Improving Buildings is Improving EU

D1.3 System Specifications and AI-based mapping concept

(T1.4 - T1.5)

Disclaimer and acknowledgements

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation program under the grant agreement No. 101092161.

Disclaimer

The content of this document reflects only the author's view and do not necessarily reflect those of the European Union or HADEA. Neither the European Union nor the granting authority can be held responsible for them.

Copyright message

Acronym **openDBL** **GA No.** ***101092161***

Full Title	ONE STEP OPEN DBL solution		
Call	HORIZON-CL4-2022-TWIN-TRANSITION-01		
Topic	HORIZON-CL4-2022-TWIN-TRANSITION-01-09	**Type of action**	HORIZON Innovation Actions
Project coordinator	CETMA		
Deliverable	[D1.3][System specifications and AI-based mapping concept]		
Document Type	[R]	Dissemination Level	[PU]
Lead beneficiary	**EM for Task 1.4 and iabi e.V. Task 1.5**		
Responsible author	[Michele Allori (EM) for T1.4; Jakob Martin (iabi e.V) for T1.5]		
Additional authors and contributors	[Michele Scafoglieri (e-metodi)] [Francesco Guerricchio (e-metodi)] [Filippo Baldazzi (e-metodi)] [Andrea Tiveron (e-metodi)] [Roberto Pazzaglini (e-metodi)] [Vito Losavio (e-metodi)] [Klaus Linhard (iabi e.V.)] [Jakob Martin (iabi e.V.)]		
Due date of delivery	22/09/23	**Submission**	[26/09/23]

Document information

Document history			
Issue	**Date**	**Comment**	**Author**
V1.0	06/09/2023	First draft	M. Allori
V1.1	11/09/2023	Draft expansion	M. Allori
V2.0	12/09/2023	T1.5 insertion	J. Martin
V2.1	13/09/2023	T1.4 expansion	M. Allori
V2.2	13/09/2023	T1.4 Sections 13 to 16	M. Scafoglieri
V2.3	14/09/2023	T1.5 expansion	J. Martin
V3.0	15/09/2023	Expansion, review, formatting, and corrections, for Project coordinator approval	M. Allori
V3.1	15/09/2023	Diagrams review	M. Scafoglieri
V3.2	16/09/2023	review, formatting, and corrections	F. Baldazzi
V3.3	18/09/2023	review, formatting, and corrections	M. Scafoglieri
V3.4	18/09/2023	Small corrections, adding T1.4 2.6.1	M. Allori
V3.5	19/09/2023	T1.4 implementation of CETMA's comments	M. Allori
V3.6	21/09/2023	review, formatting, and corrections regarding T1.5	J. Martin
V3.7	22/09/2023	Added function specification on Graph Database	M. Allori

Approved by:			
Issue	**Date**	**Name**	**Organisation**
V3.7	[26/09/2023]	Italo SPADA	CETMA

Contents

Index

List of Tables

List of Figures

Glossary of terms and acronyms used

[Guidance: Please sort alphabetically]

Table 1: Glossary of terms and acronyms used

Acronym/Term	Description
AECO Industry	Architects, Engineers, Construction Companies and Operators (Property and Facility Managers). These roles indicate the full value-chain of an industry that accounts for ca 8% of the EU GDP.
AI	(Artificial Intelligence) A field of computer science that focuses on creating systems capable of performing tasks that require human intelligence. These tasks may include problem-solving, speech recognition, and planning.
ANNs	(Artificial Neural Networks): A subset of machine learning models that are inspired by the human brain. They are used for complex tasks such as image and speech recognition, natural language processing, and more.
BIM	Building Information Modeling, a digital representation of physical and functional characteristics of a building or other facility, which serves as a shared knowledge resource for information about a facility forming a reliable basis for decisions during its life cycle.
bSDD	(buildingSMART Data Dictionary): A data dictionary that provides a reference for terminology used in the building industry, facilitating clear communication and collaboration.
CNNs	(Convolutional Neural Networks): A class of deep learning neural networks, commonly used in image and video recognition, recommendation systems, image generation, and natural language processing.
cuDNN	(CUDA Deep Neural Network library): A GPU-accelerated library for deep neural networks, which provides highly optimized primitives for performing deep learning tasks.
Design Patterns	are reusable solutions to common problems encountered in software design. They provide best practices for solving design and implementation challenges and improving code maintainability.
GBMs	(Gradient Boosting Machines) A machine learning technique used for regression and classification problems, which produces

	a prediction model in the form of an ensemble of weak prediction models.
Graph Database	a database that uses graph structures to store, map, and query data
HITL	(Human-in-the-Loop): A model of interaction where a computational system involves a human in the decision-making process, often to improve the reliability or accuracy of automated systems.
IFC	Industry Foundation Classes, a data model used to describe building and AECO industry data, is used as a common platform for sharing data across the building industry.
IoT	(Internet of Things): A system of interrelated physical devices that are connected to the internet, allowing them to send and receive data.
JSON	(JavaScript Object Notation): is a lightweight data interchange format that is easy for humans to read and write and easy for machines to parse and generate. It's widely used in web services and APIs.
KNNs	(K-Nearest Neighbours): A type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until function evaluation.
LCA	(Life Cycle Analysis): A technique used to assess the environmental impacts associated with all the stages of a product's life from cradle to grave.
NLP	(Natural Language Processing): A field of artificial intelligence that focuses on the interaction between computers and humans using natural language.
Ontology	A formal, explicit specification of a shared conceptualization that describes the concepts and relationships between them in a particular domain of knowledge.
OpenAPI	a specification for building APIs, which allows developers to describe their APIs in a standard, machine-readable format.
RDF	(Resource Description Framework): A standard model for data interchange on the web, facilitating the integration of a variety of applications using XML for syntax and URIs for naming.
RESTful API	(Representational State Transfer Application Programming Interface): is an architectural style for designing networked

	applications. It uses HTTP requests to perform CRUD operations (Create, Read, Update, Delete) on resources.
SDD	(Software Design Description): provides a detailed description of the software's architecture, design decisions, and implementation specifics. It explains how the software will be built.
SRS	(Software Requirements Specification): is a document that defines the functional and non-functional requirements of a software system. It outlines what the system should do and how it should perform.
STD	(Software Test Document): outlines the testing strategy, test cases, and procedures to verify and validate that the software meets its requirements. It guides the testing process.
Swagger	a set of open-source tools for building, testing, and documenting APIs that conform to the OpenAPI specification.
SVMs	(Support Vector Machines): A set of supervised learning methods used for classification, regression, and outlier detection.
.ttl	("turtle" Terse RDF Triple Language): A syntax and file format used to express RDF data in a concise and more readable way, compared to other RDF serialization formats.
.txt	(Text file): A file format that contains plain text, which is human-readable and can be viewed and edited using a simple text editor.
XML	(eXtensible Markup Language): is a markup language that defines a set of rules for encoding structured data in a human-readable and machine-readable format. It's commonly used for data exchange and configuration.

Executive summary

This deliverable will define the next months of software development for openDBL, setting the different functions of the platform and their interdependency. Deliverable 1.3 delves into the specifications, architectures and IT modules that will constitute the openDBL platform, its purpose is to define the various functions expected of such a platform, their possible development steps, their dependencies, and their interconnections with other functions. What follows is the result of three months of research and explorative development by the two main technical partners on the consortium, and the summary of the invaluable input from all other partners over the course of multiple meetings since the beginning of the project.

The report provides all partners with a clear view of the functioning of the future openDBL platform, bringing partners and readers on the same page on what can be expected from such an application.

Introduction

openDBL project summary

openDBL intends to integrate multidisciplinary know-how to cover the requirements of the Call and solve the issues of the current situation. The challenge of the project is to allow, through the development of openAPI, the disposal of openDBL in a unique standardized platform and create useful content, to simplify the workload of the AECO industry.

The project pursues 3 objectives: 1) create a DBL with useful content and functionalities, 2) ensure openDBL is usable and simple to use, reducing the time spent to upload, search, and process the information and data to facilitate usage and gain wide adoption, 3) ensure attractive economics, through value propositions and convenient pricing. We'll provide any user with an integrated platform for their digitization needs; ensure that information and data conform to the latest trends and needs of our target clients and support the EU's circular economy and green policies; develop automatic classification systems and data standards; facilitate the operation and maintenance activities of the buildings. This will be achieved creating an Information Delivery Manual and a Data Model and further developing our existing platform used to create a DBL for an important Italian Public Contracting Authority. openDBL will support data matching with external databases and will integrate state of-the art technologies (AI, Blockchain, IoT and VR). Our ambition is to make openDBL the platform of reference for the monitoring of building consumption, transparencies of transactions and official documents, and the positive impact on maintenance and environment.

To reach its goals openDBL is divided into 6 WPs with different goals, tasks, and deliverables. This Deliverable is part of Work Package 1: the whole WP is dedicated to the research and definition of what openDBL will effectively be, what requirements it will meet, what it will empower, and how it will be constructed. This deliverable marks the end of the 9 months period dedicated to WP1, defining the structure of openDBL and its core functionalities, and showing the results of the developing work so far. After this deliverable most of the development team efforts will be steered towards the actual development of these functionalities, rather than research or definition. The solutions offered by openDBL are numerous and innovative for this kind of project, the structure is to be intended as a backbone of which the more specific and detailed architectures and specifications will be defined only after early development is effectively ended, as is the nature of IT development.

Mapping openDBL outputs

Table 2: openDBL work description

openDBL GA Component Title	openDBL GA Component Outline	Respective Document Chapter(s)	Justification
	TASKS		
T[1.4] System Specifications]	*Goal is to define the specifications for the cloud environment, (most probably Microsoft Azure normally available as datacentres in both IaaS -Infrastructure as a Service and PaaS - Platform as a Service and for the microservices architecture managed by deployment and continuity technologies such as Kubernetes and Docker. In defining the system specifications, we will also plan how to interface and collect data from new technologies (e.g.. sensors, VR, etc.) or from existing or upcoming application, including considerations of privacy and security.*	*[1-5]*	*After an initial overview the chapters will go in details on how the functions of the openDBL web platform work, what their purpose is, and what is their added value to the end user*
T1.5 AI-Based Mapping concept	*As a starting point, this task will analyse existing standards, concepts, languages, and technology for relevant classification systems, especially under the aspect of ontologies and availability of machine-readable OWLs like BOT, SAREF, ifcOWL. Considering also relevant EU research, in particular the "Spheres" project and initiatives like bsDD-OWL, a concept for an AI-based mapping of classifications will be defined. Also available approaches of neuronal networks, machine learning and AI-technologies will be analysed for their suitability to query a wide range of classification systems to find relations of objects and attributes. In addition, existing AI-based automated translation systems like DEEPL will be analysed for their suitability to translate terms in different languages during the mapping process*	*[6-10]*	*The document emphasizes the significant role of AI in revolutionizing attribute-based mapping in the Building Information Modelling (BIM) process, particularly highlighting the advancements in data analysis and the optimization of mapping concepts in the construction and infrastructure development sectors.*

Deliverable Overview and Report Structure

- *Introduction [Page 1-12]*
    - *This section contains the disclaimer, the index, and the description of the Work package, Deliverable and tasks treated in the following document.*
- *System Specifications (Task 1.4)*
    - *Overview [Section 1-2]*

      These sections contains the software requirements and an overview of the various functionalities offered by openDBL
    - *Functionalities description [Section 3-13]*

      These sections explain more in detail the functionalities and their respective interactions
    - *Architecture and testing [Section 14-8]*

      These sections delve deeper into the architecture, testing, requirements, and solutions for the development of openDBL platform
- *AI-based mapping concept (Task 1.5)*
    - *Challenges and advantages [Section 1-5]*

      These sections provide an insightful overview of the challenges and advantages of working with an AI-enriched data model, providing also examples and case studies
    - *Challenges and advantages [Section 7-9]*

      These sections describe more in detail the concept proposed for the AI-based mapping concept in openDBL, defining the technologies and choices made so far.

T1.4 - System Specifications

1 Software Requirements specifications (SRS)

The Software Requirements Specification (SRS) describes what the software will do and how it should work.

This main section defines the purpose of the software to be built, it describes what will be built, going into the details of the project requirements. The requirements describe the functionality and quality expected from the software: therefore, incorrect, or unclear requirements will lead to the creation of software that is different from the expected one.

The purpose of this software project is to design and develop a system for managing one or more building models in a single data model repository called "OpenDBL". The system will have to provide a web interface but, above all, APIs for accessing and querying this data; this API will be based on the OpenAPI specification (an open standard), allowing developers to easily integrate this data into existing workflows and systems.

The goal is to make it easier for software construction professionals to access and use the data produced during the lifecycle of a building, while also ensuring interoperability with other systems, to meet the needs of the software construction industry. The system should be scalable, modular, and flexible, allowing the addition of new data to the system as required and should also be designed with security in mind, including appropriate access controls and authentication mechanisms to protect the integrity of the data.

2 Constraints

2.1 Assumptions

It is worth noting that the following assumptions were made in the software design process. First, the input IFC files adhere to the IFC format standard and are error-free. Any invalid files will not be processed by the system. We assume also that users accessing data via OpenAPI have a basic understanding of what RESTful APIs are, are able to understand the related documentation and, of course, know how to use the tools to access the endpoints made available by OpenDBL.

The system assumes even that API endpoints are accessible by a reasonable number of users at a time. In this context, if necessary, performance tests will be conducted to ensure optimal response times. Also, to ensure the reliability of the system, all data backups and integrity checks will be performed regularly.

Finally, it is assumed that the end users of the solution comply with all relevant regulations and laws relating to privacy and data protection.

2.2 Dependencies

The software will have dependencies on various external libraries, APIs, and frameworks to achieve its functionality. These are critical to its successful implementation and operation so these dependencies will be updated regularly and compatible with each other to avoid compatibility issues and security vulnerabilities.

2.3 Technical constraints

In illustrating some of the technical constraints related to the OpenDBL software, certainly one of the most important concerns both the size and the complexity of the data structure of the IFC files: this can lead to serious problems of efficiency in terms of both processing and extraction of the relevant information of the aforementioned files [1]. So, the software should handle this issue appropriately.

Regarding the integration in OpenDBL with the data generated by IoT sensors installed in the building, there are many more technical constraints to be addressed: for example the generation of a large amount of data in real time requires the design of data pipelines and mechanisms efficient storage; devices then use different data formats and protocols, so ensuring interoperability must be essential; security also is crucial here, so device authentication, data encryption and access controls are essential to protect the entire system. Other problems common to IoT solutions are managing a large fleet of devices, where it is important to update firmware and change configuration, minimize latency and bandwidth usage, ensure system reliability in terms of robustness and fault tolerance, optimize data transmission and processing to minimize power consumption. And all this without considering the costs (related to data storage, network usage and device management). Note that most of these issues are already being addressed and solved very well by off-the-shelf IoT solutions [2].

2.4 Regulatory and legal constraints

The regulatory and legal constraints for the software application include compliance with data protection laws, intellectual property laws, and other relevant regulations. The software should ensure the protection of personal data and privacy of users and adhere to data protection regulations such as the General Data Protection Regulation (GDPR).

Additionally, the software should ensure compliance with intellectual property laws, particularly regarding the use of third-party software libraries. The software should only use open-source libraries and tools or have proper licenses for any proprietary tools or libraries used.

2.5 Archi® software for architecture definition

In this document, and internally moving forward, the developers of openDBL will take advantage of software specifically designed for IT architecture conceptualization. In this regard, it should be considered that at the European level, within the EIRA©, has been suggested for some time the use of the ArchiMate® modelling language: an open and independent enterprise architecture standard, that supports the description, analysis, and visualization of architecture within and across enterprise domains. ArchiMate® is one of the open standards hosted by The Open Group® and is fully aligned with TOGAF®: another Enterprise Architecture framework. The ArchiMate® standard has Archi®: a free tool, open source, cross-platform tool for creating models according to the ArchiMate® standard. The tool is aimed at all levels of enterprises, companies or institutions that need to describe their organization through a standardized modelling language. Archi® is a software tool for documenting, communicating, and sharing the most salient interoperability elements of complex solutions and facilitates the sharing of solutions that are (re)usable.

3 Functionalities overview

The OpenDBL platform will be structured as a hub of services all referring to data for the life cycle of a building, provided by the owner and other collaborating actors (such as designers, builders etc.). As the array of services and functionalities are numerous, each will function through existing web libraries, modified ones, or custom-made services. Existing libraries are open source and free, to keep the cost of development as low as possible.

The following list presents all currently planned functions for the platform:

- User identification
  - Role Definition
  - Trace services
- Project Register
- Document Logbook
- Common Data Environment

- IFC Manager
  - IFC Importer and exporter
  - IfcOpenShell
  - Building data Editor
- 3D viewer
  - VR Viewer
- Point Cloud Viewer
  - Measure tool
  - 360° photos
- Maintenance tool
- Sensor Gateway
- Map Tool
- Collaboration Platform

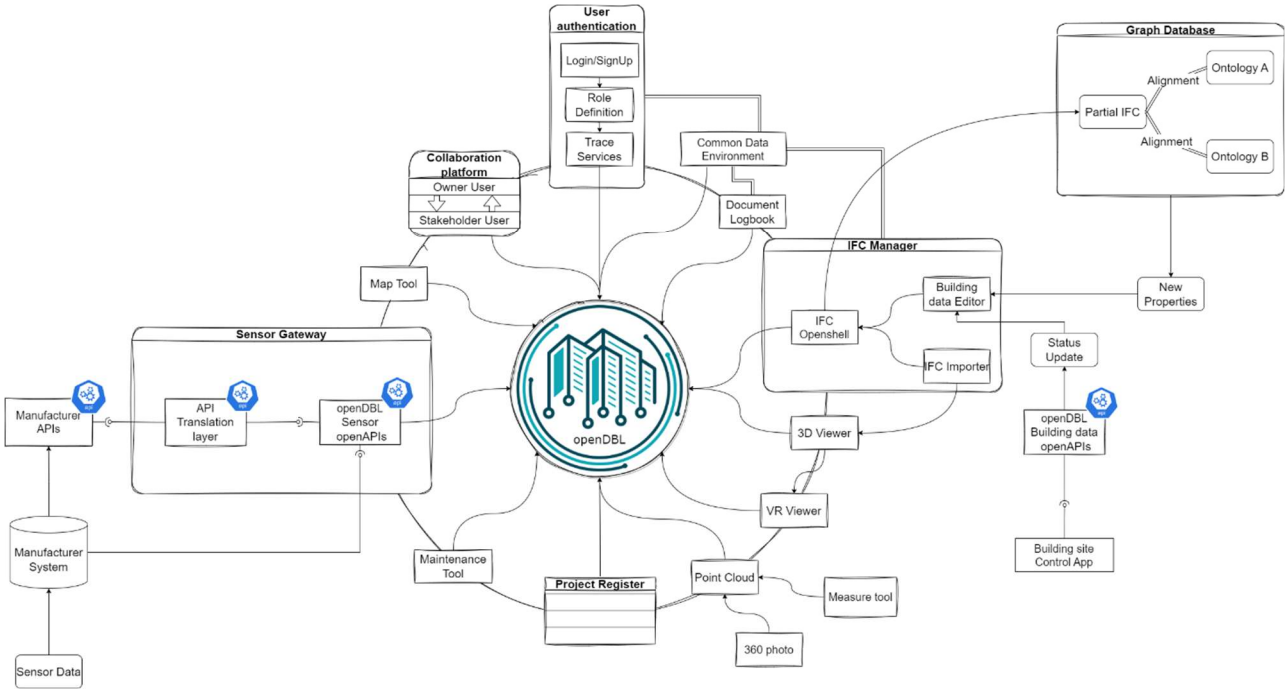


Figure 1 Overview graph of the openDBL platform and all its services

(Bigger version in Annex I at page 114)

Throughout the document there are some images depicting the user interface of a partially functioning early prototype that has been developed in July and August of 2023. These images are not to be intended as representative the final interface of

openDBL and are presented solely as visual aid to convey the full potential of openDBL functionalities.

3.1 Use case requirements

In D.1.1 a list of use cases was created. This list included use cases and technologies that would have been facilitated (or in some case would facilitate) by openDBL. All these activities stand to be improved and made easier by our developed platform and service. Follows a matrix showing what use cases might be covered by what openDBL functionalities

Table 3: Matrix of functionalities and use cases

Platform Functionalities	BIM Uses									
	Scan to BIM	Product comparison and choice	Construction and BIM to Field	Energy Management	Compliance	Health & Safety	Structural Monitoring	Facilities/ Asset Operations & Management	Energy audits & building renovation	Deconstruction and Circular Economy
User Authentication	X	X	X	X	X	X	X	X	X	X
Project Register	X	X	X	X	X	X	X	X	X	X
Document Logbook	X				X					X
Common Data Environment			X		X			X		
IFC Manager			X						X	X
3D Viewer			X						X	X
Point Cloud	X									
Maintenance tool				X		X	X	X	X	
Sensor Gateway				X		X	X		X	
Map Tool			X			X				
Collaboration Platform		X					X	X	X	X

4 User Authentication

As a platform dealing with potentially sensitive data, openDBL will require a secure Log-in service able to verify identity and ownership of data, using two factor authentication and trusted identity providers. Along with user IDs, this function comes with a number of subservices to ensure a smooth operation both for the development team and the users themselves.

4.1.1 Registering as a standard user

The user has the possibility to register as a standard user, having complete control over the buildings they own and over who collaborates on them. Once the user has registered as a standard user and confirmed their email, they can access the openDBL services anywhere with a connected device, using the login function.

4.1.2 Registering as an organization

This is the process for municipalities, companies, or organized groups of people who all work on the same projects in various capacities.

The user has the possibility to register an organization in openDBL, the creator of the organization is the administrator for the whole organization, inside openDBL. The administrator will have the possibility of creating new accounts into their organization, supplying the e-mails and relevant data and roles of any other member of the organization that is needed inside the platform. The nominated user will receive a notification and will be able to accept the invitation and start collaborating. The administrator can change roles, authorizations, and even delete or create new accounts inside the organization structure, through an intuitive interface. Internal users that are given these accounts do not have the normal control that standard users have on their own, as they cannot change their role independently, nor can they modify profile data or delete their profile.

4.1.3 Registering as a service provider

This is the process for organizations and companies that offer any kind of services in the AECO chain and want to be visible in openDBL for other users.

This type of profile is much different from the other two, as the service providers cannot create and manage their own projects, but they can only access data that has been shared with them from other users. Standard users can select a service provider to collaborate in their project and share data with them, granting them

access to what is needed for the completion of the offered service, or for a quote (more information can be found in section 13 Collaboration Platform).

4.1.4 Login process

This is the process for accessing the services of openDBL depending on the user's Role, organization, and account type.

The user navigates to openDBL login page initiates the login, they enter their email and password, or use one of the identity providers accepted by openDBL. The system verifies the user's credentials:

- If valid the system logs the user into their account and is redirected to the application's dashboard or homepage: Tracing of the activities is begun.
- If invalid the system displays an error message and the user remains on the login page: the failed attempt is registered.

At any moment, while using the application, the user can click on the log-out button or link anytime when navigating the pages of the platform, the system then logs the user out of the platform and redirects them to the login page or in a notification page: Tracing of the activities is stopped.

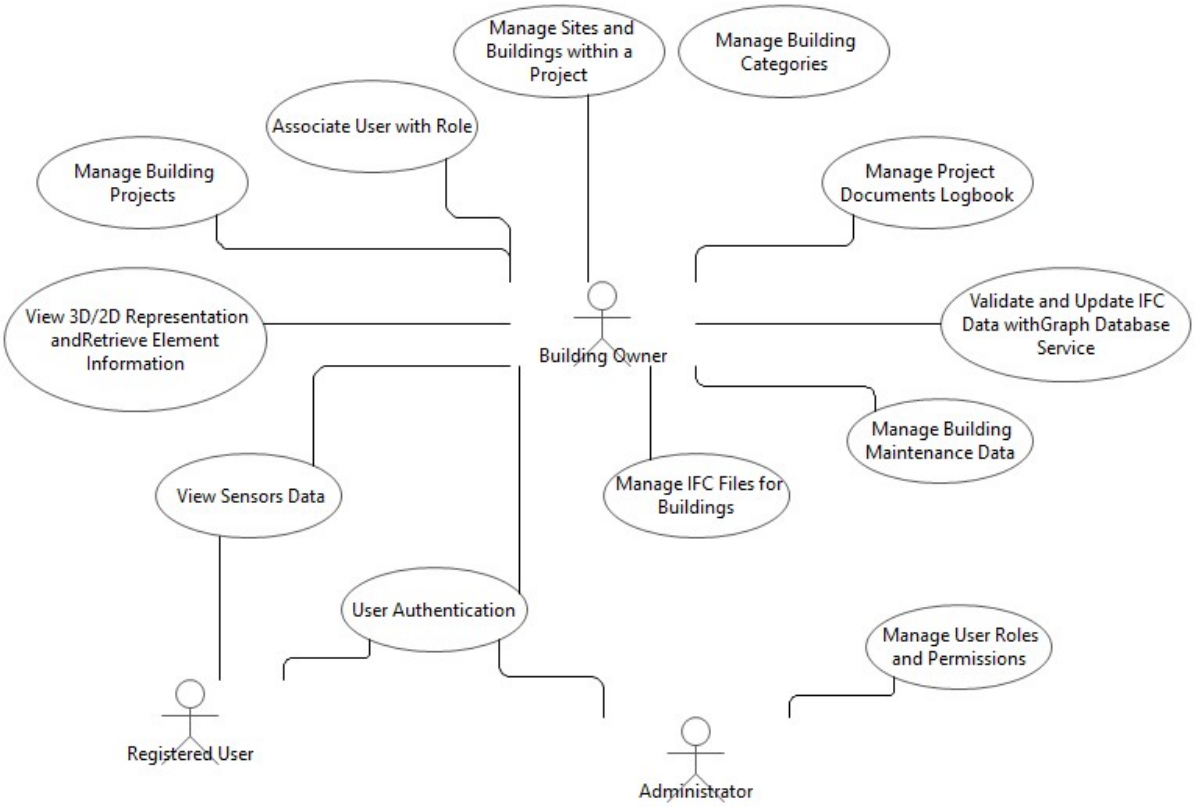
4.2 Role Definition

Users of openDBL will have different roles not only on the platform as a whole, but also in relation to the projects. openDBL will have three categories of accounts that a user can create and register as

- Standard account
- Organization Account
- Service provider

In this system, all accounts have limits on what they can do in regard to a specific project. Owners of a project will always have full control over the data regarding their project. Standard accounts can be "linked" to a project they do not own, as collaborators. When an account is linked as collaborator, they are assigned a role (usually by the owner of the project); this role defines the actions that the collaborator can and cannot take in regard to the project, the different roles available will have defined authorization levels and will come from a set already defined by the platform, rather than having the owner define them. In much a similar way, organizations could create a realm inside openDBL in which they

would create accounts for other people to collaborate on project; the main difference from the standard user account is that the account created thusly are fully controlled by the organizations, and the user assigned to them will not have



the possibility to modify, create or delete them. Service provider accounts are entirely different, as they do not create their own project, but rather are registered on the platform and might be contacted by standard users or organizations to provide a specific service: openDBL would facilitate this interaction by implementing a collaboration system (more information can be found in section 13 “Collaboration Platform”).

Figure 2 An example of allowed actions based on user Role

Such a structure will be both vertical (as an account will have a set of possible action for their project) and horizontal (as that same account will have different sets of actions depending on the project).

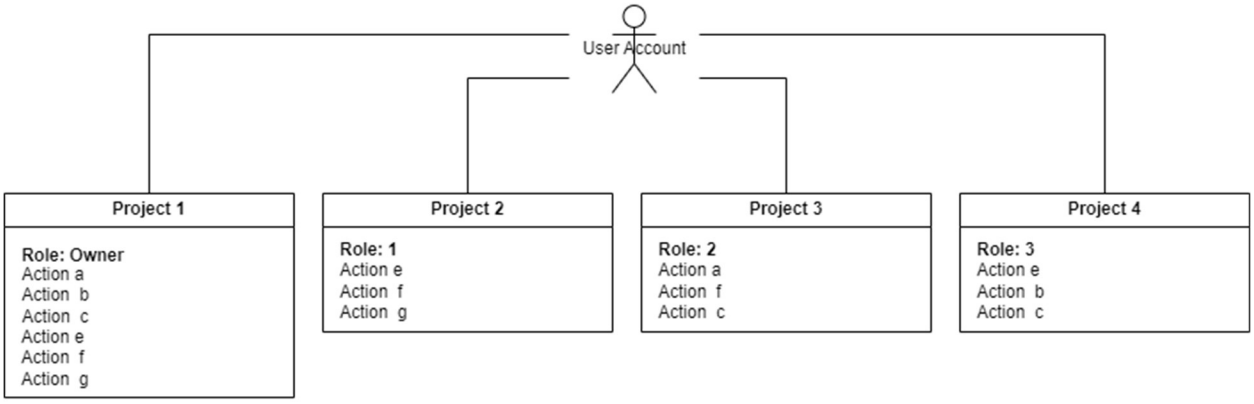


Figure 3 The vertical and horizontal structure of the role definition

4.2.1 Organization realm

The structure of organizations inside of openDBL.

All accounts created inside the organization by the Administrator after the creation of the Organization "realm" have different levels of access to different projects. When first created, these accounts need to be assigned at least one projects (more projects could be added in the future). For every project they are assigned to, they are assigned a Role. openDBL offers a list of pre-made roles; these roles grant the accounts access and editing permissions to various aspects of the project. For example, a lead architecture designer might have full approve/read/write/delete permissions for the architecture IFC file of a project but only have reading rights to other IFC files or archived documents. A user that has access to these accounts will only ever see the projects he's involved in, in their openDBL Project Register. For all intents and purposes administrators count as "owners" of the building project and will be considered equivalent from this point on in this document, Organization accounts count as contributors who do not have the possibility to create their own project, as they operate solely in the confines of their organization realm.

4.2.2 Standard user realm and contributors

How standard users can have different roles even when not using an organization account.

When a standard user creates a project inside openDBL is given the "owner" role by default. This role has full authority over every aspect of the project. Owners can approve files in the common data environment, upload, delete and update files and data of their building and involve other users in their project. Standard users can also be involved in other projects of which they are not owners; another user can nominate them as a contributor to their own project, with a role assigned to them. These contributor account work in the same way as organization accounts, the only difference being that they still have full control over their accounts, and they created their account using the e-mail address chosen by them, rather than the one chosen by the organization. Owners (and accounts with the proper role) can select a project, nominate another user as contributor and specify their role. In the project register, standard users can see both the project they own and the ones to which they collaborate. If an owner nominates as contributor an e-mail address that is not registered to openDBL, that e-mail address will receive an invitation to join both the platform and the project.

4.3 Trace Services

A common solution in collaborative IT environments, trace services are a collection of scripts and functions that trace actions taken by users inside a specific application into a dedicated archive in order to improve debugging, bug fixing, tracing the source of eventual problems, and analyse user operations. After login and before logout all actions done by an account are registered with timestamps, identification codes (id) and actions taken. A dedicated archive will speed up data access in case of consulting needs, as data volume is non-negligible.

5 Common Data Environment (CDE)

The Common Data Environment (CDE) function in openDBL is an added value feature that fosters seamless collaboration among various stakeholders involved in building projects. It ensures data consistency and reliability throughout the entire Building Information Modelling (BIM) process. This CDE will adhere to ISO standard 19650, to provide users with a well-structured framework for managing building design stages, encompassing everything from initial drafting to final execution. The CDE will interact with most openDBL services. Including a functioning common data environment in openDBL will greatly enrich the already vast set of services the project aims to deliver.

5.1 Horizontal environment

The CDE will interact with most openDBL services horizontally, meaning that most of the services will display, in their interfaces, important metadata to facilitate collaboration. However, as the planned openDBL CDE will cover only design and construction phases of a building life cycle, not all functionalities will include this.

5.1.1 In the Role definition

As the role definition already manages permissions and access rights, it already acts as a fundamental aspect for a common data environment, defining already what users can access specific project data and perform certain actions. Trace services and secure login practices ensure data security and correct permissions.

5.1.2 In the Project register

In the project register dashboard, along with an overview of their ongoing projects, users can access tasks related to those projects, key project metrics, timelines, and recent activities for quick reference. Tasks can be assigned to roles in the numerous services.

5.1.3 In the Document Logbook

By interacting with the list of documents, a user with the appropriate role might be able to assign some tasks or request updates from other roles. The document logbook will keep relevant metadata (such as uploader ID, numbered version, approval status, etc.) to ensure version control. A function of approval for uploaded documents from the relevant role will also be present.

5.1.4 In the IFC Manager

Similarly, to the document logbook, the IFC manager will display relevant metadata for version control, authorization levels. As the CDE will follow the ISO standard 19650, all IFC files will be stored using the appropriate stages, up until the project is considered built and finished.

5.1.5 In the 3D viewer

The 3D viewer will automatically display the latest approved version of a project, according to the filters set.

5.2 Notification centre

Users will receive notifications about project updates, comments, and tasks, helping them stay informed and responsive, based on their relative role in the project.

6 Project Register

The Project register is the main service presented after login as almost all functionalities of the platform are referred to specific projects and will therefore be accessible through the Project register. As a service the Register will allow Users (mainly Owners) to browse the projects they have part in, access the related functions, create, delete, rename projects, visualize changelogs, and set contributors.

6.1 Project data interactions

Inside the project register, the user is presented with all the projects they own or collaborate on. The interface includes meaningful and immediate information such as, project name, description, start date, end date (if applicable) project status, relative role of the user etc. Through this register, the user can access all openDBL services for that project, as granted by their role.

6.1.1 Creating a new building project

In the process of initiating a new project within the building project register system, the system facilitates the creation of a project by allowing users to input essential project details (i.e. name, position, etc.). Once the information is provided, the system performs validation to ensure it meets the required criteria. If the information is valid, the system proceeds to generate a new building project record. Additionally, upon successful project creation, the system automatically registers the user as the project owner, granting them the highest level of project authorization.

6.1.2 Editing an existing building project

To modify an existing building project within the project register, users first select a project for which they have ownership rights from the available list. They can then proceed to make the necessary modifications to the project information. After completing the edits, users can save their changes. The system subsequently conducts validation to ensure the changes adhere to the required criteria. If the modifications are valid, the system updates the project record with the edited information.

6.1.3 Nominating contributors to an existing building project

Users with ownership rights for a project can choose to add contributors to that project. This is done by interacting with the relevant project inside the list and accessing the form or service for adding contributors. Users then input the contributor's email address and necessary data and select the desired role for that contributor from a list. After these selections are confirmed, the system conducts the usual validation steps and updates the project record with the newly added contributor information.

6.1.4 Deleting an existing building project

Users who are project owners can delete a project from the platform, the system provides an option to delete the selected project along with a warning message to confirm this action. When the user confirms the deletion action, the system proceeds to delete the project record, along with all related documents, data, files, and database tables associated with that project. This process ensures the complete removal of the selected project from the system.

6.2 Building project front-end services

After the user decides to access the services related to a specific project the system presents that project's main interface. This front end interface grants access to all openDBL front-end services for that building project. From here, users will be able to directly access to:

- Documents
- Building Data Manager
- 3D Model
- VR Viewer
- Point cloud
- Maintenance plans
- Sensor Data History and Graphs
- Collaboration Platform portal

The list is subject to change as some services might be contained or be considered separate from others as the development of the platform goes on

7 Document Logbook

The document logbook consists of a customizable menu containing a list of document types, both pre-existent or user created, regarding the buildings and sites of the project. From administrative documents to bills, the logbook will orderly hold

all the digital documents that the owner may need for archival purposes. A selection of document types has already been implemented in an early prototype. Ideally, in the future, openDBL would recognize in which country the project resides, and consequently generate a list of document types calibrated according to the legislative needs of the country in question.

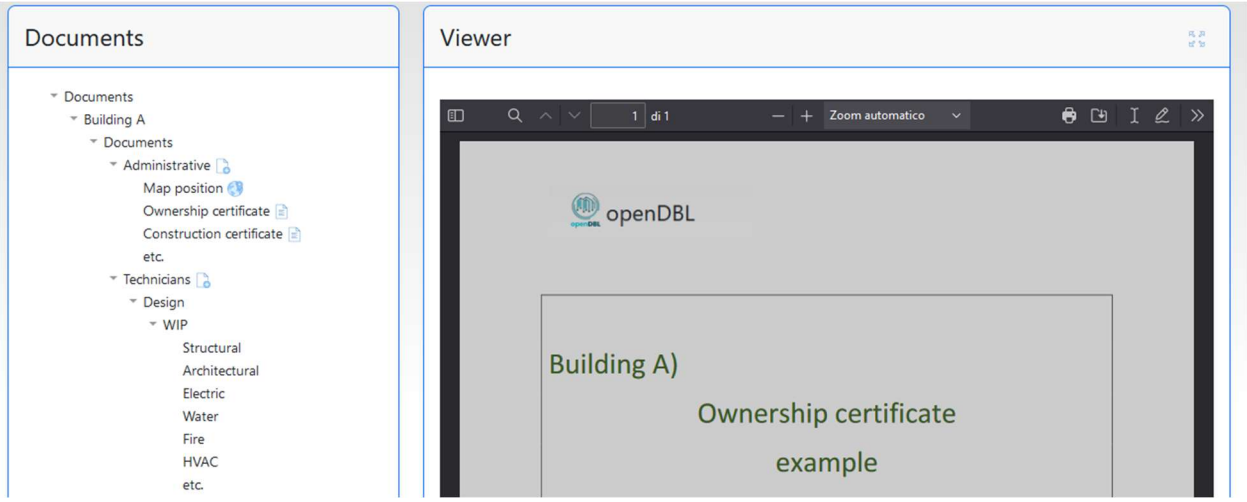


Figure 4 The document logbook interface in our early prototype

7.1 Document List

The user is correctly logged in and in a Building Project's main page, navigates to the document logbook service and is brought to the list.

This interface is presented with a set of document types and a main page space. The set of documents presents the user with a list of categories and sub-categories of documents (i.e. Administrative, Design, Ownership etc.). They can only see and interact with the documents their role allows them to. When a category or sub-category is open, openDBL would presents the option of selecting a document. The documents in the categories are pre-defined by the system, however the user can add their own, in case the list of available documents is not sufficient.

7.1.1 Uploading a new document

Within the "Document Logbook" system, users can upload documents related to their building projects. Users initiate this process by interacting with the system's interface to navigate to a document that's not present. They interact to add a document entry. Upon proceeding, they are directed to a relevant area where

they can provide and upload the desired document in an accepted format, such as .pdf. Once uploaded, the document is stored in the system's database, making it accessible for reference within the system. After this, the document appears in the document logbook list.

7.1.2 Interacting with an existing document

In the "Document Logbook" system, users can interact with documents that are already present in the system. They begin by navigating the system's interface to locate and select an existing document. Upon opening the document, they are presented with a preview of the document along with relevant information, including the date of upload and a series of functionalities. The currently planned functionalities are:

- Download: This option allows users to download a copy of the document to their local device for offline access.
- Upload: Users have the option to upload a new document, replacing the existing one if needed.
- Delete: This removes the document from the system's database, rendering it inaccessible for future reference or consultation.

These functionalities enable users to effectively manage and interact with existing documents within the "Document Logbook" system.

7.1.3 Create a new document type that is not on the list

In the "Document Logbook" system, users can also create new document types not already listed. Interacting with the system's list interface, which provides intuitive cues for adding a document, users specify the document name and upload the file. The new document is then seamlessly integrated into the list and behaves like any other document for future reference and interaction. If the user chooses so, they can delete document types created this way, so long as there is no actual document file associated with them.

8 Ifc Manager

As the vast majority of building data is exchanged using the IFC format, it being an ISO standard that is internationally adopted, openDBL will manage building information accordingly, managing, importing and possibly enriching IFC models in order to visualize and store relevant data about any specific project. The IFC manager function will obviously be tied with the common data Environment, as the upload or

update of IFC files will follow CDE standard approval procedures (in the pre-built phases), and with the Project register, as the IFC files will most likely carry the most comprehensive set of data for any specific project. At the same time the uploading of an IFC model from the user won't be essential to use openDBL but it will allow most of the functionalities planned in the platform.

8.1 IFC Importer and exporter

The first function encountered will be an import module for IFC files. The associated project would be inferred by the project selected by the user in the Project register after Log-in (eventual mismatches between projects ID in the file will be ignored), while the discipline relating to the IFC file to be loaded (architectural, structural, electrical, etc.) will be selected by a specific list.

8.1.1 Importing a new IFC file

When inside the IFC manager, the user with the adequate role may upload a new IFC file to the system. While uploading, it will be specified to which building system(s) the IFC is referring to, their working status, and what phase of the building process they belong to.

8.1.2 Exporting, editing, and deleting an existing IFC file

In a much similar way, the user would be able to interact with the existing IFC files on the platform, uploading a new version, deleting one, or even exporting one (taking advantage of possible IFC enrichment thanks to the system described later in T1.5 - AI-Based mapping concept).

8.2 IfcOpenShell

IfcOpenShell is a component that helps with the development of digital platforms for the built environment. This component allows to execute queries directly referred to IFC files without a need for conversion or translation. This would allow openDBL to store the IFC files directly into the filesystem of the server on which it operates. Most of the operations regarding IFC data will take advantage of this component, as the heart of the IFC Manager function.

8.3 Building Data Editor (BDE)

As data is imported onto the platform via an IFC file, non-geometric data needs to be visualized, and possibly edited by the user. The building data editor will present this data clearly in a folder structure where the relation between IFC subclasses and properties will be intuitive and searchable.

8.3.1 IFC Elements and Hierarchy

Elements inside the IFC files imported by the user will be presented in an easy to understand hierarchy, and through the data editor they can access all the different properties of the elements contained in the virtual building. The interface will allow users to easily interact with the various data and modify them as needed. At the same time there will be a way to visualize a specific element directly into the 3D viewer or get to the data of selected sensors.

8.3.2 Building data openAPIs

A structure of APIs will be developed and made public so that external services or software might access and write more details regarding the building project. Eventually these APIs might allow users with the appropriate data to interact with the BDE from outside the openDBL platform and add data to the IFC element that is not necessarily part of the IFC files itself (i.e. installation Status). Eventual changes in these data might even influence maintenance plans inside the platform. As the consortium (lead CEM) will develop a mobile app this functionality will be used.

8.4 Graph Database

Even though the IFC is a comprehensive classification system, not all use cases are covered by its structure, and many designers and BIM modelers tend to create their own subset of properties and classes. In IT Ontologies were developed exactly to standardize concepts between different parties and to define relationships between these concepts; openDBL aims to take advantage of the power of ontologies and relate them to the IFC schema (more information can be found in Deliverable 1.2 IDM and Data model). The usage of a Graph database would allow for easier integration between ontologies and IFC, the development of an AI system to map the various entities. More information in section T1.5 - AI-Based mapping concept. Labels originally coming from custom Psets in the IFC files might be renamed based on the mapping service and available ontologies, therefore standardized and exportable. This approach utilizes a Graph database to seamlessly integrate the IFC schema with ontologies, laying the groundwork for AI-driven attribute mapping and standardization. This integration will most likely be an automated process happening in the back-end, and not be visible to the user; an option to opt out of this functionality might be provided. For operative details on how Human-in-the-Loop systems manage this complex task, refer to section in T1.5 - 8.3 Human-in-the-Loop Systems for File Alignment in openDBL.

9 3D Viewer

Geometric data coming from the IFC import will be compiled into a single set to be viewed online as a 3D model in order for the user to navigate more intuitively the virtual space of the building and identify the relative position of every element. The 3D viewer is a powerful tool that can be expanded with multiple functionalities, most likely by branching existing JS libraries for IFC viewing (such as ifc.js). A functionality already planned is the linkage between the elements in the 3D viewer and their properties inside the building data editor, so that a user could select a 3D element and visualize either a link to that element's page inside the BDE, or a short list of important properties.

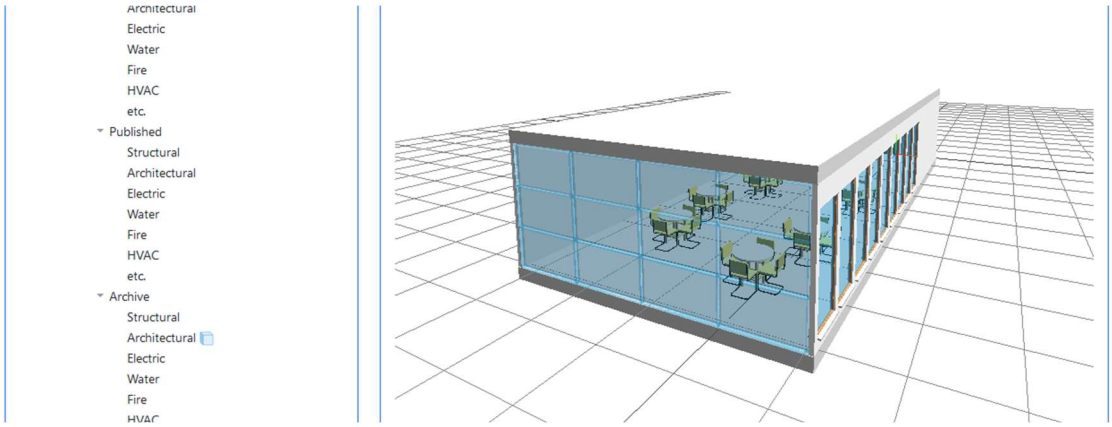


Figure 5 The 3D viewer interface in our early prototype

9.1 VR Viewer

Virtual Reality technology allows for immersive experiences and better understanding of complex architecture. openDBL aims to make the 3D viewer, and all of its functions, available for VR headsets, in order for the user to navigate in real scale the environment and better envision eventual operations.

10 Point Cloud

The point cloud visualizer, even though still a 3D virtual space, will be separate from the 3D viewer, as a point cloud model is the result of a measuring process for existing building, but it's not required to create an IFC 3D model. The point cloud visualizer will allow users to navigate any point cloud uploaded to the system, download and re-upload them for archival purposes or safekeeping.

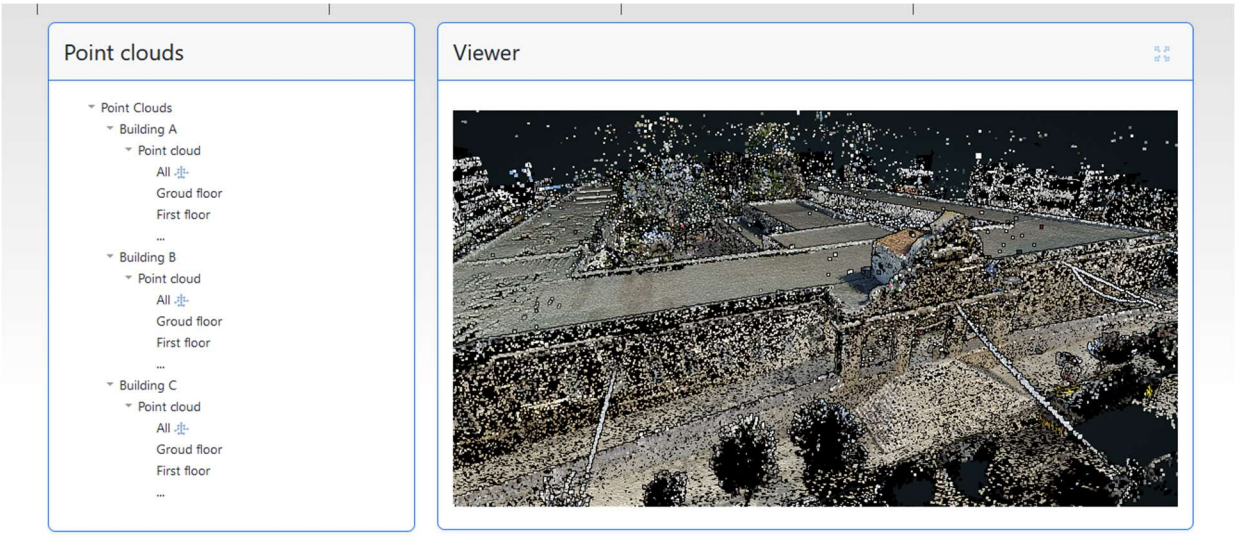


Figure 6 The point cloud interface in our early prototype

10.1 360 photos

Building survey tools with 3D laser scanners, unlike surveys generated through photogrammetry processes, generate a preview in digital format in the form of images for each acquisition station (scan position). These images are a mere 360° two-dimensional representation of the 3D survey carried out at the "observation point" of the laser scanner which are useful in the first instance for a visual inspection of what was acquired. openDBL aims to add a selection of these photographs (ideally chosen by the user) to the point cloud visualizer. These will be represented in the point cloud as special icons, or dots with a special colour, on click the user will open the selected 360photo to navigate it and check the status of the space at the date the photo was taken.

10.2 Measure tool

One of the additional functions of the point cloud visualizer would be a measuring tool. This would allow the user that uploads a point cloud to openDBL but not a finished IFC 3D model, to reference the building after the survey is completed and check for linear dimensions that might be needed afterwards. This would allow for informed decision-making without the need of re-measuring in the real world.

11 Maintenance tool

This tool works with direct linkage to the building data editor and the IFC manager. As the editor consists of a comprehensive set of digitalized elements present in the

building (after its construction), this opens up the possibilities of interaction with said data set.

A maintenance tool would allow users to individually select elements, or select them by category, and assign to them a maintenance routine, either coming from an existing set (that could be added by the developers) or by creating a custom one. This routine, containing descriptions, time needed, frequency and possibly cost, can be represented in various ways (i.e. as a colour coded Gantt chart or as a dynamic virtual calendar, or other). At the same time a function for ticketing and extra-ordinary maintenance (i.e. repairs, cleaning…) can be added with an urgency classification system, so that other users with the correct roles inside the project can take part in the active management of the building. Open DBL APIs could be added to this system so that external services could interact with the maintenance of various elements in the building.

The very nature of this kind of service means that its interface can be designed with a great amount of flexibility. This means that it could be presented as a stand-alone front-end functionality, or as integrated part of another, such the building data editor itself. Extensive testing will be performed to ensure the most intuitive and seamless user experience.

12 Sensor Gateway

Sensors are the only element in a building that actively and continuously provide data, whereas other elements have a set of properties that is unchanged after construction or installation, short of manual updates. openDBL will provide openAPIs for sensor data reception, so that potentially every sensor manufacturer or manager could send the correct data, in the correct format, towards their client's openDBL project. Given the incredibly vast amount of data produced by sensors, and the size of infrastructure needed to store raw sensor data online for multiple projects, most likely openDBL will not collect or aggregate *raw* data coming from the sensor, it will rather archive data that has already been processed by the manufacturer, manager, or vendor. The APIs exposed will express the correct format and flavour needed by the system. In case of sensor systems that already have an API structure for sensor data in place, a middle translation layer, putting the two API structures into communication, can be developed ad hoc.

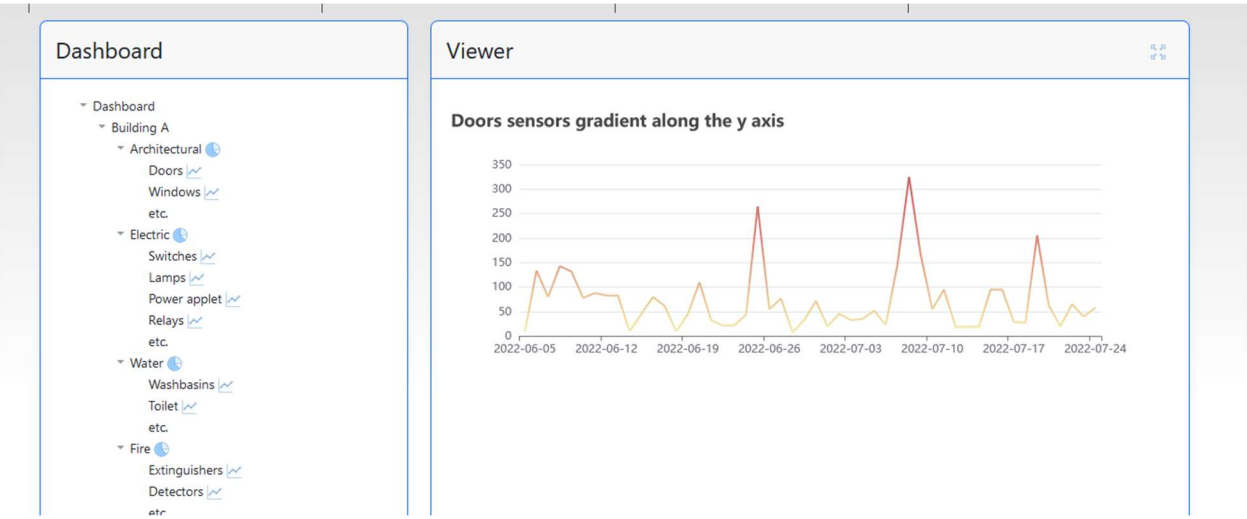


Figure 7 The IoT data visualization in our early prototype

12.1 IoT Data Visualization

Data coming from sensor allows building owners and other stakeholders to take informed decisions regarding their building. It is important then that sensor data is not only stored, but also presented in such a way that informs the user quickly on their inherent quality. This will be achieved through the use of timelines, data history, graphs, and charts. One example of this could be a graph showing the average levels of CO^2 registered in a room for every hour, informing the user that the room might need some extra ventilation in specific hours.

12.2 Events

While many sensors can provide continuous fluxes of data, often it's the fluctuation of these data or significant variations that are really important. Other times the sensors themselves instead of providing continuous data, record occurrences in a specific instant. These occurrences and important variations are called events; openDBL will track events and their time of occurrence for relevant cases (i.e. a rapid spike in temperature, the opening of a door, the passing of a CO^2 threshold, etc.). Depending on the specific case, some events might require a notification or alert to one or more users involved in the building project. One example of this could be an event pointed in the graph of the average hourly CO^2 levels, signalling a rapid spike at a specific timestamp. This would inform the user that something caused the air in the room to rapidly change and would help identify (and maybe mitigate) the cause.

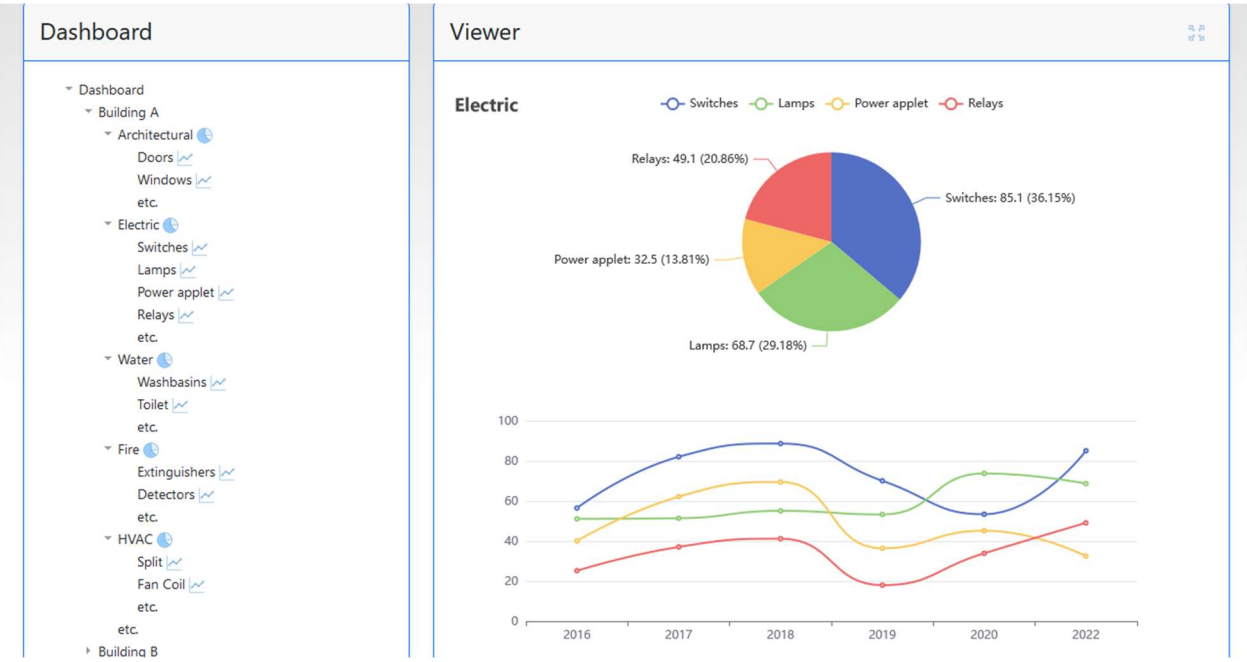


Figure 8 possible events in a consumption graph in our early prototype

12.2.1 Alerts and Blockchain

Notable events that warrant a user's attention might be too important to get lost in the system, or to be communicated wrongly. Adding a blockchain service to the logging of such events would guarantee the origin and correctness of the data received, both for safety and legal purposes. An example would be the event of a smoke detector going off: having this alert and data be exchanged via blockchain would guarantee the authenticity of the record and ensure the integrity of the data.

13 Map Tool

A simple map tool would pin the locations of different projects on a map, for a user. Upon a click a small informative panel about the project will pop up, including the role of the user regarding that specific project. This could also function as a link to the project pages just as much as the building register home page.

14 Collaboration Platform

A data exchange function, it allows the owner (or user with the right role) to authorize viewing and/or full access to specific documents and data regarding their project. In addition, users could select from a pool of service providers regularly registered onto openDBL to send them data and documents in order to request a quote for an intervention or service related to the project. This pool will be presented in an easy to understand list that is searchable and can be filtered in order to give the user the most flexibility and option in researching the perfect project partner. As this functionality would most likely have a scale comparable to the whole openDBL project, this functionality will be developed as a proof of concept, rather than a fully-fledged service. This would act as a starting point for an eventual future development after the openDBL project.

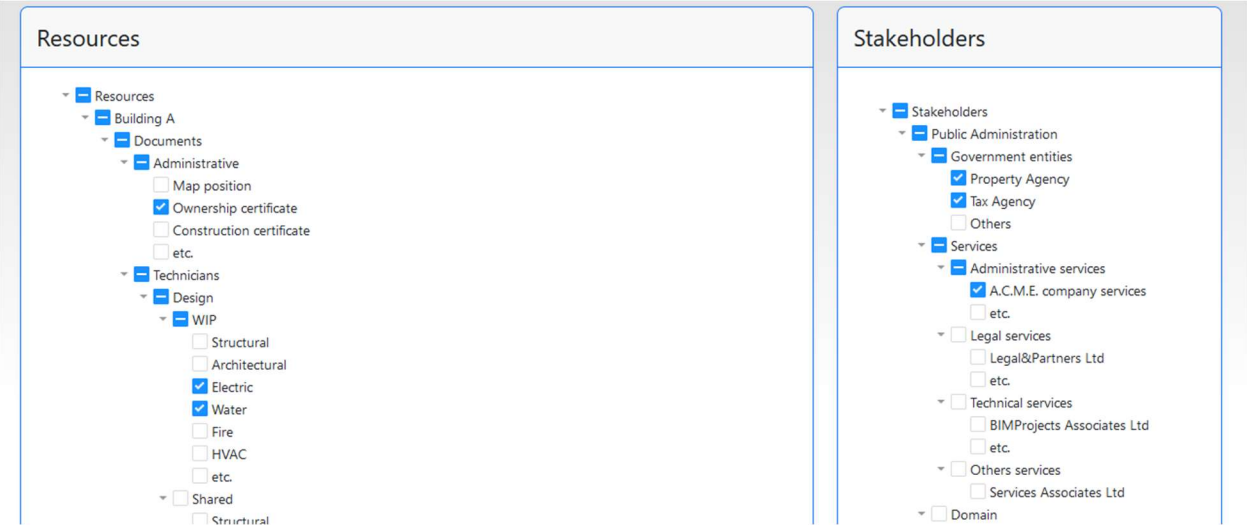


Figure 9 Data sharing and collaboration platform interface in our early prototype

15 Non-functional Requirements

Non-functional requirements (NFRs), also known as quality attributes or system qualities, are essential characteristics that describe how a software system should perform and behave, but they don't directly relate to specific functionalities or features of the software. They focus on the overall performance, usability, security, and other aspects of the system that influence its quality. Non-functional requirements help define the user experience and the system's ability to meet its functional requirements.

15.1 Performance requirements

The performance requirements of the application are especially critical when dealing with large IFC models.

The following performance requirements should be met:

- Response time: The application should provide a quick response time for user requests, even for complex queries on large datasets. The response time for typical queries should be within a few seconds, and the application should be able to handle multiple concurrent requests without significant slowdowns.
- Scalability: The application should be designed to be scalable and able to handle growing data volumes.
- Hardware requirements: The hardware resources required by the application should be reasonable and feasible. The application should not require excessive amounts of CPU, RAM memory, or storage to run efficiently.

15.2 Security requirements

OpenDBL handle sensitive information related to a building project, so it is crucial to implement these security requirements:

- Access Control: access to the application and its data should be controlled through secure authentication and authorization mechanisms. The users should be authenticated before allowing access to the system and their roles and permissions should be defined based on their responsibilities.
- Protection against attacks: the system should be protected against common attacks such as SQL injection, cross-site scripting (XSS), and cross-site request forgery (CSRF). The application should be designed to ensure that user inputs are validated and sanitized.
- Secure data transmission: the data transmitted between the server and the client should be encrypted using secure protocols such as HTTPS.
- Data Backup and Recovery: the system should have a robust backup and recovery plan to ensure that data is not lost in the event of system failure or disaster.
- Secure API access: the API should be designed to provide secure access. This can be achieved by implementing secure authentication and authorization mechanisms such as OAuth2.0.

15.3 Usability requirements

The usability requirements of the application are essential to ensure that the users can easily interact with the OpenDBL web application. So the User Interface (UI) should be intuitive and easy to use, with clear navigation and informative feedback to the user. In particular:

- User-friendly interface: the interface should be designed to provide a user-friendly experience. The users should be able to easily understand the functionalities of the application and use it without difficulty.
- Navigation: navigation should be intuitive and easy to follow. The user should be able to move through the application with ease and without confusion.
- Feedback: the system should provide clear and informative feedback to the user. The user should be able to understand what is happening in the system.
- Multilingual support: the application should support multiple languages to make it accessible to users from different countries.
-

16 Software Tests Description (STD)

This paragraph briefly introduces the Software Testing Specifications (STD) related to the OpenDBL project. Its drafting aims to facilitate the software testing procedures.

16.1 Test Criteria

Acceptance tests are a type of testing that evaluates whether a software satisfies the acceptance criteria and requirements set by SRS. These tests are conducted to gain confidence that the software is fit for production use.

Generally, we can have acceptance criteria such as:

- Functionality: the software must be able to perform at least the tasks outlined in the requirement specifications or a subset of it.
- Performance: the system must have fast response times and be able to handle large data sets. The data size should not impact the performance of the system.
- Security: the software must have proper security measures in place to ensure data privacy and prevent unauthorized access.
- Usability: user interface must be easy to use and navigate, with a user-friendly interface that requires minimal training.

- Scalability: The system should be able to scale up or down based on the needs of the users.

Therefore, even if the unit tests will still be performed internally by the development team, in this way we will be certain that the OpenDBL system will behave, in real scenarios, as expected even in anomalous situations. The team is in the process of deciding whether to perform these acceptance tests internally or not.

17 Software Design Description (SDD)

The Software Design Description (SDD) explains how the OpenDBL software will be built to meet the project goals, objectives and user requirements stated in the SRS paragraph. Below we provide a description of the OpenDBL system in terms of its software architecture and a high-level overview of the various internal components of which it is composed along with their interactions and the design decisions made to achieve the desired functionality and performance. Having this information available, it will then be possible to proceed with the implementation of the SRS requirements in a computer programming language. So, this information serves as a reference mainly for architects, developers, testers, and stakeholders who are responsible for designing, developing, and testing the software project.

17.1 Software Development Lifecycle

To build a software project like OpenDBL it is recommended to follow a structured approach to software design, development, testing and deployment. Therefore it is necessary to adopt a process, made up of well-defined phases and activities: in software engineering there are various already codified processes [3] (such as the waterfall, spiral, iterative and incremental model, v-model, etc.) and choose the one or the other depends on project size, complexity, requirements volatility, objectives and constraints.

In the iterative and incremental development process (used for example in the process called “Agile Model”) the project is divided into small, manageable iterations [4] [5]. Each iteration represents a part of the overall software development process and includes planning, design, implementation, testing, and review. Incrementally, new features are added to the software with each iteration so it normally to assumes that development starts before all the requirements are defined in detail. Some of the benefits of this approach are that in each

iteration we deliver a piece of working software, so stakeholders can see tangible progress right from the start [6]. What functionality will be developed in each iteration is driven by priorities and so by the most critical part of the OpenDBL system. Additionally, we receive feedback and validation from users during each iteration, so changes can be made in subsequent iterations to ensure that the final product actually meets user needs. Finally, breaking the project into smaller parts can lead the software team to easily manage the large software itself, identify and manage risks, and discover and resolve defects more quickly.

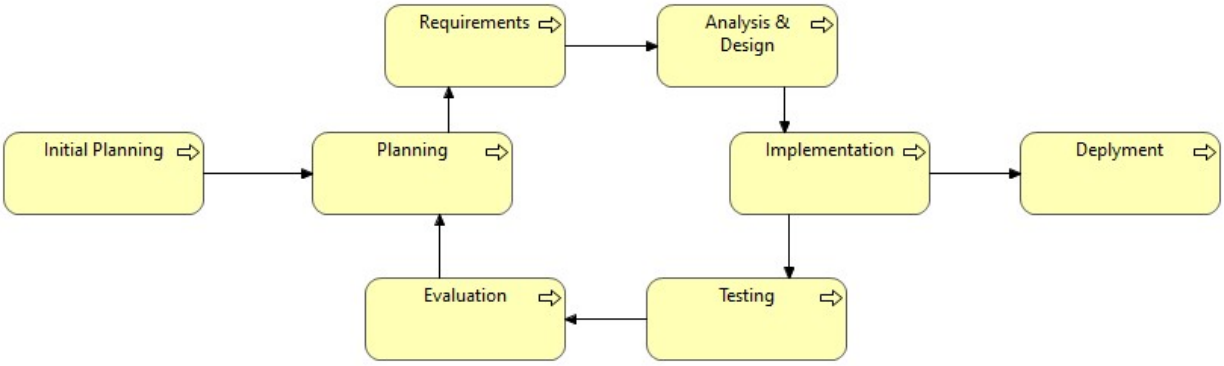


Figure 10 An Illustration of the software development lifecycle

For openDBL software project, this development process is particularly advantageous because stakeholders are involved from the beginning in providing useful feedback and above all because all interoperability challenges can be mitigated by simply moving forward with an integration gradually, adding new services, such as IoT solutions or data validation services.

18 Software Architecture

As shown in the previous paragraph, we have a phase called "Analysis and Design". This phase is particularly relevant in the software building process because it takes the requirements provided by the SRS (and by previous iterations) and produces an artifact that is the software architecture, which can then be subsequently implemented. In particular this phase emphasizes the finding of the main concepts in the problem domain of the system and defines how these objects collaborate to fulfil the requirements. As stated in the requirements, OpenDBL will be a system made of RESTful services and web applications, so is primarily a web application

which is a specialized type of client/server application that operates over the internet. In particular is a software system where one part (the client, through web browsers) interacts with another part (the server) to request and receive services or data. In this architecture, the server can concurrently run a number of services, such as other web applications, one or more database server, API services, and so on. At a higher level, a first deployment view of the system can be.

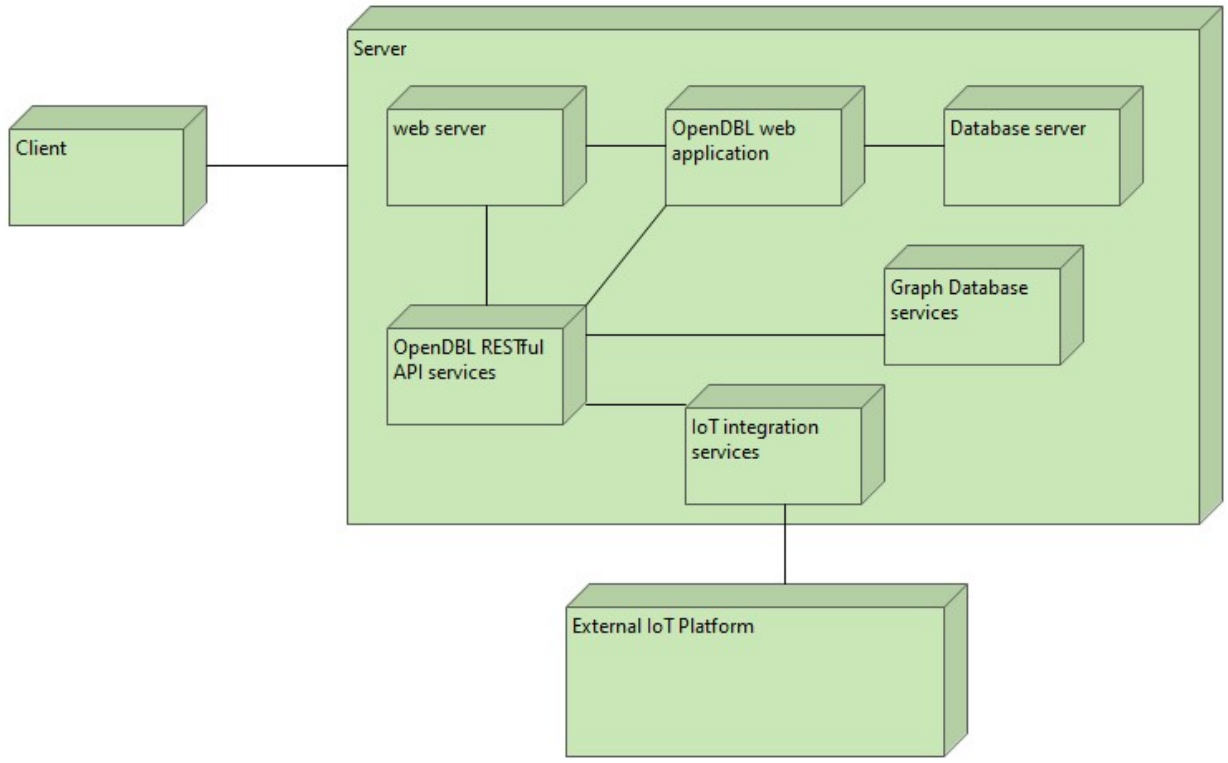


Figure 11 Structure of the interaction between Client, server and IoT services

Here client and server nodes are identified and in particular the server node shows the main pieces of the OpenDBL system as indicated by the SRS. With all this information available we can begin by describing an initial domain model of the system for the current iteration which is made up of the domain objects shown in the following conceptual diagram:

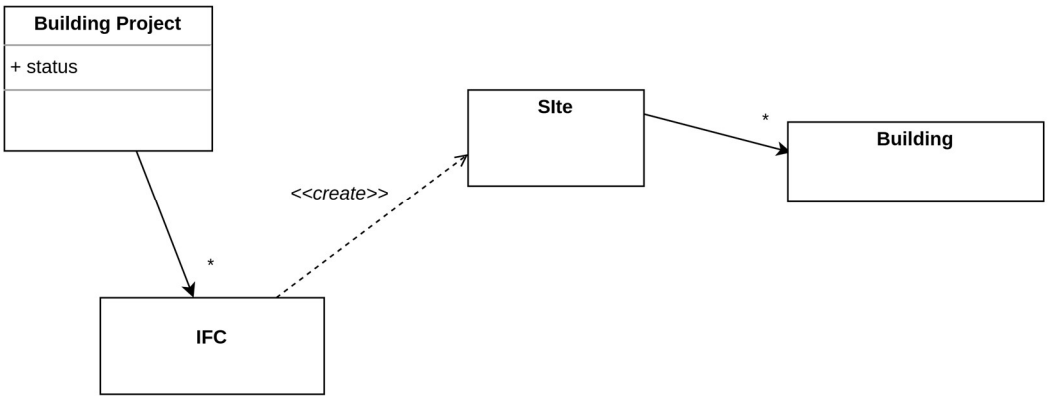


Figure 12 Conceptual diagram of the domain model

These are the main real-world concepts in the construction industry that make up the OpenDBL system. In fact, we have a construction project that can have one or more construction sites, each construction site is then composed of one or more buildings which in turn are characterized by the respective IFC files. From this high-level model, we can then go further, for example in reality each construction site can be made up of multiple sites linked in a hierarchical way, so by refining the previous conceptual model, it is possible to arrive at the following class diagram, which is more linked to non-domain software objects.

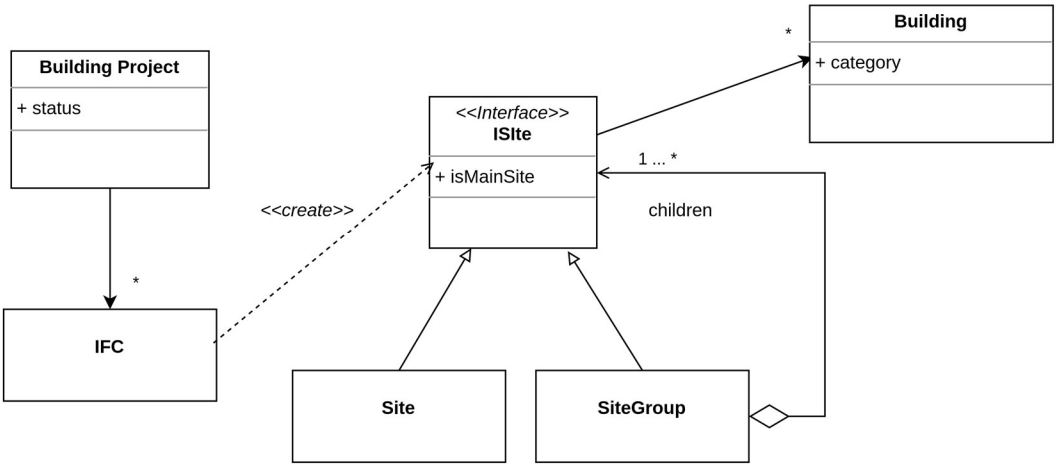


Figure 13 Domain model class diagram

Here in particular we have applied the composite pattern [7] to model the hierarchical structure of construction sites. From now on it is possible to implement this first model in a programming language and then goes on with the next iteration phase.

18.1 Architectural Design

"Architecture represents the significant design decisions that shape a system, where significance is measured by cost of change." -- Grady Booch

We can therefore begin to describe the software architecture of OpenDBL, that is, what are the internal components of this system and how they are organized. In fact, from the deployment diagram we can delve into the system and organize the logical structure of each component into levels that have a clear responsibility and a coherent separation. This means, at a lower level, breaking down the application in terms of classes, components, libraries, ensuring that design patterns are used in the right way and using frameworks where necessary. In practice, application architecture concerns the lower-level aspects of software design and code organization.

This in turn involves many design decisions. For example, we need to store data for each construction project, so we need some type of persistent storage and a common choice is to use a relational database. But to reduce the amount of work required when changing database vendors, it is best to use an Object Relational Mapping (ORM) framework (e.g. Hibernate, Entity Framework Core, SQLAlchemy, etc.). The introduction of this ORM module allows us to decouple database access from other application components and we can decide to safely change the database provider without too much trouble or with minimal effort [8].

This is a classic technique for decoupling distinct parts of a software system, promoting more flexible coupling between parts having greater cohesion.

The same applies to the authentication and authorization functionality of the application. Even in this case it is better to rely on a ready-made framework, library, or service instead of reinventing the wheel, because in the future it is possible that OpenDBL could rely on different identity providers. In any case, an additional reason is also given by the fact that security is not an easy topic to deal with.

Another example is related to the management of IFC files. Developing a library capable of reading the structure of IFC files and executing complex queries on even large files efficiently can be daunting. So, we chose to use a well-known library called IfcOpenShell written in C++ (and which has bindings in Python). So, in this case the decision made was to use this library to handle IFC files instead of creating our own.

But using IfcOpenShell in the context of the OpenDBL system means that you need an API to access and manipulate IFC files. So, we decided that it would be best to instead build an OpenAPI for this module (using FastAPI, a well-known Python web framework). A similar decision concerns the use of ready-made 3D viewers, VR, and point clouds. This therefore leads us to outline a first draft of the logical architecture of the OpenDBL system as depicted in the following package diagram:

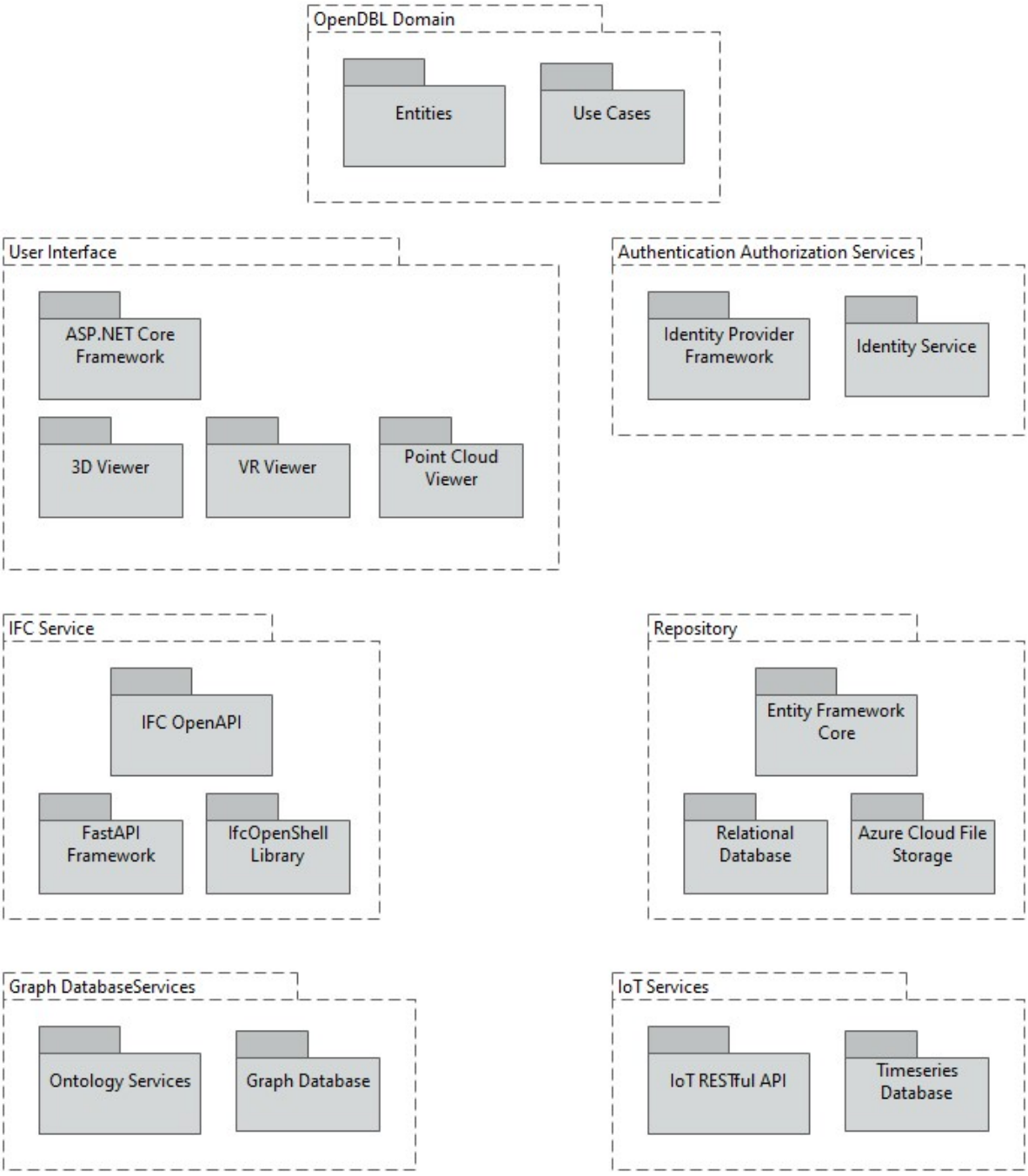


Figure 14 Conceptual diagram of the logical architecture

As mentioned, it shows us how the components are organized, but to describe how they work together we must introduce the concept of layering in software architecture, which is a design principle that involves the organization of a

software system into distinct and separate layers or levels, each with a specific set of responsibilities and functions [9]. Each layer interacts with other layers through well-defined interfaces and protocols. The most common layers are the "Presentation Layer" (responsible for managing the user interface and user interaction), "Application Layer" (also known as the business logic layer), "Data Access Layer" (which interacts with databases), "Infrastructure layer" (provides services such as logging, security, etc.). Here every component in one layer may depend on components of the layer below and this is not a good thing, because, as an example, the "Data Access Layer" which is at the bottom, finally drives the design of the system.

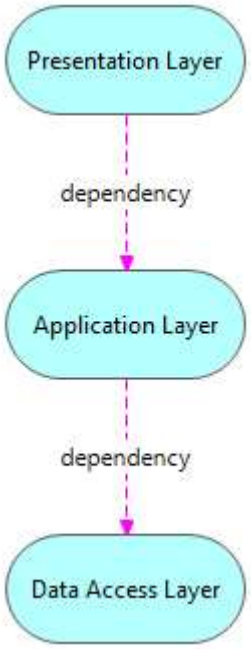


Figure 15 Dependencies through layers

So to make this architectural approach used to improve the maintainability, scalability and flexibility of software systems various architectural ideas (or patterns) have emerged and developed in the literature to push this ideas further with the such as Hexagonal [10](also known as Ports and Adapters), Onion, DCI, BCE and Clean Architecture [11] [12]. They are very similar in the objective, which is the separation of interests obtained, as we were saying, by dividing the software into tiers, where the core application logic remains isolated and interacts with

external dependencies only through adapters (which convert the data from one layer to the other) thanks to dependency inversion pattern, where the outer tiers shall only be dependent on the inner ones through the use of abstract interchangeable interfaces.

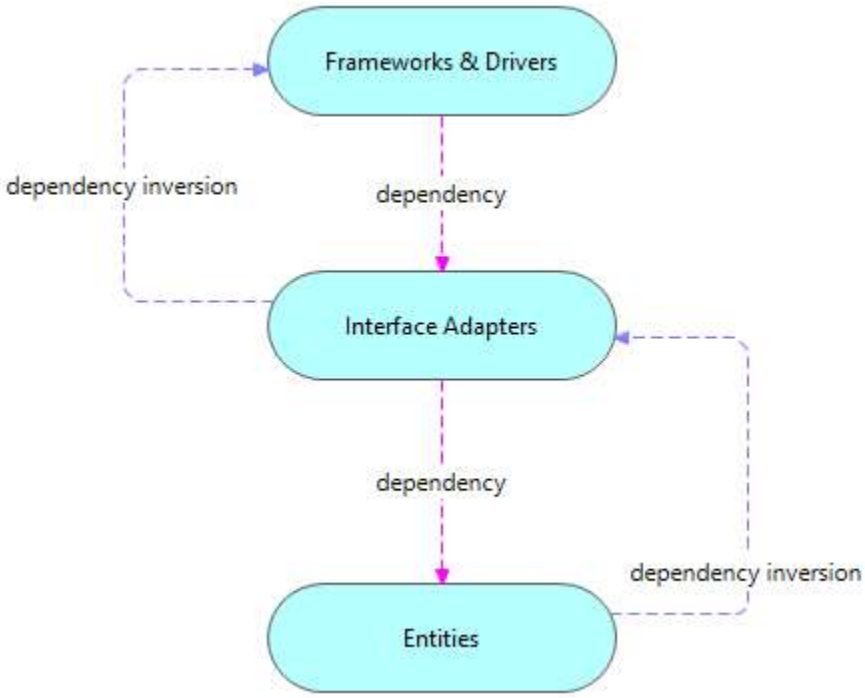


Figure 16 Dependency inversion

That leads us, from the previous package diagram to the following layered architecture for OpenDBL system.

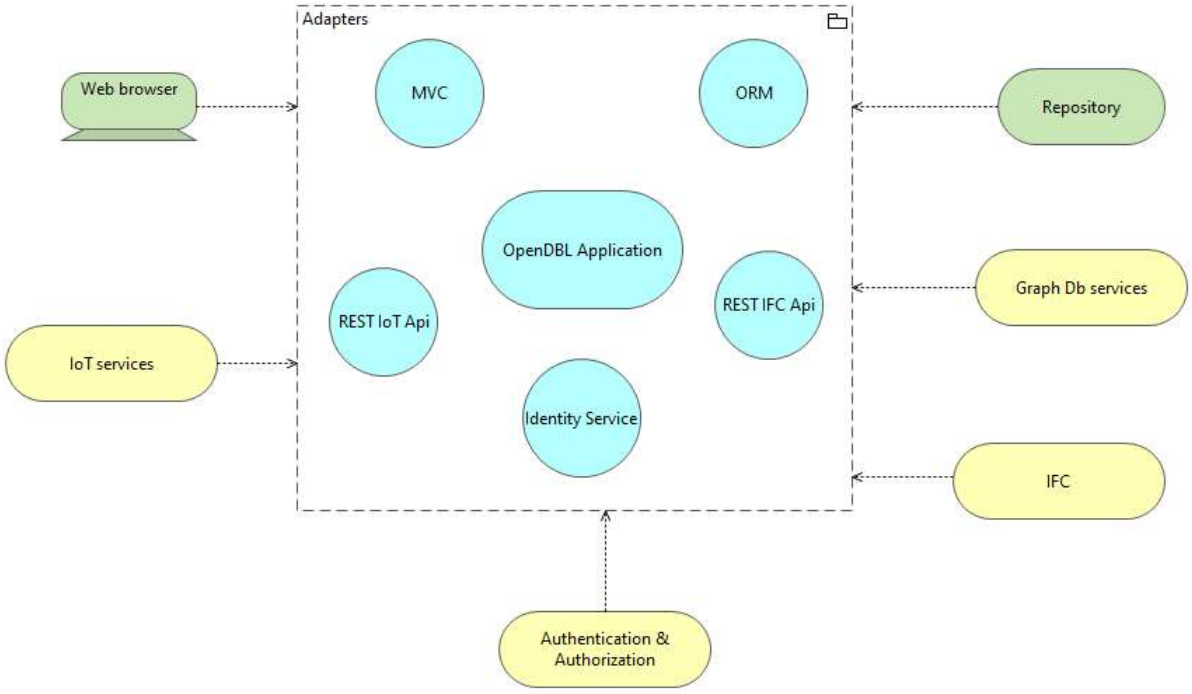


Figure 17 openDBL layered architecture

A final aspect of the architecture of OpenDBL is the way the entire application can be built. We can in fact follow a monolithic, microservices or hybrid approach. In a monolithic architecture, the entire application is built as a single, tightly integrated unit and all components (including the user interface, business logic, and data access) are combined into a single project and deployed as a single application [13]. For the microservices architecture we have to breaks the application into small, independent services that communicate over a network (usually via gRPC) and each is responsible for a specific, isolated function or requirement. Every service is then developed, deployed, and scaled independently. The drawback of this approach is that managing many services can be a complex task in terms of development and in fact is suitable for large scale applications.

So, for OpenDBL system a hybrid approach can be desirable because we can start with a monolithic application and gradually transition specific parts of the application that benefit from microservices, parts which are really independent

from the core or use primarily external services: one example over all is the “IFC service,” responsible to manage IFC files.

19 Technological Infrastructure

19.1 Physical Architecture

The physical architecture of the system refers to the deployment of the system components in a physical environment. It specifies the hardware and network infrastructure needed to support the system’s operation. The proposed system will be deployed on cloud infrastructure, which offers a scalable and cost-effective solution for hosting web services. The cloud infrastructure will be provided by a third-party cloud service provider. The physical architecture of the system will be composed of the following components:

- Load Balancer: A load balancer will be used to distribute incoming traffic across multiple instances of the application to improve availability and scalability.
- Web Role: The web Role will serve as the entry point for incoming requests and will host the web application.
- Application Server layer: The application server layer will handle the core business logic of the system, including importing IFC models, converting IFC models to the OpenDBL ontology, and updating the ontology.
- Database Role: The database role will store the OpenDBL ontology and other data required by the system.
- Cache Role: The cache role will be used to store frequently accessed data and reduce the load on the database server.
- Message Broker role: The message broker role will be used to facilitate communication between different services in the system.
- API Gateway: The API gateway will serve as the entry point for external clients to access the system via RESTful API calls.
- WAF: The web application Firewall allows to increase the safety of web applications and/or APIs, as it implements OWASP rules to protect from attack patterns that take advantage of the vulnerabilities in code, programming languages, and/or base software (such as operative systems, app-used application servers, or API-used application servers)
- Operation and Monitoring role: will provide real-time monitoring and alerts to ensure that the system is functioning correctly and meet the performance requirements.

- Logging Role: The logging role allows to acquire logs from the various systems and administrative accesses as regulated by the GDPR.

The physical architecture of the system will be designed to ensure high availability, scalability, and fault-tolerance. The cloud infrastructure will be designed to ensure that the system can handle a large volume of incoming traffic and that the system components can be scaled up or down based on demand.

In conclusion, the physical architecture of the system will play a critical role in ensuring the system's performance, reliability, and scalability. The proposed architecture will use cloud infrastructure to offer a cost-effective and scalable solution for hosting the system.

19.2 Deployment Architecture

The deployment architecture describes the physical environment in which the system will be deployed, including hardware and network infrastructure.

The system will be deployed on a cloud-based platform such as Amazon Web Services (AWS), Microsoft Azure or Google Cloud Platform or others. The deployment architecture will be based on a microservices redundant architecture, with each service deployed independently in its own container, managed by a container orchestration system.

The system will use a load balancer to distribute incoming traffic across multiple instances of the service to ensure high availability and scalability.

The data will be stored in both non-relational and relational databases, depending on their nature, performance, and scalability convenience.

The deployment architecture will be designed to comply with industry best practices and standards. This will include implementing encryption at rest (for eventual sensible data) and in transit, using secure communication protocols such as HTTPS, and implementing access control mechanisms such as OAuth2 or JWT.

Monitoring and logging will be implemented at various levels of the system to enable effective performance monitoring and troubleshooting. A centralized logging system will be used to collect and analyse logs from numerous services, so allowing the recording of administrative accesses as required by law.

Where possible, the Infrastructure as a Code (IaC) approach will be used to version the entire infrastructure and make it reproducible when needed.

T1.5 - AI-Based mapping concept

1 The Utilization of Artificial Intelligence in the BIM Process

In the rapidly evolving domain of construction science and infrastructure development, the integration of Artificial Intelligence (AI) into Building Information Modelling (BIM) processes emerges as a pivotal advancement, holding the potential to redefine the foundational paradigms of construction methodologies. This symbiotic amalgamation of AI and BIM aims not only to streamline the existing processes but also to foster a revolutionary synergy between technological advancements and human expertise, potentially ushering in a new era of efficiency, precision, and innovation. In the following sections, we explore in detail the multifaceted benefits that AI integration brings to BIM processes, highlighting its role in nurturing a progressive, sustainable, and technologically advanced construction ecosystem.

1.1 Efficiency Augmentation

The infusion of AI within BIM processes heralds a significant augmentation in operational efficiency, predominantly by introducing automated solutions for tasks traditionally characterized by labour-intensive efforts and a high propensity for human error. By automating model creation and facilitating analytical scrutiny of design alternatives, AI propels the optimization of construction sequences. This transformative approach not only accelerates project timelines but also minimizes manual errors, thereby ensuring a seamless and efficient workflow that encompasses various phases of construction, fostering an environment where resources are optimally utilized, and projects are executed with heightened precision and productivity [14].

1.2 Predictive Analysis

In the contemporary construction landscape, predictive analysis facilitated by AI stands as a cornerstone in fostering data-driven decision-making processes. AI systems are capable of aggregating and analysing data from a plethora of sources, identifying underlying patterns and trends that can be instrumental in crafting accurate forecasts vital for project planning and execution. Through advanced algorithms and machine learning techniques, AI navigates through vast datasets, offering insights that enable pre-emptive identification of potential bottlenecks, ensuring resource allocation is both strategic and informed, hence fostering a proactive approach to project management, characterized by foresight and strategic planning [15].

1.3 Cost Optimization

AI serves as a catalyst in achieving substantial cost optimizations within the BIM processes, particularly through the precise estimation of material quantities and the formulation of cost-effective workflows. It transcends traditional methods by facilitating data-driven strategies that enable accurate cost estimations and uncover potential avenues for financial savings. This approach ensures that projects maintain economic viability without compromising on quality, fostering a balanced paradigm where cost-effectiveness is harmonized with the pursuit of excellence in construction quality [16].

1.4 Quality Control

AI's role in quality control is monumental, offering automated tools capable of rigorous construction quality monitoring and swift identification of deviations from the planned model. Utilizing advancements in image recognition and data analytics, AI automates the process of quality assurance, ensuring stringent adherence to predefined standards and minimizing the likelihood of errors and defects. This proactive approach to quality control enables a dynamic response to potential issues, ensuring that quality is not compromised at any stage of the construction process [14].

1.5 Safety Management

Safety management, a critical facet in the construction sector, is significantly enhanced by AI, which aids in the early identification of safety risks and monitors compliance with safety protocols. Leveraging technologies such as machine vision, AI automates the process of monitoring construction sites, enhancing safety through the timely identification and mitigation of potential hazards. This results in a safer work environment, characterized by stringent safety protocols and a proactive approach to risk management, thereby fostering a culture of safety and compliance within the industry [16].

1.6 Sustainability Assessment

In the era of increasing emphasis on sustainable practices, AI emerges as a critical tool in evaluating the sustainability of construction projects. Through complex analyses of various facets including energy efficiency and material consumption, AI facilitates the development of strategies that promote environmental stewardship and sustainable development. This analytical prowess enables the crafting of construction strategies that are aligned with global sustainability goals,

fostering a culture of responsibility and awareness towards environmental conservation [17].

1.7 Facility Management

In the sphere of facility management, AI contributes substantially by optimizing building operation and maintenance through the development of predictive maintenance strategies. It aids in real-time monitoring of various building systems, enabling timely identification of maintenance needs and thereby ensuring operational efficiency and building longevity. This approach revolutionizes facility management by fostering a proactive approach to maintenance, characterized by data-driven decision-making and optimal resource allocation [18] [19].

1.8 Enhanced Decision-Making

AI's role in enhancing decision-making processes is unparalleled, given its ability to analyse extensive datasets and empower stakeholders to make informed and strategic choices. This data-driven approach fosters the development of strategies that are both efficient and effective, enabling project stakeholders to navigate complex decision-making landscapes with confidence and precision. Moreover, AI facilitates a culture of continuous learning and adaptation, where decisions are constantly refined based on incoming data, ensuring a dynamic and responsive approach to project management [15].

1.9 Integration of IoT

The AI-facilitated integration of Internet of Things (IoT) into BIM systems stands as a revolutionary step in building management, allowing for the collection and analysis of real-time data from infrastructure. This integration nurtures a dynamic building management system, where data from various sensors and devices are synergistically analysed to foster informed decision-making and optimal building performance. Moreover, it paves the way for the development of smart buildings, characterized by automation, energy efficiency, and enhanced occupant comfort [20].

1.10 Fostering Innovation

Lastly, the infusion of AI into BIM serves as a catalyst for innovation, opening new avenues for addressing complex challenges prevalent in the construction sector. By fostering an environment of research and development, AI encourages the exploration of novel solutions, thereby nurturing a culture of innovation and continuous improvement within the construction industry. This innovative

approach promotes the development of groundbreaking solutions, characterized by technological prowess and a forward-thinking approach to construction management [20]

2 The Use-Case in the openDBL Project

In the contemporary landscape of technological advancements, our research revolves around the identification of semantic connections between various attributes, a task earmarked for resolution through AI. This endeavour is rooted in the AI's capability to discern conceptual similarities with a degree of precision and speed that surpasses human capabilities, thereby fostering a more efficient and accurate process [21].

As we venture into the final stages of verification, our strategy pivots towards the implementation of a "Human-in-the-Loop" (HITL) system. This approach embodies a semi-automated strategy, fostering a synergistic collaboration between human experts and automated processes. The integration of human expertise not only augments the quality and accuracy of AI outputs but also serves as a vigilant overseer, scrutinizing the AI system's operations to identify and rectify potential errors and inaccuracies that might be induced by the AI systems [22].

This collaborative approach culminates in an overall enhancement in the quality of the outcomes, a testament to the harmonious integration of human intelligence and artificial prowess. The human expert operates as a sentinel, meticulously overseeing and verifying the AI system's operations, thereby ensuring a meticulous correction of errors and inaccuracies that might be propagated by the AI systems. This vigilant oversight ensures a significant elevation in the quality of the results, fostering a dynamic where the AI system's outputs are continually refined and honed to perfection [22].

Furthermore, this strategy engenders an enhanced level of trust and reliability from the user's perspective. The meticulous oversight by human experts not only ensures the reliability of the AI outputs but also fosters a heightened level of trust from the users, who can be assured of the system's accuracy and reliability. This collaborative approach, therefore, not only enhances the quality of the AI outputs but also fosters a heightened level of user trust in the AI system, paving the way for a more harmonious and effective integration of AI technologies in various domains [22].

3 Ontology-Based Data Model and its Advantages for AI Applications

3.1 Structured Knowledge Representation

Ontologies serve as a robust framework for the structured representation of knowledge, delineating concepts and the intricate relationships that exist between them [21] [23]. This structured approach facilitates AI systems in comprehending and processing complex information with heightened efficiency. By encapsulating knowledge in a well-defined structure, ontologies enable AI systems to navigate through a rich tapestry of interconnected concepts, fostering a deeper understanding and facilitating the extraction of meaningful insights from vast datasets [24] [25] [26].

3.2 Semantic Interoperability

The utilization of ontologies significantly enhances semantic interoperability, allowing AI systems to integrate and harmonize information from diverse sources seamlessly [27]. This is achieved by fostering a unified understanding of data semantics, thereby bridging the gap between disparate data formats and terminologies. Through semantic interoperability, AI systems can synthesize information from various domains, fostering a cohesive and comprehensive analytical landscape that facilitates informed decision-making [28].

3.3 Facilitated Data Integration

Ontologies act as a potent tool for data integration, aiding in the reconciliation of terminology and data structure discrepancies that often exist between different data sources [29]. This facilitation ensures a smoother data integration process, where inconsistencies are effectively addressed, paving the way for a unified data repository that serves as a rich source of information for AI applications, enhancing their analytical capabilities and output accuracy [26] [27].

3.4 Enhanced Search and Query Functions

Ontologies significantly enhance the efficiency of search and query functions by enabling semantic searches that transcend the limitations of simple keyword searches [30]. This enhancement is achieved through the incorporation of semantic relationships and hierarchies within the data model, allowing for more nuanced and context-aware search capabilities. Consequently, users can expect more precise and relevant results, fostering a more intuitive and user-friendly search experience.

3.5 Support for Automatic Inference

Ontologies provide robust support for automatic inference, empowering AI systems to derive new insights from existing data [24]. This is facilitated through logical reasoning mechanisms embedded within the ontology structure, which allow AI systems to infer new knowledge based on the relationships and properties defined within the ontology. This capability fosters a dynamic knowledge expansion, where AI systems can continually evolve and adapt to new information, enhancing their analytical depth and predictive accuracy.

3.6 Better Understanding of Context and Nuances

By modelling relationships and properties, ontologies enable AI systems to develop a deeper understanding of the context and nuances associated with data [25]. This deeper understanding facilitates more precise and nuanced analyses, allowing AI systems to discern subtle patterns and trends that might otherwise go unnoticed. Consequently, AI systems can provide more insightful and contextually rich analyses, enhancing the value and relevance of the insights generated.

3.7 Facilitated Communication Between Humans and Machines

Ontologies serve as a common language that facilitates communication between humans and machines, offering a clearly defined vocabulary and structure for information exchange [28]. This facilitation enhances the synergy between human expertise and machine intelligence, fostering a collaborative environment where complex tasks can be addressed more effectively. Moreover, it enhances the user experience by enabling more intuitive interactions with AI systems, fostering a harmonious and productive human-machine collaboration.

3.8 Modularity and Reusability

Ontology-based models often exhibit a modular structure, promoting their reusability across various applications and domains [26]. This modularity allows for the flexible adaptation of ontology structures to suit different contexts, fostering a more efficient development process. Moreover, it encourages the sharing and reuse of knowledge structures, enhancing the scalability and versatility of AI applications.

3.9 Promotion of Machine Learning

Ontologies can serve as a foundation for machine learning, providing structured, annotated data that can be utilized for model training [22]. This provision facilitates the development of more sophisticated machine learning models, as it allows for

the incorporation of rich semantic information into the training process. Consequently, machine learning models can achieve higher levels of accuracy and predictive power, fostering advancements in AI capabilities and applications.

4 Advantages of Graph Databases with Respect to AI

In the dynamic and rapidly evolving field of Artificial Intelligence, the integration and utilization of graph databases have emerged as a cornerstone in enhancing both the depth and breadth of data analysis. This section meticulously delineates the myriad advantages that graph databases confer in the realm of AI, emphasizing their pivotal role in fostering a more nuanced and comprehensive approach to data interpretation and analysis.

4.1 Complex Relationship Analysis

Standing out in their ability to adeptly model and scrutinize complex relationships between a wide array of entities, graph databases facilitate AI systems in delving deep into networks and relationship structures. This intrinsic capability not only unveils intricate insights that might otherwise remain obscured but also fosters a more nuanced approach to data interpretation. The depth of analysis is instrumental in facilitating a deeper understanding of the intricate web of relationships that define complex data structures, thereby paving the way for more informed and insightful data analysis processes [31].

4.2 Efficient Pattern Recognition

Graph databases are characterized by their proficiency in swiftly and efficiently identifying patterns and anomalies within data sets. This attribute is particularly beneficial in the domains of machine learning and data mining, where rapid pattern recognition can significantly enhance both the accuracy and speed of data analysis processes. The ability to quickly identify patterns and anomalies is a cornerstone in the development of machine learning algorithms, facilitating more accurate predictions and insights, thereby enhancing the overall efficacy of AI systems [32].

4.3 Flexibility and Scalability

Recognized for their inherent flexibility and scalability, graph databases are indispensable tools in handling large volumes of complex, interconnected data. This flexibility is instrumental in facilitating the seamless integration and analysis of data, catalysing advancements in various AI applications. Moreover, their scalability ensures adaptability to the growing demands of AI systems, providing a

robust and flexible platform that can accommodate the increasing complexity and volume of data encountered in modern AI applications [33].

4.4 Semantic Analysis

By adopting a graph-based approach to data modelling, AI systems are empowered to conduct semantic analysis, which yields a profound understanding of the underlying meaning and context of data. This analytical approach is central to discerning conceptual similarities and fostering a more nuanced understanding of data structures. Furthermore, semantic analysis facilitated by graph databases enables AI systems to delve deeper into the intricacies of data, offering a richer and more comprehensive understanding of the complex relationships and patterns that define data structures [34].

4.5 Recommender Systems

Graph databases serve as a potent tool in the development of recommender systems. These databases enable the efficient analysis of relationships and commonalities between entities, thereby enhancing the accuracy and relevance of recommendations generated by AI systems. The utilization of graph databases in recommender systems facilitates a more personalized and targeted approach to recommendation generation, leveraging the power of graph databases to analyse complex relationships and identify patterns that can help in tailoring recommendations to individual preferences and behaviours [35].

4.6 Natural Language Processing (NLP)

In the sphere of Natural Language Processing, graph databases play a crucial role in modelling words and their interrelationships. This modelling facilitates the creation of advanced language processing systems capable of understanding and interpreting complex linguistic structures. Moreover, the integration of graph databases in NLP enhances the ability of AI systems to analyse and interpret language on a deeper level, fostering the development of more sophisticated and nuanced language processing tools that can understand and interpret the subtleties of human language with greater accuracy and depth [36].

4.7 Knowledge Graphs

Graph databases are quintessential in constructing knowledge graphs, which offer a structured representation of knowledge. These graphs form the bedrock of numerous AI applications, serving as a repository of interconnected information that can be analysed to glean valuable insights. Furthermore, knowledge graphs

created using graph databases provide a rich and structured platform for the organization and analysis of knowledge, facilitating a more comprehensive and nuanced approach to knowledge management and dissemination in AI systems.

4.8 Real-time Analysis

Graph databases are equipped to undertake real-time analysis, capable of executing complex queries and analyses within a short timeframe. This real-time capability is a significant boon for numerous AI applications, where timely data analysis is of the essence. Moreover, the ability of graph databases to perform real-time analysis ensures that AI systems can respond swiftly to changing data landscapes, providing the agility and responsiveness necessary to adapt to the dynamic nature of modern data environments.

4.9 Transitive Relationships and Inference

Graph databases facilitate the modelling of transitive relationships and are adept at performing inference operations. This functionality enables AI systems to derive new insights from existing data, fostering a deeper understanding, and facilitating the discovery of novel information. Additionally, the ability to model transitive relationships and perform inference operations enhances the analytical capabilities of AI systems, enabling them to uncover new patterns and insights that can help in driving innovation and advancing the field of AI.

4.10 Improved Data Quality and Integrity

Utilizing graph databases enhances the quality and integrity of data managed by AI systems. These databases allow for the explicit modelling of relationships and dependencies between data points, thereby ensuring a higher degree of data accuracy and reliability. Furthermore, the use of graph databases in managing data quality and integrity ensures that AI systems can maintain a high level of data accuracy and reliability, fostering trust and confidence in the insights and analyses generated by AI systems.

5 Case Studies: Real-World Applications of AI Integration

In this section, we will explore a series of case studies that demonstrate the real-world applications of AI integration in various domains, highlighting the transformative impact of AI technologies and their potential to foster innovation, efficiency, and precision in diverse fields.

5.1 Construction Industry: Enhancing Project Management and Safety

In recent years, the construction industry has witnessed a surge in the adoption of AI technologies, with numerous projects leveraging AI to enhance project management and safety protocols. AI-powered predictive analytics have been instrumental in identifying potential risks and optimizing resource allocation, thereby minimizing delays, and ensuring the timely completion of projects. Furthermore, AI has played a pivotal role in enhancing safety management by facilitating real-time monitoring of construction sites to identify and mitigate potential hazards [37].

5.2 Healthcare: Revolutionizing Diagnosis and Treatment

The healthcare sector has been a fertile ground for the integration of AI technologies, with numerous applications in the fields of diagnosis, treatment planning, and patient management. AI-powered diagnostic tools have demonstrated remarkable accuracy in detecting various medical conditions, thereby facilitating early intervention, and improving patient outcomes. Moreover, AI has been utilized to develop personalized treatment plans, taking into consideration a myriad of factors to optimize treatment efficacy and minimize adverse effects [23].

5.3 Finance: Streamlining Operations and Enhancing Decision-Making

In the finance sector, AI has been leveraged to streamline operations and enhance decision-making processes. AI-powered algorithms have been employed to analyse vast datasets to identify market trends and investment opportunities, thereby facilitating informed decision-making and optimizing investment strategies. Furthermore, AI has been utilized to automate various financial operations, such as fraud detection and credit scoring, thereby enhancing efficiency and reducing operational costs [26].

5.4 Retail: Optimizing Supply Chain and Enhancing Customer Experience

The retail sector has witnessed a transformative impact of AI technologies, with numerous applications in supply chain optimization and customer experience

enhancement. AI-powered predictive analytics have been utilized to optimize inventory management, thereby reducing carrying costs and minimizing stockouts. Moreover, AI has been leveraged to develop personalized marketing strategies, enhancing customer engagement, and fostering brand loyalty [27].

5.5 Smart Cities: Facilitating Sustainable Urban Development

In the context of urban development, AI has emerged as a powerful tool to facilitate the development of smart cities, characterized by sustainable and efficient urban management practices. AI-powered solutions have been employed to optimize traffic management, reduce energy consumption, and enhance public safety, thereby fostering a sustainable and liveable urban environment [38] [39].

In the realm of sustainable building development, AI has transcended traditional applications, playing a critical role in the nascent stages of design. Specifically, automated workflows leveraging Building Information Modelling have been pivotal in computing the embodied greenhouse gas emissions in structures, allowing for the meticulous analysis of diverse design alternatives to pinpoint those with a minimized environmental footprint [40]. As delineated in your research, these methodologies employ sophisticated NLP techniques to facilitate automated semantic healing of BIM models. This, in turn, enables a robust and holistic Life Cycle Analysis (LCA), even in preliminary design phases. These avant-garde strategies signify a monumental stride towards sustainable urban development, enhancing the precision and efficiency inherent in the building design process, thereby emerging as a cornerstone in the blueprint of smart cities [40].

6 Comprehensive Concept of Our Solution

In light of the previously described aspects, the following comprehensive concept emerges (bigger version in the Annex of the deliverable):

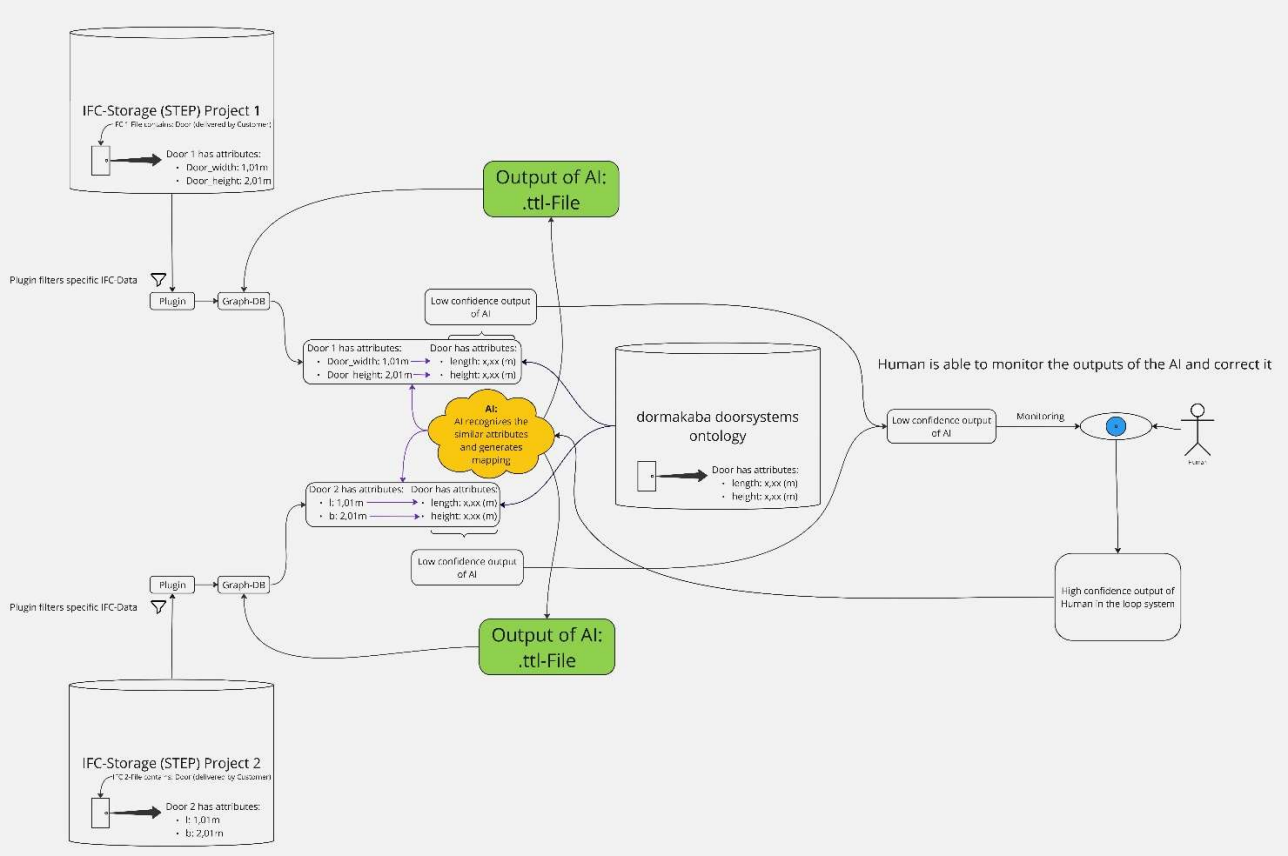


Figure 18 The proposed concept Tailored to a Specific Use Case Scenario

Within In the rapidly advancing domain of BIM, there has been an increasingly evident imperative to amalgamate innovative technological frameworks during the planning phase of construction projects. This need is not merely a function of convenience but is critical to the integration of multi-disciplinary aspects—ranging from architectural design to energy management—that together contribute to the holistic optimization of a building's entire lifecycle. The essence of the current research endeavor lies in meticulously crafting a methodological framework that not only aligns with established BIM protocols but also enhances them through the integration of Artificial Intelligence and ontology-based techniques.

The first cornerstone of our framework involves the utilization of Industry Foundation Classes storages, which are integral to separate but comparable

construction projects—referred to here as Project 1 and Project 2. These storages house IFC files in the Standard for the Exchange of Product Data (STEP) format, which is an ISO standard exchange format. Within these files, various building elements are cataloged, including, but not limited to, doors. In Project 1's IFC 1-File, doors come with a specific set of attributes; notably, the width of the door is cataloged as Door_width: 1.01m, and its height is recorded as Door_height: 2.01m. Analogously, in Project 2's IFC 2-File, doors are described with attributes coded as l: 1.01m for width and b: 2.01m for height.

To address the challenge of attribute extraction, we have developed a specialized Python plugin. This plugin has a dual functionality: first, it sifts through the IFC files to filter out door-related attributes; second, it converts these attributes into the Turtle (.ttl) format. The conversion into Turtle, a syntax for expressing data in the Resource Description Framework (RDF), is a critical step. This not only standardizes the data but also makes it compatible for integration into a graph database, such as Ontotext, thereby facilitating subsequent data manipulation and analysis.

Parallel to these IFC storages is the dormakaba doorsystems ontology—a structured framework that serves as a template for door attributes. This ontology employs a standardized nomenclature and includes attribute definitions such as length: x,xx (m) and height: x,xx (m). It serves as a benchmark against which the extracted attributes are compared.

The crux of our comprehensive concept lies in the AI-driven attribute mapping. For this, we employ machine learning algorithms specifically designed to identify semantic similarities between disparate sets of attributes. The algorithm traverses through the extracted attributes from both Project 1 and Project 2 and correlates them with the standard definitions as per the dormakaba doorsystems ontology. This algorithmic mapping, while highly efficient, is initially characterized as a low-confidence output. This is because machine learning models, especially when dealing with semantic mappings, are prone to errors and require validation.

To mitigate the risk of errors, a Human-in-the-Loop system is meticulously integrated into the framework. This system is designed to provide an additional layer of validation to the low-confidence outputs produced by the AI. Human experts meticulously review the AI-generated mappings, correct any discrepancies, and validate the attributes. These validated, high-confidence outputs are then used as training data to further refine the machine learning model, thereby incrementally

increasing the system's overall mapping accuracy and reducing the need for human intervention over time. Other aspects of HITL are also covered in Section 8.4 “Human-in-the-Loop Systems for File Alignment in openDBL”

In summation, our comprehensive concept combines the robustness of BIM with the precision of AI and ontology-based techniques. It offers an intricate yet streamlined methodological approach that significantly enhances the planning phase in construction projects. Through the synergistic interaction of automated algorithms and human expertise, the system aims for a continual improvement in accuracy and efficiency, thus rendering it an exceptionally robust, precise, and scalable solution for future applications in BIM and construction planning.

7 Description of Data Sources

In the intricate framework of our project, we are leveraging a plethora of diverse data sources to fuel the AI algorithms, thereby facilitating a seamless and comprehensive Building Information Modelling process. The Industry Foundation Classes (IFC) format emerges as a central pillar in this structure, serving as a master reference that orchestrates the integration and synchronization of various data inputs in a harmonized manner. In this complex scenario, it becomes vitally important to establish specific classifications and frameworks that can be intricately linked and integrated, depending on the particular use case, to foster a coherent and efficient data structure.

To achieve this, a deep dive into the different data formats available is essential, along with strategies to harmonize them to create a unified data repository. This repository would be designed to be easily accessed and analysed by AI algorithms, thereby enhancing the efficiency and effectiveness of the BIM process. Moreover, this section aims to elucidate the methodologies employed in data extraction and integration, providing a detailed overview of the various data sources and the techniques utilized to harmonize them, fostering a cohesive and streamlined data management process.

7.1 Bridging IFC Files, GraphDB, and Advanced Analytics

Our project adopts an AI-Driven Attribute-Based Mapping approach to establish a seamless flow of information from IFC files to advanced analytics platforms like GraphDB. This section aims to elucidate the intricate relationship between these components, emphasizing the path the data takes from IFC files to becoming actionable insights through AI algorithms.

7.1.1 The Data Journey: IFC to TXT to Turtle to GraphDB

- **IFC to TXT**: Initially, the IFC file is imported into the specialized tool "KITModelViewer," where our Python plugin extracts relevant attributes, such as door dimensions, from property sets associated with specific components. This data is saved in a text file, greatly reducing the size and complexity of the initial data set.
- **TXT to Turtle (TTL)**: A separate Python script, "txt_to_ttl_conversion_combined.py," converts the text file into Turtle (.ttl) format. This not only standardizes the data but also renders it compatible for integration into GraphDB. This modular approach allows for flexibility,

enabling the data extraction and conversion phases to be tailored by different tools or viewers.

- **Turtle to GraphDB**: The Turtle file is then imported into Ontotext GraphDB, a specialized graph-based database. This sets the stage for advanced analytics, as the data is now stored in a format that can be effectively interpreted and manipulated.

7.1.2 The Role of AI in Attribute-Based Mapping and Analytics

Once the data is in GraphDB, it is ripe for analysis through our AI algorithms. Here's how AI comes into play:

- **Semantic Mapping**: Our machine learning algorithm sifts through the GraphDB to identify semantic similarities between the extracted attributes from the IFC files and the standard definitions in the Dormakaba doorsystems ontology. This forms the basis for attribute-based mapping.
- **Low-Confidence to High-Confidence Mapping**: Initially, the algorithmic mapping is considered low-confidence. Our Human-in-the-Loop system reviews these low-confidence outputs, validates them, and feeds them back into the machine learning model, enhancing its future accuracy.
- **Dynamic Learning and Adaptation**: As the validated, high-confidence outputs accumulate, the machine learning model continually refines its mapping algorithm. This dynamic learning approach results in a system that becomes increasingly accurate, thus reducing the need for human intervention over time.
- **Actionable Insights**: Finally, AI algorithms generate actionable insights based on the attribute-based mapping. Whether it's optimizing door dimensions for energy efficiency or identifying inconsistencies in building designs, the AI engine makes real-time decisions that can be immediately implemented into the ongoing BIM process.

In summary, our AI-Driven Attribute-Based Mapping approach serves as the linchpin in the seamless integration of IFC files, GraphDB, and AI algorithms, enabling a new level of sophistication in real-time decision-making and analytics in construction projects.

While the aforementioned AI-Driven Attribute-Based Mapping process provides a general overview of how IFC files, GraphDB, and AI algorithms interact, it is imperative to delve deeper into the intricacies of the IFC data format itself. The IFC

format serves as the foundational layer in our data pipeline and a comprehensive understanding of its structure, capabilities, and constraints is crucial for effective data mapping and analytics.

In the following section (7.2: IFC), we will explore the detailed aspects of the IFC data format. This will include a closer look at the data structures, extraction techniques, and challenges associated with converting IFC instance data to Turtle format. We will also discuss the specific Python plugins and scripts used in handling IFC files, thereby giving a nuanced understanding of this integral data source in our project's ecosystem.

7.2 IFC

The Industry Foundation Classes format stands as a cornerstone in our project, facilitating the extraction and integration of essential data into the BIM process with heightened efficiency and accuracy. Utilizing Python scripts and specifically designed plugins, we can adeptly extract data from the IFC model, converting it into a more manageable and analysable format (.ttl / Turtle) for further analysis and integration. This process involves a meticulous analysis of the data structures present in the IFC model, identifying key data points and parameters that can be utilized to enhance the BIM process significantly.

Furthermore, the conversion of data into .ttl format ensures a streamlined data management process, allowing for easier integration with other data sources and facilitating a more cohesive data analysis process. This section delves deeper into the intricacies of the IFC format, exploring the various data extraction techniques and the methodologies employed in data conversion, thereby providing a comprehensive overview of the pivotal role that the IFC format plays in enhancing the efficiency and effectiveness of the BIM process.

In the current research and development landscape of computer science and data processing, the topic of data conversion and manipulation is becoming increasingly present. A special focus here is on the effective conversion of data structures into different formats in order to promote interoperability and data exchange between different systems. In this context, a specific challenge emerges in the context of converting IFC instance data into .ttl format.

IFC instance data is an essential component in the modelling of construction information, providing a detailed and structured representation of construction data models. These data structures are usually very complex and contain a large amount of information covering the different aspects of a structure. On the other hand, there is the Turtle format, a text-based serialization format for RDF (Resource Description Framework) data, which is valued for its readability and simplicity.

However, a significant issue arises when converting IFC instance data to Turtle format using traditional tools. These tools tend to integrate the entire IFC structure into the Turtle file, a process that results in an exponential increase in file size. Specifically, it can be observed that the file size of Turtle files increases by a factor of 10 to 12 compared to their IFC counterparts.

Figure 19 Size comparison between .ifc and .ttl

This significant increase in file size can have a number of adverse effects. First, it can significantly impact the efficiency of data exchange, as larger files require more storage space and entail longer transfer times. Second, it can lead to an increased load on system resources, as reading and processing larger files requires more processing power. Third, it can limit the usability of the Turtle format as a medium for data exchange, since the increased file size negates the advantages of the format's readability and simplicity.

Addressing the challenging data transfer and conversion issues posed by converting complete IFC instance files to .ttl format, an innovative approach was developed by us to improve the efficiency and practicality of the process. This approach aims not to transform the entire IFC instance file, but to focus specifically on the segments that more selective and focused data transformation.

At the beginning of this process, the IFC file is imported into the specialized tool "KITModelViewer". This tool is characterized by its integrated interface to the highly flexible and widely used Python programming language. The use of this tool is limited to the prototyping and development of the concept, as the tools used in the final platform will most likely be different, if not custom made for openDBL. This interface acts as a kind of gateway that enables seamless interaction and integration with Python, allowing a wide range of functionalities and customizations to be realized in terms of data processing.

A key component of this approach is the implementation of a Python script specifically designed to control and optimize the extraction process. In the first phase of this process, the script focuses on extracting all attributes from all property sets (PSets) associated with a particular component of the IFC file. This first phase is a critical component of the entire process, as it allows the relevant data to be effectively isolated and extracted without having to navigate through the entire data set.

By focusing on the specific portions of the IFC file that are relevant to the use case at hand, this approach allows for a significant reduction in file size, resulting in improved efficiency and speed of the conversion process. In addition, this

approach provides the ability to tailor data extraction to specific criteria and requirements, allowing for greater flexibility and adaptability to different project needs.

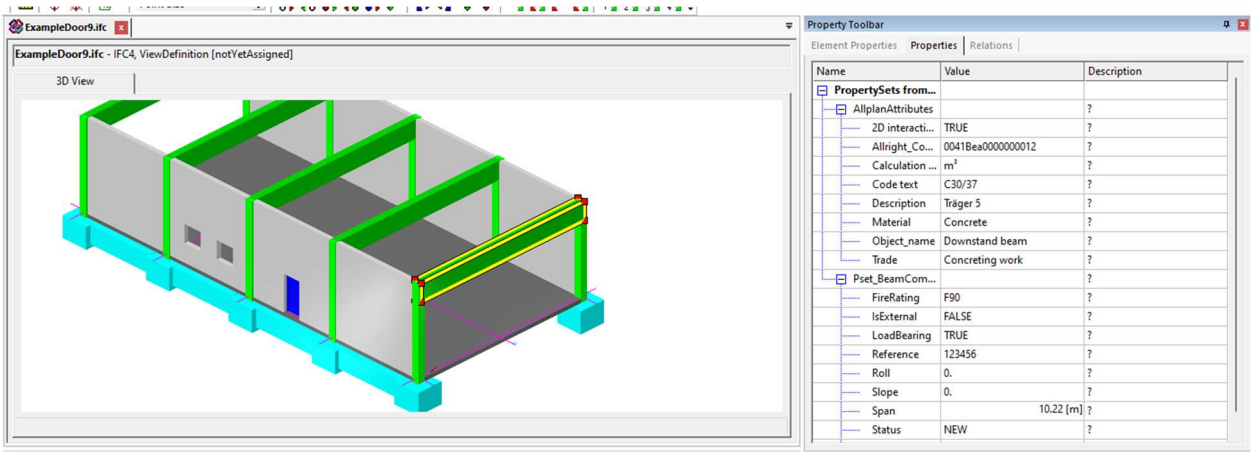


Figure 20 Selection of specific Parts of the IFC-Model in the KITModel Viewer

```
2023-09-12_13-2... INFO    PythonMessage   Starting query_selected_entities function...
2023-09-12_13-2... INFO    PythonMessage   Fetching selected entities...
2023-09-12_13-2... INFO    PythonMessage   Saving to txt...
2023-09-12_13-2... INFO    PythonMessage   Checking directory: C:\Users\moberts\Downloads\KITModelViewer_V7.1_Build-2786(1)\plugins\ifcexplorer_python3\scripts
2023-09-12_13-2... INFO    PythonMessage   Writing to C:\Users\moberts\Downloads\KITModelViewer_V7.1_Build-2786(1)\plugins\ifcexplorer_python3\scripts\result.txt...
```

Figure 21 Extracting Properties in the KITModelViewer with the Python-plugin

```
ID: 3077; Name:  ; Type:IfcBeam
Geometry Type: Extrusion
Properties: {'AllplanAttributes': {'2D interaction': ['TRUE'], 'Allright_Comp_ID':
['0041Bea0000000012'], 'Calculation mode': ['m³'], 'Code text': ['C30/37'], 'Description':
['Träger 5'], 'Material': ['Concrete'], 'Object_name': ['Downstand beam'], 'Trade':
['Concreting work']}, 'Pset_BeamCommon': {'FireRating': ['F90'], 'IsExternal': ['FALSE'],
'LoadBearing': ['TRUE'], 'Reference': ['123456'], 'Roll': ['0.'], 'Slope': ['0.'], 'Span':
['10.22', '10.22 [m]'], 'Status': ['NEW'], 'ThermalTransmittance': ['2.1']}}
```

Figure 22 The generated Output of the Python-script is stored in a .txt file

The initially generated text file (.txt) presents itself as extremely compact, with a file size that is only a few kilobytes. In our specific research, the generated file was even less than 1 kilobyte in size. This remarkable reduction in file size represents a

significant advantage, as it not only minimizes storage requirements, but also greatly improves the speed and efficiency of the data transfer process.

Following this initial phase of data extraction and preparation is a strategically designed step in which a separate Python script is applied. This script, named as "txt_to_ttl_conversion_combined.py", has the specific task of converting the previously generated text file into Turtle format. The decision to keep and implement the scripts separately was a deliberate strategy to create a flexible and modular architecture. This structuring allows the extraction processes to be designed by different viewers or tools independently of the conversion phase. This is especially relevant since the extraction mechanisms in different tools can diverge, requiring specific adaptation to each tool.

The conversion script is characterized by its flexibility and adaptability, as it works independently of the specific nature of the extraction tool, as long as the structure of the text file meets the requirements presented in the previous phase. This promotes broader applicability and reusability of the script in different contexts and projects.

```
@prefix ex: <http://example.org/>.
@prefix ifc: <http://example.org/ifc/>.
@prefix attr: <http://example.org/attr/>.
@prefix prop: <http://example.org/prop/>.
ex:3077 a ex:IfcBeam .
ex:3077 ex:AllplanAttributes ex:AllplanAttributes_3077 .
ex:AllplanAttributes_3077 ex:hasAttribute ex:2D_interaction .
ex:2D_interaction ex:hasValue "TRUE" .
ex:AllplanAttributes_3077 ex:hasAttribute ex:Allright_Comp_ID .
ex:Allright_Comp_ID ex:hasValue "0041Bea0000000012" .
ex:AllplanAttributes_3077 ex:hasAttribute ex:Calculation_mode .
ex:Calculation_mode ex:hasValue "m_cb" .
ex:AllplanAttributes_3077 ex:hasAttribute ex:Code_text .
ex:Code_text ex:hasValue "C30/37" .
ex:AllplanAttributes_3077 ex:hasAttribute ex:Description .
ex:Description ex:hasValue "Traeger_5" .
ex:AllplanAttributes_3077 ex:hasAttribute ex:Material .
ex:Material ex:hasValue "Concrete" .
ex:AllplanAttributes_3077 ex:hasAttribute ex:Object_name .
ex:Object_name ex:hasValue "Downstand_beam" .
ex:AllplanAttributes_3077 ex:hasAttribute ex:Trade .
ex:Trade ex:hasValue "Concreting_work" .
ex:3077 ex:Pset_BeamCommon ex:Pset_BeamCommon_3077 .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:FireRating .
ex:FireRating ex:hasValue "F90" .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:IsExternal .
ex:IsExternal ex:hasValue "FALSE" .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:LoadBearing .
ex:LoadBearing ex:hasValue "TRUE" .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:Reference .
ex:Reference ex:hasValue 123456 .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:Roll .
ex:Roll ex:hasValue "0." .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:Slope .
ex:Slope ex:hasValue "0." .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:Span .
ex:Span ex:hasValue 10.22 .
ex:Span ex:hasValue "10.22_[m]" .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:Status .
ex:Status ex:hasValue "NEW" .
ex:Pset_BeamCommon_3077 ex:hasAttribute ex:ThermalTransmittance .
ex:ThermalTransmittance ex:hasValue 2.1 .
```

Figure 23 Converted .txt file to .ttl

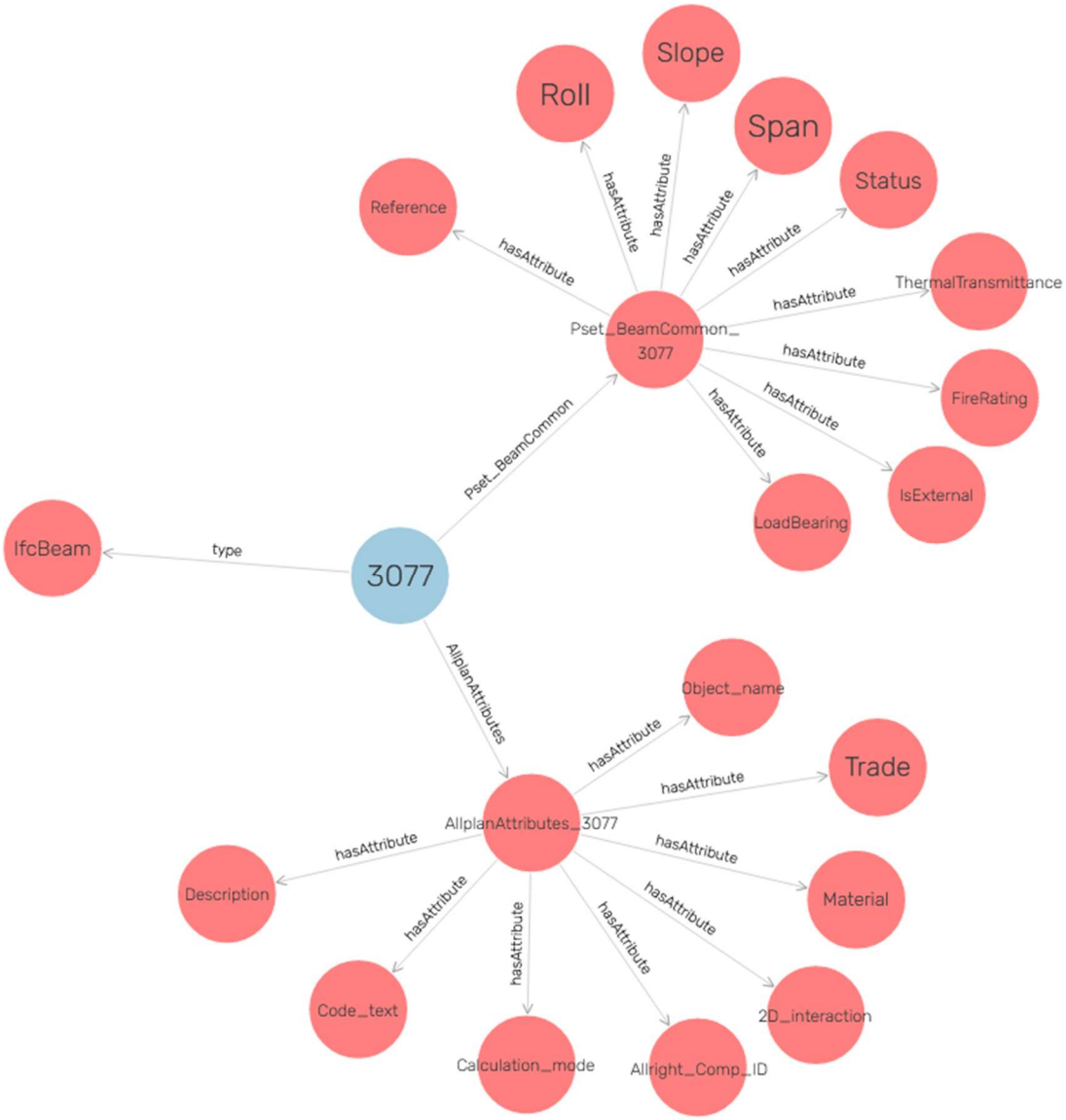


Figure 24 Import of the converted .ttl into Ontotext GraphDB

In the progressive development of technologies for data processing and analysis, the implementation of a concerted strategy for data transfer represents a critical and central aspect. In the subsequent process step, after the conversion of the data to the Turtle format has been successfully performed, a significant advancement in the data management strategy is realized, which differs significantly from a conventional data pipeline.

In this context, the converted file type is integrated into Ontotext GraphDB, a specialized database based on graph-based data processing. This marks the beginning of a phase that is not just about simply transferring data, but rather establishing a complex and multifunctional platform that has the ability to interpret and use data in an expanded and nuanced context.

In order to validate the robustness and adaptability of our innovative approach, we extended our testing to external datasets. Specifically, we utilized an .ifc file provided by in2it, detailing the "School of Ruvo." This dataset offers a unique set of challenges and complexities, serving as an ideal candidate for verifying the applicability of our data conversion and extraction methodologies.

Upon importing the "School of Ruvo" .ifc file into the KITModelViewer, the Python plugin was applied just as in our initial experiments. Consistent with our expectations, the script successfully extracted the targeted attributes from property sets associated with specific components in the .ifc model. This outcome substantiates the tool's ability to function effectively across different datasets, regardless of their source or complexity.

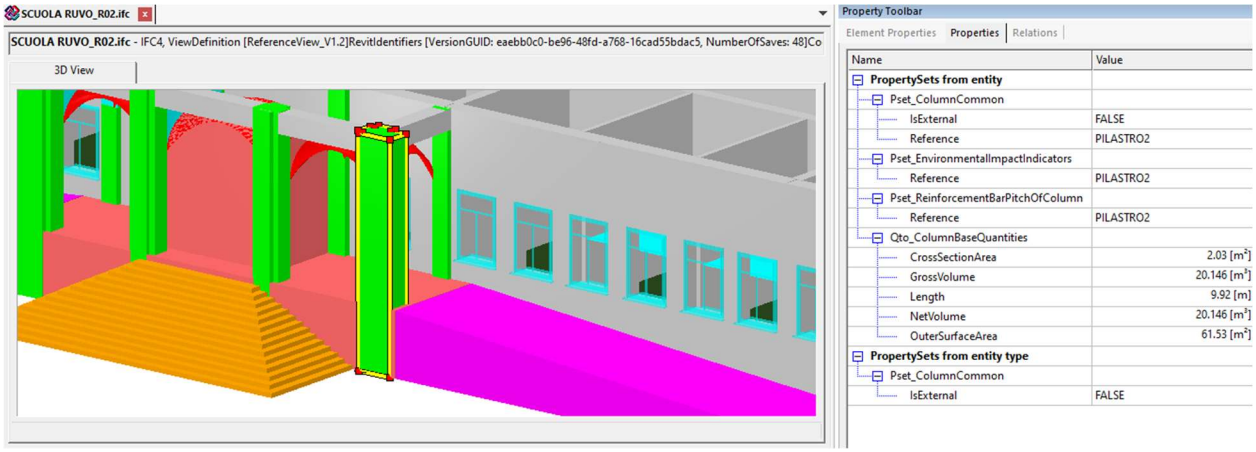


Figure 25 Screenshot of the "School of Ruvo" IFC File in the KITModelViewer

The extracted data was then saved in a text file, which, akin to our prior experiments, was extremely compact in terms of file size. The efficiency of the data extraction process was thus corroborated, affirming its general applicability.

Subsequently, the text file was converted into Turtle format using the "txt_to_ttl_conversion_combined.py" script. The transition was seamless and efficient, further validating the versatility of our Python scripts in accommodating different data structures and complexities.

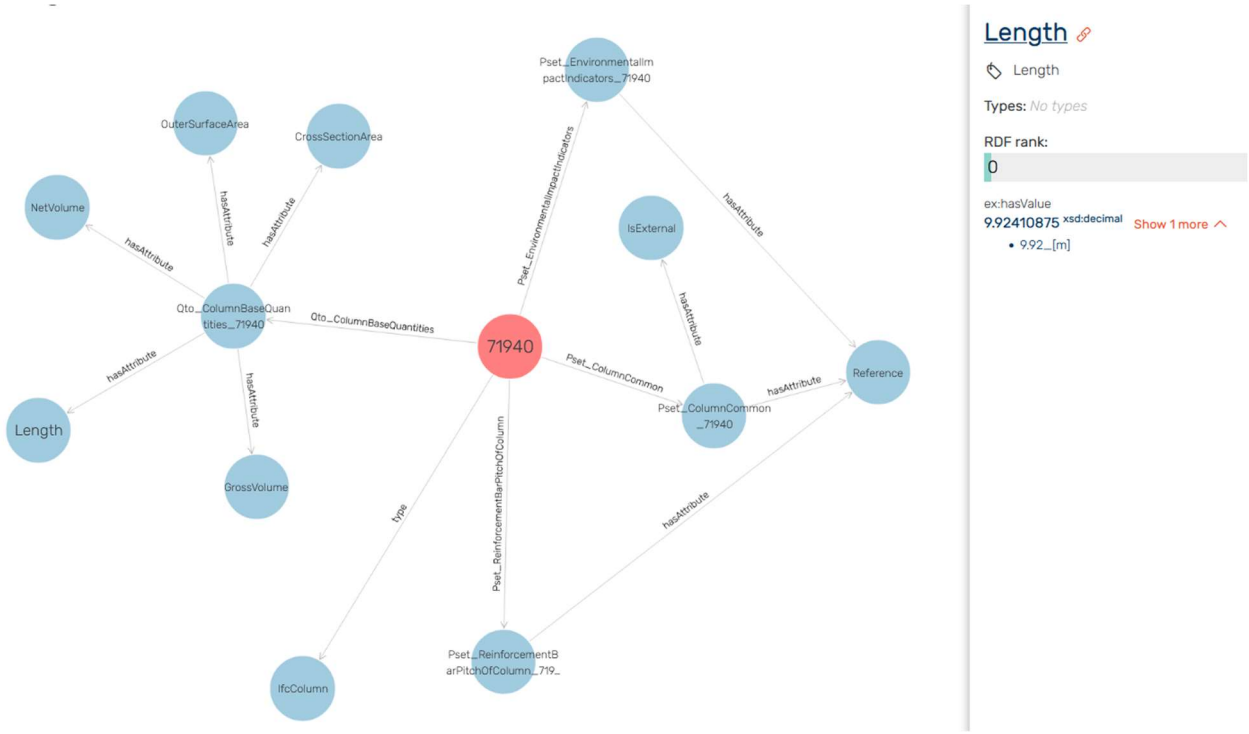


Figure 26 Visualization of Extracted and Converted Data in Ontotext GraphDB

Finally, the converted .ttl file was successfully integrated into Ontotext GraphDB, facilitating advanced, graph-based data analytics. This stage confirmed that our "communication bridge" is not only effective with our initial dataset but also capable of accommodating and enhancing the value of external datasets like the one provided by in2it.

This step thus represents a way to establish a concrete and effective link between the original IFC file, GraphDB, and an Artificial Intelligence. This is not a linear or unidirectional data pipeline, but a multifunctional interface that enables the transfer, integration, and synergy of data. This interface serves as a kind of "communication bridge" that allows the data contained in the IFC file to be integrated into a graph-based database, creating an environment where deeper and more complex analysis by artificial intelligence is possible.

This "communication bridge" opens the door to a new level of data analysis and interpretation by creating an integrative environment where the different technologies (IFC files, GraphDB and AI) can interact synergistically. This synergy allows the scope and depth of data analysis to be extended by leveraging the capabilities of each technology to gain a more comprehensive and nuanced understanding of the data.

7.3 IoTs

The inclusion of Internet of Things devices is pivotal to our project. These devices contribute a wealth of real-time data that enhances multiple aspects of Building Information Modeling. Comprising a wide variety of sensors that monitor environmental conditions, as well as smart devices integrated into building systems, Internet of Things devices offer a constant flow of information.

This information, which may come in various formats, is then standardized and incorporated into the overall structure of data. This facilitates more comprehensive and nuanced analyses. Furthermore, Internet of Things devices are invaluable for tracking and managing the performance of different building systems, as they provide data that can be used to optimize operations and improve efficiency.

7.4 bSDD (Classifications)

The buildingSMART Data Dictionary (bSDD) will be used as a vital tool in our project, offering a standardized classification system that facilitates the integration of various data sources in a harmonized manner. These classifications, which encompass a wide range of parameters and data points, serve as a guideline for data integration, ensuring that data from various sources can be harmonized and analysed cohesively. Furthermore, the bSDD classifications play a crucial role in fostering interoperability between different data sources, creating a unified data repository that can be easily accessed and analysed by various stakeholders. This approach ensures a more streamlined data management process, fostering a more collaborative and efficient working environment. The paragraph explores the various classifications available and the methodologies employed in integrating these classifications into the data management process. It also elucidates the role of bSDD in enhancing the efficiency and effectiveness of the Building Information Modeling process, providing an overview of the strategies employed in harmonizing data from various sources and fostering a cohesive data management framework.

7.5 Handling Additional Data Formats (pdf, etc.)

In In the scope of our project, additional data formats such as PDFs are integral and require systematic integration into the overarching data infrastructure. To facilitate this integration, various data extraction tools are employed based on the type and complexity of the data.

7.5.1 PDF Data Extraction:

Tools like Apache PDFBox, PyPDF2, and Tabula are leveraged for parsing and transforming data encapsulated in PDF files. These tools enable the extraction of textual data, and in the case of Tabula, even tabular data into more manipulable formats such as Comma Separated Values (CSV) or Excel spreadsheets.

7.5.2 Text and Document Formats:

For parsing HTML and XML data structures, the Beautiful Soup Python library is employed. Additionally, Apache POI is used for the extraction of data from Microsoft Office documents, and Readability.js is utilized for optimizing text readability.

7.5.3 Web Scraping:

Web data is often scraped using Python-based libraries such as Scrapy, or through browser automation frameworks like Selenium. These tools allow for the automated collection of data from web pages, which can then be parsed and standardized.

7.5.4 XML/JSON Parsing:

Data in XML format is processed using the XML.etree.ElementTree library in Python, while JSON data is handled using libraries like Jackson for Java.

7.5.5 Database Extraction:

For relational database access, SQLAlchemy and Java Database Connectivity (JDBC) are utilized. These libraries provide a robust framework for SQL queries and data retrieval.

7.5.6 API Data Extraction:

The HTTP library Requests for Python and Postman are often used for data extraction from Application Programming Interfaces (APIs). These tools are particularly useful for RESTful API calls, enabling the retrieval of structured data.

The application of these specialized tools contributes to the development of a robust data management system capable of accommodating a diverse array of data formats. Through the use of these tools, challenges associated with disparate data formats are effectively mitigated, thereby enhancing the efficiency and effectiveness of the Building Information Modeling process. This section outlines the methodologies implemented for the seamless integration of various data

formats into a cohesive data management framework, focusing on tool-specific strategies and techniques.

8 AI Technology [41]

In the rapidly advancing sphere of Building Information Modelling (BIM), the incorporation of AI technology represents a critical evolution in fostering growth in construction and infrastructure development. AI facilitates the streamlined synthesis and examination of large datasets, thereby augmenting the procedural efficiency and efficacy substantially. This synergistic interaction aligns impeccably with the data-intensive environment of BIM, as AI technologies introduce capabilities of learning and predictive analysis that hold the promise to fundamentally reshape the industry.

As we delve further into this complex domain, this introductory section intends to elaborate on the various available algorithms, identifying the optimal algorithm for our project, and ascertaining the requisite volume of training data to attain optimal results. Moreover, we aim to investigate the broad spectrum of AI technologies that can be integrated effortlessly into BIM processes, highlighting their potential applications and the myriad benefits they bestow. Through this investigation, we endeavour to illuminate the transformative capacity of AI technologies in augmenting the efficiency and effectiveness of BIM, thus heralding a new epoch in the field of construction and infrastructure development.

8.1 Overview of AI Technologies

AI is a rapidly growing field of computer science that is revolutionizing the way we interact with technology. [40] AI is often used as a blanket term for software that emulates human abilities to learn and think. AI is being used in a variety of applications, from medical diagnosis to autonomous vehicles. AI is also being used to create virtual assistants, such as Amazon's Alexa and Apple's Siri. [42]

There are several different types of AI, each with their own unique characteristics and capabilities. Since AI can be classified in many different ways, in this research we broadly classify AI between the most known types: [43]

- **Reactive Machines:** These are the simplest type of AI, and they can only react to the environment and make decisions based on past experiences. They do not have the ability to form memories or use past experiences to inform future decisions. Examples include Deep Blue, the chess-playing computer that defeated Garry Kasparov in 1997. [44]
- **Limited Memory:** These types of AI have a limited memory and can use past experiences to inform current decisions. Examples include self-driving cars,

which use sensors and cameras to detect and respond to their environment. [45]

- **Theory of Mind:** These types of AI are able to understand mental states, such as beliefs, desires, and intentions, and can use this understanding to predict the behaviour of others. [46]
- **Self-Aware:** These types of AI are capable of self-awareness and possess a sense of self. They can reflect on their own mental states and understand their own consciousness. So far, these kinds of AIs are not available yet, but it is possible that they will be available in the future. [47]
- **General AI:** Also known as Strong AI, these types of AI are capable of understanding or learning any intellectual task that a human being can. They can think, reason, and learn in a way that is similar to or even surpasses human ability. [48]
- **Narrow AI:** These types of AI are designed for a specific task, like image recognition, speech recognition, natural language processing, among others. [48]

It's worth noting that AI technology is constantly evolving, and new types of AI are being developed all the time. Additionally, many AI systems today are a combination of several different types of AI, and the distinction between them can be blurred. [49]

8.2 AI-Algorithms for attribute-based mapping

AI models have advanced greatly in recent years and have found wide use in a variety of applications. Attribute-based mapping, which maps an object's attributes to its real-world characteristics, has received significant attention. This section will examine the different AI models utilized in attribute-based mapping, and analyse their operating mechanisms, advantages, and disadvantages.

8.2.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of deep learning model that are particularly well-suited for image classification tasks. They are designed to automatically identify and extract relevant features from an image by applying multiple filters to the input data.[49] In the context of attribute-based mapping, CNNs can be used to extract attributes such as texture, colour, shape, and others from an image, which can then be used to map the object's attributes to its corresponding real-world characteristics. [50]

One of the key advantages of using CNNs for attribute-based mapping is that they are capable of handling large amounts of data and can learn complex relationships between features and target variables. Additionally, they can also learn to identify objects in images even when they are partially obscured, making them well-suited for real-world applications. Another significant advantage of CNNs is that the system does not require human supervision. [51]

One of the limitations of CNNs is that they require large amounts of labelled training data to perform well. This can be a challenge in applications where labelled data is scarce or difficult to obtain. Additionally, CNNs can also be computationally intensive, which can limit their scalability for large-scale applications. [51]

This kind of AI often finds application in Image classification and recognition. CNNs can be used to classify and recognize objects in images by mapping the attributes of a small part of an image, such as colour and texture, to the corresponding object class. CNNs are particularly effective for image-related tasks due to their ability to identify and extract local and spatial features from images. They are commonly used in applications such as image classification, object detection, and segmentation. [51]

Real-life use cases, where CNNs are implemented:

- **Google Photos:** Google Photos is a popular image and video sharing service that uses CNNs for image recognition and classification. The service automatically organizes photos and videos into albums based on the objects and people that are present in them. [52] [53]
- **Facebook:** Facebook uses CNNs for various computer vision tasks, such as facial recognition and image tagging. The social media platform also uses CNNs for recommendation systems, such as suggesting new friends to connect with. [54]
- **Tesla:** Tesla uses CNNs in its self-driving cars to detect and classify objects in real- time, such as other vehicles, road signs, and pedestrians. [55]
- **Airbnb:** Airbnb uses CNNs to automatically identify and remove duplicate listings from its platform, as well as to recommend similar listings to users. [56]
- **Nvidia:** Nvidia is a technology company that provides AI hardware and software solutions. One of their products is a deep learning software library called cuDNN, which includes optimized implementations of CNNs for use in various applications, including computer vision and natural language processing. [57]

8.2.2 Random Forest

Random Forest is a type of machine learning algorithm that is based on the decision tree algorithm. It works by building a large number of decision trees and combining their pre- dictions to make a final prediction. In the context of attribute-based mapping, Random Forest can be used to map an object's attributes to its corresponding real-world characteristics by using decision trees to make predictions about the object's attributes based on its features. [58]

One of the main advantages of using a Random Forest-type algorithm for attribute-based mapping is that it is capable of handling both continuous and categorical data, making it a flexible algorithm for a wide range of applications. Additionally, Random Forest is also relatively simple to implement and does not require a large number of computational re- sources, making it a good choice for resource-constrained applications. It can also manage missing values and still able to keep the accuracy high, when bigger amounts of data are not available. [59]

However, one of the limitations of Random Forest is that it can be prone to overfitting, particularly when dealing with large amounts of data. Which basically means, that the model fits perfectly for the training data, but when a new data input is given, the AI de- livers less optimal results. Additionally, Random Forest is also not as good as some other algorithms at dealing with complex relationships between features and target variables, which can limit its effectiveness in certain applications such as high-dimensional data, in which the ”trees” in the forest become huge; small sample sizes, in which the output of a Random Forests is imprecise compared to other AI-Models or the work with high variance, which means that the model is sensitive to changes in the training data. [60] [61]

This kind of AI is often used in the field of fraud detection. Random Forest can be used to detect fraud by mapping the attributes of financial transactions, such as amount and location, to the likelihood of fraud. Random Forest is an ensemble learning method that combines multiple decision trees to form a more robust and accurate model. It is particularly useful for tasks where there are a large number of features or where there is a risk of over fitting with a single decision tree. Additionally, Random Forest provides feature importance scores, which can be used to identify the most important factors contributing to the prediction. [61]

Real-life use cases, where Random Forests are implemented:

- **Credit scoring:** Many financial institutions use Random Forests to determine credit risk. The algorithm considers factors such as income, payment history, and debt-to-income ratio to predict the likelihood of a person defaulting on a loan. [62]
- **Healthcare:** Random Forests have been used in healthcare for various applications, such as predicting disease progression and treatment outcomes, and identifying risk factors for diseases. [63]
- **Retail:** Retail companies use Random Forests to analyse customer data and predict sales trends, as well as to identify the products that are most likely to be purchased together. [64]
- **Energy:** Random Forests are used in the energy sector for various applications, such as predicting energy consumption patterns and identifying the causes of energy outages. The algorithm can also be used to predict maintenance needs for energy infrastructure, such as power plants and transmission lines. [65]

8.2.3 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are a type of machine learning algorithms that are used for classification and regression tasks. [36] They work by finding the hyperplane that best separates the data into two classes, with the goal of maximizing the margin between the classes. [66] In the context of attribute-based mapping, SVMs can be used to map an object's attributes to its corresponding real-world characteristics by using the hyperplane to make predictions about the object's attributes based on its features. [66]

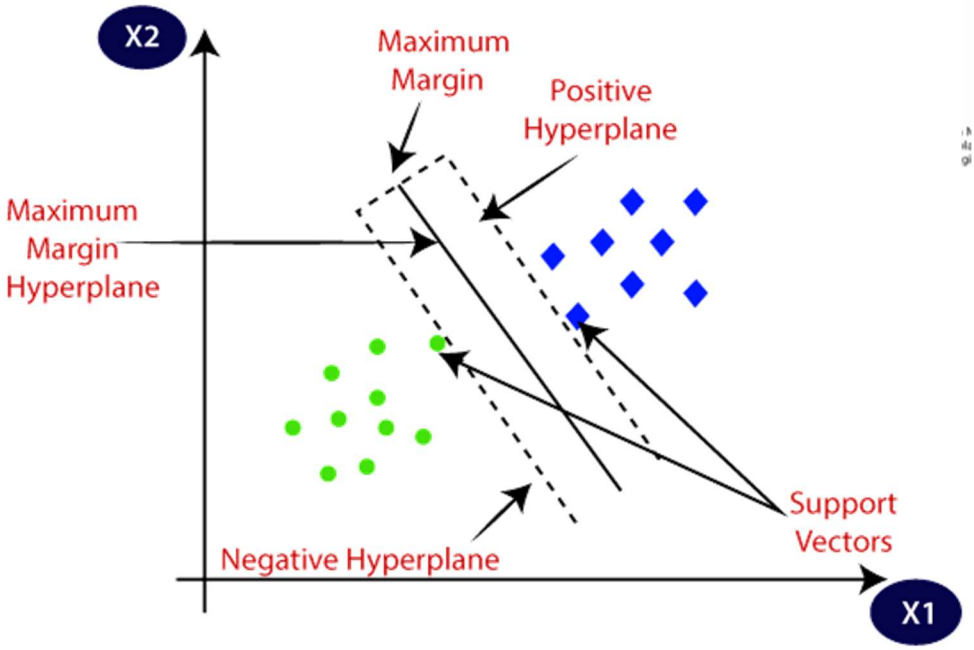


Figure 27 The Support Vector Machine algorithm [66]

One of the key advantages of using SVMs for attribute-based mapping is that they are capable of handling both linear and non-linear data, making them a flexible algorithm for a wide range of applications. Linear and non-linear data refer to the type of decision boundary that separates the data points into different classes. Linear data refers to the case where the data points can be separated by a straight line or a hyperplane itself. In this case, a linear SVMs algorithm can be used to find the optimal hyperplane that maximally separates the data points of different classes. Non-linear data, refers to the case where the data points cannot be separated by a straight line or a hyperplane in the input space. In this case, a non-linear SVMs algorithm can be used to transform the

input data into a higher-dimensional feature space where a linear decision boundary can be found to separate the data points. [66]

Additionally, SVMs are also relatively simple to implement and do not require a large number of computational resources, making them a good choice for resource-constrained applications. [66]

One of the limitations of SVMs is that they can be prone to over fitting when dealing with noisy or complex data. Another limitation is that SVMs can only work well for two-class problems and may not be suitable for multi-class problems. [67]

This type of AI algorithm is often used for handwriting recognition. SVMs can be used to recognize handwriting by mapping the attributes of each handwritten character, such as shape and stroke width, to the corresponding character class. [68]

Real-life use cases, where SVMs are implemented:

- Banks and financial institutions use SVMs for credit scoring and fraud detection. By mapping attributes such as transaction history, demographics, and behaviour to a person's credit score or likelihood of fraud, SVMs can help these institutions make informed decisions about lending and risk management. [69]
- Handwriting recognition software, such as Microsoft OneNote and Google Hand- writing Input, use SVMs to recognize handwritten characters and convert them into digital text. [68]
- Health care companies use SVMs for disease diagnosis by mapping patient symptoms and medical history to a specific disease. [70]
- NLP applications, such as sentiment analysis, use SVMs to classify texts as positive, negative, or neutral by mapping the attributes of the text to the corresponding sentiment class. [71] [72]
- E-commerce websites use SVMs for product recommendation systems by mapping user behaviour, such as browsing history and purchase history, to recommended products. [73]

8.2.4 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are a type of machine learning algorithm that are inspired by the structure and function of the human brain. They consist of a large number of interconnected nodes, called artificial neurons, that work together to perform a specific task. In the context of attribute-based mapping, ANNs can be used to map an object's attributes to its corresponding real-world characteristics by using the artificial neurons to make predictions about the object's attributes based on its features. [74]

One of the key advantages of using ANNs for attribute-based mapping is that they are capable of handling complex relationships between (artificial neurons) that make up the structure of ANNs, so it is still possible to generate output, even if some cells are corrupted. [75]

However, one of the limitations of ANNs is that they can be computationally intensive and require a large amount of labelled training data to perform well. Additionally, ANNs can also be prone to overfitting, particularly when dealing with large amounts of data. [75]

Use case: Speech recognition. ANNs can be used to recognize speech by mapping the at- tributes of speech signals, such as spectral features, to the corresponding word or phoneme. [76]

Real-life use cases, where ANNs are implemented:

- **Google Photos:** Google Photos uses ANNs for image recognition and classification, allowing users to search for specific photos based on the objects and people present in the images. [53]
- **IBM Watson:** IBM Watson uses ANNs for a wide range of applications, including natural language processing, image recognition, and recommendation systems. [77]
- **Facebook:** Facebook uses ANNs for various tasks such as facial recognition, image classification, and spam detection. [78]
- **Amazon:** Amazon uses ANNs in its recommendation system, providing personalized product recommendations to customers based on their purchase history and browsing behaviour. [79]
- **TensorFlow:** TensorFlow is an open-source machine learning library developed by Google that includes a wide range of tools for building and training ANNs. [80]
- **PyTorch:** PyTorch is another open-source machine learning library that is used by researchers and practitioners to build and train ANNs for various applications. [81]

8.2.5 K-Nearest Neighbours (KNNs)

K-Nearest Neighbours (KNNs) is a type of machine learning algorithm that is based on the idea of similarity. It works by finding the K-nearest data points to a given query point and using their labels to make a prediction about the label of the query point. The ”K” stands for the data points to the new data point and classify it based on the majority class of those 3 data points. If two out of the three closest data points are similar in their attributes, KNN can utilize this similarity to predict the real-world characteristics of a new object based on the attributes of these neighboring data points. [79] [82]

One of the key advantages of using KNNs for attribute-based mapping is that the algorithm is simple to implement and does not require a relatively robust to noisy or complex data, making them well-suited for real-world applications. [83]

Part of the limitations of KNNs is that they can be computationally intensive for large- scale applications, particularly when dealing with high-size of the data set. [83]

The general use case for KNNs is Image classification. KNNs can be used to classify an image based on its attributes, such as colour, texture, and shape, by finding the nearest neighbours in the training data set and using their class labels to make a prediction. [84]

Real-life use cases, where KNNs are implemented:

- **Healthcare:** K-Nearest Neighbours can be used for patient diagnosis by mapping the attributes of a patient's symptoms to the most likely disease. [85]
- **Marketing:** K-Nearest Neighbours can be used for customer segmentation by grouping customers with similar attributes, such as buying behaviour, into clusters. [86]
- **Finance:** K-Nearest Neighbours can be used for fraud detection by identifying patterns of transactions that are similar to known fraudulent activity. [87]
- **E-commerce:** K-Nearest Neighbours can be used for product recommendation by suggesting items that are similar to a customer's past purchases. [88]
- **Computer Vision:** K-Nearest Neighbours can be used for image classification by map- ping the attributes of an image to the most similar class. [89]

8.2.6 Gradient Boosting Machines (GBMs)

Gradient Boosting Machines (GBMs) are an ensemble machine learning algorithm that builds multiple weak learners[1] and combines their corresponding real-world characteristics by using the combination of multiple weak learners to make predictions about the object's attributes. [90]

One of the key advantages of using GBMs for attribute-based mapping is that they are capable of handling complex relationships between classification problems, making them well-suited for a wide range of applications. [90]

However, one of the limitations of GBMs is that they can be computationally intensive for large-scale applications, particularly when dealing with high-dimensional data. Addition- ally, GBMs can also be prone to overfitting, particularly when dealing with large amounts of data. [91]

[1] A weak learner is a simple machine learning model that is trained on the errors of the previous model in the ensemble. The goal of the weak learner is not to make accurate predictions on its own, but to capture the patterns that were missed by the previous models in the ensemble. [112]

Use case: Predictive maintenance. GBMs can be used to predict when a machine is likely to fail by mapping the attributes of the machine, such as operating temperature and vibration, to the likelihood of failure. [92]

Real-life use cases, where GBMs are implemented:

- **H2O.ai:** H2O.ai is an AI platform that includes GBMs algorithms for various applications, such as credit risk analysis, customer churn prediction, and sales forecasting. [93]
- **Kaggle:** Kaggle is a website for data science competitions that frequently includes GBMs models in their solutions for classification and regression tasks. [94]
- **Gradient Boosting Libraries:** There are several open-source gradient boosting libraries, such as XGBoost and LightGBM, that can be used to train GBMs models in a variety of applications. These libraries are widely used by data scientists and machine learning practitioners to solve real-world problems. [95] [96]

8.2.7 Naive Bayes

Naive Bayes is a probabilistic machine learning algorithm that uses Bayes' theorem to make predictions about the relationship between features and target variables. In the context of attribute-based mapping, Naive Bayes can be used to map an object's attributes to its object's real-world characteristics. [97]

One of the key advantages of using Naive Bayes for attribute-based mapping is that it is computationally efficient and easy to implement, making it well-suited for resource- constrained applications. Additionally, Naive Bayes is also relatively robust to noisy or complex data, making it well-suited for real-world applications. [98]

However, one of the limitations of Naive Bayes is that it assumes that the features are independent of each other, which may not always be of data. [98]

Use cases: Sentiment analysis. Naive Bayes can be used to analyse the sentiment of a piece of text, such as a customer review or a tweet, by mapping the words and phrases in the text to their corresponding sentiment polarity. [99]

Real-life Use-Cases, where Naive Bayes are implemented:

- **Spam Filtering in Email Services:** Naive Bayes is commonly used in spam filtering systems in email services, such as Gmail and Yahoo Mail. The

algorithm uses the attributes of each email, such as the sender, recipient, subject line, and content, to classify it as either spam or not spam. [100]

- **Sentiment Analysis in social media:** Naive Bayes is also commonly used in sentiment analysis of social media posts, such as tweets and Facebook posts. The algorithm uses the attributes of each post, such as the words used, hashtags, and emojis, to classify it as positive, negative, or neutral. [101]
- **Medical Diagnosis:** Naive Bayes has been used in medical diagnosis to classify a patient's symptoms into a particular disease based on their attributes, such as age, gender, symptoms, and medical history. [102]
- **Text Classification:** Naive Bayes is also used in text classification tasks, such as news categorization, where the algorithm uses the attributes of each article, such as the words used, to classify it into a particular category, such as sports, politics, or technology. [103]

8.2.8 AI Transformers

AI Transformers are a type of machine learning algorithm that uses a neural network architecture to process sequential data. They are commonly used in NLP tasks, such as language translation and text generation. In the realm of attribute-based mapping, AI Transformers correlate an object's attributes with its real-world features. [104]

One of the key advantages of using AI Transformers for attribute-based mapping is their ability to process sequential data, which is often making them well-suited for applications that require high levels of precision. [105]

The limitations of AI Transformers is their computational complexity, which can make them difficult to implement on resource-constrained applications. [106]

Use cases: Language Translation. AI Transformers can be used to translate text from one language to another by mapping the words and phrases in the source language to their corresponding translations in the target language. [107]

Real-life Use-Cases, where AI Transformers are implemented:

- **Language Translation:** One of the most common use-cases of AI Transformers is language translation. For example, Google's "Google Translate" uses an AI Transformer-based model called "Transformer-XL" to translate text from one language to another. [107]
- **Chatbots and Text Generation:** AI Transformers are also used to power chatbots, which are automated conversational agents that interact with users

to provide sup- port or answer questions. These chatbots are used in customer support, e-commerce, and many other industries. AI Transformers also find application to generate human-like text, such as news articles, product descriptions, and even creative writing. For example, the language model ”GPT-3” developed by OpenAI is being used to generate natural language text for a variety of purposes. [108]

- **Image and Video Analysis:** AI Transformers can also be used for image and video analysis tasks, such as object recognition and video captioning. For example, the ”ViT” (Vision Transformer) model developed by Google can be used for image classification tasks. [109]
- **Speech Recognition:** AI Transformers can also be used for speech recognition tasks, such as transcribing spoken words into text. For example, the ”HuBERT” model developed by Facebook can be used for speech recognition in various applications. [110]

8.3 Integration of NLP Techniques for Enhanced Attribute Mapping in openDBL

The application of AI in Building Information Modeling has garnered considerable attention for its transformative potential. While AI transformers as discussed in Section 8.2 provide a robust approach, the integration of Natural Language Processing techniques offers an avenue for nuanced, semantic-based attribute mapping. This section elaborates on the integration of NLP methodologies, inspired by the empirical research conducted in "2023_Forth_Energy&Buildings," [40] within the framework of our openDBL project.

8.3.1 Training Dataset for NLP Algorithms: Composition and Origin

A salient feature of any machine learning model, including NLP algorithms, is the quality and size of the training dataset. As delineated in the research paper "Calculation of embodied GHG emissions in early building design stages using BIM and NLP-based semantic model healing" an extensive dataset comprising attribute descriptors and contextual parameters was utilized for the training of transformer-based NLP models. Analogously, the openDBL project aims to cultivate a comprehensive dataset from multiple sources, primarily Industry Foundation Classes files and Ontotext GraphDB.

The proposed dataset is expected to encompass a multitude of attribute classes, such as geometrical dimensions, material properties, and functional parameters, specific to architectural elements like doors, windows, and HVAC systems. The dataset will also incorporate contextual relationships between these attributes,

facilitating semantic mapping. Preliminary estimations suggest that the dataset will comprise over 10,000 unique attribute entries, thus providing a statistically significant basis for training and validation.

8.3.2 Algorithmic Foundations: Transformer-based NLP Models

The architecture of the NLP model is pivotal to its performance. Drawing parallels with the transformer-based NLP models employed in "Calculation of embodied GHG emissions in early building design stages using BIM and NLP-based semantic model healing" our framework will leverage similar architectures. Transformer models are renowned for their efficacy in handling sequential data, capturing long-term dependencies, and facilitating parallel computing. These characteristics render them particularly adept at comprehending the semantics and intricacies behind diverse sets of attributes, thereby ensuring a higher accuracy rate in attribute mapping.

8.3.3 Training and Validation Mechanisms

The training regime for the NLP model will employ stochastic gradient descent algorithms with adaptive learning rates. Hyperparameter tuning will be conducted through cross-validation to optimize the model's performance metrics, such as precision, recall, and F1 score[2]. Periodic validation on subsets of the training data will be performed to mitigate the risks of overfitting, thereby ensuring a generalized model that is capable of handling unseen data.

8.3.4 Attribute Mapping Pipeline: A Computational Workflow

To elucidate the computational workflow of the attribute mapping pipeline, the journey commences with the extraction of attributes from IFC files. Subsequently, these attributes are converted into a text-based (.txt) representation, followed by a syntactic transformation into the Turtle format (.ttl) for compatibility with Ontotext GraphDB. Post-storage in GraphDB, these attributes serve as the input corpus for the NLP algorithms. The NLP model undertakes attribute classification, normalization, and mapping, the results of which can be directed towards various downstream applications, ranging from predictive analytics to real-time BIM simulations.

[2]The F1 score is the harmonic mean of precision and recall, providing a balanced measure of a mode ls accuracy. It ranges from 0 to 1, with 1 indicating perfect accuracy.

8.3.5 Conclusion and Future Directions

The integration of NLP techniques into the openDBL framework aims to not only augment the accuracy of attribute mapping but also to instill a layer of semantic understanding that is often lacking in conventional AI models. This semantic layer can prove to be invaluable in complex, multi-disciplinary projects where the cost of inaccuracies can be prohibitively high.

8.4 Human-in-the-Loop Systems for File Alignment in openDBL

In the openDBL project, Human-in-the-Loop systems play a pivotal role, especially when it comes to the complex task of file alignment and attribute mapping in Building Information Modeling (BIM). This section elucidates the intricacies of HITL systems in managing algorithmic outcomes, user interactions, and system feedback in the context of openDBL.

8.4.1 System Feedback and User Interaction

Should the algorithm work as intended, the files will be automatically aligned, and the user will receive a confirmation message. However, if the algorithm encounters an issue, the HITL system will promptly alert the user through clear, actionable feedback. This immediate response allows for a real-time review and possible manual correction by the user.

8.4.2 Handling Algorithmic Failures

In scenarios where the algorithm fails to align the files correctly, the system is designed to trigger warning messages that clearly state the nature of the error. The user will have options to either manually correct the alignment or to retry the automated process. A log file containing details of the failed operation will be generated for further analysis.

8.4.3 User Understanding and Training

For effective human-machine collaboration, it's imperative that the user understands how to interpret machine outputs. To this end, a comprehensive user guide and tutorial will be provided, walking the user through the HITL interface and explaining how to interact with it. Furthermore, a help section will be accessible at all times to address common queries and issues, aimed at facilitating a smooth user experience.

8.4.4 Ethical and Legal Implications

As HITL systems involve human judgment, it's crucial to adhere to ethical and legal standards, particularly those related to data privacy and user consent.

Transparent terms of use and a clear privacy policy will be integral parts of the user interface.

In summary, the HITL system in openDBL will be designed to be a robust and adaptive mechanism that leverages both human expertise and machine efficiency. It ensures an effective file alignment process while providing the flexibility to handle exceptions, thereby making it a reliable and ethical solution for BIM applications.

9 Conclusions

In conclusion, artificial intelligence (AI) models emerge as potent tools, facilitating organizations in the seamless and efficient integration of data from diverse sources, thereby ensuring data accuracy and consistency. The utilization of these models in automating the attribute-based mapping process significantly reduces the requisite time and resources for data management, enhancing data quality and reliability in the process. Furthermore, the adoption of AI models enables organizations to anticipate and address potential data discrepancies proactively, equipping them to navigate future data challenges and adaptations more adeptly.

In this context, the research conducted by Kasimir Forth et al. warrants particular attention, where the application of NLP AI techniques has been extensively explored in conjunction with existing BIM frameworks. This investigation corroborates the potential of NLP as a prime AI technique for our specified use-case, illustrating its capacity to augment the current landscape of data management and analytics.

However, it is critical to acknowledge that AI models for attribute-based mapping do not represent a simplistic solution to a complex undertaking. These models should operate symbiotically with other data management strategies and processes. This necessitates the persistent maintenance of stringent data quality and governance protocols, coupled with continuous monitoring and evaluation of AI model performance to ascertain optimal functioning. Moreover, organizations should manifest a proactive stance in addressing potential data privacy and security concerns that may emerge during the deployment of AI models for attribute-based mapping.

In summation, AI models harbour the potential to substantially amplify the efficiency and efficacy of attribute-based mapping. Organizations are encouraged to explore the diverse array of options at their disposal and discern the models that align best with their specific requisites. When approached judiciously, AI models can serve as a

catalyst in unlocking the latent potential of data, fostering more informed and impactful decision-making paradigms. It is imperative to emphasize that the aforementioned AI models are not prefabricated software solutions for attribute mapping; dedicated research and development are imperative for each unique use scenario to ensure the AI model yields the anticipated outcomes.

References

[1] T. Krijnena, An efficient binary storage format for IFC building models using HDF5 hierarchical data format, Automation in Contruction, Volume 113, 2020.

[2] S. Kolaric, “DBL SmartCity: An Open-Source IoT Platform for Managing Large BIM and 3D Geo-Referenced Datasets,” in *52nd Hawaii international conference on system sciences*, 2019.

[3] S. McConnell, Software project Survival Guide, Microsoft Press, 1998.

[4] B. G., Object Oriented Analisys and Design with Applications, Addison-Wesley, 1994.

[5] C. Larman, Applying UML and Patterns 3Ed, Prentice Hall.

[6] S. McConnell, Rapid Development, Microsoft Press, 1996.

[7] E. G. e. al., Design Patterns: Elements of Reusable object-Oriented Software, Addison-Wesley, 1994.

[8] M. F. e. al., Patterns Of Enterprise Application Architecture, Addison-Wesley, 2003.

[9] M. R. e. al., Fundamentals Of Software Architecture, O'Reily Media, 2020.

[10] A. Cockburn, “Hexagonal architecture,” 2005. [Online]. Available: http://wiki.c2.com/?PortsAndAdaptersArchitecture.

[11] R. C. Martin, Clean Architecture, Prentice Hall, 2018.

[12] T. Homsbergs, Get your hands dirty on Clean Architecture 2Ed, Packt Publishing, 2023.

[13] S. Brown, Sofware architecture for developers, LeanPub, 2022.

[14] K. K. a. S. Tsiutsiura, “Implementation of artificial intelligence in the construction industry and analysis of existing technologies,” 2021. [Online]. Available: https://dx.doi.org/10.15587/2706-5448.2021.229532.

[15] E. S. L. P. L. T. G. D. D. G. Mirko Locatelli, “Exploring Natural Language Processing in Construction and Integration with Building Information

Modeling," 2021. [Online]. Available: https://www.mdpi.com/2075-5309/11/12/583.

[16] I. O. J. P. H. C. B. Manzoor, "A Research Framework of Mitigating Construction Accidents in High-Rise Building Projects via Integrating Building Information Modeling with Emerging Digital Technologies," [Online]. Available: https://www.mdpi.com/2076-3417/11/18/8359.

[17] R. L. a. Y. Li, "Research on Energy-efficiency Building Design Based on BIM and Artificial Intelligence," 2021. [Online]. Available: https://iopscience.iop.org/article/10.1088/1755-1315/825/1/012003.

[18] S. S. a. V. Villa, "Trends in Adopting BIM, IoT and DT for Facility Management: A Scientometric Analysis and Keyword Co-Occurrence Network," 2023. [Online]. Available: https://www.mdpi.com/2075-5309/13/1/15.

[19] A. S. A. O. a. C. C. K. Lok, "A Sustainable Facility Management Outsourcing Relationships System: Artificial Neural Networks," 2022. [Online]. Available: https://www.mdpi.com/2071-1050/13/9/4740.

[20] A. K. R. C. M. N. J. S. R. L. Milad Baghalzadeh Shishehgarkhaneh, "Internet of Things (IoT), Building Information Modeling (BIM), and Digital Twin in Construction Industry: A Review, Bibliometric, and Network Analysis," 2022. [Online]. Available: https://www.mdpi.com/2075-5309/12/10/1503.

[21] L. R. a. S. Wallace, "Semantic Analysis in Artificial Intelligence: Tools, Techniques, and Applications," 2023. [Online]. Available: https://www.mdpi.com/2073-431X/12/2/37.

[22] A. J. S. a. P. R. Johnson, "The Role of Human Oversight in Automated Systems: Balancing AI Capabilities with Human Expertise," [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10447318.2022.2153320.

[23] A. P. D. P. Cecilia Panigutti, "Doctor XAI: an ontology-based approach to black-box sequential data classification explanations," [Online]. Available: https://dx.doi.org/10.1145/3351095.3372855.

[24] X. G. R. H. F. Z. Smaili, "Onto2Vec: joint vector-based representation of biological entities and their ontology-based annotations," [Online]. Available: https://dx.doi.org/10.1093/bioinformatics/bty259.

[25] A. P. M. W. G. S. M. Ayan Chatterjee, “An Automatic Ontology-Based Approach to Support Logical Representation of Observable and Measurable Data for Healthy Lifestyle Management: Proof-of-Concept Study,” 2021. [Online]. Available: https://dx.doi.org/10.2196/24656.

[26] D. L. M. L. A. P. R. R. Giuseppe De Giacomo, “Using Ontologies for Semantic Data Integration,” 2017. [Online]. Available: https://dx.doi.org/10.1007/978-3-319-61893-7_11.

[27] Y. K. T. M. C. N. C. N. Ö. Ö. C. S. D. Z. S. B. I. H. Y. I. S. L. R. M. E. Kharlamov, “Towards Analytics Aware Ontology Based Access to Static and Streaming Data,” 2016. [Online]. Available: https://dx.doi.org/10.1007/978-3-319-46547-0_31.

[28] P. G. Cong Peng, “Meaningful Integration of Data from Heterogeneous Health Services and Home Environment Based on Ontology,” 2019. [Online]. Available: https://dx.doi.org/10.3390/s19081747.

[29] M. O. W. a. J. P. Thompson, “Enhancing AI Outputs through Human-in-the-Loop Systems: A Synergistic Approach,” [Online]. Available: https://www.researchgate.net/publication/369614907_A_Perspective_on_The_Synergistic_Potential_of_Artificial_Intelligence_and_Product-Based_Learning_Strategies_in_Biobased_Materials_Education .

[30] L. R. A. a. K. L. Miller, “Trust and Reliability in AI-Driven Systems: The Role of Human Expertise,” 2004. [Online]. Available: https://journals.sagepub.com/doi/10.1518/hfes.46.1.50_30392.

[31] M. A. G. F. M. V. C. D. R. T. R. R. J. S. J. Serey, “attern Recognition and Deep Learning Technologies, Enablers of Industry 4.0, and Their Role in Engineering Research,” [Online]. Available: https://www.mdpi.com/2073-8994/15/2/535.

[32] C.-H. C. Ming-Fong Tsai, “Spatial Temporal Variation Graph Convolutional Networks (STV-GCN) for Skeleton-Based Emotional Action Recognition,” [Online]. Available: https://ieeexplore.ieee.org/document/9328124.

[33] W. Z. Y. X. Rui Nan, “Knowledge Graph Analysis of Digital Emergency Management Research Based on CiteSpace Visualisation: Comparative

Analysis of WOS and CNKI Databases," 2021. [Online]. Available: https://www.hindawi.com/journals/ddns/2022/4604223/.

[34] X. Q. Weiping Ji, "Analysis of the Impact of the Development Level of Aerobics Movement on the Public Health of the Whole Population Based on Artificial Intelligence Technology," 2022. [Online]. Available: https://www.hindawi.com/journals/jeph/2022/6748684/.

[35] J. U.-T. J. A. T. J. D. P. J. Ruiz-Real, "Artificial Intelligence in Business and Economics Research: Trends and Future," 2021. [Online]. Available: https://journals.vilniustech.lt/index.php/JBEM/article/view/13641.

[36] N. Yang, "Financial Big Data Management and Control and Artificial Intelligence Analysis Method Based on Data Mining Technology," 2022. [Online]. Available: https://www.hindawi.com/journals/wcmc/2022/7596094/.

[37] K. K. a. S. Tsiutsiura, "Implementation of artificial intelligence in the construction industry and analysis of existing technologies," 2021. [Online]. Available: http://journals.uran.ua/tarp/article/download/229532/2293892.

[38] Z. C. M. O. a. P. D. Zhen Liu, "Blockchain and Building Information Management (BIM) for Sustainable Building Development within the Context of Smart Cities," 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/4/2090.

[39] P. L. W. L. T. Y. a. G. I. Gaspare D'Amico, "Understanding Sensor Cities: Insights from Technology Giant Company Driven Smart Urbanism Practice," 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/16/4391.

[40] J. A. A. B. Kasimir Forth, "Calculation of embodied GHG emissions in early building design stages using BIM and NLP-based semantic model healing," 2023.

[41] J. Martin, "The use of artificial intelligence in the BIM process for mapping attributes from databases," 2023.

[42] B. Marr, "Are Alexa And Siri Considered AI?," 2021. [Online]. Available: https://bernardmarr.com/are-alexa-and-siri-considered-ai/.

[43] N. Joshi, "Types Of Arti," 2019. [Online]. Available: https://www.forbes.com/.

[44] "Deep Blue (chess computer).," 2022. [Online]. Available: https://en.wikipedia.org/wiki/Deep_Blue_versus_Garry_Kasparov,.

[45] "Limited Memory.," 2021. [Online]. Available: https://hypersense.subex.com/aiglossary/limited-memory/,.

[46] "DevTeam.Space: What is Theory Of Mind AI?," [Online]. Available: https://www.devteam.space/blog/theory-of-mind-ai/.

[47] P. Mantri, "Self-Awareness in Arti," 2022. [Online]. Available: https://blog.verzeo.com/self-awareness-in-artificial-intelligence/.

[48] Z. Larkin, "General AI vs Narrow AI," 2022. [Online]. Available: https://levity.ai/blog/general-ai-vs-narrow-ai.

[49] C. S. Smith, "A.I. Here, There, Everywhere," 2021. [Online]. Available: https://www.nytimes.com/2021/02/23/technology/ai-innovation-privacy-seniors-education.html.

[50] N. Lang, "Using Convolutional Neural Network for Image Classi," 2021. [Online]. Available: https://towardsdatascience.com/using-convolutional-neural-network-for-image-classification-5997bfd0ede4.

[51] Engati, "What are the advantages of convolutional neural networks?," 2022. [Online]. Available: https://www.engati.com/glossary/convolutional-neural-network.

[52] course, "ML Practicum: Image Classi," 2022. [Online]. Available: https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks?hl=en.

[53] M. Lewontin, "How Google Photos uses machine learning to create customized albums," 2022. [Online]. Available: https://www.csmonitor.com/Technology/2016/0324/How-Google-Photos-uses-machine-learning-to-create-customized-albums.

[54] M. A. F. A. C. G. L. Yunrong, "Social media sentiment analysis though parallel dilated convolutional neural network for smart city applications," [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0140366419320031.

[55] Y. Bouchard, “Tesla's Deep Learning at Scale: Using Billions of Miles to Train Neural Networks,” [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8711701.

[56] D. Soni, “Graph Machine Learning at Airbnb,” [Online]. Available: https://medium.com/airbnb-engineering/graph-machine-learning-at-airbnb-f868d65f36ee.

[57] V. INC, “IMPLEMENTING DEEP LEARNING USING CUDNN,” [Online]. Available: https://images.nvidia.com/content/gtc-kr/part_2_vuno.pdf.

[58] D.-I. (. S. L. /. N. Litzel, “Was ist Random Forest?,” 2020. [Online]. Available: https://www.bigdata-insider.de/was-ist-random-forest-a-913937/.

[59] H20.ai, “Was ist Random Forest?,” [Online]. Available: https://h2o.ai/wiki/random-forest/.

[60] M. L. Biologists, “Decision Trees, Random Forests, and Overfitting,” 2022. [Online]. Available: https://carpentries-incubator.github.io/ml4bio-workshop/04-trees-overfitting/index.html.

[61] N. Donges, “Random Forest Classifier: A Complete Guide to How It Works in Machine Learning,” 2023. [Online]. Available: https://builtin.com/data-science/random-forest-algorithm#uses.

[62] P. Wanyanga, “Credit Scoring using Random Forest with Cross Validation.,” [Online]. Available: https://medium.com/analytics-vidhya/credit-scoring-using-random-forest-with-cross-validation-1a70c45c1f31.

[63] P. K. K. Kumar, “A healthcare monitoring system using random forest and internet of things (IoT).,” 2019. [Online]. Available: https://link.springer.com/article/10.1007/s11042-019-7327-8.

[64] K. Prawtama, “Predicting Store Sales | Random Forest Regression.,” 2022. [Online]. Available: https://medium.com/mlearning-ai/predicting-store-sales-random-forest-regression-b77abec64c17.

[65] Y. L. C. Z. Feng, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484721006016.

[66] Javatpoint, “Support Vector Machine Algorithm,” 2021. [Online]. Available: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm.

[67] D. K, “Top 4 advantages and disadvantages of Support Vector Machine or SVM,” 2019. [Online]. Available: https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107.

[68] D. N. /. H. H. /. S. S. Yuhaniz, “SUPPORT VECTOR MACHINE (SVM) FOR ENGLISH HANDWRITTEN CHARACTER RECOGNITION,” 2010. [Online]. Available: https://www.researchgate.net/profile/Dewi-Nasien/publication/232657397_Support_Vector_Machine_SVM_for_English_Handwritten_Character_Recognition/links/6172c4243c987366c3c7bb78/Support-Vector-Machine-SVM-for-English-Handwritten-Character-Recognition.pdf.

[69] V. D. /. R. Dhanapal, “BEHAVIOR BASED CREDIT CARD FRAUD DETECTION USING SUPPORT VECTOR MACHINES,” [Online]. Available: https://ictactjournals.in/paper/IJSCV2_I4_P7_391_397.pdf.

[70] T. R. /. O. R. /. I. S. /. N. Marko, “Multilevel Weighted Support Vector Machine for Classi,” 2016. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155119.

[71] G. Bedi, “A guide to Text Classification (NLP) using SVM and Naive Bayes with Python,” 2018. [Online]. Available: https://medium.com/@bedigunjit/simple-guide-to-text-classification-nlp-using-svm-and-naive-bayes-with-python-421db3a72d34.

[72] R. V. /. M. Jayasree, “Aspect based Sentiment Analysis using support vector machine classifier,” 2013. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6637416.

[73] X. L. /. J. Li, “Using support vector machine for online purchase predication,” 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7854334.

[74] IBM, “What are neural networks?,” 2021. [Online]. Available: https://www.ibm.com/topics/neural-networks.

[75] M. M. Mijwel, “Artificial Neural Networks Advantages and Disadvantages,” 2021. [Online]. Available: https://www.researchgate.net/profile/Maad-Mijwil/publication/323665827_Artificial_Neural_Networks_Advantages_and_Disadvantages/links/5aa2c01faca272d448b5a23d/Artificial-Neural-Networks-Advantages-and-Disadvantages.pdf.

[76] J. Tebelskis, “Speech Recognition using Neural Networks,” 1995. [Online]. Available: https://isl.anthropomatik.kit.edu/pdf/Tebelskis1995.pdf.

[77] D. Pesce/ F./ Albanese / F./ Mallardi, “Identification of glomerulosclerosis using IBM Watson and shallow neural networks,” 2022. [Online]. Available: {https://link.springer.com/article/10.1007/s40620-021-01200-0.

[78] G. Templeton, “Facebook is working on ‘deep learning’ neural networks to learn even more about your personal life,” 2013. [Online]. Available: https://www.extremetech.com/computing/167179-facebook-is-working-on-deep-learning-neural-networks-to-learn-even-more-about-your-personal-life.

[79] M. Jarrell, “How Amazon Uses AI in eCommerce - Two Use-Cases,” 2021. [Online]. Available: https://emerj.com/ai-sector-overviews/artificial-intelligence-at-amazon/.

[80] D. S. /. S. Carter, “Tinker With a Neural Network Right Here in Your Browser. Don’t Worry, You Can’t Break It. We Promise.,” 2022. [Online]. Available: https://playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=circle&regDataset=reg-plane&learningRate=0.03&regularizationRate=0&noise=0&networkShape=4,2&seed=0.60587&showTestData=false&discretize=false&percTrainData=50&x=true&y=true&xTimesY=fal.

[81] PyTorch, “Neural Networks,” 2022. [Online]. Available: https://pytorch.org/tutorials/beginner/blitz/neural_networks_tutorial.html.

[82] R. Dwivedi, “How Does K-nearest Neighbor Works In Machine Learning Classification Problem?,” 2020. [Online]. Available: https://www.analyticssteps.com/blogs/how-does-k-nearest-neighbor-works-machine-learning-classification-problem.

[83] A. Soni, “Advantages And Disadvantages of KNN,” 2020. [Online]. Available: https://medium.com/@anuuz.soni/advantages-and-disadvantages-of-knn-ee06599b93366542.

[84] J. K. /. B.-S. K. /. S. Savarese, “Comparing image classification methods: K-nearest-neighbor and support-vector-machines.,” 2012. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/2209654.2209684.

[85] H. S. /. W. Cheruiyot, “Application of k-Nearest Neighbour Classification in Medical Data Mining,” [Online]. Available: https://www.researchgate.net/publication/270163293_Application_of_k-Nearest_Neighbour_Classification_in_Medical_Data_Mining.

[86] Sitta, “Customer Segmentation using K-Means and KNN Algorithms,” 2020. [Online]. Available: https://rstudio-pubs-static.s3.amazonaws.com/599866_59be74824ca7482ba99dbc8466dc36a0.html.

[87] D. R. /. S. Malekzadeh, “Combination of Deep Neural Networks and K-Nearest Neighbors for Credit Card Fraud Detection,” 2022. [Online]. Available: https://www.researchgate.net/publication/360890482_A_Combination_of_Deep_Neural_Networks_and_K-Nearest_Neighbors_for_Credit_Card_Fraud_Detection.

[88] A. Vidhya, “Movie Recommendation and Rating Prediction using K-Nearest Neighbors,” 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/08/recommendation-system-k-nearest-neighbors/.

[89] A. Rosebrock, “Your First Image Classifier: Using kNN to Classify Images,” 2021. [Online]. Available: https://pyimagesearch.com/2021/04/17/your-first-image-classifier-using-k-nn-to-classify-images/.

[90] J. Brownlee, “Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning,” 2020. [Online]. Available: https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/.

[91] V. Kurama, “Gradient Boosting In Classification: Not a Black Box Anymore!,” 2020. [Online]. Available: https://blog.paperspace.com/gradient-boosting-for-classification/.

[92] N. A. /. E. A. P. A. /. I. A. A. /. J. J. /. M. H. H. /. A. N. C. Abas, “A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems,” 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9247843.

[93] H2O.ai, “Gradient Boosting Machine (GBM),” 2020. [Online]. Available: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/gbm.html.

[94] Y. Kashnitsky, “Topic 10. Gradient Boosting,” 2020. [Online]. Available: https://www.kaggle.com/code/kashnitsky/topic-10-gradient-boosting.

[95] T. C. /. T. He, “xgboost: eXtreme Gradient Boosting.,” 2017. [Online]. Available: https://cran.microsoft.com/snapshot/2017-12-11/web/packages/xgboost/vignettes/xgboost.pdf.

[96] J. Brownlee, “How to Develop a Light Gradient Boosted Machine (LightGBM) Ensemble.,” 2020. [Online]. Available: https://machinelearningmastery.com/light-gradient-boosted-machine-lightgbm-ensemble/.

[97] R. Gandhi, “Naive Bayes Classier,” 2018. [Online]. Available: https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c.

[98] MLNerds, “Naive Bayes Classiefier: Advantages and Disadvantages,” 2021. [Online]. Available: https://machinelearninginterview.com/topics/machine-learning/naive-bayes-classifier-advantages-and-disadvantages/.

[99] J. Jorly, “DOCUMENT SENTIMENT ANALYSIS USING NAIVE BAYES,” 2020. [Online]. Available: https://medium.com/analytics-vidhya/document-sentiment-analysis-using-naive-bayes-8911f25f7c95.

[100] V. M. /. I. A. /. G. Paliouras, “Spam Filtering with Naive Bayes - Which Naive Bayes?,” 2020. [Online]. Available: https://userweb.cs.txstate.edu/~v_m137/docs/papers/ceas2006_paper_corrected.pdf.

[101] V. A. F. /. R. A. /. M. A. Hasibuan, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050919318927?ref=pdf_download&fr=RR-2&rr=8058ef2a3e43716e.

[102] S. K. /. M. Ghalib, "A naive-bayes approach for disease diagnosis with analysis of disease type and symptoms.," 2015. [Online]. Available: https://www.researchgate.net/publication/282381602_A_naive-bayes_approach_for_disease_diagnosis_with_analysis_of_disease_type_and_symptoms.

[103] S.-B. K. /. K.-S. H. /. H.-C. R. /. S. H. Myaeng, "Some Effective Techniques for Naive Bayes Text Classification," 2006. [Online]. Available: https://www.researchgate.net/profile/Sang-Bum-Kim-3/publication/3297622_Some_Effective_Techniques_for_Naive_Bayes_Text_Classification/links/56bc297a08ae7be8798bec38/Some-Effective-Techniques-for-Naive-Bayes-Text-Classification.pdf.

[104] G. Giacaglia, "How Transformers Work," 2019. [Online]. Available: https://towardsdatascience.com/transformers-141e32e69591.

[105] A. Srivastava, "What Are Transformers In NLP And It's Advantages," 2022. [Online]. Available: https://blog.knoldus.com/what-are-transformers-in-nlp-and-its-advantages/.

[106] M. Saeed, "Transformers: What They Are and Why They Matter," 2022. [Online]. Available: https://exchange.scale.com/public/blogs/transformers-what-they-are-and-why-they-matter.

[107] L. Bouchard, "Introduction to Transformer Networks | How Google Translate works | Attention Is All You Need.," 2020. [Online]. Available: https://medium.com/what-is-artificial-intelligence/introduction-to-transformer-networks-how-google-translate-works-attention-is-all-you-need-309827c9b942.

[108] G.-3. B. A. G. S. M. B. A. The Week in AI: Transformers Take Over, "Coquillo, Greg," 2022. [Online]. Available: https://exchange.scale.com/public/blogs/the-week-in-ai-transformers-take-over-greg-coquillo.

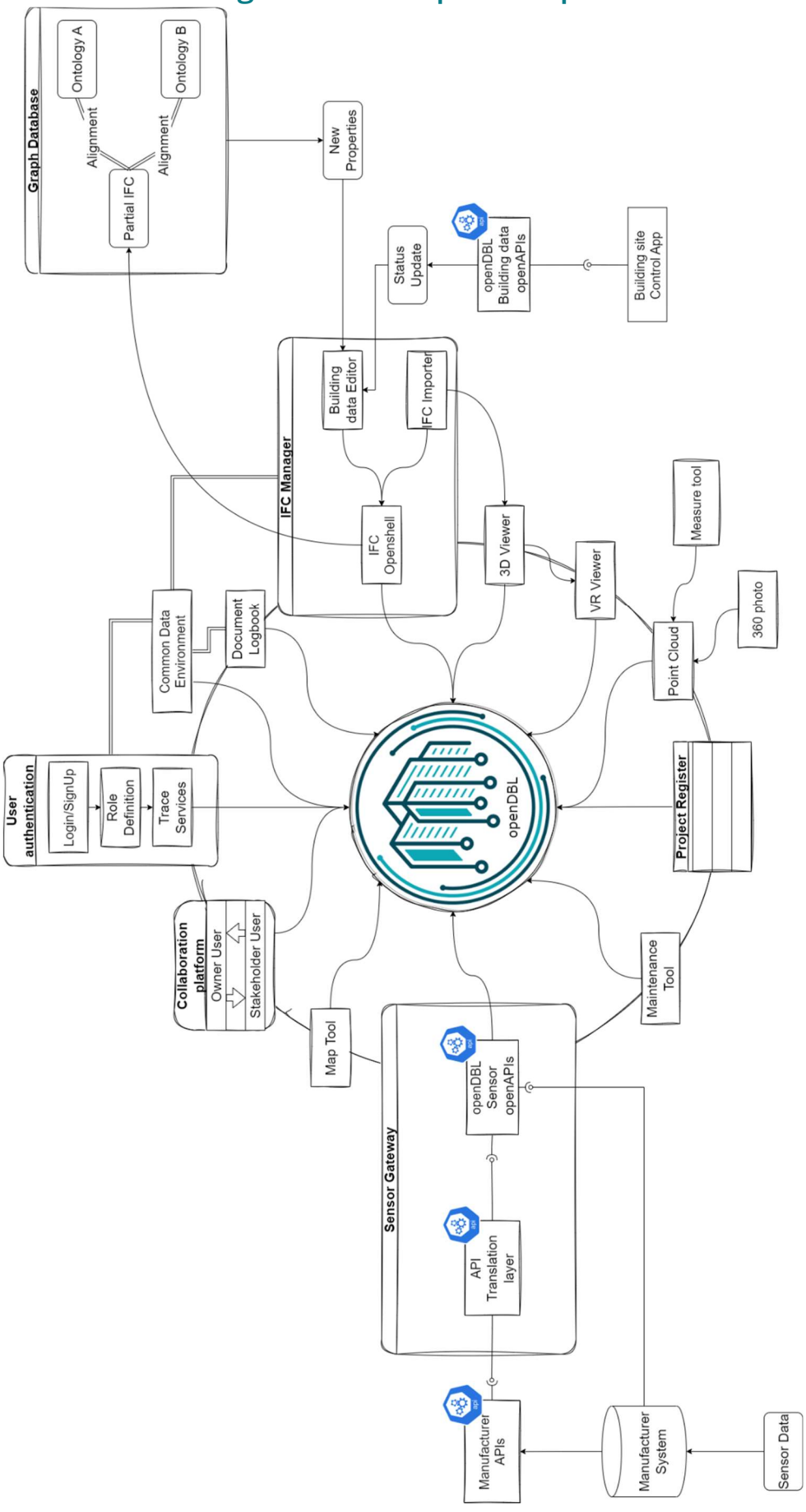
[109] A. Thamm, “Vision Transformer (ViT),” 2022. [Online]. Available: https://www.alexanderthamm.com/de/data-science-glossar/vision-transformer-vit/.

[110] M. AI, “HuBERT: Self-supervised representation learning for speech recognition, generation, and compression,” 2021. [Online]. Available: https://ai.meta.com/blog/hubert-self-supervised-representation-learning-for-speech-recognition-generation-and-compression/.

[111] Y. L. C. Z. Feng, “Enhancing building energy effciency using a random forest model: A hybrid prediction approach.,” 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484721006016.

[112] J. Brownlee, “Strong Learners vs. Weak Learners in Ensemble Learning,” 2022. [Online]. Available: https://machinelearningmastery.com/strong-learners-vs-weak-learners-for-ensemble-learning/.

Annex I: Overview Diagram of the openDBL platform functionalities



Annex II: An Illustration of Our Solution Concept Tailored to a Specific Use Case Scenario