This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957047.

**H2020-LC-SC3-EE-2020-1/LC-SC3-B4E-6-2020**
Big data for buildings

Building Information aGGregation, harmonization and analytics platform

Project Nº 957047

# WHITE PAPER II -

# BIGG: The need for harmonizing input data and AI Toolbox revolution in context of smart building management

Responsible: Inetum

Dissemination Level: Public
Version: 1.0
Date: 21/08/2023

# Executive Summary

In December 2020, the BIGG project (Building Information aGGregation, harmonization, and analytics platform) was launched to demonstrate the utilization of big data technologies and data analytic techniques on buildings's life cycle in over 4,000 buildings across six different pilot testbeds, located in Spain (Catalonia Province) and Greece (Athens, Volos, and Thessaloniki).

However, proper implementation of new technologies on big data needs standardization of input data coming from different sources. The uniqueness of the BIGG concept lies in its creation and utilization of a distinctive Ontology. This Ontology is designed to facilitate semantic interoperability of data and empower big data analytics within buildings. In this article we will explore the aim, advantages and key components of the harmonization layer of BIGG.

After harmonizing the input data, a data analytics framework is envisioned to seamlessly incorporates state-of-the-art Artificial Intelligence (AI) methods and decision support tools, facilitating the analysis of high-quality, anonymized, and interoperable building-related data gathered within the project. In order to achieve that, several innovative technologies/algorithm have emerged and are collectively named as the AI Toolbox for Buildings (AITB). This article delves into the origin, principles, and transformative potential of the AITB in driving energy efficiency and sustainability in the building sector. The AITB was conceived with a clear vision - to develop practical solutions tailored to the unique characteristics of building data. It emerged from a problem-based methodology, where applications became the starting point for the toolbox's development. Collaboration with key stakeholders in the energy sector, played a pivotal role in identifying specific energy-saving challenges and designing AI tools to address them effectively.

# Contributors Table


| DOCUMENT SECTION | AUTHOR(S) | REVIEWER(S) |
|---|---|---|
| **I. INTRODUCTION** | Sampath Mukherjee (Inetum) | |
| **II. THE BIGG DATA MODEL** | Jordi Carbonell (CIMNE)<br>Stoyan Danov (CIMNE)<br>Audrey Bouet (CSTB)<br>Guillaume Picinbono (CSTB)<br>Nicolas Bus (CSTB)<br>Polychronis Symeonidis (DOMX) | Nicolas Pastorelly (CSTB) |
| **III. AI Toolbox in BIGG** | Anthony Uhlmann (Helexia)<br>Romain Hollanders (Energis)<br>Riccardo Devivo (Energis)<br>Niels Tiben (IMEC)<br>Gerard Mor (CIMNE)<br>Sampath Mukherjee (Inetum) | Athanasios Papakonstantinou (HERON), Nicolas Pastorelly (CSTB) |

# Table of Contents

**I. INTRODUCTION .......... 5**
**I.1. Purpose and organization of the document .......... 5**
**I.2. Scope and audience .......... 5**
**I.3. Background and motivation .......... 5**
**II. THE BIGG DATA MODEL .......... 7**
**II.1. Harmonized data .......... 7**
II.1.1. The concept of Harmonized data .......... 7
II.1.2. Why is data harmonization necessary? .......... 8
**II.2. BIGG Ontology .......... 9**
II.2.1. BIGG Data Harmonization layer (GitHub) .......... 9
II.2.2. BIGG Ontology: Conclusion .......... 12
**III. AI TOOLBOX IN BIGG .......... 13**
**III.1. Why is AI needed in BIGG pilots? .......... 13**
**III.2. How AITB fulfils the objectives? .......... 14**
**III.3. The used cases that are part of AI solutions of BIGG business cases .......... 15**
**III.4. Limitations and Considerations for using AI .......... 17**
**IV. SUMMARY AND CONCLUSION .......... 18**

# Table of Figures

Figure 1 - The BIGG Data Model (Reproduced from Bouet et al. in review) .......... 10
Figure 2 - Structure of the BIGG ontology (reproduced from D4.2). .......... 11
Figure 3 - The reference architecture framework of BIGG .......... 12

# Table of Acronyms and Definitions

| ACRONYM | DEFINITION |
|---|---|
| **AI** | Artificial Intelligence |
| **ML** | Machine Learning |
| **AITB** | AI Toolbox for buildings |
| **IoT** | Internet of Things |
| **BEMS** | Building Energy Management Systems |

# I. INTRODUCTION

## I.1. Purpose and organization of the document

The core objective of this white paper is to provide a comprehensive understanding of the imperative for harmonizing data in various contexts. By delving into the intricacies of harmonization, this white paper aims to elucidate the reasons underlying the increasing need to align and integrate diverse datasets. Through insightful discussions, readers will gain a clear perspective on the challenges stemming from disparate data sources and formats, and how these challenges can hinder effective analysis, decision-making, and the realization of the full potential of data-driven technologies. Additionally, in the text we will explore the methodologies and approaches that has been employed to achieve harmonization, emphasizing the significance of standardized frameworks, data sharing protocols, and advanced technologies like AI and machine learning. The white paper also explains the motivation behind implementing AI solutions rather than traditional methods. By addressing the "why" and "how" questions related to data harmonization and AI, this document seeks to equip readers with the knowledge and insights necessary to navigate the complex landscape of data integration and thereby harness the power of data in a more unified and impactful manner. This document aims to offer a comprehensive view of the rationale behind harmonization and the intricate pathways to harnessing AI for the benefit of society at large.

## I.2. Scope and audience

This document aims to equip readers with a basic understanding of why harmonizing data is essential in our interconnected world. This document focuses on providing a simple overview of how data harmonization works and its positive impact with AI and its uses in context of BIGG use cases. No technical knowledge is needed to understand the concepts presented here.

## I.3. Background and motivation

As the world grapples with the challenges of climate change and sustainability, energy efficiency has emerged as a crucial focus area. Buildings, being significant energy consumers, present an opportunity for substantial energy savings. In the pursuit of energy consumption reduction in buildings, the actual implementation of actions encounters a complex interplay of factors. Managing the internal energy systems, striving for cost savings, and ensuring the desired comfort levels for building occupants present significant challenges that must be carefully navigated. Achieving a harmonious balance among these elements is crucial to the effective realization of energy efficiency goals in the building sector.

The primary objective of the H2020 BIGG (Building Infrastructures for Geospatial Innovation) project is to devise and experiment with a comprehensive big data solution capable of tackling a wide range of energy-building related business cases. This entails a precice analysis of the life cycle of a substantial dataset comprising various buildings, enabling the accurate assessment of the efficiency of energy-saving strategies and possible improvements.

To achieve this ambitious goal, the project centers on the development of an advanced data platform, accompanied by the aggregation of input data into the dedicated BIGG Ontology to have harmonized data. The harmonized input data is vital in the energy sector for ensuring consistency and comparability across diverse datasets. By adhering to standardized formats and units, harmonized data enables easy comparison of energy consumption patterns. It also facilitates data integration from various sources, promoting a comprehensive view of energy usage. Improved accuracy and reliability result from reducing errors and inconsistencies in analyses and decision-making. Interoperability is enhanced, supporting seamless data exchange and collaborative efforts. Scalability is enabled by employing a generic format, allowing energy initiatives to be expanded efficiently. In turn, harmonized data serves as a reliable foundation for predictive modeling and AI applications, driving data-driven energy efficiency improvements.

The integration of the components also requires a structured and scientifically rigorous approach, driving advancements in energy efficiency and optimizing buildings' energy performance. The aim of the AI Toolbox for Buildings (AITB) is to provide users with a comprehensive suite of powerful and flexible tools for building energy savings. These tools facilitate pre-processing of building data, studying relevant features, developing, and evaluating models, and delivering AI solutions targeted at specific energy-saving problems. The toolbox encompasses a diverse range of algorithms optimized for building data, offering specialized tools for handling time-series data, normalization techniques, and key performance indicators (KPIs) like degree days.

To ensure the AITB's applicability across diverse scenarios, scalability has been emphasized. The toolbox is designed to handle large and complex datasets efficiently, and users can seamlessly adapt AI pipelines to address new applications and emerging challenges. Furthermore, the AITB incorporates harmonization capabilities, allowing users to work with datasets harmonized to the BIGG Ontology. This fosters data interoperability and empowers users to share data among different stakeholders, facilitating collaborative efforts and informed decision-making.

The AITB stands as an open-source project, aimed at the spirit of collaboration and knowledge sharing for anyone to utilize the resources. This approach encourages community contributions and facilitates the rapid evolution of the toolbox through collective expertise. With user-friendliness at the forefront, the AITB ensures accessibility and ease of use. The interface on GitHub provides transparency, allowing users to access code for each module, example pipelines, and detailed documentation, thereby ensuring a smooth onboarding experience.

The principles driving the AITB's development underscore a commitment to practicality and real-world problem-solving. By tailoring AI solutions to specific energy-saving challenges and building data characteristics, the toolbox empowers users to make a meaningful impact on energy consumption. As users leverage the AI-powered capabilities of the AITB, a paradigm shift emerges in building management.

The AITB represents a significant advancement in the pursuit of energy efficiency and sustainability within the building sector. Built on principles of problem-based methodology and harmonization, the toolbox unlocks the potential of AI in delivering tailored solutions for energy savings. Its scalability, open-source nature, and user-friendliness amplify its transformative potential, enabling stakeholders to create more sustainable buildings while fostering a greener future for generations to come.

# II. THE BIGG DATA MODEL

## II.1. Harmonized data

### II.1.1. The concept of Harmonized data

Over the decades, the techniques for gathering building energy information have evolved significantly, driven by technological advancements and growing concerns about energy efficiency and sustainability. In the early days, manual meter readings and data logging were the primary methods used to collect energy consumption data, often involving time-consuming and error-prone processes. As technology progressed, the adoption of automated metering systems allowed for more frequent data collection and reduced human intervention. With the advent of smart sensors and Internet of Things (IoT) devices, real-time data monitoring became feasible, enabling building managers to access and analyze energy consumption data remotely and in near-real-time. Furthermore, the integration of Building Energy Management Systems (BEMS) and advanced analytics tools has revolutionized the way building energy information is gathered, offering valuable insights into energy usage patterns, optimization opportunities, and predictive maintenance. These technological advancements have not only enhanced data accuracy and availability but have also opened new possibilities for optimizing energy performance and driving sustainable practices in the built environment.

The rapid adoption of digital technologies and the integration of smart sensors and controllers in buildings have presented promising avenues for enhancing efficiency and optimizing energy consumption in both new and existing structures. However, a critical challenge impedes the realization of these benefits - the lack of standardization and harmonization of data definitions across diverse applications and databases.

In contemporary building environments, the plethora of sensors and systems generate vast amounts of data encompassing energy usage, temperature, humidity, occupancy patterns, and more. Nevertheless, this data often follows disparate formats, structures, and terminologies, hampering seamless data exchange, comparison, and integration across various sources. This lack of data standardization poses obstacles to effective data analysis and insightful decision-making.

The BIGG project recognizes this crucial challenge and endeavors to overcome it. This project's main objective is to develop an open-source Big Data Reference Architecture and AITB capable of supporting diverse use cases and applications throughout the building life cycle. The essence of the BIGG project revolves around harmonizing data across various systems and platforms. This necessitates harnessing the power of semantic technologies, particularly ontologies, which enable a structured and standardized representation of data meaning and interrelationships in machine-processable formats. Ontologies foster common vocabularies and shared understandings, thereby facilitating seamless data exchange and integration across heterogeneous systems.

An essential principle in ontology development is the reutilization and alignment of existing ontologies. By building upon and extending established ontologies in relevant domains, the BIGG project avoids redundant efforts, while ensuring compatibility and consistency between diverse data sources. Reusing and aligning existing ontologies create a robust foundation for connecting and relating heterogeneous data sets, preserving the original context of information.

## II.1.2. Why is data harmonization necessary?

Data harmonization, also referred to as data standardization or normalization, holds immense significance for several reasons, as it effectively addresses challenges associated with data inconsistency, quality, integration, and optimal utilization. The principal imperatives of data harmonization encompass a range of critical aspects:

One fundamental aspect is data consistency. Through harmonization, uniformity is established across diverse sources, systems, and departments in terms of representing data elements. This consistency plays a pivotal role in enabling accurate analysis, reporting, and informed decision-making. The data quality experiences a marked improvement due to the harmonization process. Data is cleansed and transformed to adhere to defined standards, resulting in the elimination of errors, redundancies, and inaccuracies, thereby elevating overall data quality.

The enhancement of data integration represents another significant facet of data harmonization. When data is harmonized, its integration into various systems and applications becomes notably streamlined. This integration aspect is crucial in creating a holistic perspective of data within an organization, fostering improved insights and collaborative efforts. Efficient data exchange and interoperability are facilitated via smooth integration. Consistent data can be exchanged seamlessly with partners, suppliers, and customers, essential for ensuring smooth business processes and collaborative endeavors within a broader ecosystem.

Harmonized data opens doors to cross-functional analysis, enabling meaningful insights to be drawn across diverse functions and departments. This analytical capability was previously hindered by inconsistent data formats and structures. Accurate reporting and compliance are also an added benefit of data harmonization. Reliable and accurate information forms the foundation of reporting and compliance endeavors, particularly pertinent in contexts involving regulatory compliance and audit requirements.

The most important and significant outcome of data harmonization is to seamlessly do advanced analytics techniques such as machine learning and AI. Clean, structured data provided by harmonization serves as the necessary foundation for these sophisticated analytics methodologies, delivering precise insights. Effective decision-making is an inherent outcome of harmonizing data. Access to consistent, accurate, and well-structured data equips organizations to make informed decisions across all levels, underpinning evidence-based approaches.

Harmonized data enables clear documentation of data lineage, which tracks the origin and alterations of data. This transparency sustains accountability and auditability. Cost and resource efficiencies are realized through data harmonization, as it curtails redundant data cleaning efforts and inconsistencies, thus conserving time, effort, and resources. Data governance establishes data quality standards, ownership, and stewardship. Harmonizing data frequently entails assigning ownership and responsibility for distinct data elements, aligning with data governance principles that advocate for clear data stewardship roles and accountable data management practices.

Long-term impact of that is into data sustainability as it simplifies data management and maintenance, aiding adaptability and scalability of data systems as organization needs to evolve. The benefits of data harmonization extend to data migration and system upgrades in the future, simplifying the process by mitigating complexities associated with data transformations.

Lastly, customer experience is positively impacted by harmonized data. Uniform, accurate customer data enhances organizations' understanding of customers, enabling the delivery of personalized services.

Within the context of the AITB, the effectiveness of data processing plays a pivotal role in achieving the project's objectives. Data pipelines serve the purpose of carrying out various data processing tasks, extracting valuable insights and predictions related to building energy consumption and performance. The success of these pipelines relies heavily on their integration with harmonized data. Harmonization ensures that the data input into the pipelines adheres to standardized formats and terminologies. Consequently, the pipelines can seamlessly use this harmonized data across various applications and analysis tasks. By establishing standardized data inputs, the pipelines can focus on their core functions of analyzing and interpreting data without the burden of handling inconsistencies or data structure discrepancies.

In conclusion, data harmonization serves as a foundational step toward realizing data-driven excellence, ensuring that data evolves from a potential 'obstacle' into an 'asset' that propels an organization's operations and growth endeavors.

# II.2. BIGG Ontology

## II.2.1. BIGG Data Harmonization layer (GitHub)[1]

As stated in D4.1 (Stoyan Danov, CIMNE, 2022), ontologies are vital for the semantic web approach defined by W3C (2015). They are formal and explicit knowledge specifications in human and machine-readable formats, enabling sharing across diverse environments. Ontologies define classes, properties, relations, and taxonomies, covering complex domains. Many ontologies follow international standards and align with each other. The BIGG ontology fully conceptualizes knowledge for BIGG Use Cases in the building domain, aligning with existing ontologies and data standards for independent sharing.

The BIGG Data Model (Figure 1) is a comprehensive representation of data structure, including UML diagrams, class descriptions, attributes, and relationships. It enables data sharing across systems, follows data standardization, and facilitates analysis of building energy consumption-related data

The BIGG Ontology serves as a meticulously structured roadmap, employing the specialized Web Ontology Language (OWL) to systematically arrange diverse concepts pertaining to buildings and their operational intricacies. This encompassing roadmap extends to encompass supplementary notions intricately linked with the effective utilization of building data analysis.

The BIGG ontology has modular structure consisting of core and extensions. The core includes three essential class groups, and the extensions provide additional classes and relations (see Figure 2).

At its core, the Ontology categorically classifies concepts into distinct strata. Notably, the category "bigg:Building" encapsulates the fundamental notion of buildings. These categories are further enriched with specific attributes, acting as descriptive markers. For instance, attributes like "bigg:buildingName" or "bigg:buildingConstructionType" provide distinct characteristics to these categories.

Beyond mere categorization, Ontology establishes vital connections, akin to well-defined pathways. These connections, referred to as ObjectProperties, establish profound interlinks between various categories. An illustrative example involves the link between "bigg:Building" and "bigg:BuildingSpace," achieved through the ObjectProperty "bigg:hasSpace." This pivotal linkage underscores the intrinsic relationship between a building and the spaces it encompasses.

The BIGG Ontology is an organizational framework, employing OWL to systematically structure and interrelate intricate concepts associated with buildings and their multifaceted operational paradigms. It not only provides a comprehensive representation but also facilitates a profound comprehension of the interplay between various building-related elements.

[1] https://github.com/biggproject/Ontology

Currently, the BIGG Ontology contains 100 classes, 181 DataProperties and 108 ObjectProperties. Table 1 provides a description of some of the main classes of the BIGG Ontology and Figure 2 provides a simplified view of their relations.

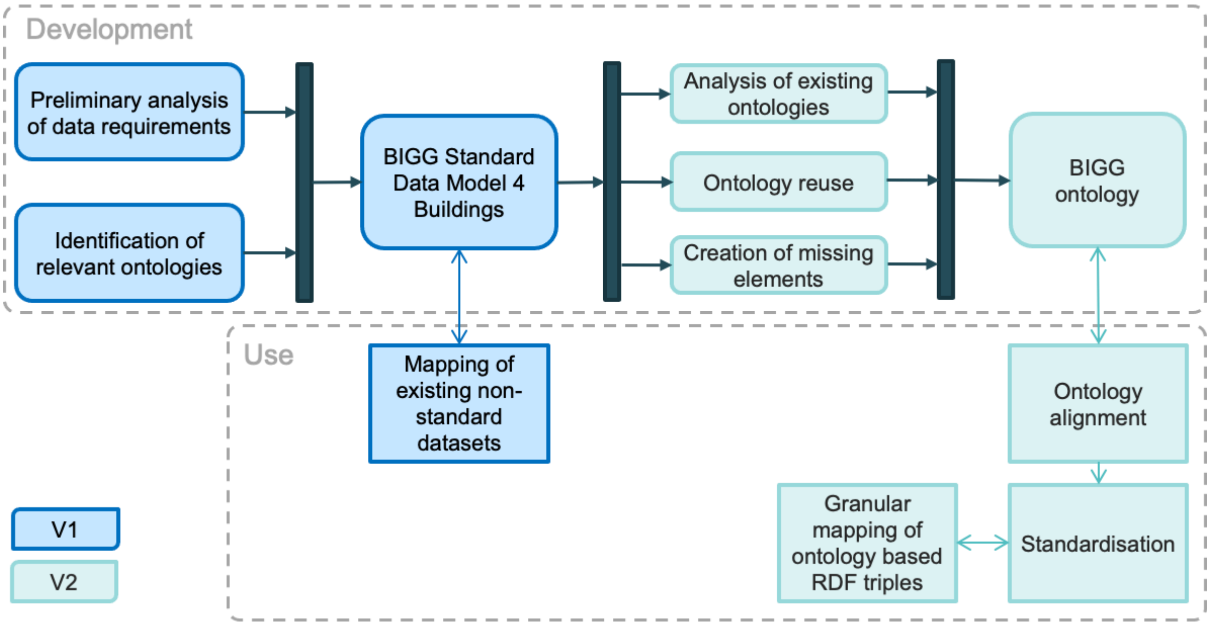


**Figure 1 - The BIGG Data Model (Reproduced from Bouet et al. in review).**

The components within the harmonization layer contribute significantly to the successful integration of AITB Business Cases. By aligning, harmonizing, and standardizing data from various sources, the layer simplifies the extraction of insights and predictions related to building energy consumption. Furthermore, it ensures data consistency and compatibility throughout the entire data processing pipeline, bolstering the overall effectiveness and reliability of the analytical process.

The Data Harmonization layer plays a crucial role in ensuring that data pipelines utilize harmonized data as their inputs. To achieve this, the input data must conform to the standardized BIGG data model. Each function block within the pipeline is specifically designed to work with this harmonized data format, ensuring consistency and compatibility throughout the entire pipeline.

The harmonization layer, illustrated in Figure 3, leverages the BIGG Standard Data Model for Buildings to establish a unified data format according to the above explained BIGG ontology. By doing so, it enables the aggregation of data from diverse sources, facilitating seamless comparison and integration. This harmonization capability proves invaluable in integrating the

AITB with both existing legacy systems and emerging technologies, as it eliminates concerns regarding compatibility issues.

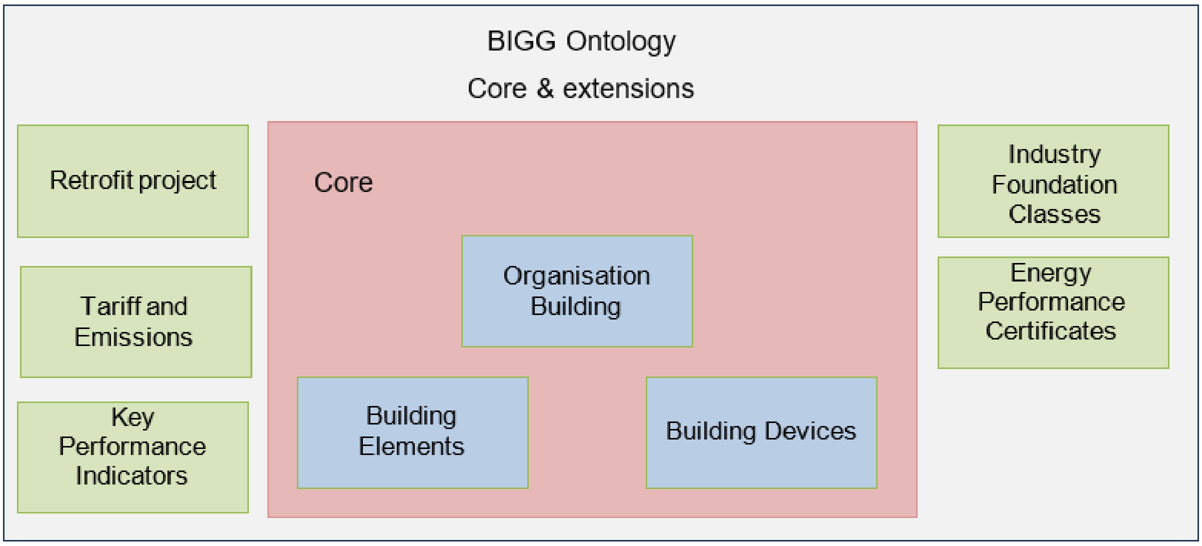


**Figure 2 - Structure of the BIGG ontology (reproduced from D4.2)**

**Table 1: Description of some of the main classes of the BIGG Ontology (reproduced from Bouet et. al. under review).**

| | |
|---|---|
| bigg:Organization | A company or institution that provides data |
| bigg: Building | A building for which data is provided |
| bigg: BuildingSpace | A space that can represent one or more rooms, floors, or zones of a Building, defined according to their use, or the necessity to separate monitoring and accounting of their energy use or performance. One BuildingSpace will be generated by default for each building, corresponding to the entire construction. |
| bigg: Element | Any generic element of the building. The type of Element can be further specified through its subclasses BuildingElement and Device. |
| bigg: Device | Any meter, sensor, or actuator that can capture a signal, emit a signal, or assume a state that can be recorded in the form of time series data. |
| bigg: Sensor | A collection of Measurements from the same Device that measure the same property in the same measurement units. |
| bigg: Measurement | Any time series record registered by a Device. |
| bigg: EnergyPerformanceContract | Energy Performance Contract the signed by the Organization |
| bigg: EnergyEfficiencyMeasure | Any measure for the improvement of the efficiency of a Building or its Elements. |

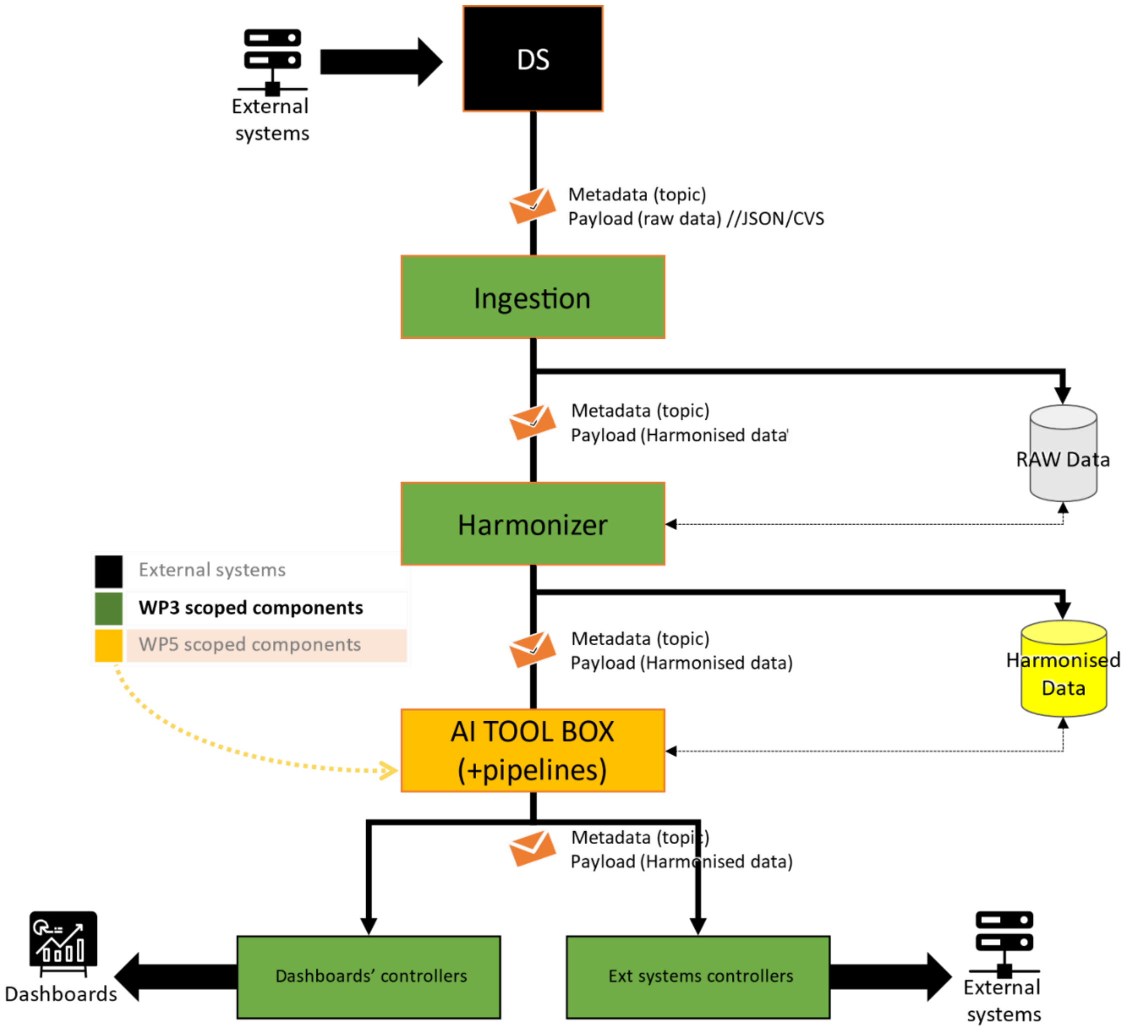


**Figure 3 - The reference architecture framework of BIGG.**

## II.2.2. BIGG Ontology: Conclusion

The BIGG Ontology has been meticulously crafted to align coherently with other existing data models, thus enhancing data interoperability and establishing a robust foundation for elevated data sharing and collaboration. The continuous maintenance of the referenced GitHub repositories guarantees ongoing support and availability of resources for the foreseeable future.

In conclusion, the vision of thisproject - to create a  standardized, and open-source Big Data Reference Architecture and AI Analytics Toolbox - holds immense promise for advancing the building industry. By addressing the challenge of data standardization and harmonization, the project lays the groundwork for a future where buildings can seamlessly share and leverage data, leading to improved energy efficiency, enhanced occupant experiences, and more sustainable building . Through the strategic use of ontologies and data pipelines, the BIGG project sets a precedent for intelligent, interconnected buildings that contribute to a greener and more sustainable world.

The following section will describe the AI toolbox that leverages input harmonized data formatted using the BIGG onthology.

# III. AI TOOLBOX IN BIGG

## III.1. Why is AI needed in BIGG pilots?

Forecasting and predictions form the bedrock of strategic planning and decision-making across a wide spectrum of disciplines. The emergence of Machine Learning (ML) and Artificial Intelligence (AI) has introduced a change in basic assumptions in how we approach forecasting, prompting a departure from conventional techniques. This essay not only explores the rationale for the pervasive adoption of ML and AI tools in forecasting and predictions but also underscores the necessity to transcend traditional methodologies.

1. **The Need to Move Beyond Conventional Techniques:**

While conventional forecasting methods have served as reliable tools for years, the complexity of modern challenges necessitates a transition to more advanced methodologies. As the world becomes increasingly interconnected and data-driven, the limitations of traditional techniques become apparent.

Firstly, the world is characterized by intricate interdependencies and nonlinear relationships that evade simplistic modeling approaches. ML and AI techniques possess the capability to unravel these intricate relationships, offering a deeper understanding of the underlying dynamics that conventional methods often fail to capture.

Secondly, the information landscape has evolved to encompass vast volumes of diverse and unstructured data. Conventional techniques struggle to harness the potential insights embedded within this deluge of information. ML and AI tools, however, excel in processing and extracting meaningful patterns from such data, opening new avenues for accurate predictions.

Furthermore, the pace of change in today's world demands agile forecasting methods. Traditional techniques, rooted in historical data, might falter when confronted with unforeseen events or rapidly evolving scenarios. ML and AI tools, with their adaptive nature, provide the flexibility required to respond to emerging challenges in real-time.

In addition, the limitations of human cognitive capacity to process and analyze large datasets hinder the efficacy of conventional techniques. ML and AI technologies have the potential to augment human capabilities by sifting through massive datasets, identifying subtle trends, and generating predictions that might be overlooked by human analysts.

2. **The Advantages of ML and AI Tools:**

ML and AI tools have gained prominence in solving forecasting problems and predictive knowledge due to their inherent strengths that set them apart from traditional methods. One of their significant advantages lies in their capacity to uncover intricate patterns within vast and complex datasets. Unlike conventional techniques, these tools can discern non-linear relationships and extract insights that might otherwise remain obscured.

Another pivotal aspect of ML and AI lies in their adaptability. These tools can learn from new data, continuously refining their predictions and adapting to changing circumstances. This adaptability is particularly valuable in dynamic environments where conditions evolve over time. The automation provided by ML and AI tools is a notable change. Once trained, these models can autonomously generate predictions, reducing the need for manual intervention. This automation not only enhances efficiency but also allows for real-time decision-making based on up-to-date insights.

Moreover, ML and AI tools excel in multivariate analysis, effectively considering a multitude of variables that might interact in complex ways. This comprehensive approach enhances the accuracy and comprehensiveness of predictions, offering a more nuanced understanding of the underlying dynamics.

The transition from conventional techniques to ML and AI-driven forecasting is not just a technological shift; it is a necessity dictated by the demands of the modern world. As we grapple with unprecedented complexities, data volumes, and the rapid pace of change, embracing these advanced tools is not merely an option—it is an imperative. By harnessing the power of ML and AI, we stand to unlock deeper insights, more accurate predictions, and a proactive approach to decision-making that aligns with the challenges of the modern time.

## III.2. How AITB fulfils the objectives?

AITB serves as an all-encompassing resource for analyzing building data and constructing AI models to optimize energy conservation within buildings. To accomplish this, we've harnessed the capabilities of both Python and R programming languages, tailoring popularly used libraries such as scikit-learn and Caret to specifically suit the usability of the toolbox.

At its core, the AI toolbox is designed for adaptability, capable of seamlessly accommodating diverse data formats, while also effortlessly navigating through harmonized data in alignment with the BIGG data model. While the toolbox itself doesn't conduct data collection, it adeptly manages a diverse array of data types, including real-time measurements, pulses, and index values—common elements in building energy data. Although the BIGG data model holds favor, it's important to note that each function block within the toolbox possesses standalone functionality, independent of this model. Essentially, the toolbox is equipped to handle a wide spectrum of data formats and structures, empowering it to seamlessly interact with various data sources.

Furthermore, while the pipelines within the toolbox are optimized for working with harmonized inputs and generating harmonized outputs, the foundational building blocks of these pipelines don't inherently demand this level of coherence. In essence, the toolbox's adaptability to different data formats positions it as a potent instrument for effective building data management, regardless of the specific presentation.

These function blocks are engineered with a versatile approach, offering applicability across a spectrum of building-related scenarios. They are defined by their inputs, functions, and outputs, forming the cornerstone of distinct modules. These modules, in turn, are thoughtfully categorized into module blocks, encompassing four primary classifications:

1. Data preparation modules: Encompassing function blocks that synchronize with initial data management stages, such as quality assessments, outlier identification, and timestamp management.

2. Data transformation modules: These function blocks delve into data categorization and secondary dataset management, involving elements like calendar and weather data.

3. Modeling modules: The focus of these function blocks revolves around constructing, evaluating, and testing data models.

4. Reinforcement learning modules: Lastly, function blocks in this category pertain to the creation and training of reinforcement learning agents—a specialized facet of machine learning.

The creation and enhancement of the AITB marks a substantial stride towards achieving energy efficiency and sustainability objectives within the building sector. AITB places a focused emphasis on updating and organizing the function blocks based on specific business cases and seamlessly integrating them within the BIGG Data Model. The result is a more efficient and streamlined tool capable of swiftly and accurately predicting energy usage, pinpointing areas for energy savings, and optimizing energy performance in buildings.

In conclusion, the AITB's development, accompanied by its harmonization capabilities and successful application in Business Cases, constitutes a significant contribution towards advancing energy efficiency and sustainability in the building sector. Its powerful and flexible tools provide a scalable solution, benefiting a diverse user base and serving as a pivotal component in the BIGG project's endeavors to create more sustainable buildings. The AITB represents a groundbreaking leap in utilizing AI to optimize energy consumption in buildings, and its potential impact on the building sector is undeniably significant.

## III.3. The used cases that are part of AI solutions of BIGG business cases

- **Energy benchmarking of buildings (GitHub)[2]**

This application focuses on the benchmarking and monitoring of energy consumption in buildings. It achieves two primary goals:

First, it evaluates the energy usage of a single building by analyzing its historical consumption and weather data, a process known as longitudinal benchmarking. This involves breaking down total consumption into three components: baseload, heating, and cooling. These components, combined with static building information, lead to the estimation of various Key Performance Indicators (KPIs). These KPIs provide insights into usage patterns, costs, savings, and emissions associated with energy consumption over time.

Second, the application compares the KPIs generated from the longitudinal benchmarking process with those of similar buildings in terms of characteristics and weather conditions. This comparison, referred to as cross-sectional benchmarking in literature, helps identify how a building's energy consumption performance compares to its peers.

Thus, the application can provide valuable insights for optimizing energy usage, reducing costs, and minimizing environmental impact. This contributes to more informed decision-making and improved energy efficiency in buildings.

- **Energy Efficiency Measures (EEM) assessment (GitHub)[3]**

This pilot evaluates the energy, cost, and emissions savings resulting from the implementation of specific EEMs. This assessment hinges on data-driven modeling of a building's energy consumption, comparing consumption before and after EEM implementation. This approach mirrors the energy benchmarking concept, constituting a longitudinal evaluation of EEM effectiveness.

- **Baseline identification for Energy Performance Contracts (GitHub)[4]**

The aim of streamlining the Measurement and Verification (M&V) process, which is a crucial step for Energy Service Companies (ESCOs) in managing Energy Performance Contracts (EnPCs), is pursued by the Energy Performance Contract (EnPC) management application. Typically, the M&V process is manually managed with Excel sheets, leading to time-consuming and error-prone outcomes. This makes the process inefficient and labour-intensive.

To standardize the M&V process, a solution has been developed by the BIGG consortium using the AI toolbox, with the goal of accurately and flexibly identifying baselines. The core modules of the toolbox were employed to construct a pipeline that can adapt to the requirements of any EnPC contract and facilitate the identification of a consumption baseline regression model from historical data, using input data such as weather, occupancy, and calendar information. High flexibility is provided by the developed pipeline regarding the types of models employed for baseline identification, allowing users to customize the resulting models to their needs while ensuring interpretability for non-expert users. The models that are identified can then be used as building blocks for global solutions, such as those developed within the context of the BIGG projects, where the entire M&V process, including savings tracking, assessment of financial benefits for both ESCOs and building owners, reporting, and more, is managed comprehensively from start to finish.

---

[2] https://github.com/biggproject/A1-Benchmarking

[3] https://github.com/biggproject/A2-EEM-assessment

[4] https://github.com/biggproject/A3-EPC-baseline-identification

- **Occupancy pattern detection for Comfort and energy optimization (GitHub)[5]**

Addressing the needs of both energy efficiency and occupant comfort presents a challenge for energy experts. Existing building management systems focus on comfort but neglect the growing demand from building owners to reduce energy expenses. Additionally, these systems often fail to consider factors like occupancy, weather forecasts, and special events. To overcome this, HVAC controllers need to incorporate new information sources such as weather data and occupancy patterns to decrease energy consumption without compromising comfort. Notably, accurately identifying occupancy patterns is crucial for determining optimal energy-saving strategies, but manual identification can be intricate.

The AITB gives a solution designed to automatically detect occupancy patterns within buildings. It ingests the activity of people of households via the IoT sensors and considers the holidays. The AITB lightens the workload for energy experts while enhancing the accuracy of collected data. As a result, this toolbox has become an indispensable asset for energy experts striving to enhance energy efficiency and create comfortable building environments.

- **Energy consumption forecasting (GitHub)[6]**

Forecasting energy usage plays a vital role in effective energy management and planning. Precise predictions benefit both energy providers and consumers by guiding well-informed choices. Introducing an intelligent algorithm for energy consumption forecasting significantly enhances forecast accuracy, fostering the growth of sustainable and cost-efficient energy systems.

This algorithm aids energy suppliers in orchestrating production and distribution, optimizing consumption, and preventing power outages. For energy consumers, precise predictions enable efficient energy management, cost reduction, and leveraging off-peak periods. Moreover, the algorithm identifies consumption trends, informing efficient energy policies and systems that work towards a dependable and sustainable energy future. The result is a more resilient and eco-friendly energy system.

- **Gas demand response (GitHub)[7]**

This pilot is dedicated to designing an ingenious demand response (DR) plan by tapping into the adaptability of gas consumption in residential heating. This collaborative approach engages customers, allowing gas providers to sidestep unwarranted expenses and diminish CO2 emissions. The pilot's primary goal is to hit a specific gas consumption target, effectively fine-tuning energy efficiency, curtailing gas usage, and yielding economic savings along with environmental advantages.

The pilot case realizes its ambition through collective management of multiple boilers, orchestrating their gas consumption to meet the stipulated objective. In this orchestration, user comfort remains paramount – residences must not swing to uncomfortable temperature extremes

To translate this vision into reality, a reinforcement learning (RL) approach is adopted, shaping a demand response controller policy. This RL agent learns from historical or simulated data, discerning optimal actions based on raw input data. The distilled policy is subsequently applied in real-world scenarios to execute the demand response strategy. The specific actions of the policy pivot on input data and the DR objectives, whether it involves tweaking boiler operations to align with target gas consumption or refining energy usage in response to evolving conditions.

---

[5] https://github.com/biggproject/A4-Occupancy-pattern-detection

[6] https://github.com/biggproject/biggpy/tree/usecase14/UC14

[7] https://github.com/biggproject/biggpy/tree/main/gas_demand_response

## III.4. Limitations and Considerations for using AI

While ML and AI tools offer numerous advantages, they are not without limitations. These tools heavily depend on high-quality and representative data. Poor or biased data can lead to inaccurate predictions or reinforce existing biases, making data quality a paramount concern. The complexity of ML models also raises concerns about interpretability. Unlike traditional methods, where the reasoning behind predictions is often clear, some ML models operate as "black boxes." This lack of transparency can be problematic, especially in applications where accountability and understanding are crucial. Another challenge is the risk of overfitting. If not managed properly, ML models might perform well on the training data but struggle to generalize to new, unseen data. Balancing model complexity to prevent overfitting is a delicate endeavor.

Additionally, the resource-intensive nature of training sophisticated ML models demands substantial computational resources and time, potentially limiting access for smaller organizations or individuals. Privacy and security concerns also arise when these models are trained on sensitive data, necessitating robust data protection measures. Thus, while AI brings forth unparalleled opportunities, it also raises complex ethical questions that cannot be overlooked. Addressing bias, ensuring transparency, safeguarding privacy, and establishing accountability are just a few dimensions of the multifaceted ethical landscape surrounding AI. A collaborative effort among researchers, policymakers, industry leaders, and ethicists is crucial to strike a harmonious balance between technological advancement and the well-being of individuals and society.

# IV. SUMMARY AND CONCLUSION

In summary, the H2020 BIGG project has developed a data platform having two key components in the form of a harmonizer and a highly promising AI Toolbox for Buildings (AITB) that has significant implications for enhancing energy efficiency and sustainability in the building sector. The development of an advanced data platform is centered upon the aggregation of input data into the dedicated BIGG Ontology to create harmonized data. The importance of harmonized input data in the energy sector lies in its role in ensuring consistency and comparability across diverse datasets. Easy comparison of energy consumption patterns is enabled by adhering to BIGG data model format and units. The data integration from various sources is facilitated via a harmonization layer promoting interoperability to support seamless data exchange and collaborative efforts.

The AITB, grounded in rigorous scientific principles and practical problem-solving approaches, offers a robust framework for addressing these critical challenges. Through its six distinct use cases, the AITB provides a comprehensive toolkit for optimizing energy consumption and improving building performance. From energy benchmarking and efficiency measures assessment to baseline identification for Energy Performance Contracts, occupancy pattern detection, energy consumption forecasting, and gas demand response, the AITB covers a wide range of energy-saving scenarios. The harmonization capabilities ensure its adaptability and scalability to diverse situations and user requirements.

By harnessing the capabilities of artificial intelligence, the AITB introduces a data-driven approach that has the potential to revolutionize building management. It offers the prospect of informed decision-making, cost reduction, reduced environmental impact, and enhanced occupant comfort. The AITB represents a significant stepforward in the pursuit of enhanced energy efficiency and sustainability. It aligns with our shared commitment to address pressing global challenges and promote responsible resource management.

In a world where climate change and sustainability concerns are paramount, the BIGG project provide a promising glimpse into a future where technology and innovation are leveraged to create more intelligent, efficient, and environmentally responsible buildings.

BIGG