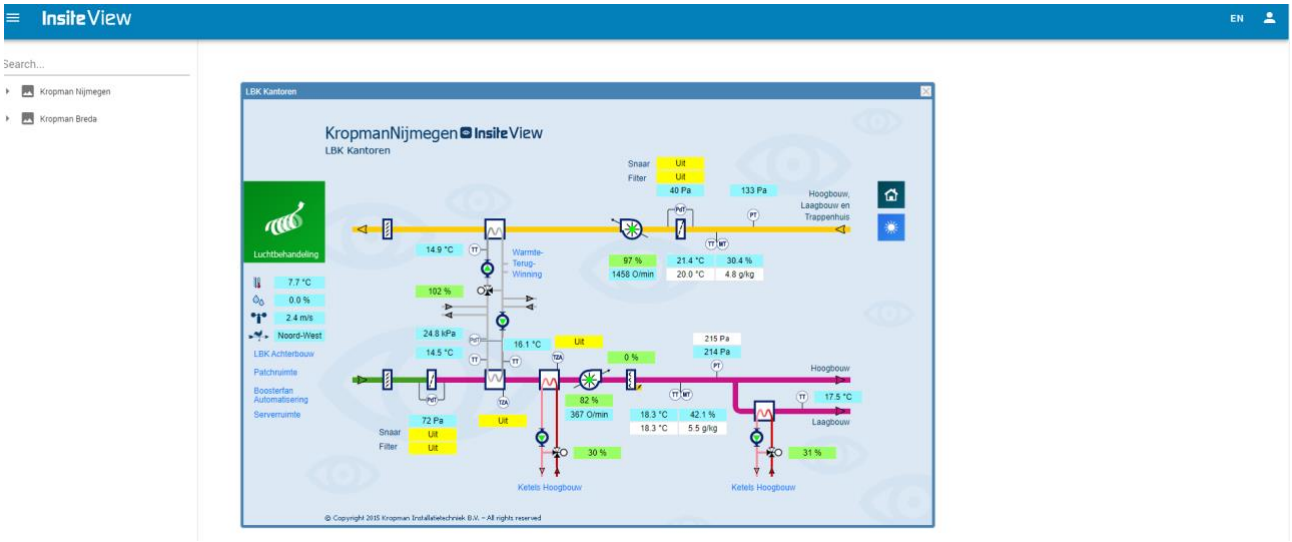

Figure 7 Snapshot of Kropman InsiteView SCADA System

5.1.2 Smart devices and IoT systems

There are many smart devices in buildings with the widespread adoption of IoT technologies. For example, smart energy meters, air quality sensors, smartphones, and watches collect and communicate data using Thread, WiFi, Bluetooth, Zigbee, BLE etc. IoT devices communicate data acquired from sensors and actuators, and communicate these data to the cloud IoT platform using various communication standards like GPRS, WiFi, CoAP, MQTT and HTTP. Matter is the latest protocol to connect to IoT devices. It is an initiative of the Connectivity Standards Alliance, and developed through collaboration amongst all the leaders of the IoT industry. Fig. 8 shows the ecosystem of IoT including communications, data storage and analytics.

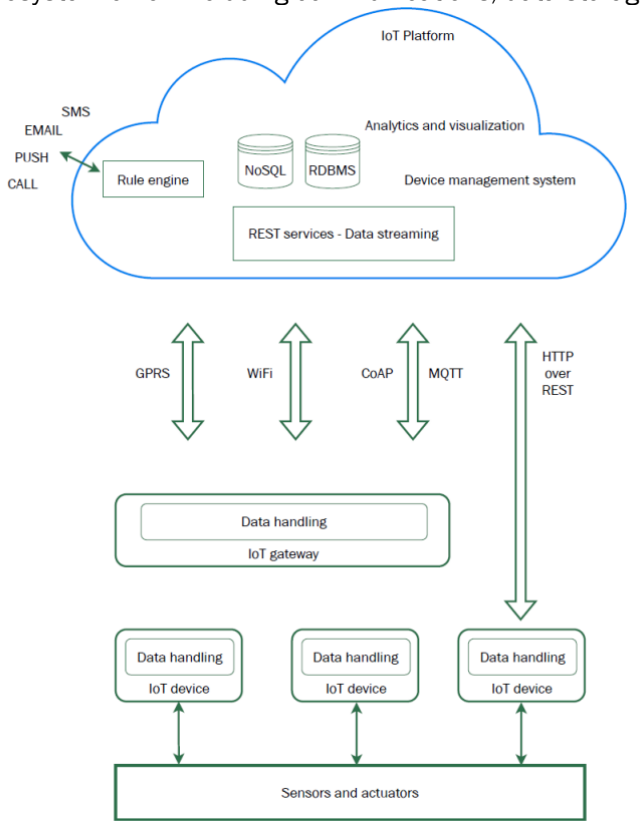

Figure 8 Smart devices in buildings (Pethuru Raj, 2017)

Other systems include asset management systems, maintenance management systems, access control systems etc. However, BMS and IoT devices make the most part of metered sensor data and they are valuable in developing data driven methods, especially related to use cases 1, 2, 3 & 4.

5.2 Building contextual data/ Semantic information

5.2.1 BIM Models

While the operational data included above mainly concerns time series data, another invaluable source of information is contextual information, which is one of the main sources of information in use case 5. Traditionally, the construction industry is known to be "document-centric" (Pauwels & Petrova, 2020). Building information such as structural, architectural and MEP drawings, materials schedules, device characteristics, regulations, specifications, are usually available in the form of 2D drawings and documents. The building Information Model (BIM) (C. M. Eastman et al., 2019) is now an essential component of the Architecture Engineering and Construction (AEC) industry. An example can be seen in Fig. 9 for part of the Atlas building in TU Eindhoven.

Figure 9 BIM Model of Atlas Building TU/e

A BIM model can be defined as the digital representation of a building that contains semantic information about the building elements (Pauwels & Petrova, 2020). A properly developed and managed BIM model is a data source of high fidelity. It includes geometry, spatial location, and a broad representation of metadata about the properties of the building, its subsystems, devices, MEP equipment, etc. (C. Eastman et al., 2011). These may be available in a closed data format (e.g., Revit) or neutral data format (e.g., IFC standard). What does not fit in these BIM models are time-series data (dynamic data), so the BIM model is primarily used as a static source of data.

5.2.2 Sensor Metadata

In a sensor measurement, its *value* and *timestamp* are the key components. However, its metadata is equally important in many aspects. Metadata means data about data. Properties of different equipment, sensors and controls used in buildings are represented using metadata. For example, a sensor's device id, function, location, type, units, manufacturer, relationship with other systems, and any description of the measurement are its metadata. An example of such BMS sensor metadata is shown in Figure 10 for the BMS of Atlas Living Lab.

ItemName	ItemDescriptionDutch	ItemDescriptionEnglish
11NR009LT-030PIRTM	AANWEZIGHEID 9_Z01	PRESENCE 9_Z01
11NR009TE-033CPA	CPA VIA WANDMODULE RUIMTE 9_Z01	CPA VIA WALL MODULE ROOM 9_Z01
11NR009LT-033PIRTM	AANWEZIGHEID 9_Z01	PRESENCE 9_Z01
11NR009TE-030CPA	CPA VIA WANDMODULE RUIMTE 9_Z01	CPA VIA WALL MODULE ROOM 9_Z01
11NR009LT-104PIRTM	AANWEZIGHEID 9_V04	PRESENCE 9_V04
11NR009QT-104CO2	*CO2 METING 9_V04	*CO2 MEASUREMENT 9_V04
11NR009TE-104CPA	CPA VIA WANDMODULE RUIMTE 9_V04	CPA VIA WALL MODULE ROOM 9_V04
11NR009FT-104FLW	FLOW TOEVOER VAV 9_V04	FLOW SUPPLY VAV 9_V04
11NR009LT-102PIRTM	AANWEZIGHEID 9_V01	PRESENCE 9_V01
11NR009QT-102CO2	*CO2 METING 9_V01	*CO2 MEASUREMENT 9_V01
11NR009TE-102CPA	CPA VIA WANDMODULE RUIMTE 9_V01	CPA VIA WALL MODULE ROOM 9_V01
11NR009FT-102FLW	FLOW TOEVOER VAV 9_V01	FLOW SUPPLY VAV 9_V01
11NR009QT-305CO2	*CO2 METING 9_445	*CO2 MEASUREMENT 9_445
11NR009TE-305CPA	CPA VIA WANDMODULE RUIMTE 9_445	CPA VIA WALL MODULE ROOM 9_445
11NR009LT-305PIRTM	AANWEZIGHEID 9_445	PRESENCE 9_445
11NR009FT-305FLW	FLOW TOEVOER VAV 9_445	FLOW SUPPLY VAV 9_445
11NR009FT-304FLW	FLOW TOEVOER VAV 9_442	FLOW SUPPLY VAV 9_442
11NR009LT-304PIRTM	AANWEZIGHEID 9_442	PRESENCE 9_442
11NR009QT-304CO2	*CO2 METING 9_442	*CO2 MEASUREMENT 9_442
11NR009TE-304CPA	CPA VIA WANDMODULE RUIMTE 9_442	CPA VIA WALL MODULE ROOM 9_442
11NR009TE-059CPA	CPA VIA WANDMODULE RUIMTE 9_421	CPA VIA WALL MODULE ROOM 9_421

Figure 10 BMS Sensor Mapping Table

It has been widely discussed in the literature that the available metadata is ambiguous, building and vendor-specific and not machine-readable in many cases (Bhattacharya et al., 2015; Mishra et al., 2020; Stinner et al., 2019; Suzuki, 2015). This makes it difficult to maintain the uniformity of the sensor and hinders the development of applications for analytics without expert prior knowledge of existing systems (Bhattacharya et al., 2015). Further, due to the lack of meaningfulness in the labels, the knowledge that can be inferred is limited.

Brains for Buildings project primarily aims at developing solutions to improve the overall performance of buildings by developing solutions for fault detection and diagnosis (FDD), energy flexibility and improving user experience. More often than not, these applications need knowledge across multiple systems in a building, for example, data from HVAC, fire, lighting, access control system etc. Above systems are usually installed by different vendors, and therefore follow different naming conventions, making it difficult to understand the data. If the metadata is not properly documented, then a lot of interpretation work needs to happen. And even if the metadata is well-documented, the different buildings and systems use diverging standards and conventions, leading to lots of manual work to adapt procedures and systems from building to building. An example naming convention for a room temperature sensor in Atlas Living Lab is 11NR008TE-001TRL which is a temperature sensor (TRL) on the 8th floor (8). This information is available from the mapping table shown in Figure 10, yet often needs to be manually assembled and used.

5.3 Data formats

Most common data formats for exchanging operational time-series data are CSV (Comma Separated Values), JSON (JavaScript Object Notation), Excel (xlsx), XML and pdf. It is much more common that data extracted from a local BMS follows xlsx or CSV format. The source data is typically stored in relational data stores (SQL), yet those stores are seldom directly available from the system provider. These SQL-compliant databases are at best accessible through an API owned by the vendor; hence, users are often limited to these CSV and XLSX file formats. An example data set extracted from the Atlas Living Lab at TUe is shown in Figure 11. The first row shows the sensor labels that need to be matched with the labels in Fig. 10.

Timestamp	11NR008LT-059PIRTM	11NR008LT-301PIRTM	11NR008LT-302PIRTM	11NR008QT-013CO2	11NR008QT-038CO2	11NR008QT-039CO2	11NR008QT-040CO2	11NR008QT-301CO2
1/1/2021 0:02	2	2	2	437	435	436	449	425
1/1/2021 1:02	2	2	2	440	442	445	451	427
1/1/2021 2:02	2	2	2	439	441	444	449	425
1/1/2021 3:02	2	2	2	433	438	443	459	427
1/1/2021 4:02	2	2	2	439	442	442	459	428
1/1/2021 5:02	2	2	2	435	437	445	463	432
1/1/2021 6:02	3	3	3	449	443	447	456	430
1/1/2021 7:02	3	3	3	436	447	444	451	426
1/1/2021 8:02	3	3	3	442	448	448	461	433
1/1/2021 9:02	3	3	3	434	452	458	456	431
1/1/2021 10:02	3	3	3	433	454	452	463	438
1/1/2021 11:02	3	3	3	446	452	450	458	433
1/1/2021 12:02	3	3	3	448	450	450	466	428
1/1/2021 13:02	3	3	3	443	448	447	457	440
1/1/2021 14:02	3	3	3	449	444	449	462	437
1/1/2021 15:02	3	3	3	448	440	447	454	429
1/1/2021 16:02	3	3	3	443	442	443	444	441
1/1/2021 17:02	3	3	3	446	437	446	450	440
1/1/2021 18:02	3	3	3	444	440	448	453	438
1/1/2021 19:02	3	3	3	444	442	445	456	442
1/1/2021 20:02	3	3	3	446	439	448	452	440
1/1/2021 21:02	3	3	3	445	441	449	450	441
1/1/2021 22:02	3	3	3	447	444	447	451	439
1/1/2021 23:02	3	2	3	448	440	449	455	447
1/2/2021 0:02	2	2	2	447	436	448	450	441
1/2/2021 1:02	2	2	2	439	442	448	457	439
1/2/2021 2:02	2	2	2	443	442	446	463	442
1/2/2021 3:02	2	2	2	445	439	441	466	446
1/2/2021 4:02	2	2	2	438	435	447	456	441
1/2/2021 5:02	2	2	2	438	437	451	453	439

Figure 11 Atlas Living Lab data extracted from BMS in excel format

It is very common for an API to return data in JSON format. JSON files are machine-readable and human-readable and transport data in key-value pairs. For example, a JSON response (for a request of data sources) from Kropman InsiteView takes the format shown in Figure 12.

Get Data sources Request - GET

https://ivs.kropman.nl/insitereports/data.php?action=getdatasources&token=c41b0-nzyzzje0ndcynd&sort=id

Get Data sources Response

```
{
  "success":true,
  "datasources":{
    "26":{
      "name":"Kropman Breda [IV32]",
      "type":"InsiteView History",
      "datapoints":{
        "1":"Buitentemperatuur NO [Weerstation] 1-2TT2 [\u00b0C] [3.1.1537]",
        "2":"Buitentemperatuur ZW [Weerstation] 1-3TT1 [\u00b0C] [3.1.1538]",
        "3":"Aanvoertemperatuur [C.V. ketel] [\u00b0C] [3.1.1539]",
        "4":"Retourtemperatuur [C.V. ketel] [\u00b0C] [3.1.1540]",
        "5":"Ber.aanvoertemperatuur Ketel [C.V. ketel] [\u00b0C] [3.1.53021]"
      }
    }
  }
}
```

Figure 12 JSON response from Kropman InsiteView API

Some API responses take XML format, if they are implemented using the older SOAP web service technique. For example, Figure 14 shows the Envision Manager API response for energy consumption in a particular location. Envision Manager is the smart lighting control system in the Atlas Living Lab, TU Eindhoven. It provides a SOAP-based web service for getting and applying illumination levels, correlated colour temperature (CCT) etc. In addition to the SOAP API (XML) or the REST API (JSON), these applications often provide also a web interface where users can log in and visualise the data in dashboards, as shown in Fig. 13.

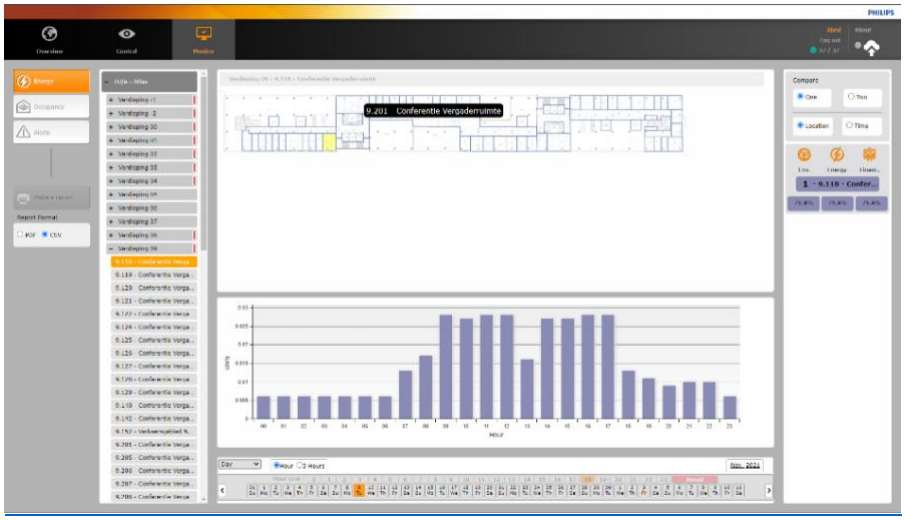

Figure 13 Envision Manager User Interface


```
<locationID>00000000-0000-0001-0000-000000004A44</locationID>
<metricReportList>
<fromDateTime>2021-11-09T00:00:00+01:00</fromDateTime>
<toDateTime>2021-11-09T00:59:59.999+01:00</toDateTime>
<avgKW>0.006</avgKW>
<endAccKWH>67.536</endAccKWH>
<maxKW>0.006</maxKW>
<minKW>0.006</minKW>
<startAccKWH>67.53</startAccKWH>
</metricReportList>
```


Figure 14 API response of the Envision Manager Lighting control system in XML

5.4 Integrating heterogeneous systems and data sets

Smart building applications rely on different systems such as HVAC, lighting, alarms, energy meters, utility, weather services, smart devices, etc., that normally generate time-series data. Other than the time-series data, Building Information Models (BIM) (C. M. Eastman et al., 2019), IFC files, and product databases also contribute to the data in the built environment. However, there is little interaction between these islands of data (Curry et al., 2013; Hu et al., 2016) due to their heterogeneity. Lack of interoperability has been identified as a technical barrier that hinders the research and development of smart building applications. Therefore, a significant body of research and commercial parties investigate how these data silos can be integrated. To that end, various techniques have been developed, such as semantic tagging, data schemas and ontologies for AEC and BAS industry. The below sections provide some documentation and indications in these technological directions.

5.5 Data Models and Schemas

Due to the heterogeneity of building data, a data model is useful to define data points in an understandable and reusable manner. For example, IFC (*Industry Foundation Classes (IFC) - BuildingSMART Technical*, n.d.) is a standardised data model that is intended to describe architectural, building and construction data.

5.5.1 IFC

IFC is an open international standard for BIM data that are exchanged and shared among software applications used by the various participants in the construction or facility management industry sector (*ISO - ISO 16739-1:2018 - Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries — Part 1: Data Schema*, n.d.). It is also a platform-neutral open file format specification.

However, not all aspects of the built environment can be modelled in IFC. For example, a crucial part of the operational phase of a building is the dynamic sensor data, which is ideally not represented using IFC. Another point is that, although the IFC standard describes many common HVAC, lighting, and sensor devices, they are

not capable of representing the context of the devices contained within (Fierro, 2021). IFC is relatively unpopular in the BAS industry and is more used in the context of BIM and the AEC sector for engineering and construction operations. Therefore, several other data models have been developed, and some of them are discussed in what follows.

5.5.2 Other Data models

- iotschema.org (Full Hierarchy - Iotschema.Org, n.d.) is an extension of schema.org that provides vocabularies to enable semantic interoperability for connected things across diverse IoT ecosystems. These vocabularies are developed by "W3C Community Group Schema Extensions For IoT".
- Google Smart Home provides a schema for IoT devices and their actions for a smart home.
- openHAB is an IoT platform for Smart Homes. It proposes a semantic model to logically structure the home and equipment (Semantic Model | OpenHAB, n.d.)

A plethora of other data models are also available, especially for IoT applications, which are not included in the above list for simplicity. The landscape of IoT is diverse in terms of communication protocols and data models for data exchange, making it a challenging situation for developers. Web of Things (WoT) aims to increase flexibility and interoperability for cross-domain IoT applications by introducing a simple interaction abstraction based on properties, events, and actions (Documentation - Web of Things (WoT) (W3.Org), n.d.).

"Smart data Models" (Smart Data Models, n.d.; Smart Data Models - FIWARE, n.d.) is an initiative that aims to develop harmonised data models to support the adoption of compatible common data models. This smart data model includes three elements: the schema (technical representation of the model defining the technical data types and structure), the specification (a written document for human readers) and examples.

5.6 Semantic Tagging

5.6.1 Project Haystack

Project Haystack is an open-source initiative to develop tagging conventions and taxonomies for building equipment and operational data. Haystack is a popular metadata standard among many BAS vendors. Haystack corporation includes industry partners from Honeywell, Siemens, etc. Haystack developed standardised tags for describing data giving the data more meaning by adding metadata to it. This makes data more discoverable for other applications. These tags describe the metadata, and the tags can be applied at the device level, controller level or server level. A full list of tags is available at Project Haystack (project-haystack.org).

The following example shows Haystack tags applied to a temperature sensor in an air handling unit. A set of *name: value* pair (such as unit: "°F") is known as a *dict* (dictionary). Tag names without a value (such as discharge, air, temp) are called *markers*. These marker tags can also be combined (such as temp-sensor).

A Haystack representation can be encoded in file formats Zinc, JSON, Trio, CSV or RDF. An Air Handling Unit (AHU) described in Trio format is shown in Figure 15. In this example, an element is described with identity id:@a-0001. The description of the equipment is given by dis:Alpha Airside AHU-2. Two tags “ahu” and “chilledWaterCooling” describe the chilled water system. Reference to a chilled water plant is made using the chilledWaterRef:@a-07b8 line. Tags that follows “elec”, “equip” and “hotWaterHeating” describe the hot water system which is referred by using the reference hotWaterRef:@a-07da.

```
id:@a-0001
dis:Alpha Airside AHU-2
ahu
chilledWaterCooling
chilledWaterRef:@a-07b8
elec
equip
hotWaterHeating
hotWaterRef:@a-07da
hvac
singleDuct
siteRef:@a-0000
vavZone
---
```

Figure 15 Haystack example of tagging of an Air handler Unit in Trio format
(Source: Project Haystack – Examples)

5.7 Ontologies in building domain

Ontologies are used for integrating heterogeneous databases due to their independence from lower-level data models (Gruber, 1993). In contrast to data models, these ontologies are independent of particular applications since they consist of generic knowledge that can be reused by different applications (Spyns et al., 2000). Several vocabularies and ontologies such as BASont (Ploennigs et al., 2012), SAREF (Daniele et al., 2015), SSN (Haller et al., 2018), RealEstateCore (Hammar et al., 2019), Haystack (*Project Haystack - Tags*, 2020), and Brick (*Brick Schema Building Blocks for Smart Buildings*, 2019), BOT (Rasmussen et al, (2017)), PRODUCT, and PROPS etc. have emerged to fulfil different information domains in different phases in the building lifecycle. However, to the best of the authors' knowledge, Haystack, Brick and REC ontologies are the ones that are most commonly recognised by BAS practitioners and therefore, they are discussed in brief in what follows.

5.7.1 Haystack Ontology

The Haystack ontology provides standardised modelling of site, space, equipment, point, device and weather entities (Points – Project Haystack, 2022), which are described briefly below.

- site: single building with its own street address
- space: location or zone within a site
- equip: physical or logical piece of equipment within a space
- point: sensor, actuator or setpoint for an equip
- weatherStation: weather station observations
- device: computers, controllers, networking gear

An example AHU representation using Trio encoding was shown earlier in Figure 15. When modelled using the Haystack Ontology, the result is the data representation shown in Figure 16. The Haystack ontology uses the ph:hasTag relationship to represent the tags associated with the equipment. To define the relationships with other equipment, it uses phIoT:chilledWaterRef, and phIoT:hotWaterRef. Similar other relationships and properties can be represented as preferred, leading to a graph of data and metadata.

```
_:a-0001 a phIoT:ahu ;
    ph:hasTag phIoT:ahu,
        phIoT:chilledWaterCooling,
        phScience:elec,
        phIoT:equip,
        phIoT:hotWaterHeating,
        phIoT:hvac,
        phIoT:singleDuct,
        phIoT:vavZone ;
    rdfs:label "Alpha Airside AHU-2" ;
    phIoT:chilledWaterRef _:a-07b8 ;
    ph:dis "Alpha Airside AHU-2" ;
    phIoT:hotWaterRef _:a-07da ;
    phIoT:siteRef _:a-0000 .
```

Figure 16 Haystack example of representing an Air handler Unit using Haystack Ontology in Turtle format (Source Project Haystack – Examples)

5.7.2 Brick Ontology

The Brick ontology is designed with a focus on supporting energy applications based on BMS in commercial buildings. It is capable of representing physical, logical and virtual assets in buildings and the relationships between them. Brick uses OWL and RDF standards to define the ontology. Figure 17 shows an example that represents common HVAC components using Point, Location and Equipment classes and their relationships in Brick schema.

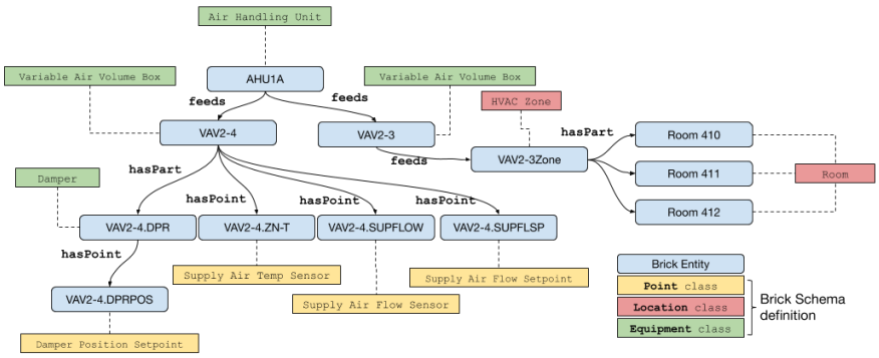

Figure 17 Modelling with Brick (Brick, 2022)

Figure 18 shows a real-world example of representing an AHU using Brick.

```
soda_hall:ahu_A1 a brick:AHU ;
    brick:feeds soda_hall:vav_C180,
        soda_hall:vav_C300,
        soda_hall:vav_C300B,
        soda_hall:vav_R784 ;
    brick:hasPoint soda_hall:ahu_occpy_SODA1____OCCPY,
        soda_hall:ahu_start_stop_SODA1______S_S,
        soda_hall:curtl_SODA1____CURTL,
        soda_hall:override_event_SODA1____EVENT,
        soda_hall:rat_SODA1_LOW_RAT2,
        soda_hall:rat_SODA1_LOW_RAT3,
        soda_hall:rat_SODA1_LOW_RAT4,
        soda_hall:rat_SODA1__LOW_RAT,
        soda_hall:smoke_alarm_SODA1_SMK_ALM2,
        soda_hall:smoke_alarm_SODA1_SMK_ALM3,
        soda_hall:smoke_alarm_SODA1_SMK_ALM4 ;
    brick:isLocationOf soda_hall:supply_fan_S11 .
```

Figure 18 Brick representation of an example AHU (Source: Brick Soda Hall)

A comparison of Haystack and Brick ontologies is done by Quinn & McArthur (2021), revealing the following important facts.

- Both ontologies are machine-readable since they are serialised in standard formats.
- Freedom to use flexible semantics allows Haystack to represent a variety of buildings that falls outside the norm. However, this flexibility compromises the portability of smart building applications across buildings.
- Since Brick represents concepts in a prescribed manner, it allows smart building applications to be portable across buildings.
- Haystack uses the concept of Tags to represent semantic concepts, and any inconsistency of representations leads to difficulty in reading and relating the tags to real-world meaning. When the ontology is not human readable, it becomes difficult for a user to query.
- Brick supports interoperability with other ontologies such as BACS, SAREF because they all use the same ontology language (RDF format). This allows for making more holistic smart building applications. Haystack is still in the process of serialising to RDF.

5.7.3 RealEstateCore Ontology (REC)

RealEstateCore is an OWL ontology that enables data integration for smart buildings. REC is developed by a consortium including some of the largest real estate companies in Northern Europe. REC focuses on merging and bridging four domains: Business administration, Digital representation of the building's elements – BIM, Control and operation of the building BMS, and IoT technologies.

The example in Figure 19 shows three VAV devices VAVL1.01, VAVL1.02, and VAVL1.03 that belong to an AHU, AHUL1.01. Their associated sensing is defined using the *hasCapability* relationship.

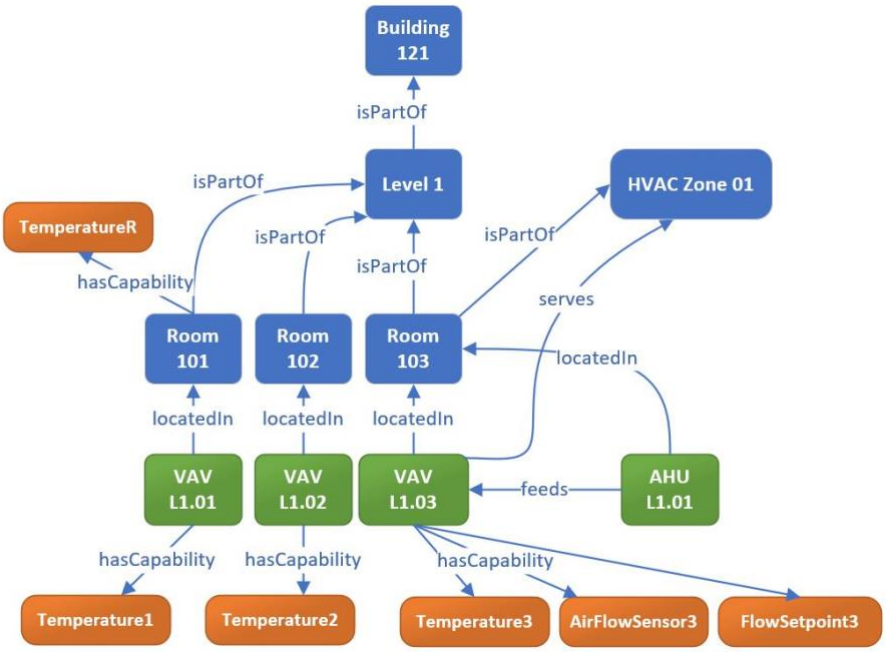

Figure 19 RealEstateCore Ontology example usage in Azure Digital Twin (Source : realestatecore.io) (Hammar et al., 2019)

Above three ontologies as well as other ontologies that apply to the Smart Building domain are shown in Figure 20. This diagram shows that IFC and Haystack take an object-oriented approach while other ontologies such as SOSA, SSN, OSPH, BOT, ifcOWL, SAREF and Brick take a Descriptive Logic approach.

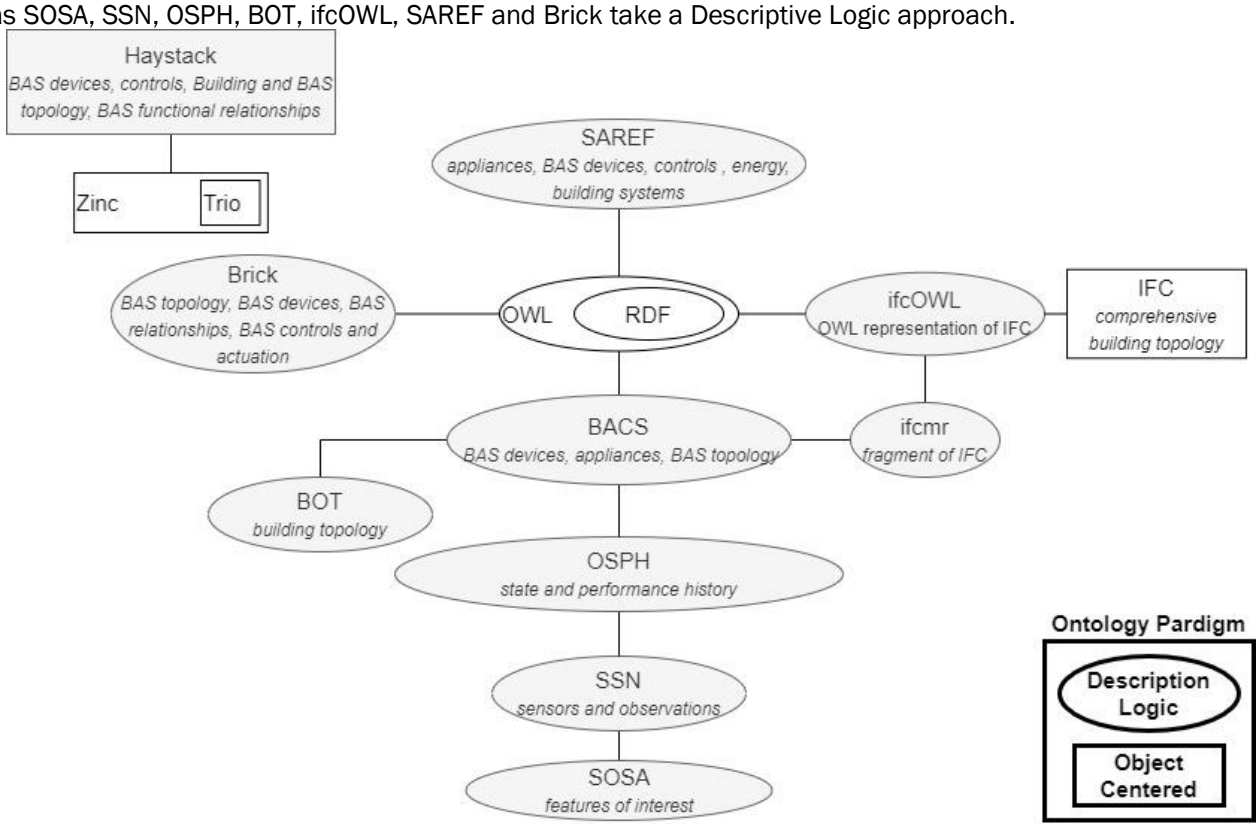

Figure 20 Ontologies in the smart building domain (Quinn & McArthur, 2021)

5.7.4 Industry Alignment with Metadata Schemas

Literature reviews and overall market studies indicate that the Brick ontology and Haystack standard have gained traction in the building industry and are the most popular metadata standard in the building industry.

- According to the Brick Schema website, Johnson Controls and Schneider Electric are commercial members of the Brick consortium.
- According to the Project Haystack website, its board members include Conserve It, J2 Innovations, Legrand, Lynxspring, Siemens and SkyFoundry. They also have 23 associate members, including KNX Association and Tridium.
- ASHRAE's BACnet Committee is collaboratively working with Project Haystack and Brick Schema (ASHRAE, 2018; *Haystack Connect | The Place for the Project Haystack Community to Network, Share, Create Syner-Gy, and Generate New Business Opportunities*, n.d.). According to the ASHRAE website, *"the collaboration will result in unification and standardisation enabling interoperability on semantic information across the building industry, particularly in building automation. The unified effort is meant to ease the exchanging of data over established communication protocols like Haystack web services or BACnet, and when applied on data stored in databases and cloud applications, will support machine interpretation of the semantics of such data"* (ASHRAE, 2018).

RealEstateCore also has an ecosystem certification program for companies, organisations, and individuals that provide applications based on the RealEstateCore API (CERTIFICATION - RealEstateCore, n.d.).

5.7.5 Semantic graphs

A semantic graph can be used to describe a building and its systems using standard syntax and semantics. This semantic graph can then be used to identify the relationships and properties of each component in the building. An example of a semantic graph of Atlas Living Lab in TU/e is shown in Figure 21 Part of the graph representation of Atlas Living Lab. This graph has been described using Brick ontology and BOT ontology.

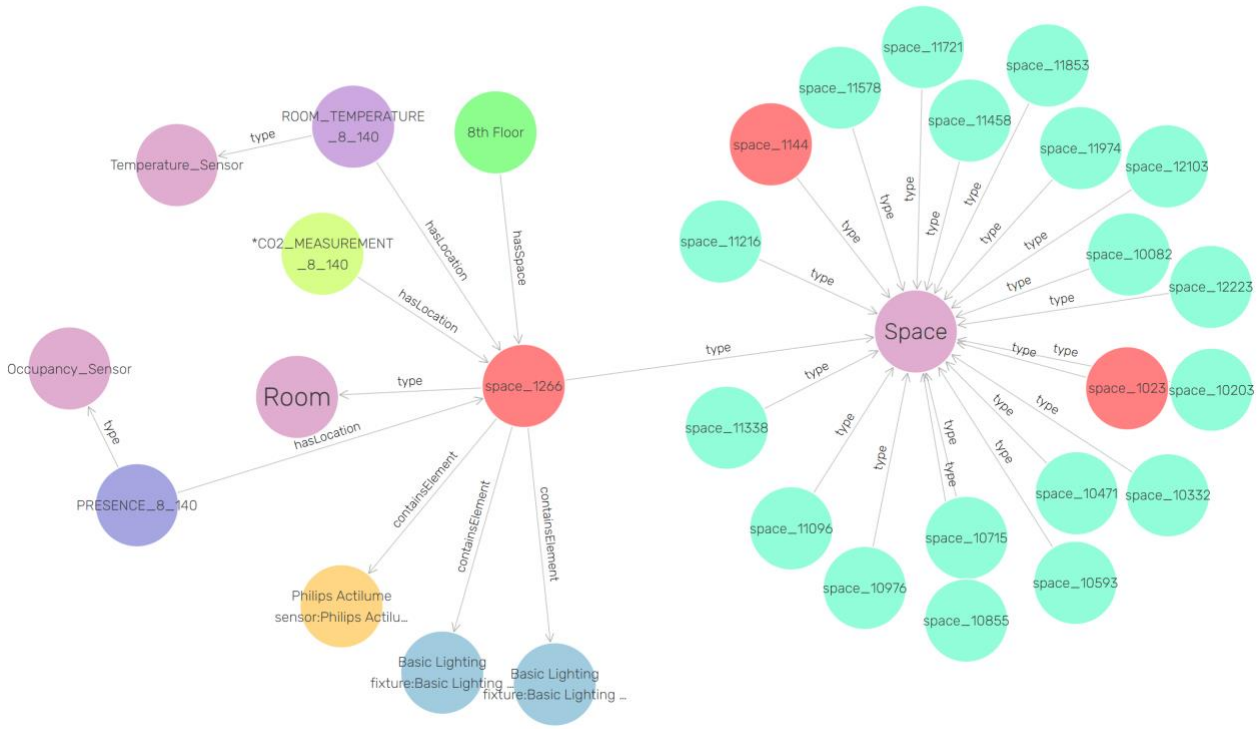

Figure 21 Part of the graph representation of Atlas Living Lab

5.8 Data platforms for smart buildings

Penetration of Digitalization, the Internet of Things, Digital Twin, Big Data, Cloud Computing, and Artificial Intelligence in the built environment is increasing. There are numerous heterogeneous connected devices and controllers within a building that need to be coordinated for the above advancements. This coordination is beyond the capability of traditional BMSs, and therefore, buildings rely on other software solutions, referred to

as middleware. Middleware resides between the hardware devices and other data sources that produce data and the applications using these data (Prakash et al., 2020).

- Vendor-specific software:

  Some of these solutions originate from well-known building industry market leaders such as Honeywell, Schneider, Johnson Controls, Siemens etc. One example is EcoStruxure™ Building Advisor from Schneider. Vendor-specific software solutions primarily allow integrating their own devices and applications.

- Vendor-neutral Platforms:
  Some other companies provide software and analytics that are interoperable with hardware and sensors from other vendors. One such widespread solution is the Tridium Niagara Framework. It facilitates an open, no vendor lock-in framework to connect different systems and is supported by a community of System Integrators and Developers. Other vendor-neutral IoT Platforms include ThingSpeak, The Things Network, FIWARE, openHAB and Tuya, which allow the integrating smart sensors and meters from different manufacturers.
- Open-source platforms:
  Furthermore, some efforts aim to create open data platforms, making such data platforms available to the public. Three such examples of open-source data platforms are BEMServer, XBOS and BEMOSS. In what follows, we try to briefly identify their main features according to authors, but it is not the intention of this writing to confirm their functionality.

BEMServer (Bourreau et al., 2019; *Home - BEMServer, the World's Premier Open Source Building Energy Management Platform*, n.d.) is an open-source platform with,

- Storage techniques for the building data model, time-series data and events.
- A data model based on IFC and Haystack called BEMOnt
- Data pre-processing 9unit conversion, time series resampling, data cleansing)
- A REST API
- Data visualisation tool based on Grafana

According to XBOS (Fierro & Culler, 2015; *XBOS Overview - XBOS Docs*, n.d.), this open-source building operating system provides the following features.

- Hardware Presentation Layer
- Canonical Metadata Definition, Storage, and Usage
- Control Process Management
- Building Evolution Management
- Security: Authorisation and Authentication
- Scalable UX and API

BEMOSS (*BEMOSS*, 2015) is also an open source software designed for small and medium sized commercial buildings featuring an open source architecture, interoperability, scalability, and security. It supports Zigbee, Modbus, BACnet, HTTP, openADR and Smart Energy protocols.

IEA EBC - Annex 81 - Data-Driven Smart Buildings, **Subtask A also aims at developing Open Data and Data Platforms** to develop methodologies that support the implementation of real-time Open Data sharing in buildings.

6 MARKET STUDY ON EXISTING DATA AND SMART BUILDING DATA PLATFORMS

6.1 Introduction

Building owners continuously seek better monitoring and controlling mechanisms, analytic algorithms, and intelligent devices for delivering services that make buildings more energy efficient. These functionalities cannot be met by using legacy BMS systems due to their limited integration capabilities; their analytics are inadequate for the job, they lack the required software applications, and they have legacy user interfaces (Sinopoli, 2016). Therefore, most buildings use services from building platform providers that have **solutions to integrate multiple systems and provide useful information via analytics, reporting and data-driven applications**.

Building solution platforms have been developed with a variety of features and complexity to handle the complexity of modern buildings. The global smart building market is dominated by giants such as Honeywell, ABB, Johnson Controls, Schneider Electric, Siemens etc. Priva is also dominant in the Netherlands. These systems provide a variety of data in diverse formats and approaches. In this Section 6, the aim is to evaluate these data sources and find out to what extent they can respond to the data needs and requirements listed in Section 3 and 4.

In chapter 3 and 4, we identified what the data needs and requirements in buildings are. To investigate to what extent software solutions and systems are able to supply the needed data, we interviewed three platform providers in the Netherlands that are part of the B4B project: Kropman, Cloud Energy Optimizer and Simaxx Technologies. Hence, the results in this Chapter are composed using a method that is identical to the method used in Chapter 4, only with different participants and different questions. Semi-structured interviews were held with the participants. A questionnaire was prepared initially as a guideline for the interviews (see Appendix 2). These questions aim at understanding existing and emerging data needs, data types, data collection and analysis methods, security, etc. Upon receiving the consent from the participants (See Appendix 3 for the consent form), several interview sessions were conducted. After the interviews, a qualitative analysis was again performed, similar to what was documented in Section 4, to understand the state of the art, future directions and barriers when using data in the smart building sector.

This section summarises the needs arising from the platform providers from the perspective of the three parties that were interviewed.

6.2 Solutions available in smart building platforms

The main functionality of a smart building platform provider is to collect, analyse and provide meaningful information to the end-user of the building. Some of the functionalities identified by the platform providers who participated in the interviews are described below according to the five use cases identified from the literature and market study in Chapter 3.

6.2.1 Use case 1 - Fault detection and diagnostic analysis and alerts

Detecting and diagnosing problems in buildings can be done with the help of people with expert knowledge. But this is time-consuming and error-prone work. Automated diagnostics can use rule-based techniques such as maximum/minimum limits and other machine learning algorithms to detect faults. These applications use algorithms that search for anomalous data that represent faults and suggest possible causes and remedies. They use alerting via email or alarm systems to indicate the faulty situation.

Another fault detection case is automatic checking and reporting of energy usage during the weekend and after office hours provide insights into anomalous energy consumption in a building. When the energy consumption is outside the predefined boundaries, the system gives a warning.

Cloud Energy Optimizer uses Machine Learning to predict the future indoor temperature in the building and then generate a warning if the building is to become too cold. Here, the temperature setpoint is not fixed and is determined based on the outside temperature.

The reliability of the data is ensured by mainly checking its magnitude. If a reading remains the same throughout the day or if the value exceeds set limits, that could indicate a faulty reading. Detection of faulty meter reading is often performed using Machine Learning.

6.2.2 Use case 2 - Occupancy and occupant behaviour predictions

No applications were found under this use case. Collecting and analysing occupant related data was not very common among the platform providers. The practitioners that we interviewed mentioned that it is due to the lack of such requirements from the building owners. Further research would need to look into existing systems to support this use case.

6.2.3 Use case 3 - Energy flexibility and Demand Side Management

Forecasting services predict future energy consumption values based on historical data. Energy consumption forecasts use outside weather data, historical energy usage data, occupancy, and indoor climate data. One of the primary uses of forecasting is to manage energy resources cost-effectively. When an installation has multiple energy sources like Solar PV, Battery Storage, and thermal energy storage, these forecasts can be used to plan the energy mix most economically. Figure 22 shows an example from Kropman where they use ten weeks' average energy consumption data to predict and benchmark future energy consumption.

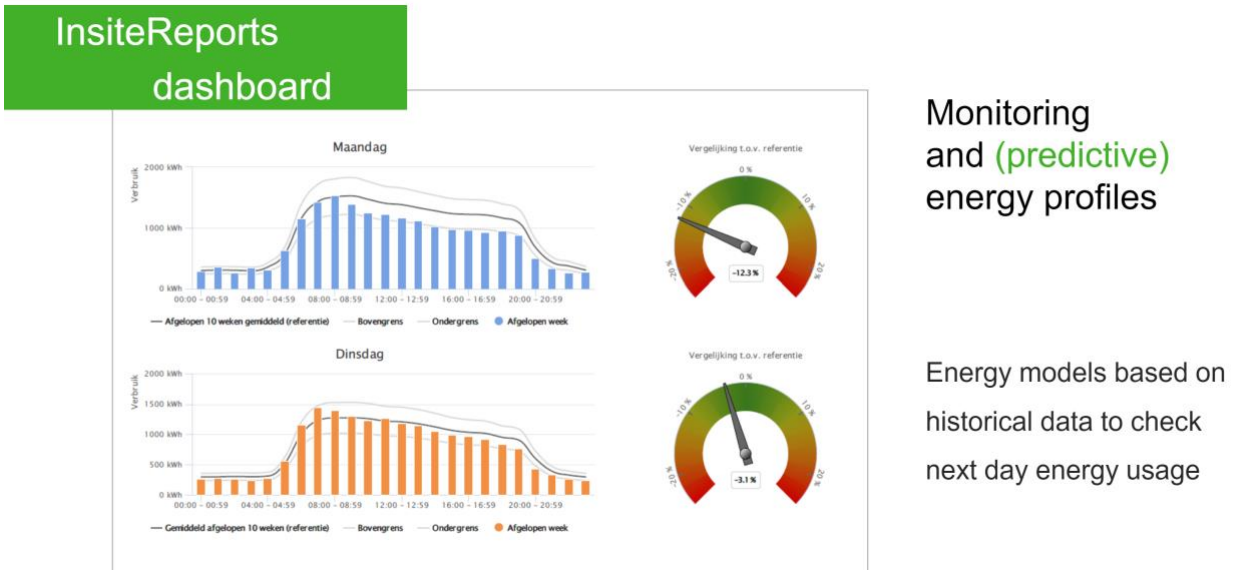

Figure 22 Kropman uses the past ten weeks' average energy consumption to benchmark and predict future energy consumption.

6.2.4 Use case 4 - Efficient HVAC control

Depending on the season and the outside temperature, comfortable indoor temperature can be different. Controlling the HVAC system efficiently needs collecting this data and fine tuning the indoor temperature setpoints to optimize the energy consumption (example in Fig. 23). This gives an insight into whether the building is overcooling or undercooling, thereby making room for improving comfort and energy usage.

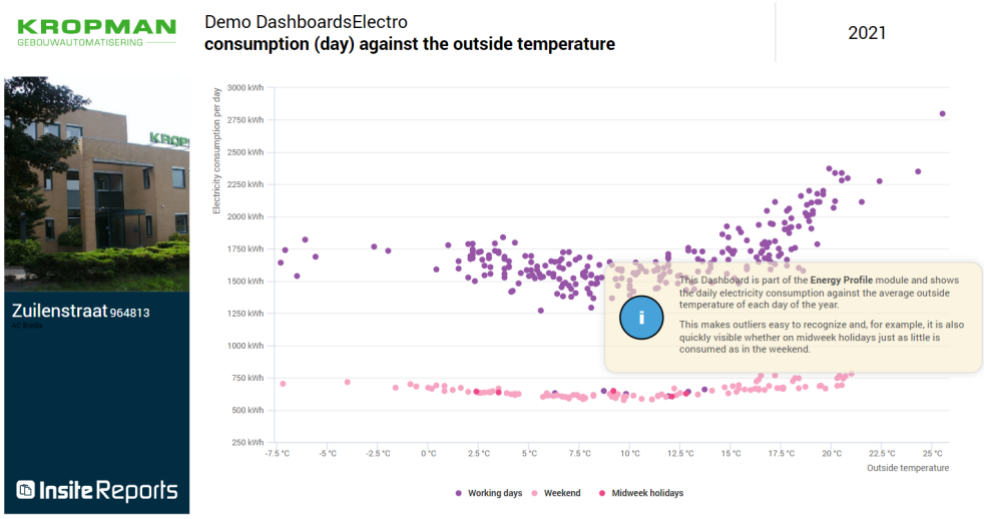

Figure 23 Electricity consumption based on the outside temperature

Other than the weather data and indoor temperature data, important information such as location of the sensors, orientation of the building, area of the building, building material, and volume of concrete are considered essential inputs by the platform providers to calculate the optimum control strategy for HVAC systems. Figure 24 shows how Cloud Energy Optimizer estimates the amount of thermal energy stored in walls and floors. The amount of 'stored' thermal energy in the building mass determines how much heating is required in the upcoming period. This is an important input because the thermal energy embedded in building materials change according to the outside temperature.

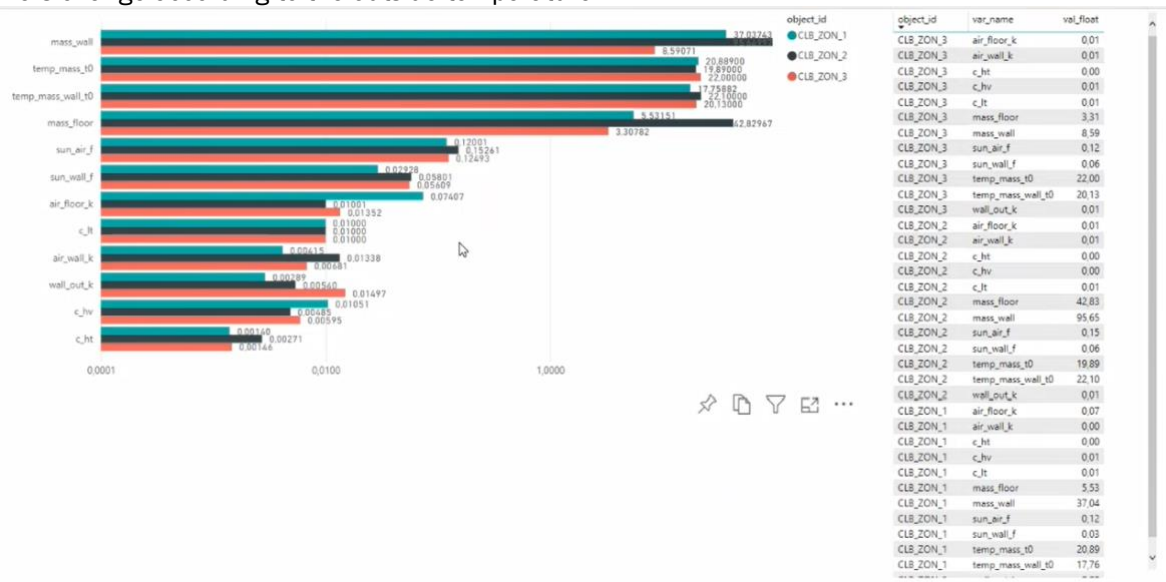

Figure 24 Using thermal energy embedded in walls/floors to predict next-day energy usage. This allows taking seasonal variations n the temperature into the predictions. Source Cloud Energy Optimizer

6.2.5 Use case 5 – Facility management, monitoring and controlling

These smart building platforms provide data analysis, reporting and dashboards to show compliance with regulations. Functions that they offer include, monitoring and reporting of Legionella, monitoring and reporting for energy compliance requirements such as periodic reporting to comply with government regulations on the usage of ATES (heat and cold storage), system, energy regulations from RVO to report energy usage, BREEAM-NL and Reports for social housing to comply with warmteWET rules on the supply of heat to consumers. Figure

25 shows a dashboard for monitoring Legionella, where the warm water temperature is continuously monitored for its maximum, minimum and average temperatures.

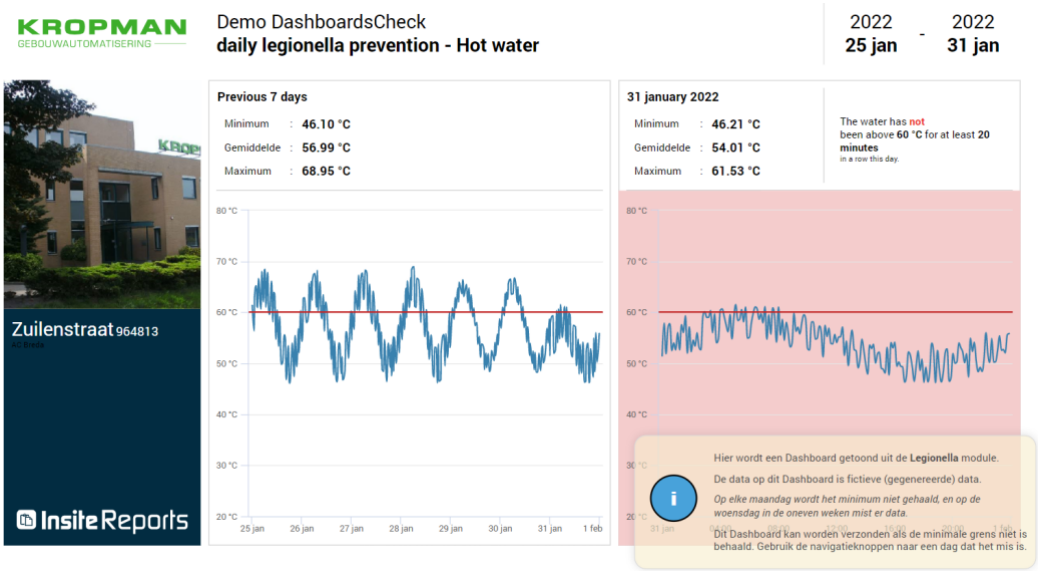

Figure 25 Dashboard for monitoring Legionella by Kropman

Aquifer thermal energy storage (ATES) is the storage and recovery of thermal energy in the subsurface. ATES is applied to provide heating and cooling to buildings. Storage and recovery of thermal energy is achieved by extraction and injection of groundwater from aquifers using groundwater wells. Buildings need to monitor the extraction of this thermal energy by law, and they are monitored as shown in Figure 26 in the case of Kropman.

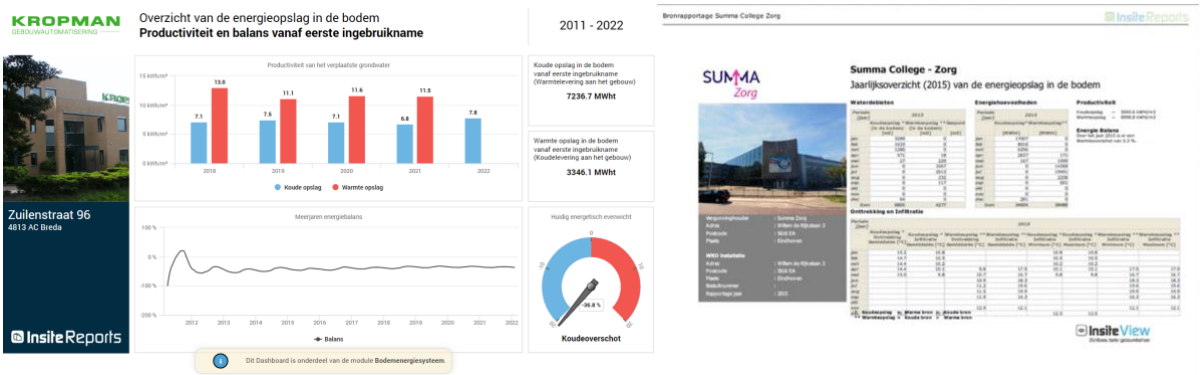

Figure 26 Monitoring and reporting Aquifer Thermal Storage by Kropman

Monitoring and reporting of indoor environmental parameters such as CO2 levels and temperature to comply with standards for indoor climate is much common in buildings. Figure 27 shows a dashboard for monitoring CO2 levels and Figure 28 shows the indoor temperature monitoring dashboard.

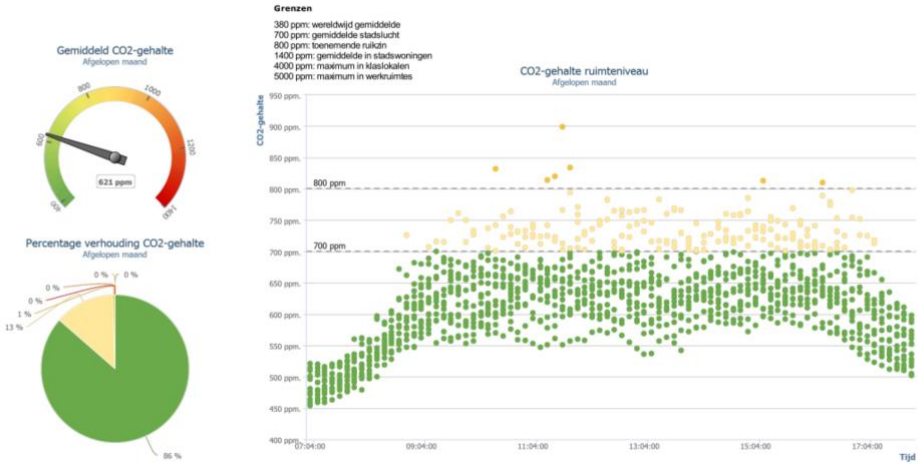

Figure 27 Monitoring CO2 levels by Kropman

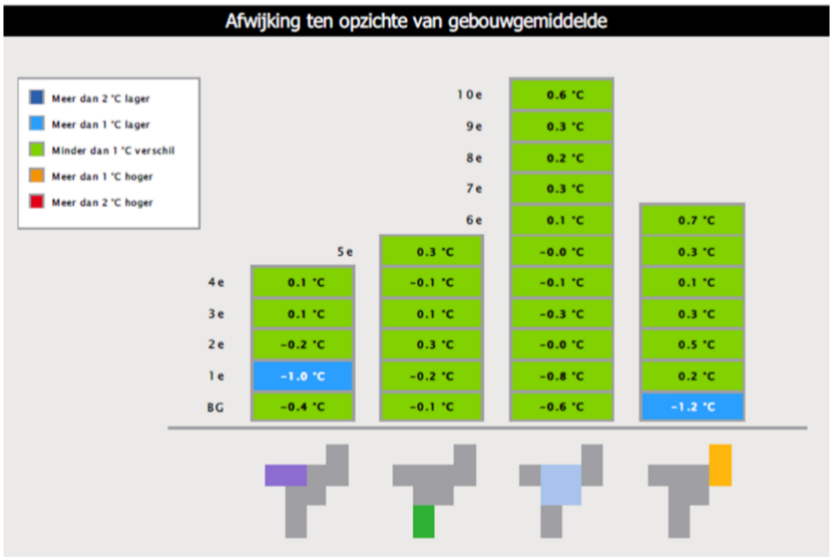

Figure 28 Aggregation of room temperature in a dashboard by Kropman

Displaying historical average energy consumption of the building on a yearly, monthly, daily and hourly basis is shown in Figure 29. This data is available from the BMS or EMS.

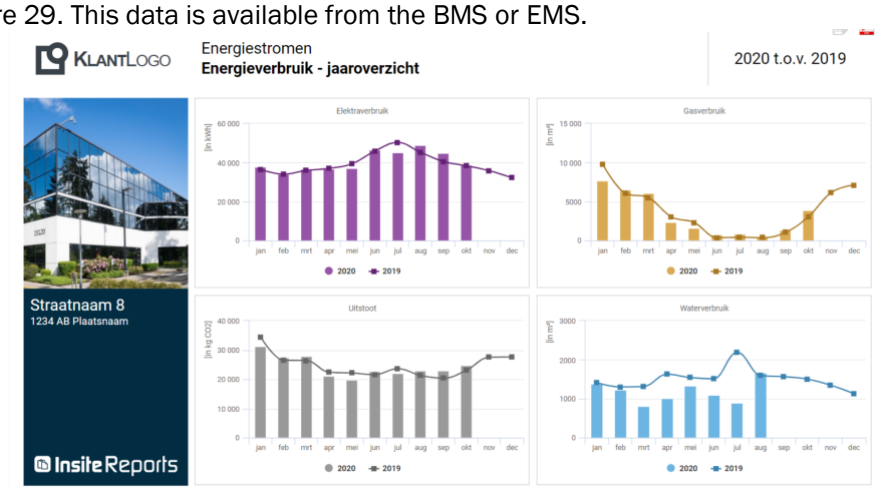

Figure 29 Usage of electricity, water, gas and CO2 emission by Kropman

Collecting the data from the grid operator, EV charging stations (such as newMotion), and solar PV generation companies via their APIs periodically gives an overview of how much energy the solar panels are producing and how much energy the EVs are consuming (Fig. 30).

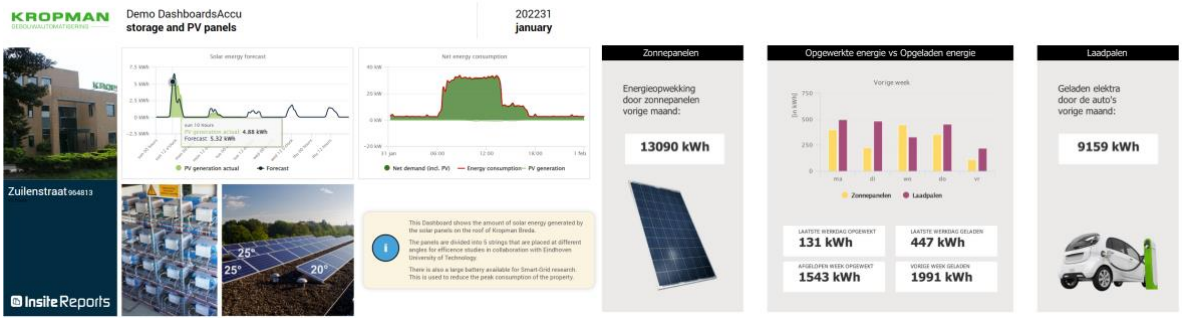

Figure 30 Renewable energy and battery storage dashboard by Kropman

The dashboard in Figure 31 shows monthly KPIs for temperature, humidity, air quality, electricity, gas and CO2 production over the past 12 months. Each KPI is determined based on adjustable limit values that are agreed upon with the customer.

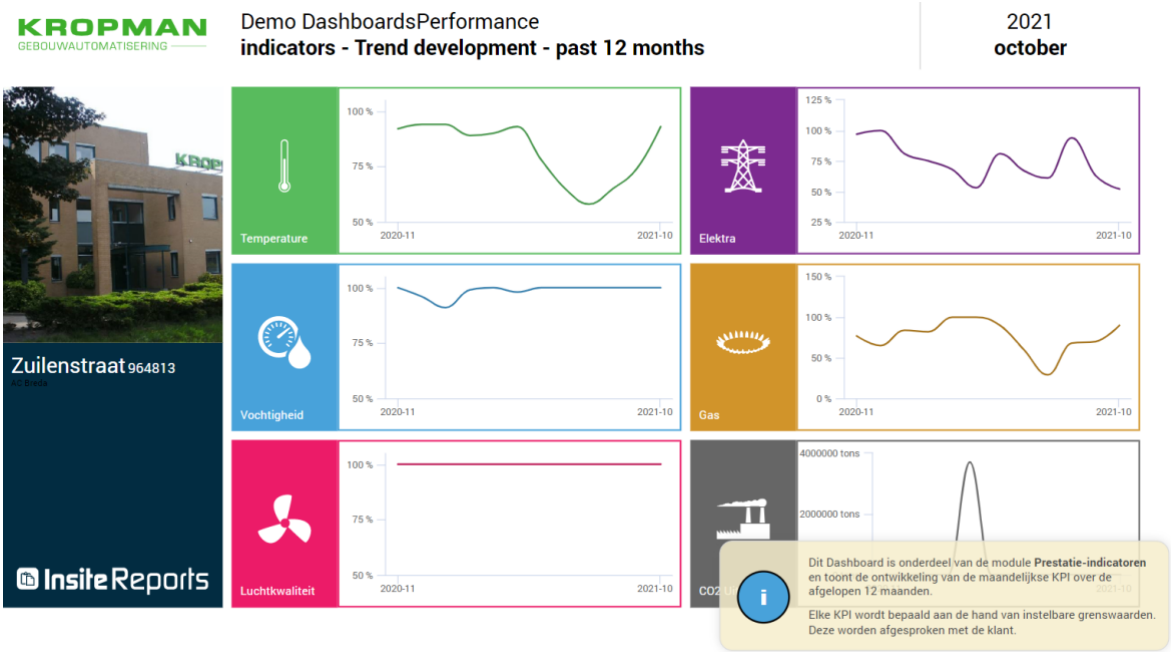

Figure 31 KPI monitoring dashboard by Kropman

In conclusion, the majority of the applications available from the platform providers tend to be on the side of monitoring and reporting (use case 5). We also note that they are currently using MPC and Machine Learning techniques for forecasting energy demand and maximising energy flexibility (use case 3). Using Machine Learning for fault detection and diagnosis of HVAC systems (use case 1) is actively being investigated at the moment. While the available data from BMS are sufficient for improving performance, there is a lack of data related to the occupant (Use case 2). More data about the occupant behaviour and indoor environmental parameters can be used to improve the indoor environment quality, predict occupancy, and use spaces optimally.

6.3 Data Sources and data collection

Smart building applications depend on data collected from the building and external services. Data is collected from local building management systems, EMS, IoT devices etc., using various protocols such as BACnet, Modbus, MQTT, and OCPI (see Figure 32). For other closed protocols used in BMSs like PRIVA, Schneider, and Johnson Controls, connectors and plugins are used to extract data. One platform provider (Simaxx) mentioned that although they used their drivers, now they are using libraries provided by Niagara Framework®, a software platform to manage and control diverse systems and devices regardless of manufacturer or protocol. Using

these libraries allows faster onboarding, and their developers can focus more on developing applications rather than integration problems. Extracted data are usually transferred to the platform provider's cloud to be used in other applications.

With emerging information needs such as energy predictions, optimum renewable energy usage and energy flexibility, data from external sources such as weather forecasts and predictions of prices in the Dutch utility market (day-ahead prices) are also linked to the data collection.

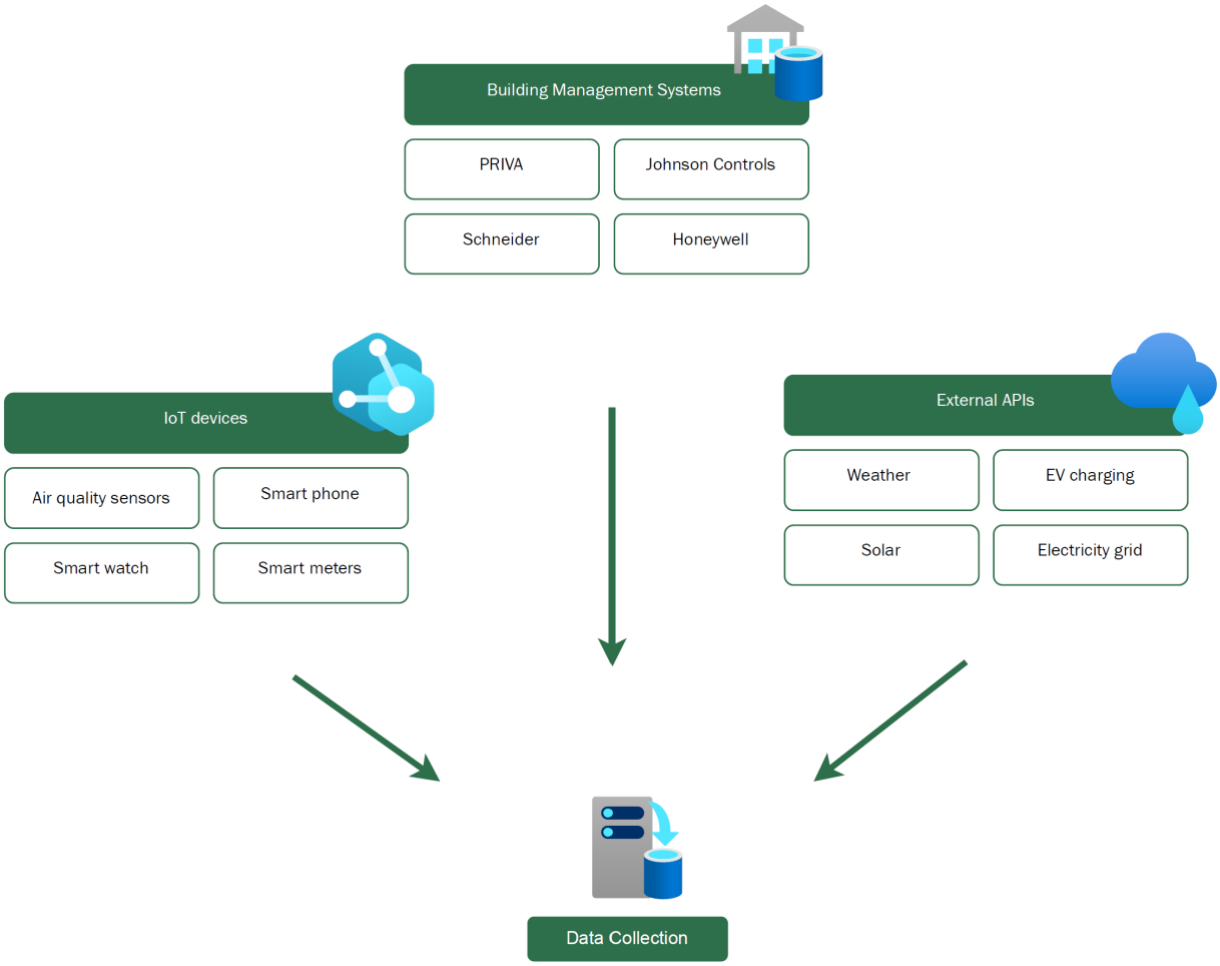

Figure 32 Data collection from different sources

6.4 Data Storage

Two data storage techniques are commonly used in smart building platforms. They are relational and non-relational databases. Most BMS systems store time-series data in an **SQL database**. Some common SQL databases include Microsoft SQL Server, MySQL, Oracle Database, and PostgreSQL. **NoSQL databases** are popular in cloud-based solutions. Some examples are MongoDB, Cassandra and Cosmos. NoSQL databases overcome the scalability issue and complexity and processing overhead in relational databases. These NoSQL databases are schemeless, meaning there are no tight structures and relationships defined.

Other than those two prevalent techniques, there are emerging data storage solutions, namely time-series databases and graph databases. A **time-series database** (TSDB) is a database that is optimised for storing and serving data with a timestamp. Sensor data from IoT devices is a very good example of time series data where the datasets are composed of a timestamp and a value. InfluxDB, Prometheus and MongoDB Timeseries are some commonly used time-series databases. They provide significant performance improvements in terms of storage compared to general-purpose databases.

With the addition of linked building data and ontologies, **graph databases** are emerging in the smart buildings field for storing and querying the RDF data in the database. Examples are GraphDB and RDF4J. Though storing time series is possible in the RDF graph, it has proven to be inefficient (Pauwels et al., 2015) and also not suitable for list-oriented analysis algorithms (Pauwels et al., 2022). However, it is valuable and useful to

include the links to online identifiers where time-series data can be retrieved in the graph (Pauwels et al., 2022). This feature is available in the BRICK Schema where *brick:time-series* relationship relates a Brick point to the TimeseriesReference that indicates where and how the data for this point is stored (Timeseries - BrickSchema, n.d.). An overview of the mentioned data storage solutions is given in Figure 33.

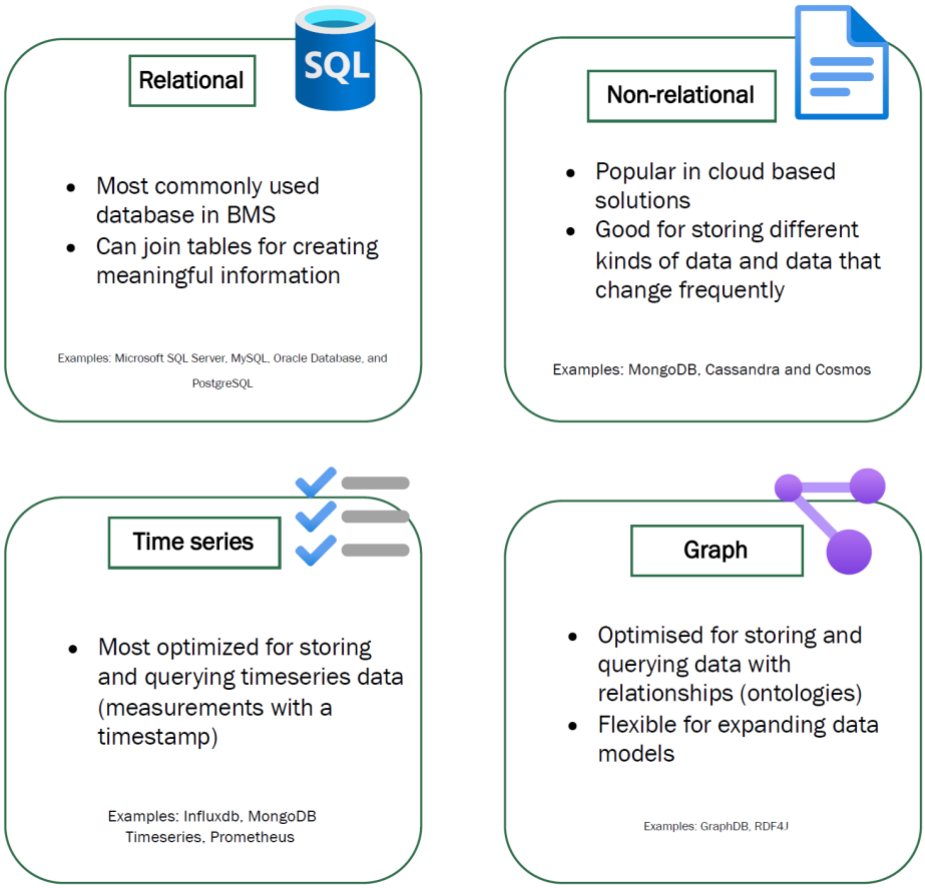

Figure 33 Databases for storing different data types

6.5 Future needs and requirements for smart building platforms

Using data-driven methods for optimising building performance – We observed that using MPC and Machine Learning techniques for forecasting energy demand and maximising energy flexibility are already initiated by the smart building platforms. Machine Learning will also be used for fault detection and diagnosis of HVAC systems in the future. While the available data from BMS are sufficient for improving performance, there is a lack of data related to the occupant. More data needs to be collected, including occupant feedback and indoor environmental parameters, to improve the indoor environment quality, predict occupancy, and use spaces optimally.

Cloud-based services, privacy and data security - We observed that every solution provider we interviewed uses cloud services for data storage and analytics. This overcomes the limited storage and processing capabilities in existing BMS systems. Also, offering web-based access to remotely manage single or multiple buildings located at geographically different sites allows building managers and owners to view and benchmark the performance of each building in terms of energy usage and services and detect any unusual behaviour. While VPN, authentication and authorisation, and encryption provide a secure environment, an in-depth review of data policies and security measures is needed. This topic is discussed in Deliverable 4.01 in detail and is kept out of scope here.

Metadata standardisation - Implementing standard semantic descriptions like Brick or Haystack is widely discussed within platform providers. Their decision on selecting which ontology, vocabulary or method to choose is determined by choosing a particular standard by building controllers' giants like Schneider, Honeywell, etc., and the continuous support provided by academia and industry.

Extending the information model to improve maintenance - Extending the information model to include a product's manufacturer's information, product datasheets, and link to the repair and maintenance team will improve workflow management by including maintenance information. The possibility of using BIM models to extend existing information models is being investigated. It is expected here that graph databases and ontology-based RDF graphs provide useful avenues for further research and development in this direction.

6.6 Problems and limitations to implementing smart building platforms

The investigated smart building solutions and their implementations in buildings often face several barriers and obstacles. The following section describes the main barriers identified according to platform providers.

- **Insufficient data** - Newer applications are being introduced based on data-driven methods. Although available data is sufficient to improve energy performance, more data is required to improve indoor air quality since there is no sufficient building data available for indoor air quality improvement.
- **Data Quality** - Data quality is low in the sense that there are missing and erroneous data that need to be cleaned up before performing data analytics. Sometimes, specific values do not have the correct magnitude with reporting or with the SCADA systems. Cloud Energy Optimizer and Simax mentioned that they use Machine Learning algorithms to check data accuracy when preparing data for algorithms.
- **Out of date data** - Out of date documentation and drawings make it challenging to find out the current status of the building and its components.
- **Unstructured naming conventions** – It is often difficult to find data points because of the ambiguous naming conventions used. Data is not self-descriptive, and this makes it tedious to identify the data type, its provenance, and intended usage for practitioners and data scientists. This demands an extra technical expert to give context to the data. Usually, a translation of available points to the desired format by the platform provider is required. Also, the technical know-how of how a building and its controllers operate is a prerequisite for the smart building application developer. In the current scenario, understanding the building and its controllers is extremely hard for, for example, a data scientist. Therefore, methods should be developed such that the data is self-descriptive and can be understood by an inexperienced person.
- **Data discovery** from legacy systems- Although BAS suppliers claim they deliver open systems, participants mentioned that they are not fully open. It is difficult to discover data points because they are hidden and transported in a scrambled manner.
- **Data ownership and security** - In some cases, there is an ambiguity about the ownership of data, privacy and policy for sharing. This topic is discussed in detail under Deliverable 4.01 of this B4B project.

7 DEFINITION OF AN IDEAL BEST-CASE SCENARIO

This last section lists a few key take-aways for data collection and integration plan from the literature and market study.

7.1 Key take-aways for data collection

The development of smart building applications (five use cases) relies heavily on data acquired from multiple systems. These systems are,

- Building Management systems (BMS)
- Energy Management systems (EMS)
- Smart meters and IoT systems
- Asset Management Systems
- Building Information Models (BIM)
- ICT systems and equipment
- External services such as weather, grid
- Post occupancy surveys

And the data acquired from the above sources can be categorised as,

- Metered time-series data from sensors
- Data from external services
- Occupant feedback
- Domain expert knowledge
- Contextual data/ metadata

From Chapter 3 through Chapter 6, we discussed the smart building use cases, their data needs, requirements and also the limitations which hamper the development of applications and systems in support of such use cases. To summarise, developing and implementing smart building applications encounter following problems related to data.

- Unavailability or partial availability of required data
- Low-quality data (accuracy, low resolution, timeliness, missing samples)
- Unstructured data (naming conventions, metadata standards, data formats)
- Limitations due to accessibility, privacy and ownership

Therefore, a proper data collection method must resolve the above problems and facilitate the development of smart building applications.

7.1.1 Improve data availability

There can be several reasons behind the unavailability or partial availability of data. The main reason is the absence of a reliable data collection and management principle. Often, a lot of parameters are being measured, but they are not systematically collected. Therefore, it is important to collect and manage data in a reliable way. Data replication and backup can be used to ensure that the collected data is not lost.

If a particular sensor is absent from the BMS, other methods like using IoT devices such as smartphones, watches, thermostats, and Indoor Air Quality (IAQ) sensors can be used to collect data from the building. These devices' ability to communicate using WiFi and Bluetooth provides ample opportunities for collecting data without intervening the available systems. However, it is also important to store and manage these data, because the IoT platforms usually do not maintain historical data for a longer period of time by default.

Another economical way of collecting data is via virtual sensors (Ran et al., 2020; Reppa et al., 2014) when physical sensors are unavailable or costly to install. Virtual sensors derive an estimated value for a reading using the data available from other devices. Also, when in lack of historical data, data can be collected by simulating the required condition (such as faults) when a building is not in full use.

7.1.2 Improving data quality

By using multiple redundant sensors, data from these redundant sensors can be used to compare and validate the readings. This gives an opportunity to correct any incorrect values. It is also important to calibrate the sensors properly to avoid erroneous readings. Details about calibration can be stored using a metadata

storage technique for verifying. Sometimes, datasets are available from buildings, but they are not in good quality due to uncalibrated sensors and meters, making that data useless.

Data should be accurate, complete, up to date and consistent in order to provide valuable information. Since there can be many discrepancies in data, selection of a suitable data pre-processing technique should be made by the researcher or user according to their requirements. It is important to note that the techniques used for treating erroneous and missing values vary from one application or researcher to another, such as replacing with average, median, arbitrary value, the previous value, interpolation etc.

7.1.3 Structuring the data

Naming conventions differ significantly from building to building. Some naming conventions are documented in a mapping table, while others are not. Using a self-descriptive naming convention makes it easy to understand the data and its meaning.

To avoid the inconsistencies brought by naming conventions, data must be semantically annotated using a well-established standard. Using widely accepted standards to express syntax and semantics of data can make it easier to understand them, as well as integrate with other systems. Many such data models are available in the literature: e.g. Brick, Haystack, SSN, SOSA, iotschema. These existing data models should be used where possible.

Data must be associated with its human- and machine-readable metadata. Making them machine-readable allows reasoning and validation. Since it is not practical to have all those information in the name itself, tags that follow a convention must be used. Also, relevant metadata should be stored in a database separately (Capehart & Brambley, n.d.).

7.1.4 Accessing data adhering to privacy and ethics guidelines

Collecting and using data from privacy invasive sensors requires additional security measures. Other than that, there are data policies that govern the sharing of data between parties. Accessibility and privacy-related issues are a very broad topic and are covered in detail in the already published *Deliverable 4.01 Literature study and market study of existing regulations and approaches regarding privacy, ethics, and security, including GDPR constraints*, available on www.brainsforbuildings.org.

7.2 Data Integration reference architecture

Data sources or systems communicate using different protocols and exchange data in different formats standards. Therefore, these data are often siloed, and they are not available via a single platform, meaning that there is little to no interaction between them. Therefore, it is difficult to produce useful information and insights based on them. However, we understood that more and more data-driven smart building applications seek data from multiple systems to make informed decisions about the energy, maintenance and occupant. These applications also need to write back their outputs to the building controllers. Therefore, developing methods to integrate those silos is important to allow users to access all the data they need, develop applications and write back into the building controllers.

The ability to integrate systems from multiple vendors is a major requirement for building owners as well. Building owners are struggling to integrate all the data they receive from different vendors (assets, energy, maintenance), to understand the energy usage, maintenance, and overall performance of their buildings. Platform provider companies also need to integrate different systems in their solutions such as BMS, BIM, external weather services, EV charging stations, and utility companies. But those systems use different data models and APIs, making it difficult to integrate them without dedicated coding.

Above findings suggest that a reference architecture should have following features.

- Provide methods for data acquisition, storage, manage and backup to improve data availability
- Provide standardisation to data originating from different systems using standard schemas and well recognised data models
- Provide access to historical and real time data for monitoring and controlling applications via APIs, reducing the complexity of accessing data to the user
- Support bidirectional communication between the smart building applications and building controllers. This requires connection with control APIs of building control systems.
- Provide scalable solutions that can accommodate growing number of data sources and users.

– Data ownership, security and privacy measures should be considered when sharing data among different applications and users. Other than building information, there are many other private and sensitive information such as financial, personal records, occupant's information, etc. As more and more buildings use cloud solutions, these data become vulnerable to external attacks and unauthorised access. Many security and privacy measures have been developed to protect this information in different stages like transmission, storage, and processing. For more information, please refer to the *Deliverable 4.1 Literature study and market study of existing regulations and approaches regarding privacy, ethics, and security, including GDPR constraints*, available on www.brainsforbuildings.org.

Considering these needs and requirements, the ideal best-case scenario for a data collection and integration plan for smart building applications can be drawn up as shown in Figure 34.

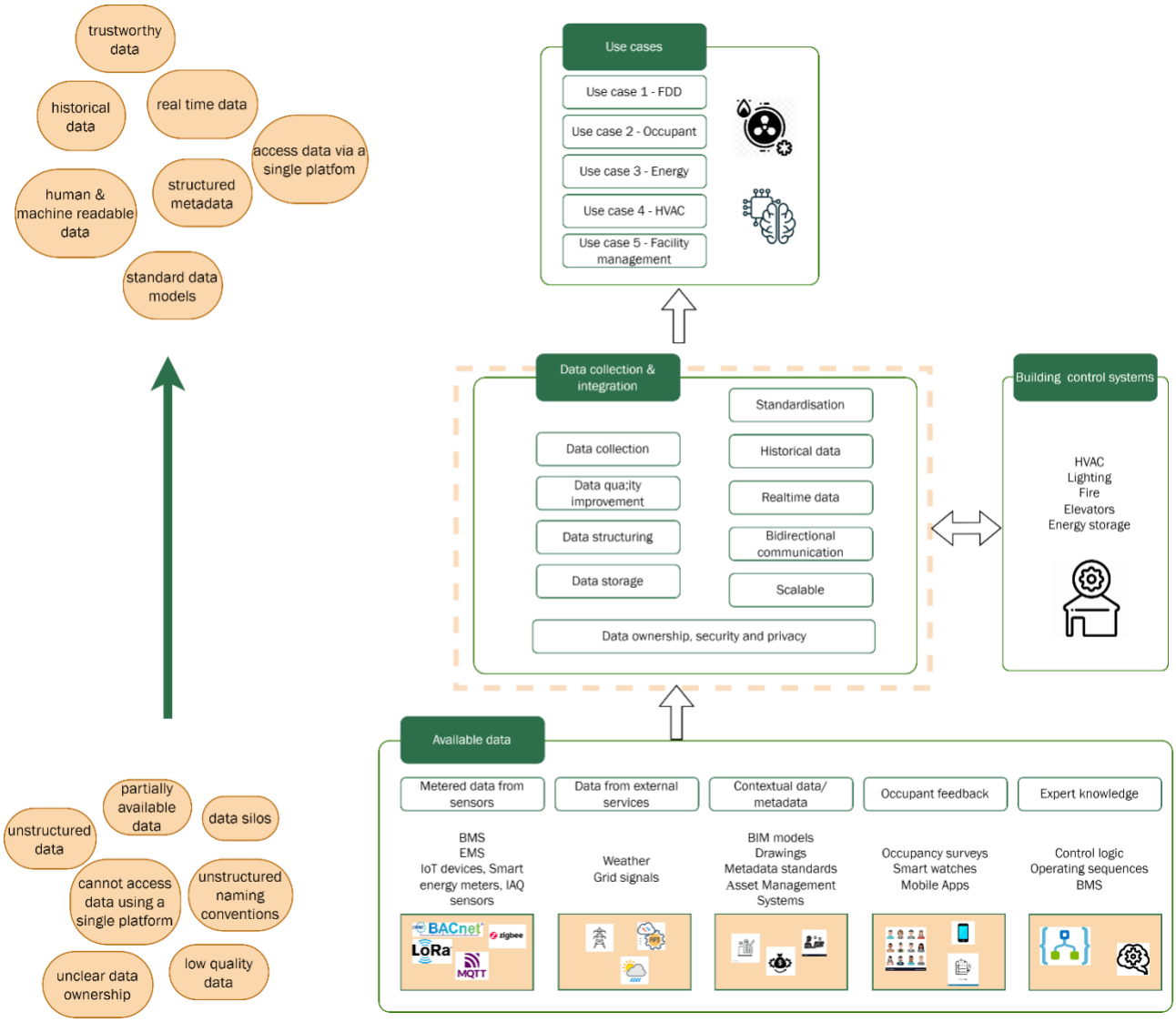

Figure 34 Ideal best-case scenario for data needs and requirements of smart building applications

REFERENCES

Ardakanian, O., Bhattacharya, A., & Culler, D. (2016). Non-intrusive techniques for establishing occupancy related energy savings in commercial buildings. *Proceedings of the 3rd ACM Conference on Systems for Energy-Efficient Built Environments, BuildSys 2016*, 21–30. https://doi.org/10.1145/2993422.2993574

ASHRAE. (2018). *ASHRAE BACnet Committee Works with Other Organizations on New Standard*. https://www.ashrae.org/news/esociety/ashrae-bacnet-committee-works-with-other-organizations-on-new-standard

BEMOSS. (2015). https://www.bemoss.org/

Bhattacharya, A. A., Hong, D., Culler, D., Ortiz, J., Whitehouse, K., & Wu, E. (2015). Automated metadata construction to support portable building applications. *BuildSys 2015 - Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built*, 3–12. https://doi.org/10.1145/2821650.2821667

Bolchini, C., Geronazzo, A., & Quintarelli, E. (2017). Smart buildings: A monitoring and data analysis methodological framework. *Building and Environment*, *121*, 93–105. https://doi.org/10.1016/j.buildenv.2017.05.014

Bourreau, P., Chbeir, R., Cardinale, Y., Corchero, A., Lafréchoux, J., Frédérique, D., Salameh, K., Calis, G., Constantinou, R., & Kallab, L. (2019). BEMServer: An Open Source Platform for Building Energy Performance Management. *Proceedings of the 2019 European Conference on Computing in Construction*, *1*, 256–264. https://doi.org/10.35490/ec3.2019.176

BREEAM - Sustainability Assessment Method. (2015). https://www.breeam.com/

Brick. (2022). *Home - BrickSchema*. https://brickschema.org/

Brick Schema Building Blocks for Smart Buildings. (2019). *March*. https://brickschema.org/docs/Brick-Leaflet.pdf

Capehart, B. L., & Brambley, M. R. (n.d.). *Automated Diagnostics and Analytics for Buildings*. Fairmont Press, Inc.

Central Government Real Estate Agency. (2018). *Handbook RVB BOEI (RgdBOEI) | BOEI inspections | Central Government Real Estate Agency*. https://www.rijksvastgoedbedrijf.nl/documenten/richtlijn/2018/03/01/handboek-rvb-boei-inspecties-deel-1-algemeen

CERTIFICATION - RealEstateCore. (n.d.). Retrieved February 7, 2022, from https://www.realestatecore.io/certification/

Chen, Y., Lin, G., Crowe, E., & Granderson, J. (2021). Development of a unified taxonomy for hvac system faults. *Energies*, *14*(17). https://doi.org/10.3390/en14175581

Corry, E., Pauwels, P., Hu, S., Keane, M., & O'Donnell, J. (2015). A performance assessment ontology for the environmental and energy management of buildings. *Automation in Construction*, *57*, 249–259. https://doi.org/10.1016/j.autcon.2015.05.002

Cox, R., Walker, S., van der Velden, J., Nguyen, P., & Zeiler, W. (2020). Flattening the electricity demand profile of office buildings for future-proof smart grids. *Energies*, *13*(9). https://doi.org/10.3390/en13092357

Curry, E., O'Donnell, J., Corry, E., Hasan, S., Keane, M., & O'Riain, S. (2013). Linking building data in the cloud: Integrating cross-domain building data using linked data. *Advanced Engineering Informatics*, *27*(2), 206–219. https://doi.org/10.1016/j.aei.2012.10.003

Daniele, L., den Hartog, F., & Roes, J. (2015). Created in Close Interaction with the Industry: The Smart Appliances REFerence (SAREF) Ontology. *Lecture Notes in Business Information Processing*, *225*, 100–112. https://doi.org/10.1007/978-3-319-21545-7_9

Documentation - Web of Things (WoT) (w3.org). (n.d.). Retrieved February 28, 2022, from https://www.w3.org/WoT/documentation/#web-of-things-in-a-nutshell

Dong, J., Winstead, C., Nutaro, J., & Kuruganti, T. (2018). Occupancy-based HVAC control with short-term occupancy prediction algorithms for energy-efficient buildings. *Energies*, *11*(9), 1–20. https://doi.org/10.3390/en11092427

Eastman, C. M., Tang, S., Shelden, D. R., Pishdad-Bozorgi, P., Gao, X., Tang, S., Shelden, D. R., Pishdad-Bozorgi, P., & Gao, X. (2019). A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. *Automation in Construction*, *101*(January), 127–139. https://doi.org/10.1016/j.autcon.2019.01.020

Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). BIM Handbook, a Guide to Building Information Modelling 2nd ed. In *John Wiley & Sons, Inc, Hoboken*.

European Commission. (2008). *Energy efficiency: delivering the 20% target*.

Fierro, G. (2021). *Self-Adapting Software for Cyberphysical Systems*. http://www2.eecs.berkeley.edu/Pubs/TechRpts/2021/EECS-2021-159.html

Fierro, G., & Culler, D. E. (2015). Poster abstract: XBOS: An eXtensible Building Operating System. *BuildSys 2015 - Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built*, 119–120. https://doi.org/10.1145/2821650.2830311

Finck, C., Li, R., & Zeiler, W. (2019). Economic model predictive control for demand flexibility of a residential building. *Energy*, *176*, 365–379. https://doi.org/10.1016/j.energy.2019.03.171

Full Hierarchy - iotschema.org. (n.d.). Retrieved February 28, 2022, from http://iotschema.org/docs/full.html

Gellings, C. W. (1985). The Concept of Demand-Side Management for Electric Utilities. *Proceedings of the IEEE*, *73*(10), 1468–1470. https://doi.org/10.1109/PROC.1985.13318

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Aquisition*, *5*(2).

Haller, A., Janowicz, K., Cox, S. J. D., Lefrançois, M., Taylor, K., Le Phuoc, D., Lieberman, J., García-Castro, R., Atkinson, R., & Stadler, C. (2018). The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. *Semantic Web*, *10*(1), 9–32. https://doi.org/10.3233/SW-180320

Hammar, K., Wallin, E. O., Karlberg, P., & Hälleberg, D. (2019). The RealEstateCore Ontology. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11779 LNCS*. https://doi.org/10.1007/978-3-030-30796-7_9

Haystack Connect | The place for the Project Haystack community to network, share, create syner-gy, and generate new business opportunities. (n.d.). Retrieved February 7, 2022, from https://www.haystackconnect.org/wp-content/uploads/2019/05/Proposed-ASHRAE-Standard-223P-Bernhard-Isler.pdf

Home - BEMServer, the world's premier open source building energy management platform. (n.d.). Retrieved February 24, 2022, from https://www.bemserver.org/

Hu, S., Corry, E., Curry, E., Turner, W. J. N., & O'Donnell, J. (2016). Building performance optimisation: A hybrid architecture for the integration of contextual information and time-series data. *Automation in Construction*, *70*, 51–61. https://doi.org/10.1016/j.autcon.2016.05.018

Industry Foundation Classes (IFC) - buildingSMART Technical. (n.d.). https://technical.buildingsmart.org/standards/ifc

ISO - ISO 16739-1:2018 - Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema. (n.d.). Retrieved January 6, 2022, from https://www.iso.org/standard/70303.html

Lazarova-Molnar, S., & Mohamed, N. (2016). Challenges in the data collection for diagnostics of smart buildings. *Lecture Notes in Electrical Engineering*, *376*(November), 941–951. https://doi.org/10.1007/978-981-10-0557-2_90

Lazarova-Molnar, S., & Mohamed, N. (2019). Collaborative data analytics for smart buildings: opportunities and models. *Cluster Computing*, *22*, 1065–1077. https://doi.org/10.1007/s10586-017-1362-x

Lazarova-Molnar, S., Shaker, H. R., Mohamed, N., & Jorgensen, B. N. (2016). Fault detection and diagnosis for smart buildings: State of the art, trends and challenges. *2016 3rd MEC International Conference on Big Data and Smart City, ICBDSC 2016*, 344–350. https://doi.org/10.1109/ICBDSC.2016.7460392

Lymperopoulos, I., & Jones, C. N. (2018). Experimental Implementation of Frequency Regulation Services Using Commercial Buildings. *IEEE Transaction on Smart Grid*, *9*(3), 1657–1666.

Ma, Y., Kelman, A., Daly, A., & Borrelli, F. (2012). Predictive control for energy efficient buildings with thermal storage: Modeling, stimulation, and experiments. *IEEE Control Systems*, *32*(1), 44–64. https://doi.org/10.1109/MCS.2011.2172532

Martani, C., Lee, D., Robinson, P., Britter, R., & Ratti, C. (2012). ENERNET: Studying the dynamic relationship between building occupancy and energy consumption. *Energy and Buildings*, *47*, 584–591. https://doi.org/10.1016/j.enbuild.2011.12.037

Maslesa, E., & Jensen, P. A. (2019). Drivers for IWMS implementation in real estate management. *Journal of Corporate Real Estate*, *21*(1), 72–87. https://doi.org/10.1108/JCRE-08-2018-0025

Mayne, D. Q., Rawlings, J. B., Rao, C. V., & Scokaert, P. O. M. (2000). Constrained model predictive control: Stability and optimality. *Automatica*, *36*(6), 789–814. https://doi.org/10.1016/S0005-1098(99)00214-9

Mishra, S., Glaws, A., Cutler, D., Frank, S., Azam, M., Mohammadi, F., & Venne, J. S. (2020). Unified architecture for data-driven metadata tagging of building automation systems. *Automation in Construction*, *120*, 1–20. https://doi.org/10.1016/j.autcon.2020.103411

Mugnini, A., Coccia, G., Polonara, F., & Arteconi, A. (2020). Performance assessment of data-driven and physical-based models to predict building energy demand in model predictive controls. *Energies*, *13*(12). https://doi.org/10.3390/en13123125

O-Prognose. (n.d.). Retrieved February 27, 2022, from http://www.nen-2767.nl/index.html

Palensky, P., & Dietrich, D. (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics*, *7*(3), 381–388. https://doi.org/10.1109/TII.2011.2158841

Pauwels, P., Costin, A., & Rasmussen, M. H. (2022). *Knowledge Graphs and Linked Data for the Built Environment*. Springer International Publishing. https://doi.org/10.1007/978-3-030-82430-3_7

Pauwels, P., & Petrova, E. (2020). *Information in Construction*.

Pauwels, P., Terkaj, W., Krijnen, T., & Beetz, J. (2015). Coping with lists in the ifcOWL ontology. *EG-ICE 2015 - 22nd Workshop of the European Group of Intelligent Computing in Engineering*, *June 2015*, 111–120.

Pethuru Raj, A. C. R. (2017). The Internet of Things. Enabling Technologies, Platforms, and Use Cases. In *Angewandte Chemie International Edition, 6(11), 951–952.*

Ploennigs, J., Hensel, B., Dibowski, H., & Kabitzsch, K. (2012). BASont - A modular, adaptive building automation system ontology. *IECON Proceedings (Industrial Electronics Conference)*, 4827–4833. https://doi.org/10.1109/IECON.2012.6389583

Points – Project Haystack. (2022). https://project-haystack.org/doc/docHaystack/Points

Prakash, A. K., Zhang, K., Gupta, P., Blum, D., Marshall, M., Fierro, G., Alstone, P., Zoellick, J., Brown, R., & Pritoni, M. (2020). Solar+ optimizer: A model predictive control optimization platform for grid responsive building microgrids. *Energies*, *13*(12). https://doi.org/10.3390/en13123093

Project Haystack - Tags. (2020). https://project-haystack.org/tag.

Quinn, C., & McArthur, J. J. (2021). A case study comparing the completeness and expressiveness of two industry

recognized ontologies. *Advanced Engineering Informatics*, *47*(October 2020), 101233. https://doi.org/10.1016/j.aei.2020.101233

Ran, F., Gao, D. ce, Zhang, X., & Chen, S. (2020). A virtual sensor based self-adjusting control for HVAC fast demand response in commercial buildings towards smart grid applications. *Applied Energy*, *269*(April), 115103. https://doi.org/10.1016/j.apenergy.2020.115103

Rasmussen, M. H., Pauwels, P., Hviid, C. A., & Karlshøj, J. (2017). *Proposing a Central AEC Ontology That Allows for Domain Specific Extensions*. *July*, 237–244. https://doi.org/10.24928/jc3-2017/0153

Reppa, V., Papadopoulos, P., Polycarpou, M. M., & Panayiotou, C. G. (2014). A distributed virtual sensor scheme for smart buildings based on adaptive approximation. *Proceedings of the International Joint Conference on Neural Networks*, 99–106. https://doi.org/10.1109/IJCNN.2014.6889976

Roth, K. W., Westphalen, D., Feng, M. Y., Llana, P., & Quartararo, L. (2005). Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and Energy Savings Potential. *U.S. Department of Energy*. https://www.researchgate.net/publication/242718245_Energy_Impact_of_Commercial_Building_Controls_and_Performance_Diagnostics_Market_Characterization_Energy_Impact_of_Building_Faults_and_Energy_Savings_Potential

RVO. (2022). *Long-term energy efficiency agreements (MJA3/MEE) | RVO.nl | National*. https://www.rvo.nl/onderwerpen/duurzaam-ondernemen/energie-besparen/mja3-mee

Schumann, A., Hayes, J., Pompey, P., & Verscheure, O. (2011). Adaptable fault identification for smart buildings. *AAAI Workshop - Technical Report*, *WS-11-07*, 44–47.

Semantic Model | openHAB. (n.d.). Retrieved February 28, 2022, from https://www.openhab.org/docs/tutorial/model.html

Sinopoli, J. (2016). Advanced technology for smart buildings. In *Artech House power engineering library*.

Široký, J., Oldewurtel, F., Cigler, J., & Prívara, S. (2011). Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, *88*(9), 3079–3087. https://doi.org/10.1016/j.apenergy.2011.03.009

Smart Data Models. (n.d.). https://smartdatamodels.org/

Smart Data Models - FIWARE. (n.d.). Retrieved February 7, 2022, from https://www.fiware.org/smart-data-models/

Spyns, P., Meersman, R., & Jarrar, M. (2000). STAR Lab Technical Report Data Modelling versus Ontology Engineering Data modelling versus Ontology engineering. *Building*, *31*(4).

Stinner, F., Neißer-Deiters, P., Baranski, M., & Müller, D. (2019). Aikido: Structuring data point identifiers of technical building equipment by machine learning. *Journal of Physics: Conference Series*, *1343*(1). https://doi.org/10.1088/1742-6596/1343/1/012039

Suzuki, L. R. (2015). *Data integration approaches for smarter operation of large commercial buildings*.

Tang, H., Wang, S., & Li, H. (2021). Flexibility categorization, sources, capabilities and technologies for energy-flexible and grid-responsive buildings: State-of-the-art and future perspective. *Energy*, *219*, 119598. https://doi.org/10.1016/j.energy.2020.119598

Tang, R., & Wang, S. (2019). Model predictive control for thermal energy storage and thermal comfort optimization of building demand response in smart grids. *Applied Energy*, *242*(December 2018), 873–882. https://doi.org/10.1016/j.apenergy.2019.03.038

Taylor, P., Katipamula, S., & Brambley, M. R. (2005). *Methods for Fault Detection , Diagnostics , and Prognostics for Building Systems*. *11*(December 2012), 37–41.

timeseries - BrickSchema. (n.d.). Retrieved March 4, 2022, from https://brickschema.org/ontology/1.2/relationships/timeseries/

TNO. (2017). *Legislation - TNO Arbo in Europa*. https://www.arboineuropa.nl/en/legislation/

Wagner, A., O'Brien, W., & Dong, B. (2017). Exploring occupant behavior in buildings: Methods and challenges. In *Exploring Occupant Behavior in Buildings: Methods and Challenges*. https://doi.org/10.1007/978-3-319-61464-9

Wang, W., Hong, T., Li, N., Wang, R. Q., & Chen, J. (2019). Linking energy-cyber-physical systems with occupancy prediction and interpretation through WiFi probe-based ensemble classification. *Applied Energy*, *236*(November 2018), 55–69. https://doi.org/10.1016/j.apenergy.2018.11.079

WELL Building Institute. (2020). *WELL Certified | International WELL Building Institute | IWBI*. https://www.wellcertified.com/

XBOS Overview - XBOS Docs. (n.d.). Retrieved February 26, 2022, from https://docs.xbos.io/

APPENDIX 1

Brains for Buildings – WP4 Data Integration

Questionnaire for building owners

Name :

Organization :

1	Question	Please describe your role in the company, how you are related to building sector and your experience?
	Answer	
2	Question	What are the types of buildings that you own? And who manages them?
	Answer	
3	Question	What are the issues you currently face in the operational phase of these buildings that leads you to seek solutions from this project?
	Answer	
4	Question	What data do you have about your assets and how (where) do you store, access and update that information?
	Answer	
5	Question	What are the naming conventions, classifications and procedures you use for accessing this information?
	Answer	
6	Question	What are the devices and systems that generate data about the building operation and the occupants?
	Answer	
7	Question	What data are usually monitored and collected?
	Answer	
8	Question	What is the purpose of collecting above data?
	Answer	
9	Question	Do you have all the data that you need, or what additional data do you think will be included in the future and why?
	Answer	
10	Question	Do you use naming conventions for the data? What are they?
	Answer	
11	Question	How is data stored? And for how long?

	Answer	
12	Question	In which formats the data is stored?
	Answer	
13	Question	There are many parties involved in building operation like platform providers, building owners, renters, occupants, etc. What level of control/ownership each party has on data?
	Answer	
14	Question	Do you perform data analysis? (If the answer is "yes", skip Question 15 & 16 and go to Question 17. If the answer in "no", answer Question 15, 16 and go to Question 25)
	Answer	
15	Question	Do you try to continuously improve any of the building performance aspects: operation cost, comfort, energy, user feedback? If yes, in what ways? If not, why?
	Answer	
16	Question	How do you measure or benchmark the performance of a building (for example in terms of its energy efficiency and occupant satisfaction)?
	Answer	
17	Question	What data do you analyze and what do you expect by doing it?
	Answer	
18	Question	What is the procedure for data analysis and who is responsible for this task?
	Answer	
19	Question	What platforms and solutions do you use for data analysis?
	Answer	
20	Question	Where do you apply the results from above data analysis?
	Answer	
21	Question	Do you perform the same analysis for each building that you own?
	Answer	
`22	Question	What can you say about the visualization and reporting ability of the existing systems? What value do you extract from the visualizations?
	Answer	

23	Question	What limitations do you encounter when analyzing data using the platforms and solutions you mentioned?
	Answer	
24	Question	What do you think about the complexity of the data analysis procedure? Is it efficient and easy to use or cumbersome and needs lot of manual work?
	Answer	
25	Question	What are some critical requirements of occupants and how much control do occupants have over those requirements on the building systems?
	Answer	
26	Question	Do you collect and analyze occupants' feedback? How? If not, why?
	Answer	
27	Question	Do you operate more than one system (from different vendors) for managing your assets? If so, what are the challenges that you run in to when operating multiple systems from different vendors?
	Answer	
28	Question	What are the solutions that you currently use to solve above problems?
	Answer	
29	Question	What external constraints do you experience for data collection and analysis?
	Answer	
30	Question	What security features are used in your systems? How do you prevent data tampering or misuse?
	Answer	
31	Question	What other drivers are there for these security measures?
	Answer	
32	Question	How do you apply and maintain the General Data Protection Regulation?
	Answer	
33	Question	What functionalities would you like to have in a future smart building platform and how do you expect your buildings to perform differently in future by using such platform?
	Answer	

APPENDIX 2

Brains for Buildings – WP4 Data Integration

Questionnaire for platform providers

Name :

Organization :

1	Question	Please describe your role in the company, how you are related to building sector and your experience?
	Answer	
2	Question	What types of services/solutions that you usually provide for buildings?
	Answer	
3	Question	What are the solutions that are mostly requested by building owners from you? And are you able to provide those solutions?
	Answer	
4	Question	Which systems and protocols available in buildings can be handled in your solutions?
	Answer	
5	Question	What issues do you encounter when discovering and extracting information from different systems, specially from an existing building?
	Answer	
6	Question	What data is usually monitored and collected?
	Answer	
7	Question	What is the purpose of collecting above data? Do you think all the data collected in buildings are utilized in some way or are there unnecessary data that never serve any purpose?
	Answer	
8	Question	Do you include building information in your solutions? How?
	Answer	
9	Question	In which formats you find the data? And what desired formats do you convert them in to?

	Answer	
10	Question	Do you use naming conventions for the data? What are they? How do you convert client's naming convention to your desired format?
	Answer	
11	Question	What are your data storage techniques? (Type of database, on-premises or cloud, retention period)
	Answer	
12	Question	What platform do you provide to the client for data consumption and how do you charge for it?
	Answer	
13	Question	What are the typical functionalities of the above platforms?
	Answer	
14	Question	How sensitive are your solutions for a certain upgrade in the building?
	Answer	
15	Question	What security features are used in your systems? How do you prevent data tampering or misuse?
	Answer	
16	Question	What other drivers are there for these security measures?
	Answer	
17	Question	There are many parties involved in building operation like platform providers, building owners, renters, occupants, etc. What level of control each party has on data?
	Answer	
18	Question	Do you have all the data that you need in buildings, or what additional data do you think will be included in the future?
	Answer	
19	Question	Do you perform data analysis? If so, what are the typical data analysis cases available in the solutions provided by you?
	Answer	
20	Question	What kind of data do you analyze? (Historical, live)
	Answer	
21	Question	What platforms do you use for data analysis?
	Answer	

22	Question	What is the normal procedure for data analysis?
	Answer	
23	Question	How do you validate whether the data that you collect is correct and how do you handle missing/ incorrect data?
	Answer	
24	Question	After analyzing, what kind of insights do you provide about the buildings and to whom?
	Answer	
25	Question	What benefits have you brought to the building owners and occupants from data analysis?
	Answer	
26	Question	Are your solution packages generic, specific for building category or different for each building?
	Answer	
27	Question	How easy or hard it is to reuse or scale one of your solutions from one building to another, given the variety of systems in different kinds of buildings?
	Answer	
28	Question	Do you provide solutions to collect and analyze occupants' feedback? If so, how do they work?
	Answer	
29	Question	Do you refer to any standards and/or regulations when you are providing solutions?
	Answer	
30	Question	What external constraints do you experience for data collection and analysis?
	Answer	
31	Question	What is your policy regarding privacy and ethical usage of occupant and other data in the building? How do you apply it in practice?
	Answer	
32	Question	Have privacy and ethical usage of data ever become a barrier to implement any of your solutions? How and where?
	Answer	
33	Question	What other functionalities would you like to add further in your solutions?
	Answer	

APPENDIX 3

Subject information for participation in scientific research

Project : Brains for Building's Energy Systems
Work Package : WP4 - Data integration
Task : T4.1-Mapping data needs and requirements

Introduction

Dear Sir/ Madam,

Thank you for agreeing to be interviewed as part of the above research project. Participation is voluntary. Participation requires your written consent. Before you decide whether you want to participate in this study, you will be given an explanation about what the study involves. Please read this information carefully and ask the investigator for an explanation if you have any questions. You should explicitly agree to being interviewed and how the information contained in their interview will be used. This consent form is necessary for us to ensure that you understand the purpose of your involvement and that you agree to the conditions of your participation. Please read this information sheet and then sign this form to certify that you approve the conditions.

1. General Information

This interview is being carried out as a part of Brains for Building's Energy Systems (B4B). The study has been designed by TUe and TNO and will be carried out by two investigators. Details of the investigators are given in Appendix A. This interview is aimed at two parties, platform providers and building owners. It is planned to interview three platform providers and two building owners.

2. Purpose of the study

Brains for Building's Energy Systems (B4B) is a multi-year, multi-stakeholder project focused on developing methods to harness big data from smart meters, building management systems and the Internet of Things devices, to reduce energy consumption, increase comfort, respond flexibly to user behaviour and local energy supply and demand, and save on installation maintenance costs. This will be done through the development of faster and more efficient Machine Learning and Artificial Intelligence models and algorithms. The project is geared to existing utility buildings such as commercial and institutional buildings.
There are four key work packages (WP), each dedicated to a research theme as below.
WP1: Self-diagnostic for EE (Energy-Efficiency) and Smart Maintenance
WP2: Energy Flexibility
WP3: User centric interfaces and feedback
WP4: Data Integration
This interview is carried out as a part of WP4 Task to map data needs and requirements. Interviews with platform providers are used to identify the state of art commercially available data platforms developed for buildings, their functionalities, constraints, privacy and security measures and emerging technologies. Interviews with building owners are used to identify the state of art technologies, devices and sensor that are used for operations, barriers to implement energy efficient strategies, problems in the operational stage etc.

3. What participation involves

During the study, you will be called for an interview (in Microsoft Teams environment) with the two investigators. There will be two one-hour interviews. Investigators will send you a questionnaire for preparation beforehand. Since the questions are multidisciplinary, the questions are sent to you beforehand for preparation.

4. What is expected of you

In order to carry out the study properly it is important that you follow the study instructions.
It is important that you contact the investigator(s):

- if you no longer want to participate in the study
- if you have any complaints
- if your contact details change
- if you want to change the time for interview

5. If you do not want to participate or you want to stop participating in the study

It is up to you to decide whether or not to participate in the study. Participation is voluntary.
If you do participate in the study, you can always change your mind and decide to stop, at any time during the study. You do not have to say why you are stopping, but you do need to tell the investigator immediately.
The data collected until that time will still be used for the study.

If there is any new information about the study that is important for you, the investigator will let you know. You will then be asked whether you still want to continue your participation.

If you wish to file a complaint regarding the interview, please contact work package leader mentioned in the General Information section.

6. End of the study

Your participation in the study stops when

- you choose to stop
- the end of the entire study has been reached
- the investigator considers it best for you to stop
- Brains for Buildings consortium, the government, or Ethical Review Board, decides to stop the study.

The study is concluded once all the participants have completed the study.

7. Usage and storage of your data

Personal data

Personal information about the participant (name, company, professional background) will be collected to contact you and conduct the interviews.
To protect your privacy, we will not share your name, the raw interview recordings and other information that can directly identify you to render the data anonymous. The data cannot be traced back to you in reports or any publications resulting from the study.

Confidential data

If the participant wants to omit certain information, they can avoid answering the specific questions. Also, the participant can request not to share the given answers within public domain or even within other work packages in the B4B project. Further, we will not share the raw interview recordings or information that it contains with any party, without your approval.

Access to your data for verification

Some people can access all your data at the research location. Including the data without a code. This is necessary to check whether the study is being conducted in a good and reliable manner. Persons who have access to your data for review are Associate Professor Pieter Pauwels, TU Eindhoven and Assistant Professor Ekaterina Petrova, TU Eindhoven. They will keep your data confidential. We ask you to consent to this access.

Retention period

Recordings of the interviews (.mp4) and transcripts (.txt, .docx) will be stored in TUe OneDrive personal cloud storage until 30 April 2025, after that they will be permanently deleted. Access to above files is restricted to Lasitha Chamari. Recordings of the interviews (.mp4) will be stored in TNO OneDrive personal cloud storage of Elena Chochanova until 30 April 2025, after that they will be permanently deleted. Access to above files is restricted to Elena Chochanova.

Storage and usage of data

Recordings of the interviews (.mp4), transcripts (.txt, .docx) will be analysed and used in the deliverables of the B4B project, unless the participant specifically requests some parts of the answers not to be shared. Before publishing any of the deliverables, we will send the draft to the participant, so that he/she can review the content and request not to publish any content that he/she finds confidential.

Withdrawing consent

You can withdraw your consent to the use of your personal data at any time. This applies to this study and also to storage. The study data collected until the moment you withdraw your consent will still be used in the study.

More information about your rights when processing data

For general information about your rights when processing your personal data, you can consult the website of the Dutch Data Protection Authority.

If you have questions about your rights, please contact the person responsible for the processing of your personal data. For this study, See Appendix A for contact details.

If you have questions or complaints about the processing of your personal data, we advise you to first contact the investigators. You can also contact the Data Protection Officer of the institution [see Appendix A for contact details] or the Dutch Data Protection Authority.

5. Any questions?

If you have any questions, please contact the investigator(s).
If you have any complaints about the study, you can discuss this with the investigator. If you prefer not to do this, you may contact the Work Package leader, See Appendix A for contact details.

6. Signing the consent form

When you have had sufficient time for reflection, you will be asked to decide on participation in this study. If you give permission, we will ask you to confirm this in writing on the appended consent form. By your written permission you indicate that you have understood the information and consent to participation in the study. The signature sheet is kept by the investigator. Both the Investigator(s) and yourself receive a signed version of this consent form.

Thank you for your attention.

Appendix A: contact details

1. For general questions	
Research investigators	Lasitha Chamari, Technology University Eindhoven l.c.rathnayaka.mudiyanselage@tue.nl Elena Chochanova, TNO elena.chochanova@tno.nl
2. For complaints	
Work package leader	Pieter Pauwels, Technology University Eindhoven p.pauwels@tue.nl
Data Protection Officer of the institution	Bart Schellekens (data protection officer TU/e) functionarisgegevensbescherming@tue.nl
3. For more information about your rights	privacy@tue.nl or functionarisgegevensbescherming@tue.nl

Appendix B: Subject Consent Form

Please (X) the appropriate boxes to indicate Yes or No	Yes	No
Taking part in the study		
- I have read the subject information form. I was also able to ask questions. My questions have been answered to my satisfaction. I had enough time to decide whether to participate.		
- I know that participation is voluntary. I know that I may decide at any time not to participate after all or to withdraw from the study. I do not need to give a reason for this.		
- I understand that taking part in the study involves recording (audio and video), transcripts (.txt, .docx), and these will be destroyed once the results are published. I give permission for the collection and use of my data to answer the research question in this study.		
Risks associated with participating in the study		
- I agree to inform (during or after the interviews) if the discussions contain confidential data that are not to be included in publications/ reports/ deliverables.		
- I agree to review the results sent by the researchers and inform them if there is any confidential or sensitive information that should be removed before publishing		
Use of the information in the study		
- I understand that information I provide will be used for making deliverables and publications of B4B project. Results will be analysed by the two researchers mentioned above.		
- I understand that personal information collected about me that can identify me, such as name, profession, company will not be shared beyond the two investigators.		
- I understand that any summary interview content, or direct quotations from the interview, that are made available through academic publication or other academic outlets will be anonymized so that I cannot be identified, and care will be taken to ensure that other information in the interview that could identify myself is not revealed.		

- I know that some people may have access to all my data to verify the study. These people are listed in this information sheet. I consent to the inspection by them.		
Storage of interview data		
- I give permission to store recordings (.mp4) and transcripts (.txt, .docx) of the interviews in the TUe OneDrive personal cloud storage of Lasitha Chamari until 30 April 2025. Transcripts will be anonymised. - I give permission to store recordings (.mp4) of the interviews in the TNO OneDrive personal cloud storage of Elena Chochanova until 30 April 2025.		

Signatures

_____	_____	_____
Name of participant	Signature	Date

I hereby declare that I have fully informed this study subject about this study.

If information comes to light during the course of the study that could affect the study subject's consent, I will inform him/her of this in a timely fashion.

Lasitha Chamari	Lasitha	02/06/2022
Name of investigator	Signature	Date
Elena Chochanova	Elena	04/03/2022
Name of investigator	Signature	Date