POLITECNICO DI MILANO

Master in

Building Information Modelling

European Master in
Building Information Modelling

Data flow from BIM to Digital Twins

………………………………………………………………..

Supervisor:

Alberto Pavan (PoliMI)
Jose Oliveira (DiRoots Lda)

Author:

André Malheiro

Co-funded by the
Erasmus+ Programme
of the European Union

a.a. 2019/2020

AUTHORSHIP RIGHTS AND CONDITIONS OF USE OF THE WORK BY THIRD PARTIES

ACKNOWLEDGMENTS

I wish to express my gratitude to José Oliveira, founder of DiRoots, for his support and for giving me the opportunity to develop my research and use case. Also, to all the team in the company that contributed with their knowledge, especially Gopinath and Ayodeji.

I will also extend this gratitude to professor and architect Alberto Pavan, as a tutor from Politecnico di Milano and the BIM A+ board.

Last, to my parents, Alberto Malheiro and Isabel Azevedo, that always made me want more no matter what, to my sister, Sónia Malheiro, that always supports and respects my mad ideas, to my nephew Tomás that made me understand the importance of time, and to my future wife, Mariana Bessa, that never stops encouraging me to look for something new, despite all the hard times and long nights of working.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of ethics and conduct of Politecnico di Milano.

SOMMARIO

I Big Data sono ampiamente utilizzati in diversi settori dell'attività umana, dalla ricerca aziendale alla ricerca scientifica, tuttavia l'industria dell'architettura, dell'ingegneria, della costruzione e dell'esercizio (AECO) non ha ancora raggiunto la stessa capacità di raccolta, gestione e analisi dati (Loyola, 2018). BIM è diventato sinonimo di migliore efficienza, comunicazione e collaborazione nella progettazione, l'costruzione, funzionamento e manutenzione degli edifici. Pertanto, il BIM è un contenitore di dati che raccoglie informazioni durante tutte le fasi di costruzione attraverso il contatto con altre tecnologie, come i sistemi di Informazione Geografica (GIS), le Identificazioni Di Radiofrequenza (RFID), l'Internet of Things (IoT), la Realtà Virtuale (VR) e l'Augment Reality (AR) (Farghaly et al., 2017).

Mentre la gestione del ciclo di vita del prodotto (PLM), tra gli altri, è in grado di comprendere e scatenare il valore aziendale nel mondo manifatturiero, AECO sta ancora lottando per definire il quadro e gli standard giusti che rilasceranno il valore della progettazione basata sui dati. Il contenuto generato durante il progetto è un valore da analizzare e utilizzare per automatizzare il processo decisionale. Per evitare lo spreco di informazioni / dati, il settore dovrebbe concentrarsi sulla definizione di un percorso chiaro per utilizzare i dati per la creazione di ambienti meglio costruiti con esperienza utente e valore aziendale migliorati.

Con il processo Plan-Build-Operare-Integrare (PBO-I) in sostituzione di Plan-Build-Operare-Decommission (PBOD), verrà distribuito più contenuto di dati grazie alla sua integrazione con i gemelli digitali (DT). Ciò potrebbe generare un ecosistema in cui l'integrazione dei dati mostrerà non solo dati statici di progettazione e costruzione, ma anche dati reali di sensori che contribuiranno alla simulazione e all'ottimizzazione di un ambiente costruito / attivo. La combinazione di più DT e il loro posizionamento in un ambiente dati connesso (CDE) aumenterà potenzialmente la capacità di prevedere e adattare la pianificazione in base alle condizioni di progettazione e sito, considerando che la gravità, i raggi di uva, la luminosità e altri parametri avranno un impatto su l'asset automaticamente. Questa diventerà una prospettiva predittiva che influenzerà i progetti futuri e presenti attraverso l'uso di Big Data, Machine Learning (ML), Intelligenza Artificiale (AI), ecc. Nel contesto degli ambienti BIM e DT.

L'attuale lavoro accademico tiene traccia del flusso di dati sin dall'inizio del processo BIM, anche prima dell'inizio della progettazione, e percorre l'intero processo fino a raggiungere la fase operativa in cui è possibile valutare nuovi dati. La scalabilità dei dati si evolverà ulteriormente nella prospettiva della città, che sfrutterà i processi su scala ridotta integrati in un ambiente in cui lo scambio di dati avvantaggia il processo decisionale globale, contribuendo a una risposta migliore e adeguata da parte dell'AECO.

Parole chiave: Big Data, BIM, Digital Twins, Gestione delle informazioni, Internet delle cose (IoT)

ABSTRACT

Big Data is documented to be used widely in several fields of human activity, from business to scientific research, however, Architecture, Engineering, Construction, and Operation (AECO) industry has not yet achieved the same capacity to collect, manage and analyse data (Loyola, 2018). BIM became a synonym of better efficiency, communication, and collaboration in design, construction, and building operation and maintenance. Hence, BIM is a data container that gathers information during all building phases through the contact with other technologies, such as Geographical Information Systems (GIS), Radio Frequency Identifications (RFID), Internet of Things (IoT), Virtual Reality (VR) and Augment Reality (AR) (Farghaly et al., 2017).

While Product Lifecycle Management (PLM), among others, is capable of understanding and unleashing business value, AECO is still struggling to define the right framework and standards that will release the value of data-driven design. The content generated during the project is a value to be analysed and used to automate the decision-making process. To avoid information/data waste, the industry should focus on the definition of a clear path to use data for the creation of better-built environments with enhanced user experience and business value.

With the Plan-Build-Operate-Integrate (PBO-I) process in replacement of Plan-Build-Operate-Decommission (PBOD), more data content will be deployed due to its integration with Digital Twins (DT). This could generate an ecosystem were data integration will not only display static data from design and construction but also real data from sensors that will contribute to the simulation and optimization of a built/asset environment. Combining multiple DTs and placing them in a Connected Data Environment (CDE) will potentially increase the capacity to predict and adapt planning according to the conditions of design and site, considering that gravity, uva rays, luminosity, and others parameters, will take impact on the asset automatically. This will become a predictive perspective that will influence future and present projects through the use of Big Data, Machine Learning (ML), Artificial Intelligence (AI), etc in the context of BIM and DT environments.

The current academic work tracks the data flow from the very beginning of the BIM process, even before the design starts, and walks through the entire process until we reach the operational stage where new data can be assessed. The scalability of data will further evolve to the city's perspective, which will take advantage of smaller-scale processes integrated into an environment where data exchange benefits the global decision making, contributing to a better and suitable answer from the AECO.

Keywords: Big Data, BIM, Digital Twins, Information Management, Internet of Things (IoT)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This academic work starts from the technological rigidity that characterizes the Architecture, Engineering, Construction, and Operation (AECO) industry. BIM as a methodology is the principle of something but it is not believed to be the end in itself. Like in other industries, we should be capable of founding in our projects as architects, engineers, etc, the will to increase performance, profit, and user experience while contributing to society and cities' sustainable growth.

Allen (2016) in his conference for Autodesk University called "The Future of BIM Will Not Be BIM – and it's Coming Faster than You Think", starts a discussion about the future of the industry in the next 3 to 10 years. Generically, what can be extracted is that even Building Information Modeling/Management (BIM) needs to adapt, evolve, and gain maturity through the interconnection with other fields of knowledge. This idea of Allen (2016), is the reflex of the adaptation delay of our industry. Even if we constantly talk about data, or more commonly information, we were not able to properly use it and in the process, we are still losing business value due to the reduced technological maturity of most of the key players, from the design to construction and operation. Even after several years of BIM integration, there is not yet a consistent data utilization despite its availability.

The purpose of this research is to identify the data value and show data wasted without any turnback for any of the players involved in the process. In the end, the reader should be able to understand the potential of the data generated through the building lifecycle and that the full benefit of Big Data, from the physical or digital environment, will promote the evolution of the AECO industry in parallel to other industries.

As for the objectives, this work should identify the data extracted from BIM and digital twins and possibilities to optimize future projects, namely in concern to the user experience.

From what can be told, most of the research found was part of papers or journals isolated or integrated into published books. As for the content, considering the integration between different fields of knowledge (e.g., architecture, BIM, PLM, computer science, etc), it requested parallel reading because some terminology is still not common among all the expertise areas. Bilal, Oyedele, Akinade, *et al.*, (2016), was found to be the authors with more complete information, showing an advanced knowledge on different subjects important for this academic work. Loyola, (2018) worked as a guideline of important topics that were connected to other papers and journals that seem important not only for the purposed work but also for further study.

Regarding the research methods, Walliman states that written data is supported by secondary data that has been gathered and archived effectively by different sources (Walliman, 2010). Secondary data can be put away either in statistical or descriptive format and it is more affordable in contrast with directing essential research (Naoum, 2004). Along these lines, new gathered information will be utilized to contrast with the current secondary data with further demonstrate the reliability and validity of the information. This sort of information can offer huge saving money on assets, is subtle and great information (Saunders et al., 2009), and is extremely point by point and for the most part, allowed to access and use whenever (Olsen, 2011). Cautious determination of secondary data is fundamental as the information utilized will influence the ultimate result of the research.

During the development of the research, the author utilized the secondary data collection strategy to review literature (e.g., diaries, journals, books, articles, and so on.), some of them to enlarge the hypothetical contentions and writing blend. Utilizing this information requires distinguishing proof of the first authors by making the best possible references and suggestions to prevent any potential written falsification. Kumar (2013) focused on that utilizing this secondary data can be temperamental because of the source beginning, distribution date, and author's understanding. Be that as it may, for this academic work, the information gathered through this device will be deliberately chosen through the most proper databases (e.g., ITcon, Emerald, Solar Library, ScienceDirect, Elsevier and so on.) to maintain a strategic distance from complexities, for example, diminished reliability and validity.

As for the results and data analysis, there are two fundamental techniques for data collection– qualitative and quantitative. The primary attributes that separate qualitative and quantitative strategies for data collection are words and numbers individually (Saunders et al., 2009). Information gathered by the qualitative technique is ordinarily elucidating and assembled from review and meeting information, it is a lot harder to examine than quantitative data. Qualitative research is increasingly abstract, and it centers around the importance, assessing points of view or sentiments to acquire confirmation and backing towards a specific article (Naoum, 2004). The nature of qualitative data relies upon the methodological aptitude, affectability, and uprightness of the specialist (Patton, 2005). Qualitative research regularly plans to examine the "why" and "how" issues of research, for the most part to assemble a top to bottom comprehension of human conduct and the reasons that administer such conduct (Kothari, 2009).

On the other hand, information gathered using the quantitative strategy (e.g., through questionnaires utilizing 'closed' questions) empowers 'categorizing data' (Saunders et al., 2009; Olsen, 2011) and can be utilized to deliver insights or graphs (Gray, 2004; Saunders et al., 2009). Further upheld by Naoum (2004) quantitative data collection is not conceptual and is hard and solid as the information can be estimated, it is unmistakable, countable, and ready to be broken down with measurable methodology. This sort of information will be helpful for the researchers to gathering or arrange and subsequently investigate. Analysis tools will regularly be utilized to streamline this procedure and make visual translations, for example, outlines or charts which help to expand the readers' understanding.

Due to the social and health environment in which this academic work will be developed, with clear limitations of traveling and assessing information near the industry representatives, the research will be based on a qualitative research method.

Being said, after the introduction chapter, the second chapter will inform the reader about the types of information currently identified as being part of the BIM environment. In the first part of the first chapter, structured and unstructured data will be presented with examples and according to the common definitions. This will be the first approximation to other fields of knowledge. The second part of this chapter will highlight the data/information produced by each project stage from RIBA 2020. The overall objective will be to identify the vast amount of information generated in different formats during all the project stages and how they stand in the relation between the information produced and the information needed.

The third chapter will add a new layer of complexity, introducing the concept of a Common Data Environment (CDE). The first objective is to explain to the reader that BIM data is not all about single models but also multiple models from different disciplines and other documents with different formats. The reader will be later introduced to three case studies that were explored during the contact with one of the biggest pharmaceutical companies in the world. The challenge was to propose an integration for data from different sources and disciplines. This chapter will cover the integration of PLM and BIM in one single source of information through the use of connectors between software platforms with or without data transfer or the creation of a unique environment where data should be deposited.

The AECO industry generates data throughout the entire lifecycle of a facility and from different sources of information. BIM captures not only all the dimensions of Computer Aided Design (CAD) information generated during the design and construction stages but also all the data created during the communication and collaboration process between the stakeholders (Yu, Liang and Wang, 2020). Each stakeholder, especially the ones that have a direct influence on the construction, generates data silos that are later combined in a Federated Model that includes both As-Designed and As-Built data. In this process the data is merged, exponential increasing the vast amount of information to be processed. Later, during the operation & maintenance stage, a new layer of data is added resulting in a more complex and deep BIM data repository. This accumulation of data pushed the construction into the Big Data age (Bilal et al., 2016).

In the fourth chapter, Big Data will be documented has to be used in several fields of human activity, from business to scientific research, however AECO industry has not yet achieved the same capacity to collect, manage and analyze data as other industries (Loyola, 2018). BIM will be proposed as a data container that gathers information during all building phases through contact with other technologies (Farghaly et al., 2017). The reader will be introduced to data attributes that characterize information and consequently will learn the general principles of Big Data Engineering (BDE) and Big Data Analytics (BDA). The end of the chapter, will introduce Machine Learning (ML) technology and the relation with Artificial Intelligence (AI). Due to the short time available for the current research, this subject will be presented in an intelligible way leaving space for further development as this could be one of the technologies with a bigger impact in the AECO industry like it is already happening in other industries.

The fifth chapter will introduce Digital Twins (DT) value. The reader will benefit from the recent literature review concerning not only the potential of BIM and DT in the context of industry 4.0 but also the adoption of these technologies in the CapEx and Opex. UK will be represented through the Gemini principles that will evidence the guiding values of information management across the built environment.

The recent literature from buildingSmart International (bSI) will support the idea of a DT ecosystem, showing, the potential relation between DT and Plan-Build-Operate-Integrate (PBO-I) in replacement of Plan-Build-Operate-Decommission (PBOD). As normal, bSI will also provide insight about future standards for data models, for data management and integration, and concernings about data security and privacy. The bSI ecosystem will identify the possible value of multiple assets DT integration. Complementary, the concept of Connected Data Environment (CDE) and its potential will be presented as possible uses in design and construction future works.

The last part of the chapter will propose the concept of scalability factor associated with DT CDE to promote bigger digital models capable to represent de real-time city-data, also known as “smart cities”.

The last and sixth chapter will present a use case where BIM and DT will be combining IoT and data sensors. The purpose is to stimulate and test the integration of BIM data and DT data into the same environment. The simplicity of the process should work as a principle to showcase what can be achieved through the use of real-time data.

The investigation has been developed in partnership with DiRoots, a company with extensive knowledge of BIM processes and digital solutions. The implementation of processes and technologies side by side with PLM based industry, namely the pharmaceutical/manufacturing, has exposed different maturity levels between AECO and other industries but also revealed similar basic and structure principles for both methodologies. The prototype and testing was the result of the combined effort from DiRoots team and resources and my will to break boundaries between fields of knowledge.

2. BIM – INFORMATION AND TECHNOLOGY

This chapter aims to provide some key knowledge about BIM methodology basic concepts and prepare the reader to the possibilities that outcome from the control of the information needed and the information unknowingly created as part of the process.

For context comprehension, first is necessary to mention the BIM model description as specified on BIM dictionary, "BIM model is an object-rich, data-driven 3D digital model generated by a project participant using a BIM software tool". Second, the model should be considered in its abstract form as an intention to unconstrain the concept from disciplinary relations and be able to highlight the importance and presence of geometric and non-geometric information. As for the asset term, it should be understood as the object used as the link between the model and an "external database to operate and maintain a facility or a portfolio of facilities" (BIM Dictionary, N.D.)

According to ISO 19650, Building Information Modeling (BIM) is the value obtained by improving specifications and defining the information needed at the design, construction, operation, and maintenance stages of buildings and infrastructure. It is also important to mention that appropriate technologies are an important part of adequate information delivery (UK BIM Alliance, 2019a). Concepts like generating, storing, managing, exchanging, and sharing, are also well spread across the industry but is when they are combined with the management of information in an "interoperable and reusable" that they form a more clear concept (Vanlande, Nicolle and Cruz, 2008).

The misunderstanding of the real value of information may lead to a construction of fragmented artifacts that serve no other purpose than a commercial illustration of what was pretended. Rather than focusing on scope with defined deliverables, the information model could, due the access to correct and "powerful" tools, summarize an intention not fully achieved. The "powerful" under this context can only mean that it is easy to create an adaptive component (family/object type available on the authoring tools) that represents a geometry capable of graphical support an intention. Nonetheless, to create real value it is important that the designer does not sculpt objects without understanding the purpose and the characteristics that will result in added value for the client and the project, and from the client's side, the specifications and requirements that will assure, for instance, the capacity to operate and maintain a facility must be defined.

Achieving a successful BIM adoption with the desire outcomes hinges on not only the methods, processes, schedules, and protocols that are used to control and verify the project development but most important on the clearness of the project/asset by the client or asset proprietor (UK BIM Alliance, 2019b). If the quantity and quality of information answer the project/asset needs and it is possible to transmit between the supply chain branches, the consequent reduction of information waste may result in more efficient use of the model all along the project lifecycle.

When compared to the traditional Computer-Aided Design (CAD), BIM exponentially surpasses the 2D CAD representation by line, polygons, and surface. Furthermore, BIM potential as a repository for different types of information (Lu and Lu, 2018), boosts the application and benefit from the utilization of BIM through buildings and infrastructures lifecycle. The model is the result of components combination, with geometric and non-geometric information, combined in a structured interrelation system. The complexity of the model depends on components information, and that is “expandable” and contain information like “carbon emission figures, thermal properties, and associated construction waste” (Lu and Lu, 2018), besides what is considered to be “common” model properties.

Nevertheless, delivering just the right amount of information may not be possible. Even though information considered “As-Needed” may be requested as a deliverable, the outcome may not be fully beneficial due not only to the lack of real consciousness of what is the information value, but also, to the lost or unknown potential of parts of the information created that may be locked down with the handover of the project. As a matter of fact, the vast amount of information inside BIM may difficult the access to important information that lays under different coatings of information, only reaching its potential if combined and structured. Due to the complexity of the information being generated during digital models’ creation and exchange, the dispersion of information may lead not only to the delivery of inconsistent and incomplete information, but also to the occlusion of important information.

2.1. Information Model

The establishment of the information importance, lead to the need for clarification regarding the terms information and information model, and also, what are the sources of information that contribute to the stratification and maturity of an “information model” according to industry standards. For that purpose, the terms are defined according to the UK BIM Alliance and to ISO 19650 Part 1: Concepts and Principles definition.

“Information” as a “reinterpretable representation of data in a formalized manner suitable for communication, interpretation or processing”, this information can be processed by humans or automatic means (ISO 19650-1, 2018). Even though it is not mentioned, it is understood that automatic means can be referred to as the use of coding languages to query and filter data or even to machine learning/artificial intelligence processes. The mention of the concept “data” for the first time is important because it shows a semantic approach to other fields of knowledge like computer science. However, it is not clear why the ISO suggests the replacement of the concept of “data” by “information”, and on the description of the word “information” by “data” as stated on point 3.1.4 on the IEC 82045-1:2001 for Document Management (Internation Organization for Standardization, 2001).

The concept of the “information model” is then defined as “(…) not just a single or federated geometrical model but a collection of information containers (…)” (UK BIM Alliance, 2019b). The addition of new layers of information that transcends the geometric value of a federated model may suggest an increasing complexity through the introduction of different substrates of data.

“Information container” is, therefore, the last concept to be defined on the ISO 19650 part-1 definition: “named persistent set of information retrievable from within a file, system or application storage hierarchy”. The ISO 19650 part-1 also mentions that structured information containers include geometrical models, schedules, and databases, while unstructured include documentation, video clips, and sound recordings (ISO 19650-1, 2018).

The information model is at the end, the combination of multiple information containers, which may suppress different types and file formats. The complexity of the information model will be intrinsically attached to the substance of the data inside the containers.

For the purpose of semantic coherency with the other fields of knowledge, “information” will be from now on mentioned as “data”, as it was not clear the reason why ISO 19650 replaced the concepts defined on the IEC 82045-1:2001 Document Management - Part 1: Principles and methods. The concept of “data” is broadly used, and it is in line with computer science’s own concepts and terms used by most of the research articles consulted for this dissertation. Nevertheless, both “data” and “information” are used to define each other and when referred should be globally considered as the same concepts.

2.1.1. Structured and Unstructured Data

Structured data is commonly used around the world by many companies like Amazon and Google. The first uses specific data templates to define the properties of a certain product in a complete and standardized way, allowing buyers to navigate, compare products, and access information regarding the product fulfillment (UK BIM Alliance, 2018). The second uses structured data to recognize the content of a certain page and, as mentioned on the google developers page, to collect information about the web and the world in general (Gooogle, N.D.).

In the AECO as shown in table 1, what is called structured data can be displayed as geometric models generated through algorithms and schemas (e.g. Industry Foundation Classes – IFC); schedules like Quantity Takeoff - QTO or Bill of Materials - BoM accessed through Microsoft Excel Spreadsheet - XLSX or Portable Document Format – PDF; and, databases like Codebook - a database application used to capture and manage requirements (e.g. University of Montreal Health Centre project by Cannon Design and NEUF architect(e)s), NBS Chorus – an online specification platform for construction, or, BIMObject – a platform that allows users to search and select materials and products.

Table 1 – Structured Data

	Geometrical Models	**Schedules**	**Databases**
Examples	Architectural Structure MEP	QTO(Quantity Takeoff) BoM (Bill of Materials)	Codebook NBS Chorus BIMObject
Formats	IFC, RVT, NWC, DGN, PLN, etc	XLS, PDF, TXT, etc	Cloud-based, XML

The unstructured data, according to ISO 1950, is shown in table 2. Normally this type of data includes text and multimedia files, and its structure does not follow a pre-defined data schema which means it is not easily searchable as structured data (Taylor C., 2018). The most common file examples include virtual video and audio files but also email and mobile data like text messages and location.

Table 2 – Unstructured Data

	Documentation	**Video Clips**	**Sound Recordings**
Examples	Exchange Information Requirements (EIR) BIM Execution Plan (BEP) Asset Information Requirements (AIR)	Virtual simulations Meetings Records	Meetings Records
Formats	Docx, PDF, etc	MPG, AVI, WMV, etc,	WAV, MP3, etc

The batch of files that are shown under documentation are directly or indirectly related to what information needs to be mentioned at the beginning of this chapter. Therefore, it is important to briefly explain the content of the documents that compose the batch of information requirements' container, according to ISO 19650, as shown in figure 1.

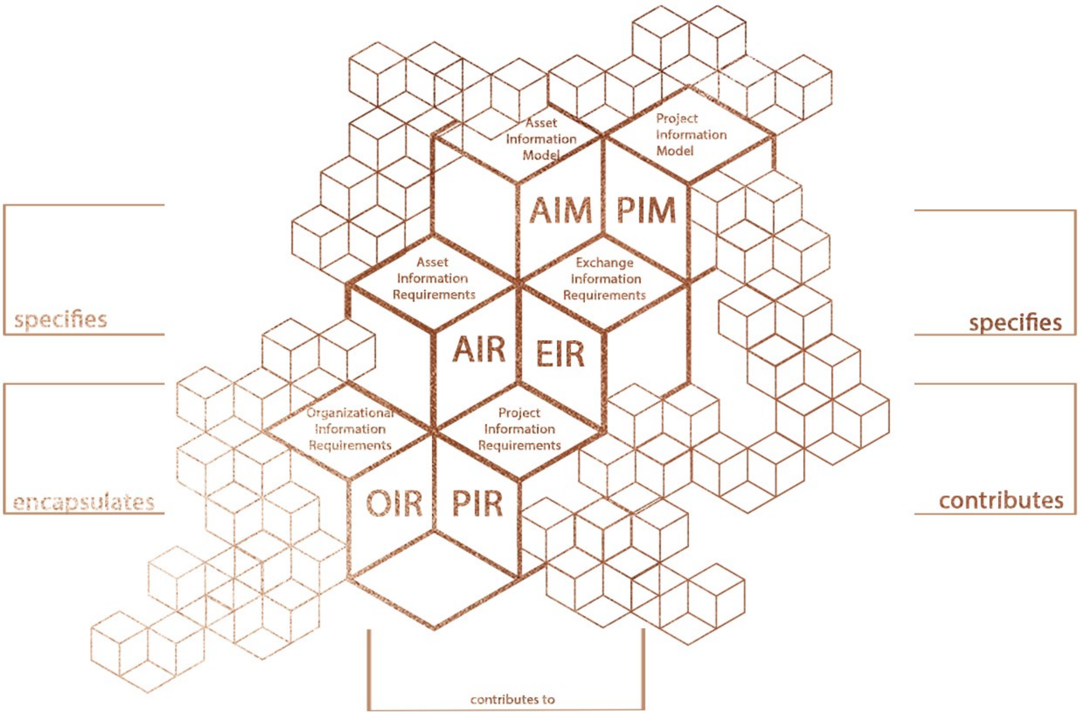


Figure 1 – Relation between information requirements

The Organization Information Requirements (OIR) presents the information needed to meet the high-level objectives of the information receiver (also known as the appointing party). These documents may refer to the asset management or business operation strategy and are the first generic definition of how to deliver the Asset Information Model (AIM). Deeper detail is achieved with the Asset Information Requirements (AIR) when defining the "managerial, commercial, and technical aspects of producing asset information". The document complements the information from OIR and adds "information standards and production methods and procedures to be implemented by the delivery team"(ISO 19650-1, 2018). The data created with both documents will define the AIM specifications information and could be used to check compliance and isolate the asset parameters that are important for the data receiver.

The Project Information Requirements (PIR) expresses the necessary information to answer the top objectives regarding a specific project or asset, as it is referred both on the project management process and asset management process (ISO 19650-1, 2018). Typically, the client adopts a generic version of this document to be used in multiple projects with the possibility of adding amendments.

The Exchange Information Requirements (EIR), defines the technical aspects that should increase the information needed to answer the PIR and should be aligned with key project milestones associated with the completion of part or complete project stages. Like the AIR, the EIR defines a set of "managerial, commercial and technical aspects" but in this case related to the project information instead of asset information (ISO 19650-1, 2018).

As shown in figure 1, the remaining documents are related to the AIM and the Project Information Model (PIM) deliverables resulting from the AIR and EIR specifications, respectively. The AIM can contain 'equipment register, cumulative maintenance costs, records of installation and maintenance dates, property ownership detail', while the PIM can contain details of "project geometry, location of equipment, performance requirements during project design, method of construction, scheduling, costing and details of installed systems, components, and equipment, including maintenance requirements, during project construction" (ISO 19650-1, 2018). The deliverables are, therefore, the consequence of what is defined on the documentation mentioned above, containing fundamental data that will determine the final state of the data at the handover stage. To summarize, figure 2, shows how the different information requirement documents succeed each other.

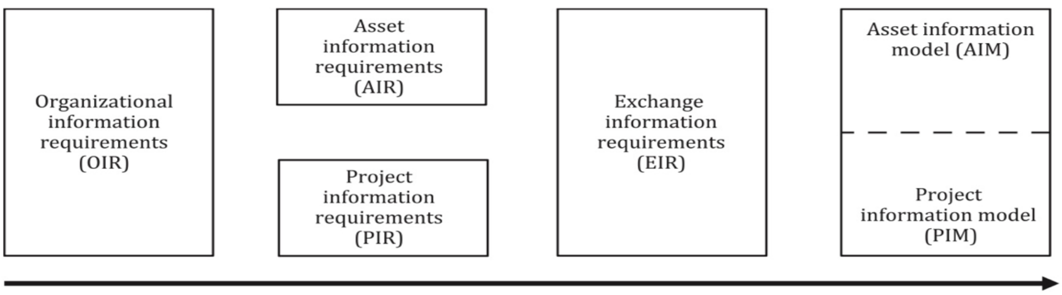


Figure 2 – Simplified illustration of the progression of information requirements (Source: ISO/FDIS 19650-3:2020

Additional documents related to the Construction Project Information Exchange (CPIX) are responsible for adding new layers of information that will contribute to the increase of variety and volume of the project data.

The next document's descriptions are according to ISO 19650-2 (2018)[1]:

The **Project Implementation Plan[2]** is used by the appointing party to assess the lead appointed party/supplier competencies regarding capacity to produce and exchange quality data, project experience, training to create and deliver the project/asset, and IT capability and competence.

The **Task Information Delivery Plan (TIDP)**, which "set out each team's responsibility for delivering information" (Richards, 2013), identifies and assigns responsibilities for the information that each lead appointed party/supplier delivers. In this document, the roles and responsibilities of the team members shall be defined, such as data ownership, responsibility, and authority (Richards, 2013).

The **Master Information Delivery Plan (MIDP)** organizes the roles and responsibilities set in TIDPs alongside with the construction program, defining the time for the project's information preparation and what protocols and procedures should be incorporated (Richards, 2013). It follows the contract award and it shall be used by the project delivery manager (PDM) to initiate the project and to define the list of information deliverables (e.g. models, drawings, specifications, equipment schedules, etc) managed by change control process (Richards, 2013).

The **BIM Execution Plan (BEP)**, first submitted as a pre-contract addresses the EIR subjects and validates the "approach, capability, capacity and competence to meet the EIR" (Richards, 2013). As a post-contract document, it explains the lead appointed party's methodology to deliver the project/asset using BIM, and completes EIR information with the roles, responsibilities, authorities, standards, methods, procedures, major project milestones aligned with the project program, survey strategy, etc.

In a high-level analysis, according to British Standards Institution - BSI (2019) the information delivery process starts with the "Assessment and need" through the EIR, previously explained and referred in figure 1. As for the procurement phase the BEP, pre and post-contract, is the document identifying the required information, while the MIDP considers the information after the contract is awarded. This is only an overview of the succession of documentation and despite not being directly related to the scope of this academic work, it presents the value of data that is generated even before the beginning of the project. It is important to mention that the multiple documents that compose the BIM process may include or be included in another document.

From what was shown, the BIM information delivery results from the combination of multiple structured or unstructured data sources that are deposited inside information containers. The enrichment of the BIM environment is, therefore, the result of the variety and volume of information that a single project

[1] Despite following the ISO 19650 standards, some references may reflect the knowledge within PAS 1192-2:2013. However, only the concepts maintained on the ISO 19650-2: 2018 were considered.

[2] According to British Standards Institution - BSI (2019), like other terms, there is no collective term for project implementation plan on ISO 19650-2: 2018. However, considering the importance of listing important documentation for the project the term is considered important for the scope of this academic work.

can generate. Moreover, the project lifecycle development is proportional to the data maturity that composes and shapes the deliverables.

2.2. Level of (BIM) data definition according to project stages

The information created and defined before the project's beginning already represents a vast amount of data to be considered in a more advanced phase of this academic work. Once the requirements and deliverables are defined and assigned, the project starts. The model's development follows the objectives defined for each key milestone and the owner/author is responsible for adding the geometric and non-geometric information needed to answer each project stage's requirement. In this process, the project/asset matures and increases its data complexity, variety, and volume, meaning that more data is generated, but some is wasted and left on previous stages, or it is purely neglected.

The proposed analysis of data generated by project stage is thus important because it will help to identify the information needed, and more importantly the one that can feed the algorithms that help establish patterns. Furthermore, putting the data generated in perspective with the project stages intends to prove that the AECO industry can be associated with the concept of Big Data.

The EIR is used to define the key milestones in which data is gathered and shared with the client or project/asset owner. The building information model data submitted at specific project stages are called data drops, and the data content should be enough to promote the next steps on project/asset development (Design Buildings, 2020).

The data drops' content is adapted to the project stages' sequence and may include IFC models and native project files, Construction Operations Building information exchange (COBie) spreadsheets, and QTO schedules, and, reports. To define information requirements to be used at different project stages, the ISO 19650-2 (2018)[1] is a very important documents, as it deals with the construction (CapEx) phase and the requirements to achieve BIM level two maturity, while PAS 1192-3:2014 deals with operational phase (OpEx) oriented to the use and maintenance of Asset Information Model for Facilities Management (NBS, N.D. / Designing Buildings, 2019). The "level of model detail and model information is part of this document and can be used to understand how data is distributed and developed during seven different stages.

2.2.1. Data by Project Stages

Due to the maturity of the Royal Institute of British Architects' (RIBA) plan of work, and its broader influence among the worldwide BIM community, this subchapter is substantiated by the RIBA project stages. When compared to other similar plans, like Architects' Council of Europe (ACE) "Scope of Services", the content of "Plan of Work 2020" from RIBA presents itself as a more intelligible document with a 2020 version release. Therefore, RIBA project stages are more suitable to identify the data generated at the following data drops:

[1] Replaces PAS1192-2:2013.

Stage 1 (Preparation and Briefing) feasibility studies might need architectural skills, engineers, and/or surveyors to assess key project risks. They can come up with a project programme and an execution plan. The project brief preparation includes project and sustainability outcomes, quality aspirations, and spatial requirements. At this stage the project budget is agreed, the site information is gathered and deployed, the project programme and project execution plan are prepared (RIBA, 2020). Stage 0 is by default responsible for providing data with less maturity, but that will work as the foundation for the project and shall be considered during the entire project lifecycle, namely for budget control purposes.

Stage 1 (Preparation and Briefing) feasibility studies might need architectural skills, engineers, and/or surveyors to assess key project risks. They can come up with a project programme and an execution plan. The project brief preparation includes project and sustainability outcomes, quality aspirations, and spatial requirements. At this stage the project budget is agreed, the site information is gathered and deployed, the project programme and project execution plan are prepared (RIBA, 2020). This stage represents the information's definition and preparation even before any design process has begun, however, there are already a set of documents that start to constraint and define the project's future steps, namely health and safety pre-construction information. These documents are already part of the global strategy and should be taken into consideration in further stages.

Stage 2 (**Concept Design**) underlies the architectural conceptual design approved by the appointing party/client following the necessities defined in the initial project brief. This stage integrates the strategic engineering requirements, cost plan, project strategies, and outlines specifications. At this maturity level, the appointing party/client should expect to be granted with the pre-application planning advice and building regulations compliance (RIBA, 2020). It is normal to return to the preparation and briefing during this stage as it ought to be refreshed and delivered as the final project brief toward the finish of Stage 2.

Stage 3 (**Spatial Coordination**) is generally about testing and approving the architectural conceptual design. Design studies are undertaken to clarify decisions made in stage 2 and to create more layers of detailed design, engineering analysis, and Cost Exercises to test the Architectural Concept resulting in a spatially coordinated design that must be aligned to updated Cost Plan, Project strategies and outline specification. At this stage, the change control procedures are initiated and the preparation of the design programme takes place, whilst the design proposal is reviewed and aligned with the building regulations for the preparation and submission of the planning application (RIBA, 2020).

Stage 4 (**Technical Design**) involves the preparation and coordination of design specialists and all information needed in the manufacturing and construction of a building. Stages' 4 and 5 information may overlap in most of the projects. The core processes and documents at this stage are the building regulations, the planning conditions, and the construction phase plan (RIBA, 2020). The architectural, building services, and structural engineering's design are furthermore enhanced to give a specialized value to the project, and the design work of specialized subcontractors is created and closed.

Stage 5 (**Manufacturing and Construction**) is a non-design phase except when responding to site queries. It should be realized through the finish of the manufacturing and construction process hence the completeness of the commissioning process. The core tasks are related to the assembly and production of the building systems according to the construction programme, ensuring the construction quality, while solving the site queries and preparing the building manual for delivery. This stage combines all the information generated that has a direct or indirect impact on the production of the building systems. Therefore, even if not mentioned, all the requirements and specifications are part of the asset/building to be delivered (RIBA, 2020).

Stage 6 (**Handover**) is the data drop stage where the building is "delivered" and culminates with the end of the construction team and those liable for administrating and finishing off the building contract work. During handover, the building is verified to confirm that it is aligned with the building contract and its delivery should match the plan. The project team in this stage holds priority handing over the whole project in line with project requirements' to be fulfilled. Other services may be required and are to be addressed by project-specific schedules of services, which ought to be lined up with the procurement and handover strategies. Additionally, the project performance is reviewed as part of the maintenance tasks that need to be completed and include a post-occupancy evaluation (RIBA, 2020).

Stage 7 (**Use**) is where the building is "used, operated, and maintained efficiently" (RIBA, 2020). The design and construction team will be keen on getting continuous input, to assist with how they may improve the performance of future structures. Most of the tasks will have no stage 7 obligations to fulfill. At this stage, facility and asset management processes are implemented taking into consideration the post-occupancy evaluation aligned with the building's performance and the verification of project outcomes that includes sustainability outcomes. In the future, a Digital Twin might be used to optimize the operation and maintenance of the building and to compare predicted performance with actual performance (RIBA, 2020).

While the finish and handover of a building or asset culminates in the user's experience at stage 7, all the information generated since stage 0 structures and constraints the outcome. The information needed and wasted generated in a project that goes through all stages could help define future buildable assets with more informed and complete processes.

2.2.2. Level of Information Need

It is important to mention that in comparison to PAS 1192 terminology, ISO 19650 suggests that the level of definition/development, that is constituted by the model graphical information (Level of Detail, LOD) and also the non-graphical (Level of Information, LOI), should be replaced by the global term level of information needed (British Standards Institution - BSI, 2019). Therefore, the use of the term Level of Information Needed shall be adopted instead others LoX terminology.

The intent is to show that the information/data generated should be according to the needs of the project/asset and its different stages and the incapacity to control the data journey may result in information/data waste.

According to ISO 19650-1 (2018), the level of information needed defines the "extent and granularity of information" being its main purpose to prevent its waste or, as stated, avoid the delivery of "too much information". The degree of information needed is an expansive concept that communicates the quality, quantity, and granularity of different deliverables. The determination of the information needed along the project/asset lifecycle should be defined within the OIR, PIR, AIR, or EIR (ISO 19650-1, 2018).

The fundamental step required on the definition of the level of information needed is to understand what the minimum quantity of information compulsory is to answer the main requirements and to enable appointed parties' actions. Normally, the information that travels between different appointed parties contains data that transcends the fundamental project/asset's needs, meaning that not all the content will be used or is important, and consequently will be wasted.

There are various ways to communicate the degree of information needed, including the richness of geometrical details and the wastefulness of datasets. At each stage, within the design and construction phase, the objects that compile the model should contain geometrical and non-geometrical information. For that purpose, an Information Delivery Plan (IDM), should state the flow of information across the project stages mentioned previously and the data exchange (Arayici *et al.*, 2017).

The IDM defines the information needed for the project and the information enrichment according to each project stage's requirements. Establishing the deliverables and its attributes are fundamental to understand what type of information container (e.g., a model) will be used, to expose parts of information, and also which format will be used to exchange data (Arayici *et al.*, 2017). Relying on native formats may result in a bigger amount of information traveling between stages, while using IFC schemas and its subsets, like COBie, can result, due to the Model View Definition (MVD), in the reduction of the information waste and more selective and conscient data exchange.

The volume of information on a model can be defined in different levels, and varies from country to country, as shown in figure 3.

Source	LoX system	Whole Model	Model Element	Geometric data/info	Non-Geometric data/info
BIPS 2007	Information Levels	x	x	x	x
CRC 2009	Object Data Levels/Level of Detail		x	x	x
Department of VA 2010	Level of Development (LoD/LOD)		x	x	x
Vico Software 2011	Level of Detail (LOD)	x	x	x	x
NATSPEC 2011	Level of Development (LOD)		x	x	x
HKIBIM 2011	Level of Detail		x	x	x
NYC DDC 2012	Model Level of Development/ Level of Development (LOD)	x	x	x	x
	Model Granularuty		x	x	x
PennState University (PSU) 2012	Level of Development (LOD)		x	x	x
USC 2012	Level of Detail (LOD)		x	x	
US Army Corps of Engineers (USACE) 2012	Level of Development (LOD)		x	x	x
	Element Grade/Grade		x	x	x
AIA E203™ 2013	Level of Development (LOD)		x	x	x
BCA 2013	Level of Detail		x	x	x
PAS 1192-2 2013	Level of model Definition		x	x	x
	Level of model Detail (LOD)		x	x	
	Level of model Information (LOI)		x		x

CIC BIM Protocol 2013	Level of Detail (LOD)	x		-	-
BMVBS 2013	Level of Development		x	x	x
BIM 2014	Information Level	x	x	x	x
AEC (CAN) 2014	Level of Development	x	x		
Le Moniteur 2014	Level of Detail/ Level of Development (LOD)		x	x	x
BCPP 2014	Level of Development (LOD)		x	x	x
	Level of detail (LOd)		x	x	
	Level of accuracy (LOa)		x	x	x
	Level of information (LOi)		x		x
	Level of coordination (LOc)	-	-	-	-
CBC 2014	Level of Detail (LOD)	x	x	x	x
BIM Taiwan 2014	Level of Development	x	x	x	x
	Level of Completeness	x	x	x	x
	Level of Detail	x	x	x	x
ABEB-VBA 2015	Level of Development (LOD)		x	x	x
D&R 2015	Level of Development (LOD)	x		x	
BIMForum 2015	Level of Development		x	x	x
	Element Geometry		x	x	
	Associated Attribute Information		x		x
NBS BIM Toolkit 2015	Level of Detail (LOD)		x	x	
	Level of Information (LOI)		x		x
AEC (UK) 2015	Level of Definition		x	x	x
	Level of Information (LOI)		x		x
	Grade/Level of Detail (LOD)		x	x	
SZGWS 2015	LOD	x		x	x
USIBD 2016	Level of Development				
	Level of Accuracy	x	x	x	

Figure 3 – Comparison of the intended coverage of LoX systems (Source: BIM Think Space, 2016)

Initially, the LOD was created to measure geometric and non-geometric data's consistency and then evolved to more geometric focused criteria. LOD can also be used to define the Level of Development or even assume other concepts like "Level of Information" (BIM Think Space, 2016). The levels differ by country, classified from 0 to 6 (e.g., Denmark), from A to E (e.g., Australia CRC), from 100 to 500 (e.g., United States of America), from 1 to 7 (e.g., United Kingdom), or from -100 to 500 (e.g., Germany). For certain levels there is still some lack of clarity about how to accomplish it and may not even be possible considering AECO's industry maturity.

As mentioned by Marzia Bolpagni on BIM Think Space website, at this point, the number of "LoX" increased, which is related to the progression of the information models' definition (Bolpagni, 2016). Even though in 2004 Vico software started working on a standard to facilitate the management of information, the management of BIM content is still not clear. The lack of consistency could be associated with the existence of different definitions and acronyms and its changes across the years and from country to country (Bolpagni and Ciribini, 2016).

The absence of a common definition for data management introduces new levels of information complexity and diversity. Despite the attempt to establish a correspondence between different LoX's, there is still some confusion that raises difficulties when trying to select and structure the information/data. Even Though the geometrical model is important for the BIM process, the non-geometrical information is rather important to accomplish the requirements present on the EIR.

Considering that previously it was done a RIBA project stages' review, for the clarity of the subject, it is proposed a synthetic review of the UK LOD levels.

LOD 1 may not include a geometrical model (also named as graphical model), but if it exists it was probably generated from an existing asset information model. Other information may be related to other buildings or schedules.

LOD 2 may present mass schematic and bidimensional symbology used to represent generic elements to be used as the foundation of the next levels.

LOD 3 the object has a generic representation, specification, and attributes, which grants to the client the basic elements for the decision and selection of the best solution/product.

LOD 4 should represent 3D objects with attached operation, access, maintenance, installation, and replacement spaces' specifications.

LOD 5 replaces the generic object by the manufacturer's objects combining the minimum information from the original object with the manufacturer's information.

LOD 6 represents the as-constructed project/asset and the necessary information for maintenance and operation, commissioning records, health and safety requirements, etc.

This subchapter tries to relate all the information and the model's data under the definition of Level of Information Needed (or level of detail/development as mentioned in the PAS 1192-2). It was important to explore the concept of information needed to understand that extra information does not have a direct relation with a more complete or flawless project/asset. The minimum should permit a more coherent, accurate, and precise version of the deliverables and requirements. In the end, the use of RIBA LOD levels and the generic comparison with other countries only served to correlate the granularity of geometric and non-geometric information to the defined project stages. It is believed that there is a lot more detail to be introduced to this subject, but for this academic work, further discussions won't benefit the defined scope.

3. COMMON DATA ENVIRONMENT - CDE

This chapter will focus on the repository of all BIM geometric and non-geometric data produced by the unique players at the different stages and in the transition of information between states. The goal is to close the global analysis of the data generated and collected in a BIM process.

The Common Data Environment (CDE) is the established source of information where the collection, management, and exchange of each information container involving a unique project/asset, takes place (ISO 19650-1, 2018). Fundamentally, the overall structure provides access to four main stages: Work in Progress (WIP), Shared, Published, and Archived. The information/data inside these areas can be managed, stored, approved, and exchanged with and between stakeholders at defined project stages or data drops (Comiskey *et al.*, 2017).

From the previous chapter it is now clear that the project lifecycle is responsible for generating a variety and volume of data in different formats like models, drawings, and documents. These project/asset information containers are managed inside CDE, which improves the reliability, efficiency, coordination, and consequently the quality of the outcome. Quality compliance is ensured by the definition of workflows, coding standards, metadata, roles, responsibilities, and authorization rights (Comiskey *et al.*, 2017). The control of the main structure and all the information spread along with the different states is fundamental to the audit trail and version control of the information.

There are also concerns about data security that should be addressed when establishing a CDE. The PAS 1192-5: 2015 identifies and proposes the implementation of measures to promote security awareness within the organizations. The security controls are applied through the project/asset lifecycle to ensure safety, authenticity, availability, confidentiality, integrity, possession, resilience, and utility (Designing Buildings, 2019).

3.1. Case Study - data from multiple CDE

While developing the academic work, DiRoots, the company that supported the author, started a project as BIM consultants. That project, for a pharmaceutical company, should align two domains that, at this point and at our best knowledge, were not commonly combined: the PLM and BIM. The processes defined should be applied to any future project, but the current scope was an expansion project of one of the pharmaceutical's pavilions. Due to the disclosure agreement, neither the name of the pharmaceutical company, its country or the PLM company's name will be mentioned in this academic work.

Research shows that PLM is now more related to the automotive and aerospace industries, which means it is more suitable for the manufacturing sector where products are created in a more controlled environment. BIM is globally assumed to be more adequately applied to the AEC(O) industry, namely civil structures and modular building construction (Bertozzi, 2009). Nevertheless, it is important to look at stage three of bim maturity (figure 4). A closer examination reveals that, despite being in stage two, with mixed manual and automated information management processes, the stage three will achieve go through different layers (standardization, technology, information, and business) to implement enhancements that will promote a globalized standardization process, a CDE based information processes, the trigger of Big Data, the federation of information models, and object-based information models. As for the business layer, it will benefit from the digital boost, unblocking new processes ISO 19650-1 (2018).

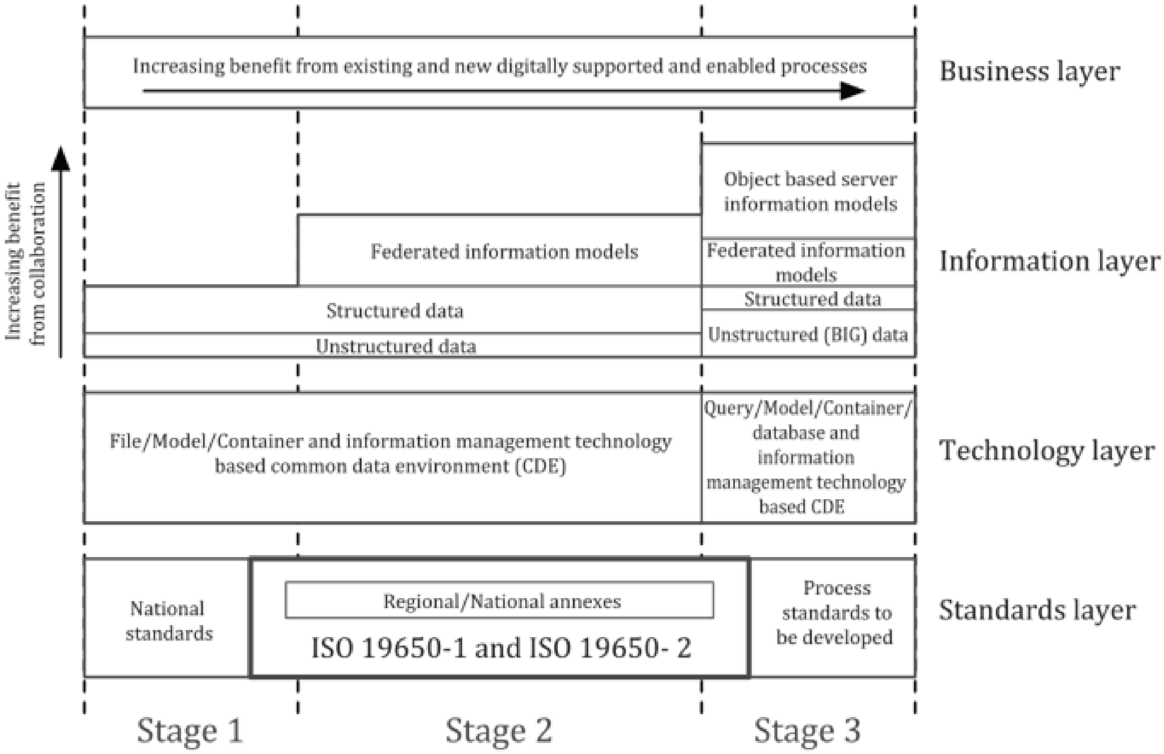


Figure 4 – Analogue and digital information management stages (Source: ISO 19650-1, 2018**)**

The combination of PLM and BIM prompted a group of possible interactions and exposed some differences between the two industries. Table 3 created by Bertozzi (2009), highlights the major characteristics of each industry, to help the reader understand the challenges associated with a process where the information from two domains has to be deposited in an area that has to be accessible by multiple lead appointed suppliers.

Table 3 –Industry characteristics (Bertozzi, 2009)

Sector Type	Firms dimension	Data management and tool	IT adoption	Industry organization	Product type	Supply chain management and industry structure
			BIM			
AEC Industry	Small and Medium	Largely separate application modules, 3D CAD, CAM,	Low and medium	Highly medium fragmented	Complex project, components, and process,	Project-Based

		CAE, 4D-5D BIM		Localized Multidisciplinary and heterogeneous team, Slowness to changes	Individual nature of the project	High variation in project structures and delivery methods Short Term and more isolated relationships with client Lack of process commonality, standardization, and integration
PLM						
Complex Manufacturing Sectors and in detail Automotive and Aerospace Industry, Shipbuilding Industry, AEC Industry	Big, globalized and consolidated	Higher Levels of integration. PLM application modules, PDM, 3D CAD, CAM, CAE. Information modeling architectures, development toolkits, business app	Long experienced PLM and ERP use, but with different levels of adoption	Partially integrated “islands of information” Globalized and consolidated	Complex product	Product-based Long-Standing and collaborative relationships with client Lack of a holistic view of users of information Engineering methods. Support decision- making from whole life cycle perspective [

The fundamental step was to identify the potential similarities between BIM and PLM and try to create the guidelines that would help the companies from different disciplines to work collaboratively. In both cases, the major value was in the data control and information management, its repository, and exchange during the building/product lifecycle (Bertozzi, 2009).

Considering the multidisciplinary of the project, even if in the end a single CDE is available, the data from all CDE’s should be ready for edition. To avoid versioning conflict and data duplication, the client wanted to have access to the single source of information but in a dynamic ecosystem where input from one of the “secondary” CDE’s would activate a gateway to establish access to the data inside the originator CDE. In this scenario, a group of “ghost CDE’s” will be used to generate, share and exchange information according to established requirements, and once the operational phase was reached the main CDE will work as the access point to the deposited information. This would be the client's shared area, however, the communication with other CDE’s will enhance the use of data considering that it won’t remain static after its creation. The final result will be the connection with a bidirectional gateway of multiple CDE’s into a platform that will be used to manage the information workflow as needed.

As was mentioned before, the information needed is even more important when we try to combine multiple domains. The real value is not in the capacity to send and receive information between different CDE’s but rather in the capacity to exchange the trigger data that will promote the development of other parts of the project through sustainable and coherent data exchange. Moreover, not being able to track the correct version of a specific document or model will compromise the entire workflow and the completeness of the building/asset.

Once the data has been identified the important step was to create a flow diagram to grant control over the information in different project stages or other data drops. Additionally, the PLM disciplines were working with the 3DEXPERIENCE platform from Dassault Systémés, which has specific internal procedures that have been taken into consideration during the proposal of the information process.

PLM disciplines use the User’s Requirement Specification (URS) to describe the business needs and are defined early on the validation process. “The URS is generally a planning document, created when a business is planning on acquiring a system and trying to determine specific needs.” (OfniSystems, n.d). The URS includes programme requirements (the workflow), data requirements (the type of information), and lifecycle requirements (how the system will be maintained, and users trained) (OfniSystems, n.d). Putting URS in BIM perspective it includes the type and amount of information that normally relies on OIR, PIM and AIR (figure 5).

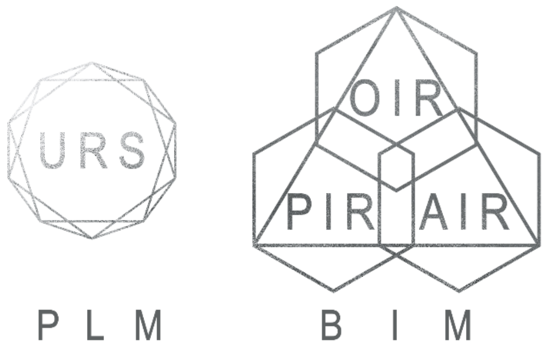


Figure 5 – PLM and BIM equivalent documentation

The URS was the main document used as the project trigger since it was responsible for defining requirements, specifications, and deliverables. The particularity of working in parallel with BIM and PLM was that in the pharmaceutical industry the PLM disciplines have a more consolidated position than those in the BIM side. With years of development in an industry with a more homogenous group of suppliers that are familiar with the technology and traditional workflows, such as change management, the PLM industry adds some new considerations and presents a more mature collaboration and data management process. Nevertheless, despite some considerations about BIM incompleteness, BIM and PLM are being combined to transfer manufacturing qualities into the construction industry (Bertozzi, 2009).

Due to this last reason, the beginning of data management started with the creation of the URS inside the Dassault Systémés platform, in a "module" called ENOVIA. As shown in figure 6, asset creation is triggered by the requirements that will be associated with a specific deliverable which could be related to the construction or the manufacturing side. This process could have been made automatically using URS native document format (e.g., docx) and identifying the chapters that have correspondence to certain deliverable requirements.

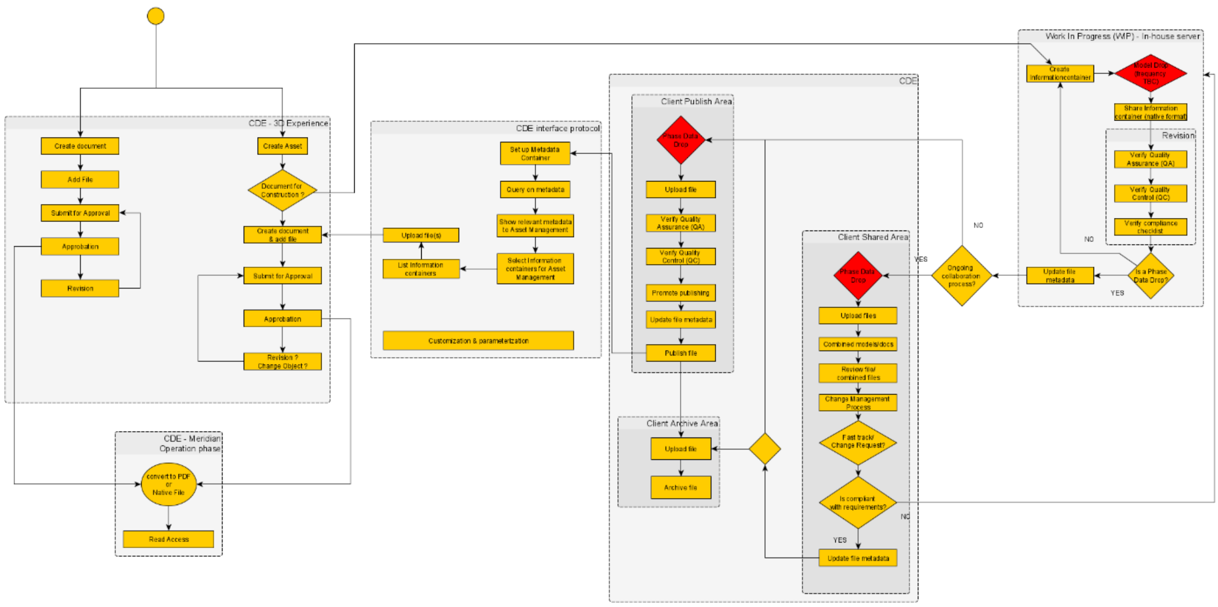


Figure 6 – Flow diagram explaining the information/data flow between multiple CDE's

If the asset needed more information from the BIM disciplines, namely the 3D model, the list would remain incomplete and not sharable until future inputs. Meanwhile, the assignment to the correct lead appointed party/supplier occurs, and, despite the lack of technical information and business support at this stage, communication through different CDE's is granted through an Application Programming Interface (API) automation.

The communication on the CDE for the lead appointed parties/supplier was updated with the requirements, deliverable specifications, and data drops. The work in progress (WIP) state refers to information being developed by the task team. The information container should not be accessible to other task teams (ISO 19650-1, 2018). This state may occur on the in-house server of each lead appointed party/supplier that also may be using different authoring tools. The quality assurance and quality control verification will define the compliance of the requirements and may only be used for internal control to share or publish important information/data. To control the data exchange, information container's naming, including status and revision metadata must be updated according to the naming convention as shown in figure 7.

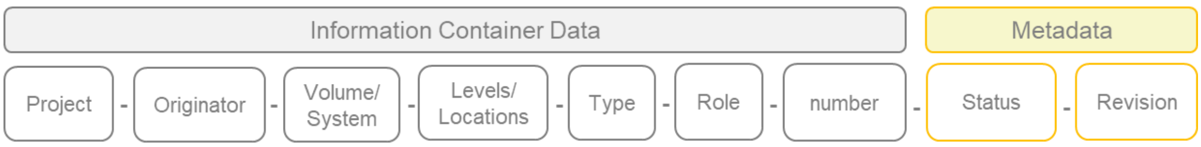


Figure 7 – Information container naming convention (Source: ISO 19650-2, 2018**)**

The need for collaboration may guide the data to its shared state, where the federation model can be created and the change management process is initiated. The change management process may require some changes on the BIM task team and/or on the PLM task team. For that to happen a link between both CDE's is created to identify the impacted elements. At this state, the constructive and collaborative enhancement of the data should occur through the exchange of information between delivery teams (ISO 19650-1, 2018).

After the compliance assessment, the documents, model, or drawings on the shared state are pushed into the published state and the information container is again updated accordingly. The information at this state has been authorized for use. It is used for the PIM at the end of the project and in the AIM for asset operations. Information on the archive can also be used seen as it includes shared and published information containers and its audit trail during the information management process (ISO 19650-1, 2018).

Situations, where the information goes directly from WIP to the Published state, may occur if a piece of information is not relevant for sharing or is not available for any kind of change, for example, an equipment's specification.

The CDE interface protocol represents the moment where the transition of the information container to the PLM CDE happens. The naming is updated to match the asset list and all metadata is organized to be used inside 3DEXPERIENCE. The IFC schema should be managed to establish the data structure supported by the Dassault Systémés (therefore DS) platform. The mapping of IFC attributes should be properly documented to be used in similarity to the 3Dxml file format data tree structure.

IMPARARIA plugin would have been another possibility to import Revit native files to 3DEXPERIENCE, but the possibility was not explored due to limitations identified by the technical teams and due to the costs and time needed.

After traveling from WIP state on the different task teams' in-house server to the shared and/or published state inside the client's CDE, the data should be combined with the asset's requirements and deliverables, which are already inside DS CDE. If some new deliverables or specifications are created, the asset is updated with the new information through a revision or change object process, but if some part of the deliverable is incomplete or does not comply with the requirements it should be sent back to the client's CDE and from there to the task team's WIP state.

Once everything is declared as approved it moves on to Meridian's Engineering Document Management System (EDMS)/CDE from Accruent which is already used by the client for the operation phase. This means that all the documents needed for operational purposes will be present and can be accessed by the assigned roles, according to defined responsibilities. The Meridian platform grants to the client the possibility to "maintain engineering data integrity" allowing navigation and maintenance of the relationship between documents, CAD models and assets; "ensure successful control and change management", maintaining the master data up-to-date; "reduces the handover cost" while managing the exchange of engineering documentation with external contractors and identifying data inconsistencies; and, "breakdown information silos between departments" through the creation of a single source of truth for all engineering information (Accruent, 2020). This last property, the capacity of dealing with multiple CAD formats, and the fact that it was already in use by the client were the conditions that defined Meridian as the best suitable option for the client CDE.

If the documentation is not BIM related, the creation of the document and all the associated processes would have taken place inside the 3DEXPERIENCE CDE and then pushed to the EDMS/CDE.

This process results from a very complex data exchange between different platforms that could be managed by multiple delivery teams. It requires automation protocols using algorithms and API's management to be able to glue the different sources of information into one last document. However, to ensure that all the documents, models, drawings, among others can be changed, the relational link has to be bidirectional and the information remains in multiple CDE's at the same time.

3.2. Case Study - data in a single CDE's

The need for a less complex information management system, where the data duplication could be avoided, resulted in the schema on figure 8. The process starts in the same way as previously exposed, generically called PLM CDE. However, the focus will be on the relation between delivery teams, not the software itself. Hence, the WIP state will be the area where all the delivery teams create and edit the data. The existence of some platforms that would allow real-time collaboration between task teams, the goal was to avoid software constraints. Therefore, and following the open BIM standards, the IFC schema will be used for collaboration and coordination purposes. The MVD will be used to exchange the information needed with other delivery teams and information manager (buildingSmart, N.D.).

When the information containers, which are in the shared state inside the client's CDE, need to be modified, a version is copied to the archive state and the information manager does the download of the information to proceed with the 3D coordination, 4D simulation, or any other process required. After the information container is uploaded directly to the same shared state, from where it has been downloaded. This means that the different processes may take place on the WIP state but the output documents, model, etc, that are important to the project lifecycle, will be placed and named accordingly inside the shared state.

If there is no need for revision the information container is published and the relevant data to the PLM delivery team is accessed. In this step the asset information is combined with the PLM data and is then uploaded to the published state as a new information container. At this point the information is ready to be accessed and used in the operation & maintenance process and also by the facility's management.

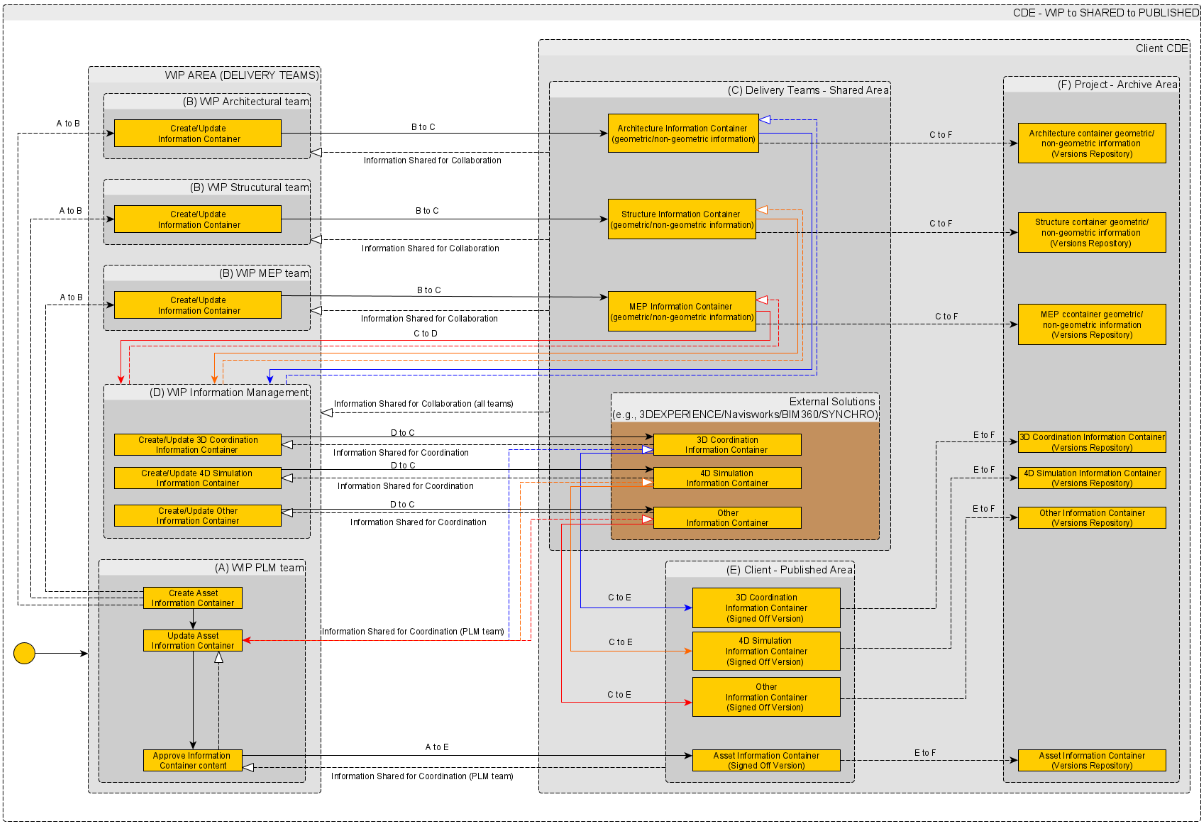


Figure 8– Flow diagram explaining the information/data flow single CDE

The proposed workflow simplifies the building/asset project, avoiding the duplication of information, and creating a single source of information to be accessed by different delivery teams according to responsibilities and security protocols.

The possibilities are even bigger if you align this proposal with the Open CDE standards from buildingSmart International (bSI). On their 25th of May 2020 webinar, they focused on the integration between different application vendors and cloud vendors. Instead of using the API of different cloud vendors, in a bespoke solution, bSI proposes a standardized API that will promote the interaction between application vendors and cloud vendors. Currently, bSI is working on the standards that will feed the Open CDE and the integration with IFC and BCF. The group understands the potential of the BCF API to accomplish OpenCDE API and they are proposing to use it in the interactions between author and cloud vendors while avoiding extra time and cost downloading and uploading files. As mentioned in the webinar video, a commercial version of this process is not yet available, constraining further exploration on the subject.

3.3. Future exploration - Data from interconnected CDE's

Even though the possibility of interconnecting multiple CDE's was not explored, the assessment of the two previous workflows raized some possibilities regarding a future gateway between multiple CDE's. The information will remain on the CDE of the multiple delivery teams and will be accessed through the API or connectors of the different third party companies.

The project for the pharmaceutical company involved multiple domains from both PLM and BIM, as well as the Food and Drug Administration (FDA) requirements and certified software, among others. Altogether the volume, variety, and velocity of data will easily increase day by day and at a certain point the audit trail will be too complex and mistakes due to data duplication, accessing the wrong version, or corrupted data, may occur.

3DEXPERIENCE from Dassault Systémés and Meridian from Accruent already showed some advances in dealing with data from other third-party sources, for instance, 3DEXPERIENCE, as mentioned before, uses Impararia to import Revit files to its environment, being able to structure data to be accessed and managed (Dassault Systemes, N.D.), and Meridian is capable to integrate Revit and Navisworks formats through its API and even using Autodesk Forge to explore the data visualization (Meridian Enterprise, N.D.).

Considering the know-how that the mentioned third-parties' software developers, and others that may not have been referred, already have, a future exploration may include a framework where each domain will have its independent CDE structure. However, to ensure the stability of the interconnected CDE's, every domain should follow the same workflow states and the same standards for information containers' naming and metadata.

The information will be released and made available using the CDE's different states (WIP, Shared, Published), however, when needed, the delivery team from PLM will access a specific state of the BIM CDE to analyse and link the necessary information to accomplish specific and predetermined requirements. The main difference is that in this case, the information remains within the author's CDE and a link is used to promote a bidirectional information flow. With the change management process in mind, the way to analyse and communicate the impact a change will have on other elements and users from different delivery teams will be greatly improved.

The lack of test and the need for processes' automation may increase the time and cost of the project, nonetheless, it is believed that this workflow will mean a major step on information management and will annulate the duplication of data, and, consequently, the errors and time spent searching for the exact version of information or importing all information from a model when the information necessary is only, for instance, a subset of the IFC schema.

This page is intentionally left blank

4. BIM AND BIG DATA – STATE OF THE ART

What has been shown so far is that a building/asset is responsible for generating a vast amount of information from the brief to the use stage. Multiple disciplines, from architecture to the MEP, are responsible for creating deliverables according to the requirements, and even after the handover stage these resources are used to operate and maintain buildings/assets, which means that through the asset management system the information is optimized and edited through its entire lifecycle.

The volume of data increases when the analyses moves from singular documents or models to a common environment where the data from different delivery teams is combined to create more complex projects with a higher level of reliability and sustainability. More task teams working simultaneously in a collaborative ecosystem also represent a quicker data dispersion and with it the possibility of data waste. Creating a standard process to connect the data from different sources will be a key milestone in the process of solving the interoperability challenges between industries that will enormously decrease the costs of not having it, as shown by the US National Institute for Standards and Technology (Howard and Björk, 2008).

Where BIM transcends its unique and isolated niche and embraces a broader collaborative ecosystem is on the relation with other industries like manufacturing, aerospace, and automotive. The pharmaceutical project unveiled that BIM should not act isolated, but instead be a part of layers of information that gather and analyse automation processes through API's, algorithms, ML, AI, and even BlockChain (BC).

This chapter has the goal of verifying if BIM information can be considered Big Data, if the information generated can match the needs of Big Data Engineering (BDE) and Big Data analytics (BDA), what are the main problems, and the future possibilities.

4.1. Data-driven design

The building design is the result of data constraints like EIR/client requirements, structural calculations, and building codes. This data is normally assigned to the engineering side of the work, while aesthetics are linked to the intangible side of the architects' expertise (Binnekamp, 2019). Despite the dissimilarities between the two fields, the evolution of the AECO industry promoted a higher use o "objective, verifiable, and quantitive data" to forge a new decision-making paradigm called "data-driven design" (Deutsch, 2015).

Following the application of the Evidence-Based Design (EBD) approach to the data-driven design, it is suggested the use of data to make key project decision points (Hamilton & Watkins 2009). Furthermore, the approaches to data-driven design are assuming the use of reliable and trustful data to better answer the EIR/client requirements while optimizing the design outcomes with fewer risks (Loyola, 2018).

It is known that designs are normally the result of the client's requirements and the designers' skills and interpretation. The suitability of the design regarding the user's needs, among others, is usually questionable, and may not be supported by gathered and trustful information. Rarely the data collected (energy consumption, migration patterns, population densities, etc.) is considered during design development, which blocks the design progression through a more accurate answer to the client and environment needs (Bilal, Oyedele, Qadir, *et al.*, 2016).

The controlled and trustworthy use of data expands BIM prosperity and productivity - enhancing the project management process. The real worth of BIM data is dependent on the relevant structured information which is incorporated into the models. There is an expanding enthusiasm inside the act of architecture in "data-driven design", which refers to the integration of simulation into the design process through the use of qualitative and quantitative data to inform designers throughout the design process. The result is a design that is more engaged and tailored to the user's preferences, goals, and behaviors (Benner and McArthur, 2019).

4.1.1. Data use challenges and possibilities

Most of the data can be collected through different technological devices or systems like web, sensors, mobile phones, other smart devices, etc., but it is also available and may be reflected in the documents mentioned in the initial chapters. For instance, the Project Implementation Plan (PIP) may reflect previous project experiences and require new IT capabilities to surpass some issues identified on the lessons learned, or in the MIDP case, new specifications about the deliverables may reflect the experience where the information outcome was insufficient or inoperable, and because of that, the model should be delivered in both native and IFC.

Data can also emerge from various sources, for example, analytics, client testing, research (primary, secondary, generative), and ease of use tests. It tends to be utilized in an assortment of ways all through the structure procedure as there is no one accepted way that works for everybody in each circumstance.

Data-driven design has been a holy phrase in architecture for some time now. The ability to refine and apply information on any range of topics, from movement to sun paths to air quality, holds enormous potential to positively impact design, not just for one party but for all. Decisions can be made faster, buildings can be built better, inhabitants can be made more comfortable

Notwithstanding the knowledge about the existence of all of this data, no tool, at the moment, is capable of enabling the integration and use of data by designers. Leveraging the data within the design process will potentially result in a new model for data-driven design and its integration with BIM authoring tools and with the different suppliers since the start of the design (Bilal, Oyedele, Qadir, *et al.*, 2016).

Currently, BIM data is partially locked under unstructured documentation (UK BIM Alliance, 2018). There is a need for a clear definition of information requirements even before the project starts, as shown in the subchapter 2.1.1. (structured and unstructured data). Throughout the BIM process stages, data can be divided into multiple components as mentioned in the UK BIM Alliance (2018): "general product information" - mainly the type/instance naming and classification (e.g., COBie); "design information"- the dimensions and performance requirements (e.g., fire or acoustic ratings); "manufacturers' information" – the product information; "field information" - the information collected on-site (e.g., installation date); "real-time data" - the actual benefit of on-site product (e.g., performance in service); and, "computer-generated information" - information not directly controlled by the user (e.g., name and version of the software).

The variety of information, and the lack of control, with multiple deliveries and task teams copying only parts of the information from models to avoid the need of change in the case, for instance, an equipment needs to be replaced by a more economic version, may contribute for an incoherent data model with parts of information spread across the project (UK BIM Alliance, 2018). Despite all the potential of data-driven design, to be able to properly use data is important to establish a list of recommendations, otherwise, no matter the volume of data generated, the model will remain inconsistent and most of its attributes may never be accessed, downgrading the model's potential.

As mentioned by UK BIM Alliance (2018) on its report "A Fresh Way Forward For Product data", to promote a data-driven design the data contained in the BIM model should contain only the information needed; a database should be created containing the manufacturer's information (normally exists in pdf as unstructured information); the 3D objects should be created/used according to the lead appointed party/supply product requirements, avoiding the increment of the model weight; information should be created and maintained by a "single source"; standards and methodologies should dictate the way information is shared and connected; if the information published does not answer the user's needs then it should be possible to access other information subsets according to requirements; and, correct data needs to be available. Despite some major recommendations, there is still a long way to reach a place where standards for data are well established in the AECO industry.

The main **challenges of data-driven design** can be related to restricted individual abilities to judiciously break down and deal with a high quantity of data in a brief timeframe. In this sense, a viable information design enables the catch of patterns and relationships, that cannot be conveniently interpreted (Meier, Roy and Seliger, 2010).

Information completeness is another worrisome challenge of data-driven design. This issue features a fundamental exchange between the scope of the information gathered and the need to give a viable analysis. As such, the information needs to incorporate multidimensional viewpoints, yet it should be reasoned by the client, not giving a misguided feeling of precision. The challenge for data-driven design techniques is to offer steady help for choice while adapting to the characteristic incompleteness of data (Bertoni, 2018).

The nature of data is not left out. The information accessible during design can both have a numerical and a nominal nature. The ability to converge in a single model both nature types (or quantitative and subjective evaluation) is viewed as a challenge for the advancement of data-driven design (Bertoni, 2018).

Data-driven design techniques need to address the challenge of not underestimating how the designers independently decipher and "weight" the information gotten from the data. Data-driven design models should be justifiable and straightforward for clients that will have the option to comprehend the need and the importance given to the information (Bertoni, 2018).

As for future **possibilities**, according to Chastain A. (2017) the data-driven design has the potential to use the vast amount of data to improve the planning and delivery of services by using the data to identify and address complex problems. It is imperative to know that by employing the data-driven design approach to design, data will be more publicly available making it easier for designers to get critical information and develop faster. The **automation of most of the planning processes**, through the data-driven design, will help reduce time wastage and potentially save money by testing designs before the construction process. The improvement of the **relationship between materiality and building/asset location** using data that considers the materials' availability and sustainability. Improvement in **user experience** can be enhanced. Utilizing collectivized information on client input can assist engineers in improving the experience for present and future users. There is an opportunity for **continuous learning from previous projects**, considering that for each building built, a massive amount of information is created. Designers can take this data, examine it, and appreciate what worked and what did not. By dissecting data from the project (and from the surrounding environment), structural and architectural designers can decide or estimate what external or internal forces made the structure not settle as arranged, and they can utilize their discoveries to forestall something like this to reoccur (Chastain A., 2017).

4.2. Big Data

The concept of data is not always clear in the BIM context. Most of the time, "data" is confused or mixed with "information". There is a difficulty when trying to mention one or the another, because it is considered that most, or even all the time, both words are describing the same content. This may result from the segregation of information between different fields or simply by semantic purposes. It is rather important to make sure that a common language is reached and both AECO and computer science are using the same terminology. To be able to align and explore the potential of Big Data in AECO industry it seems fundamental to agree on the use of term data. This is the main goal of this subchapter as it tries to explain the concepts that are common to computer and data science while creating the parallelism with the BIM and AECO industry's reality.

Cambridge Dictionary defines data as "information, especially facts or numbers, collected to be examined and considered and used to help decision-making or information in an electronic form that can be stored and used by a computer". What is therefore to be defined is the complete concept of "Big Data": "is the systematic processing and manipulation of data to uncover patterns, relationships between data, historical trends and attempts at predictions of future behaviors and events." (NIST, 2014).

Regarding the definition of Big Data, one perspective, constrained by the commercial vision, defines the boundary between data and Big Data by the impossibility of accessing datasets (collection of data) through traditional methods, while the academic perspective, describes it as datasets with three characteristics: **volume**, which refers to the quantity of data from tera to zillion bytes; **variety**, which means it is represented by multiple formats; and, **velocity**, which refers to the frequency of data stream, with reduced data latency (Laney, 2004) (NIST, 2014)(Loyola, 2018). Whenever talking about Big Data another important factor is the need for a scalable architecture to manage, store, and process the amount of data. With this in mind, the industry and the companies can prepare their IT systems to manage the exponential growth of data (Cross, 2018).

The three v's identified in the domain of Big Data are also clearly recognized in the data generated during BIM processes. The data is typically generated by multiple appointed parties/suppliers which produce a considerable amount of information in different formats, from EIR to federated models, during the project lifecycle, in a dynamic collaborative environment.

Furthermore, other v's have been proposed: **veracity** represents the trustworthiness on the data available; **virtual** or digital representation (self-explanatory); **variability** takes into consideration the inconsistency on the data flow; **validity** is proved by the relation between the data generated and its expected use; **venue** referring to the collection of data from multiple sources; and, the **volatility** that considers the amount of time data should be stored (Borne, 2014)(Tsai *et al.*, 2015)(NIST, 2014)(Loyola, 2018).

Notwithstanding the blurred definition of the Big Data conceptual definition, it is assumed by the authors of the articles considered in this academic work, that two main areas of development are to be taken into consideration. The Big Data Engineering (BDE) and Big Data Analytics (BDA) (Loyola, 2018)(Bilal, Oyedele, Qadir, *et al.*, 2016)(De Mauro, Greco and Grimaldi, 2014).

4.2.1. Big Data Engineering (BDE)

Notwithstanding the blurred definition of the Big Data conceptual definition, it is assumed by the authors of the articles considered in this academic work, that two main areas of development are to be taken into consideration. BDE and BDA (Bilal, Oyedele, Qadir, *et al.*, 2016)(NIST, 2014).

Commonly, in a database, the information is structured in a table-based model composed of rows and columns known as Relational Database Management Systems (RDBMS), and a Structured Query Language (SQL) commands are used to access and manage items. The combination of RDBMS and SQL is suitable to structure data, however, it is not capable of handling unstructured data in the same way because in large datasets it is not easily scalable (Loyola, 2018) (Hu *et al.*, 2014).

To deal with unstructured data, the non-related databases and NoSQL are considered to be more suitable as they offer multiple storing possibilities, they are more scalable, and have better performance than RDBMS (Loyola, 2018).

Normally data platforms are divided into two groups: Horizontal Scaling Platforms (HSPs) are those that use multiple servers and perform scalability by increasing the number of machines to the existent group, and, the Vertical Scaling Platforms (VSPs) which increase the processor, memory or disk capacity to improve hardware performance (Bilal, Oyedele, Qadir, *et al.*, 2016).

BDE can be divided into Big Data processing and Big Data storage. The first refers to "parallel and distributed computation" as the foundation of BDE, while the second is "provided either by distributed file systems or emerging NoSQL" (Bilal, Oyedele, Qadir, *et al.*, 2016).

Regarding the Big Data processing, and according to Loyola (2018), Apache Hadoop positions itself as the currently elected Big Data infrastructure platform, which may be due to its open-source framework based in Java. With Hadoop Distributed File System (HDFS), even bigger files can be stored, indexed, and spread by multiple machines. The **MapReduce (MR)** process is another virtue of this platform, allowing the division of complex data into small units to be processed in parallel (figure9) (Bilal, Oyedele, Qadir, *et al.*, 2016)(Loyola, 2018).

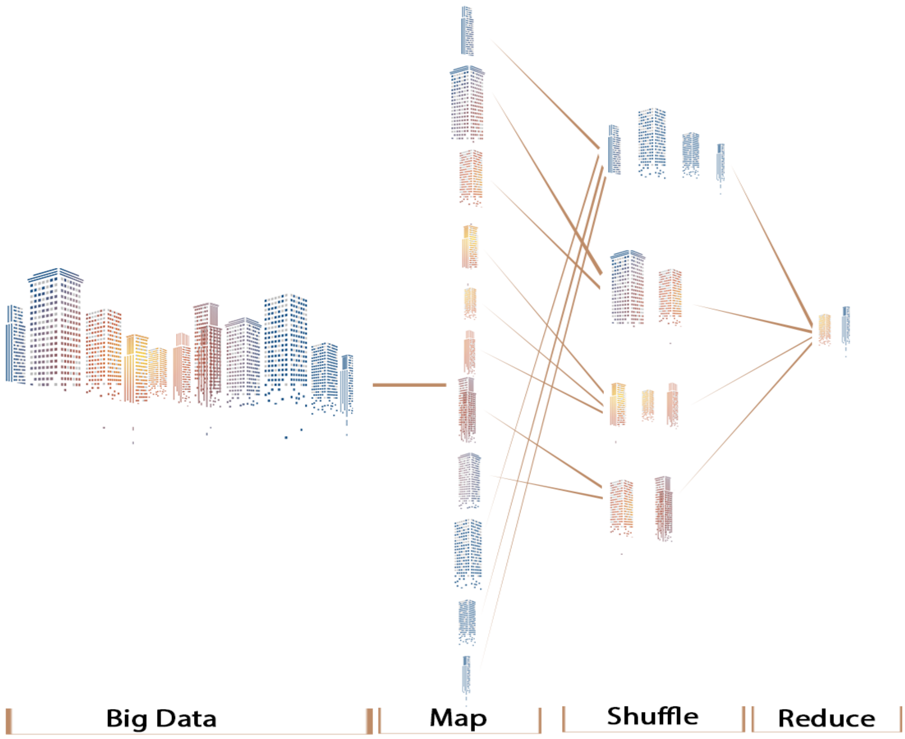


Figure 9 – MapReduce Processing

Nevertheless, when real-time or iterative processing is needed the MR is not powerful enough. That being said, **Directed Acylic Graphs (DAG)** is an alternative for processing Big Data. Berkeley Data Analytics Stack (BDAS) is an evolution of the common process of "map-then-reduce" and supports DAG while providing a dynamic response to intricate computation on Big Data (Stoica I., N.D). BDAS relies on Spark, a component capable of better "in-memory computation and high expressiveness" than MR (Bilal, Oyedele, Qadir, *et al.*, 2016).

What is to be noted is that these technologies, processes, and systems are already being used in other fields of knowledge. Their maturity may help the AECO industry integration, namely when speaking about that generated by and in the model and also by external sensors. The way data is processed is a key step that defines data's analyses benefit and future implementations.

As for the Big Data storage, distributed file systems, like HDFS and Tachyon (BDAS file system), they are capable of tolerating hardware breakdown while working with larger datasets, and to use in-memory data to enable performance and also works with Spark and MR, respectively. Also, NoSQL technology, whose architecture is capable of leading with fragmented information from the AECO industry, is pointed as the technology that better suits the data storage from BIM databases (Bilal, Oyedele, Qadir, *et al.*, 2016).

Despite all the technical information about the multiple systems and platforms that can be related to BDE, no further details will be considered as they won't benefit the scope of this academic work. Further information can be accessed in the articles "Big Data in the construction industry: A review of present status, opportunities, and future trends" and "Big Data architecture for Construction Waste Analytics (CWA): A conceptual framework" from Bilal *et al.* (2016).

4.2.2. Big Data Analytics (BDA)

According to Correa, (2015) "Data Analytics is a technique where mathematical models from given phenomena are not available, but where patterns that characterize them can arise from available data and that, therefore, could be used instead of a formal model".

Valuable patterns can be identified even beneath layers of structured and unstructured data. The vast amount of data can be used to discover significant information used in the decision process and is described as the **"Knowledge Discovery in Databases" (KDD)** method (Bi and Cochran, 2014) (Loyola, 2018).

BDA is the key to access the information generated and to promote its full potential. This symbiotic integration of diverse data sources with BIM will ultimately lead to the next-generation designs that can meet the wider requirements of sustainability, user experience adaptation, environmental concerns, and even broader infrastructures of the emerging concept of smart cities (Bilal, Oyedele, Qadir, *et al.*, 2016). It was with the growing interest in the data value and the complexity and multiplicity of skills needed to process the amount of information that data science field of knowledge was raised (Loyola, 2018). The exponential expansion of this field of knowledge with particular interest from multiple disciplines led to a new educational paradigm with curriculums that combine computer science, statistics, machine learning, data visualization, and communication, among others (Provost and Fawcett, 2013).

The integration between different fields of knowledge and disciplines is a challenge. Finding common ground and a language that suits multiple needs is a complex task, that if succeeded may define new ways of data-driven designs. It is clear that data waste is a commonly accepted consequence of the design processes, however, the goal is to evolve into a more stable and clear organization where information management explores the full potential of data.

The reliance on data and the incentive to use it in BIM processes have only contributed to the use of Big Data by a few people in the AECO industry. Big Data Analytics has a rich intellectual tradition and borrows from a wide variety of fields (figure 10). There have been many related disciplines with essentially the same core focus: finding useful patterns in data (Bilal, Oyedele, Qadir, *et al.*, 2016).

In the end, it does not matter if it is called **business intelligence**, used to analyse large datasets to help in the making of informed business decisions (Chen *et al.*, 2012); **predictive analytics**, uses statistics to predict future or unknown events (Boire, 2013); **data mining**, used to in multidisciplinary fields to "mine" knowledge from data (Han and Kamber, 2001) (figure 10); but the overall concept is always related to the capacity of processing and analysing large datasets to predict unexpected patterns and to assist with decision making.

In the AECO industry there are already some examples of the use of statistics to identify the causes for construction delays (Kim, Soibelman and Grobler, 2008), and of the data mining to gather and analyse data from previous projects, and implement improvements on future ones (Carrillo, Harding and Choudhary, 2011).

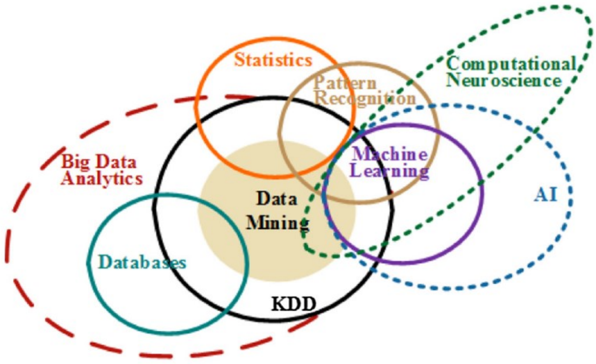


Figure 10 – Big Data Analytics multidisciplinary (Source: Bilal, Oyedele, Qadir, *et al.*, 2016**)**

4.3. Machine Learning (ML)

Before any considerations regarding ML, it is important to clarify the relation between ML and AI. With AI actions and decisions are automated taking advantage of a "set of tools for making computers behave intelligently" (Bown-Anderson, 2020). The author also mentions the difference between **Artificial Narrow Intelligence (ANI)** that includes all the algorithms used to perform one task correctly (e.g., self-driving cars), and, **Artificial General Intelligence (AGI)**, which is a hypothetical AI capable to do what humans do. What is important to understand is that AI has a broader range then ML and for the purpose of this academic work, ML is considered a tool to enable AI.

Being said, ML is considered a branch of AI that gives computers the capacity to figure out how to reach a result without any programmed guideline. Generically, algorithms are used to inspect datasets and to identify connections between data, without having previous knowledge about them (Loyola, 2018). BDA strategies depend on different types of data processing going from basic counting on to progressively expound measurements and complex ML methods, such as Deep Learning. It has accomplished exceptional outcomes given enormous datasets for issues identified with images, speech, and natural language processing (Holzwarth *et al.*, 2019).

ML approaches are methodically ordered into two classifications: supervised and unsupervised. In the first, a set of labeled data is accessed allowing the ML to create categorization hypotheses dependent on the training data labels, while the second finds shrouded examples in unlabeled data and infers a function to describe data structure (Loyola, 2018). According to Krijnen and Tamke, (2015) unsupervised methodologies work without the reason of referenced labels. More theoretically, it is a strategy to lessen the costs in the AEC industry by hailing extraordinary circumstances that may require extra checks or coordination. Supervised machine learning for example can arrange floor plans as indicated by its intended functionality. Such framework can be found in the light of a huge documented system for building models in which data relating to proposed capacity and use is frequently divided and inadequate. It may also be utilized to finish missing qualities in such a dataset (Krijnen and Tamke, 2015).

According to Loyola, (2018) some of the most used and efficient algorithms used in building information modeling are the following:

Artificial Neural Networks (ANNs) are an arrangement of interconnected "artificial neurons" that process and break down information in a non-direct design, continually refreshing their loads dependent on iterative outcomes. ANNs show complex connections among sources of input and output, to discover startling examples or structures in the information. As ANN models grow in multifaceted nature, they become a kind of "black box" with scientific procedures that are uninterpretable by people. Their outcomes, albeit significant, become less reasonable. This is a key detriment in situations where it is expected to see how the frameworks reveal themselves for a specific outcome, for instance, when attempting to foresee occupants' activities;

Genetic Algorithm (GA) is a metaheuristic search algorithm reliant on the procedure of normal determination. Given a population and an optimization issue, the calculation tests an irregular population from assessed cases, and afterward, stochastically, recombines them after a few iterations until a targetted workgroup is reached depending on the essential parameters;

Decision trees are supervised algorithms that build a model to predict the estimation of an output (leaves) considering a few data inputs (branches). At the point when the output is a class, the tree is called a classification tree, and when the output is a real number, it is called regression tree;

Support Vector Machines (SVMs) "are a set of related supervised learning algorithms used for classification and regression". From a named dataset, an SVM acquires an ideal n-dimensional hyperplane which is utilized to order new cases. They are often used to recognize if a case resembles a class, for which, outstandingly, they do not require preparing training samples.

The complexity walks side by side with the desire of making AECO industry more efficient and profitable. Big Data and technologies like ML and AI should be capable of pushing the industry towards the place where other industries already are and started to build new paradigms.

4.4. Big Data in AECO industry - a critical review

Notwithstanding the challenges and possibilities accruable from Big Data in this industry, some difficult issues survive from concern. According to Bilal *et al.*, (2016), even though the AEC(it does not mention the Operation part) industry creates huge amounts of information for the duration of the existing pattern of a structure, the utilization of Big Data benefits misses the advancement made in other fields of knowledge.

Nevertheless, data-driven decision-making at the design stage is uncovered to bring a transformation for forestalling a significant proportion of construction data waste. This propels a move from the static thought of waste analytics (Ekanayake and Ofori, 2004). Big Data leads to an analysis of disaggregated and massive datasets to reveal non-evident relationships identified with design, procurement, materials selection, etc which could lead to squandering during the construction stage. This requires the utilization of Big Data, ML/AI for effective data management. Especially robust waste generation estimation models, BIM-based materials' choices during design specification and comprehensive waste minimization system are key research regions that require the use of these Big Data technologies.

Performance prediction models have been of wide relevance in different spaces of the AECO industry and BIM in general. Designers are encouraged to make the right choices while constructing, maintaining, and rehabilitating structures (Kargah-Ostadi, 1993). Big Data technologies are extremely relevant and can assist through the use of real-time computation, reliable model development, and enhanced visualization.

One of the key difficulties that AECO industry's face is that a significant part of the project management related procedures keeps on being manual and does not have the required automation that different enterprises have embraced. Amazingly, the AECO industry is utilizing innovation to plan and update activities and CAD/BIM to structure the projects, however, it abstains from utilizing the correct innovation tools and platforms to capture project and contract management processes information. This not only influences the quality and credibility of captured data, but also implies that a significant part of the data gets squandered and neither gets captured nor shared. It is crucial to note that most BIM software are independent frameworks in which a single PC is utilized for the dominant part of the computations, presenting extreme limitations to enormous storage, effective management, sharing, and synchronization of the BIMs that are developing in size and multifaceted nature step by step (Chen, Chang and Lin, 2016). It also creates the issue with different information sources, however, a portion of those applications could, in the long run, be dropped by the organization or even decommissioned by the product merchant.

It is vital to understand how Big Data can affect BIM at various stages. Data capturing and collection is used at the various levels in a Design, Building, and Operation type project, as explained below according to Burger R. (2019):

Design – Previously collected data from other projects can be used to identify patterns and align the new projects to better suit its function, and can also be considered at the tendering/bidding stage avoiding time and cost slippage.

Build - Large information from climate, traffic, network and business can be investigated to decide the ideal staging of construction exercises. Geolocation of equipment additionally permits coordination of maintenance and substitution of spare parts.

Operation – Building or structures' performance can be monitored through sensors. Traffic pressure data and levels of flexing in extensions can be recorded to distinguish any beyond the field of play occasions using structural health monitoring systems. This information is sent back into BIM frameworks to plan maintenance exercises as required. Energy conservation in different building types and structures can be followed to guarantee the fit within planned objectives and to report for future use on the design stage (to be further developed in the next chapter).

Despite all the potential of relating Big Data and BIM, there is also the possibility that BIM does not have enough scale to align with Big Data. It was shown before that a vast amount of information can be created during all project stages, nonetheless, authors like Correa (2015) argue that even large projects' BIM model could generate files with 1 GB, which is not enough to take advantage of Big Data. Certainly, not all design offices or projects will take advantage of Big Data. The volume property mentioned before, that characterizes Big Data, should not be ignored, and that is why the scope of the academic work does not focus on the idea of independent BIM models but rather data from a CDE that may include different BIM models and documentation. Still, for predictive analysis and data mining, the creation of a repository of multiple project documentation may be the correct path to the proper data usage. Furthermore, considering that 60% of BIM data is geometry and the remaining 40% are attributes, properties, and relations between objects in a schema (e.g., IFC) (Eastman C.; Teicholz P.; Sacks, 1886), BIM data can be used in Facility Management (FM) combining BIM information with data from the real world (e.g., sensors, employees log in about daily tasks) (Correa, 2015).

To summarize, it is proposed the integration of all data, structured and unstructured, whose maturity proves to be adequate enough to be published inside the CDE and into a Big Data platform that would allow data analysis to predict future issues/challenges while identifying patterns. For that purpose, data should be gathered, treated, and displayed in a useful and intelligible way. From this point, data that has been mobilized during design and construction will be available for FM through the creation of a digital twin. ML will be put in place to run the entire project lifecycle and validate design and construction options or predict future outcomes during the decision making.

This page is intentionally left blank

5. DIGITAL TWINS

Up to this point data generated is mainly produced during the design and construction stages. It is with Big Data, and all the associated technologies and processes (e.g., machine learning, artificial intelligence, etc), that a turning point is reached. This new chapter aims to cover the operational phase where FM can be enhanced by digital twins.

Only recently the massive quantity of data generated by digital twins became accessible to companies, due to digital technology limitations (Parrott and Warshaw, 2017). The reduction of cost and performance improvements allowed the combination of Information Technologies (IT) and Operation Technologies (OT) to promote the use o digital twins (Mussomeli, Gish and Laaper, 2015).

From a certain viewpoint, a digital twin is the continuity of what has been developed until the handover stage. It gives the owner the complete knowledge about the asset from its design and construction but also promotes quicker solutions for anticipated problems with a high level of accuracy, meaning better design and construction/production of future assets/buildings (Parrott and Warshaw, 2017).

The concept was used first by NASA in the rescue mission of Apollo 13, first as a "mirrored system" and later, in the 2000s as an evolution of PLM. Since then, the concept evolved to an "integral digital business decision assistant" (BSI Group, 2019).

5.1. Digital Twins value

The global digital twin market was valued at USD $3.8bn in 2019 and is expected to reach USD $35.8bn by 2025. It I predicted that by 2021 about half of all large establishments will use some form of Digital Twin resulting in an improvement by 10% in effectiveness (BSI Group, 2019)

According to some authors, a Digital Twin (DT) is an integrated as-built product that includes a batch of data collected to investigate potential design or operation adaptations to more effectively and safely manage individual products (Reid and Rhodes, 2016).). Others may suggest that DT can be described as a digital model enabled by the use of sensors that simulates a physical object in a live setting (Grieves, 2014). As mentioned by Parrott and Warshaw (2017), the foundation of DT is the "massive, cumulative, real-time, real-world data measurements across an array of dimensions". The digital version of a real object can give a valuable contribution to the physical object's behaviour, providing important information regarding performance and simulate the outcome of design or manufacturing/construction changes.

Digital twins are the link between the digital and real-world. In the digital world, CAD, which can include BIM platforms and tools, is considered to be enclosed in a computerized environment, while the real world is captured by sensors that commonly measure the position and other properties of a single component, not being capable of establishing interactive relations between multiple components (Parrott and Warshaw, 2017).

It is a mistake to consider DTs a sensor-enabled model that works as a repository for data. Instead, digital twins are dependent on data from other domains, for example, procurement teams may contribute to digital twins' insight in costing and supplier selection (Autodesk, 2019). This is an important point as the digital twin cannot be considered as an isolated model, but rather a model that communicates with other users to provide them the complementary information for the inputs previously generated. As mentioned in the Autodesk (2019) article, the current level of maturity does not promote that analysis, and implementation of the data gathered by the digital twins is done automatically, meaning that the DT is not used for taking active decisions. Instead, nowadays digital twin still works as a silo of information to elevate the human made decision (figure 11).

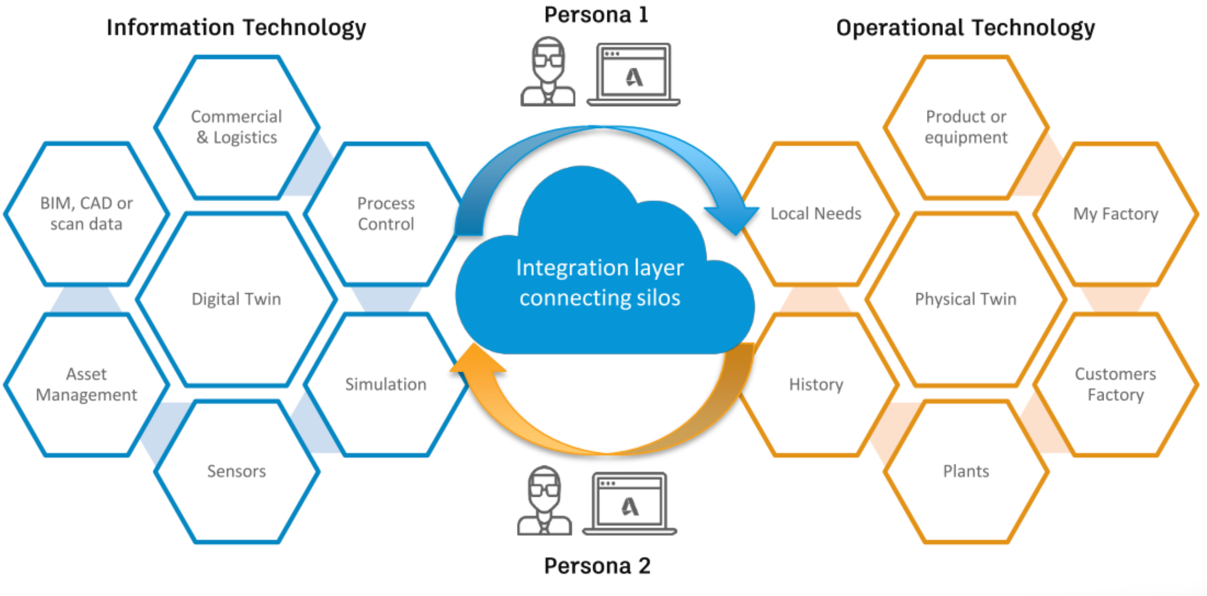


Figure 11 – Information and Operation Technology as data silos (Source: Autodesk, 2019**)**

As mentioned before, the AECO industry is not at the same level as other industries, for example manufacturing where PLM enjoys full business benefits from the use of digital twins. Following the same principles of a manufacturing process example, it is proposed an adaptation of the high level process presented by Parrott and Warshaw (2017):

Sensors – the equipment distribution through defined rooms creates signals enabling DT to capture environmental data concerning users' occupation, UVA light, and temperature in the real world;

Data – real-world data captured from the sensors is assembled with data from the BIM model, such as design specifications and bill of materials (BOM);

Integration – the bidirectional communication between the real and digital world is established by the sensors through integration technology;

Analytics – data is analyzed using "algorithms simulation and visualization routines" used by digital twins to generate insights;

Digital twin – exposes the data values that are far from the optimal conditions, which may trigger an opportunity to save costs or achieve a better performance rating;

Actuators – more easily associated with the manufacturer's side, they can be used to promote action in the real world that may be done automatically (e.g., turning off the lights on an empty room using proper electrical devices) or manually (e.g., cutting down the electrical power of a certain building part).

The value of DT is measured by the business value it represents, however, the value can also be considered for the user and designer. The principle imported from the PLM side is important to perceive, even with the needed adaptations: the importance of a structure that transforms collected data into an automated action that promotes building or product answers to a specific problem as a result from a collective analysis of multiple elements' properties.

5.2. From BIM to Digital Twins (DT) – From CapEx to OpEx

The digital transformation of industry 4.0 is fostering the development of Big Data, DT, IoT, AI, and BIM, among other technologies. BIM isn't simply about CAD drawings, it is about the administration of steady, discernible information that follows structured information management. Digital Twin helps organizations to have a complete replica of their product covering from design to the entire lifecycle of the product, enabling them to understand the product completely, how its systems work and how to optimize them. When BIM is strengthened by the IoT its utility is enhanced, becoming a more mature information delivery tool that englobes details not considered in the construction phases.

BIM is considered to bring benefits during the entire building lifecycle through the expansion of datasets. The BIM model of an asset is intended to be used during design, construction, and operation, and contains part of the information a DT would require. When software enables the complete support of all building processes it will be possible, DT will become the natural evolution of the BIM model (Allplan, 2020).

As mention in Trimble Inc. (2019), DT and BIM integration can lessen some of the BIM challenges, such as:

Asset insight – DTs are analytical tools that offer insight to conditions and advises about future operations;

Team connectivity - DTs can embed component dimensions, model details, etc to facilitate the project collaboration through construction phases;

Optimized fabrication - error reduction through the up-to-date information embedded in the DT;

Ongoing support – in the project handover, DTs can be shared with the client to support enhancements.

According to Robins (2019), DT will surpass BIM, spanning the asset lifecycle. For CapEx, DT will provide a design optimization and work as a way to replicate construction, logistics, and fabrication sequences, while for OpEX it will enable the representation of the real-world conditions into the asset's DT through the continuous surveys, photogrammetry, LiDAR and sensors.

It is not clear for the time being if DT will replace BIM, if they will merge, or even if DT is the natural evolution of mature BIM standards. No proper tool or platform to represent and merge datasets from both ecosystems was found, which leaves space for future enhancement on this topic. Nevertheless, it is important to look for the current industry state of the art to identify the principles that, like in BIM history, can be broadly used and evolved to reach new levels of maturity.

5.3. The Gemini Principles

Only in the UK shared data could release an extra 7bn £ per year in the infrastructure sectors (Deloitte, 2017). With that in mind, UK Gemini principles seek the foundation's definitions and guiding values of information management across the built environment (Bolton *et al.*, 2018).

The UK National Digital Twin (NDT) aims to improve digital built Britain planning, delivery, and whole-lifecycle management of buildings and infrastructures. According to Bolton *et al.* (2018), the representation of realism depends on the data quality that is used to create the digital twin, the trustworthiness of algorithms and code competence in which the model is based, and, the visual quality of the output. Furthermore, the author defines the variety of the DT ecosystem in:

"Variety of purposes" – **potential futures**, if it works on a predictive and preventive environment; **current state**, if it focused on real-time monitoring to optimize asset performance; and, history, if intended to record and to deploy lessons learned.

"Variety of spatial scales" – asset/building scale; network or building scale; system, city or regional scale; and national scale.

"Variety of temporal scales" – operational timescale; reactive maintenance timescale; planned maintenance timescale; and capital investment timescale.

"Variety of approaches to modeling" – geometric and geospatial modeling; computational/mathematical/numerical modeling; and artificial intelligence and machine learning.

The NDT defines what can be a principle of classification for the DTs and can help to align the digital model with its purpose, avoiding information, time, and cost waste. This is similar to what was mentioned about BIM data and shows the evolution of the principle of the reduction of information waste while promoting the assessment of the context in which the DT will be implemented.

The purpose, trust, and function are the three principles of the framework proposed by the NDT. The first must support public timelessness benefit, create value, and improve performance, and furnish awareness into the built environment. The second must grant security, be as open as possible, and rely on appropriate quality data. The last must be based on a federation environment, have clear ownership, governance, and regulation, and, be as adaptable as the technology and society (Bolton *et al.*, 2018).

The built environment key players should be driving effective information management while promoting value for the organizations, the economy, and society. The information must be considered regarding the extent of its purpose and must aim for concrete objectives to bring benefits to users, businesses, and industries. The Gemini principles are therefore highlighting important considerations about information management adaption to the AECO industry.

5.4. The ecosystem of Digital Twins

The increasing data in the digital models, being from BIM, DT or both, raises the need for an ecosystem where information management can be boosted. The digital transformation brought by industry's 4.0 process is facilitating the advancement in mobile and cloud-based technologies, IoT, AI, sensors, robotics among others, that are changing the industry's business models (buildingSMART, 2020).

DT and PBO-I are two concepts that buildingSMART International (bSI) integrated into their open and international digital solutions and standards for the built asset industry (figure 12). bSI looked into other industries' success cases and tried to bring to the PBO-I the benefits of improved data management and control, aligned with standardization. The goal is to promote a data-driven digital version of a built asset unlocking the value for the multiple players involved in the process.

Figure 12 – PBO – I schema (Source: buildingSMART, 2020**)**

While technology allows dynamic and real-time data exchange between physical and digital models, bSI (2020) points to the exchange at the "right-time intervals", which should promote the "just-in-time" action. Nevertheless, the real potential is in the integration of multiple DTs in what is called the "ecosystem of digital twins". The sustained answer to a proposed challenge or issue will be sustained by having access to the right information instead of relying on a heuristic model.

PBO-I replaces the PBOD and focuses on an interconnected system where integrated data updated within the right timeframe unleashes the unknown information adding value and promotes informed decision making. DT should be allied to ISO 55000:2014 for asset management to make the connection of the built asset industry and the connected data streams (buildingSMART, 2020).

To map information to enable interoperability BuildingSMARTInternational (2020) identifies three areas that need further development:

Standards for data models – despite the existence of data models standards, there is the need to extend the scope and deploy data models according to the different levels of information: geospatial information for urban planning, digital engineering for construction, or operational data for asset management. Furthermore, the link between data models is needed to ensure the comprehension of data content between them;

Standards for data management and integration – map the knowledge between data science and information management to create a resilient ecosystem for DT where a trustworthy and sustainable data system is created through semantic precision;

Data security and privacy – considering the data exchange between projects, lifecycle phases, levels of scale, and tools, there is a need for further development regarding the hosting of data, data ownership, and privacy.

The bSI proposal for DT ecosystem is based on the possibility to release the value associated with the integration of multiple assets DTs. The connection of the data that flows between digital models will add value to the industry. Like the past, bSI's initial work about DT raises concerns and promotes the further development of standards and data management that have the power to influence future development regarding this topic.

5.4.1. Openness

The bSI ecosystem misses the subject of openness, which seems important in continuity with what has been investigated in chapter 3. The siloed data that characterizes much of the information produced in the CAD/BIM tools won't allow the scalability of DT technologies. Unless data can be used and synchronized in a federated environment with other DTs, the information will not have either veracity or fidelity (Robins, 2019).

Connected data environment (CDE) will be responsible for federating multiple live data, releasing architecture premises to be used during 'design and construction work. The real value will be in the high quality of the open-source data libraries that will provide bespoke digital twins solutions (Robins, 2019).

The connected data environment works for the DT reality as the Common Data Environment for BIM. If the software's native formats were ignored, the open standards would work in both cases to promote the data exchange between multiple disciplines and stakeholders. However, the DT CDE will promote a broader span of the data which will enlarge the amount of data flow through multiple DTs in multiple environments with different scales.

5.5. Smart cities through digital twins' use

The DT CDE has the scalability factor needed to promote bigger digital models like the ones that represent real city-data. Urbanization depends not just on a city's enrichment of physical capital but also on the accessibility and nature of communication, monitoring, and infrastructure (Caragliu, del Bo and Nijkamp, 2011). The ability for smart cities to incorporate the use of sensors would enhance their ability to be monitored by perceiving the environment and communicating with the model to provide continuing information. An example is the "Virtual Singapore" which is the foremost digital twin in the world of an already existing city (Farsi et al., 2020).

DT will help in monitoring performance, making quality check, and improving interactions thereby bringing a drastic reduction in resources' use and consumption. It will also help in reducing contact between residents and government leading to a reduction in strikes and activism. The development of smart cities is a process of urbanization that can additionally improve the effectiveness, dependability, and security of a city. The coordination of ICT with the IoT and AI strategies will be useful for the urban/metro smart city advancement which would be used by its end users for overall management. The DT will enhance interactions with the communities and this will make the city infrastructure accessible and ensure law and order. Furthermore, every event happening in the city could be monitored. The information and data will be gathered by utilizing the sensors incorporated with the real-time monitoring systems (Tao et al., 2019).

Smart City incorporates a progressively all-encompassing examination and visualization approach into the real-time knowledge discovery process from heterogeneous city-data. A spatiotemporal knowledge discovery framework is used as the initial approach towards using the city's data to make decisions based on the DT. The accessibility of spatiotemporal sensors' data is opening significant chances to exploit its interesting potential to help improve actions and better administer smart cities towards sustainability , giving extra insights to city's dynamics (Mohammadi and Taylor, 2020).

Smart cities are the evolution, together with the increased growth of data. The data generated is considered Big Data and it is supposed to continue its exponential expansion as the cities have their own digital assets.

This page is intentionally left blank

6. USE CASE

While reproducing an entire DT ecosystem would be unachievable for the present academic work, due to time and support limitations, it is explained how, at a reduced scale, the integration of environmental sensors and IoT will benefit and shape the future of BIM. The use case showcases an intention and a real example of how data can be gathered and utilized inside a BIM model. Furthermore, a simulation of a complex system of interconnections would benefit from a more advanced comprehension of disciplines like physics, maths, and computational design.

Consequently, what is presented is a combination of two devices, one with Wi-Fi connection and another that reads the room’s environmental conditions; an IoT platform with a dashboard; a Google sheets connection through a webhook; and a dynamo script that will push the information into the digital twin positioned inside the Revit model.

6.1. Arduino MKR WIFI 1010 and MKR ENV SHIELD

Arduino is an open-source software and hardware company. The range of hardware covers a big variety of possible integrations of sensors in small projects. The selection of the brand and its products was the result of a market research with a solution that could be combined with an IoT platform.

The MKR WIFI 1010 (figure 13) is the entry ‘mainboard’ for the basic IoT and was selected due to its direct connection to the Wi-Fi router. This will grant that values from the sensor will be constantly updated on the platform.

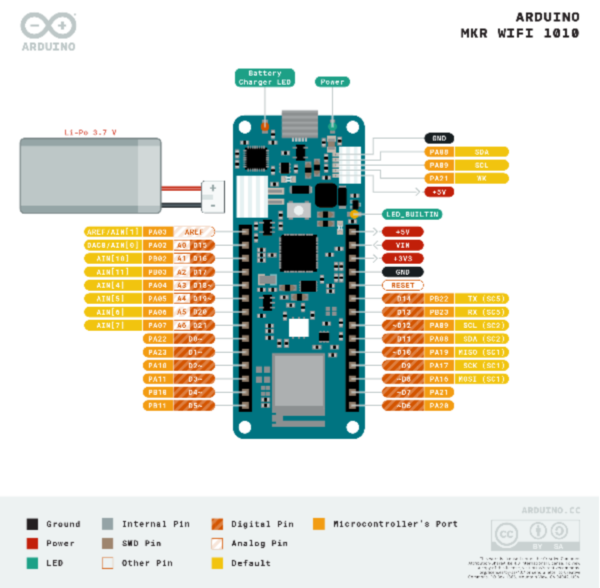


Figure 13 – Arduino MKR WIFI 1010

The MRK ENV SHIELD (figure 14) allows an MKR board like the MKR WIFI 1010 to acquire the data from the sensors embedded on the sensor board. The board sensors will be responsible for gathering information about atmospheric pressure, temperature, humidity, ultraviolet UVA intensity, and light intensity (Lux). UVB intensity and UV index are also supported but will not be part of this use case as the IoT platform limits the maximum number of properties to five.

.

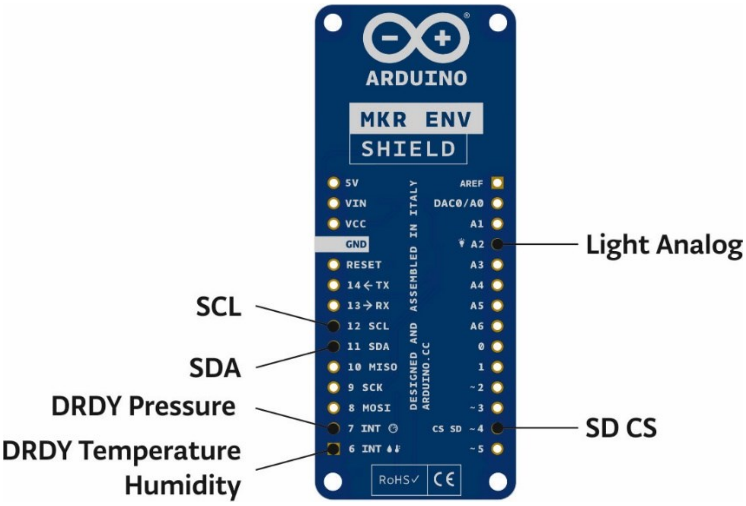


Figure 14 – Arduino MKR ENV SHIELD

Both boards are physically connected and communicate with each other. However, to read the values extracted from the sensors it is necessary to connect the mainboard to the Wi-Fi and then create the script with the parameters to be displayed on the Arduino IoT dashboard.

6.2. Board configuration and IoT connection

After the connection between the two boards, it was necessary to create the configuration that will establish the connection between the MKR Wifi 1010 board and the Arduino IoT platform. To better understand the importance of this platform it is necessary to perceive that IoT is used to describe a system where mechanical and digital devices take advantage of technology to communicate and interact over the internet (Trimble Inc., 2019).

The IoT is the system that allows the DT's real-time data updates through the use of sensors and allows communication through the internet breaking the necessity of using a hard drive as the main repository for the data.

Once the concept is clarified, and the board is registered in the Arduino webpage, to establish the connection to the Arduino IoT platform it was necessary to edit the code to transmit the desired environment's parameters. As shown in figure 15 (see Appendix 3 for more detail), the line 18 of the code refers to '<Arduino_MKRENV.h>' which is the name of the board with the sensors, and the lines 43 to 51 are responsible for publishing the values of humidity, luminosity (lux), pressure, temperature, and UVA in a loop system that will show the values in a defined interval of time:

'void loop() {

ArduinoCloud.update();

// Your code here

humidity = int(ENV.readHumidity());

lux = int(ENV.readIlluminance());

pressure = int(ENV.readPressure());

temperature = int(ENV.readTemperature());

uva = int(ENV.readUVA());

delay(1000);’

Furthermore, to grant the access to the Wi-Fi connection, the tab ‘Secret’ was filled with the SSID code and the password for the Wi-Fi connection. This code suffered minor changes from the one available at the GitHub.

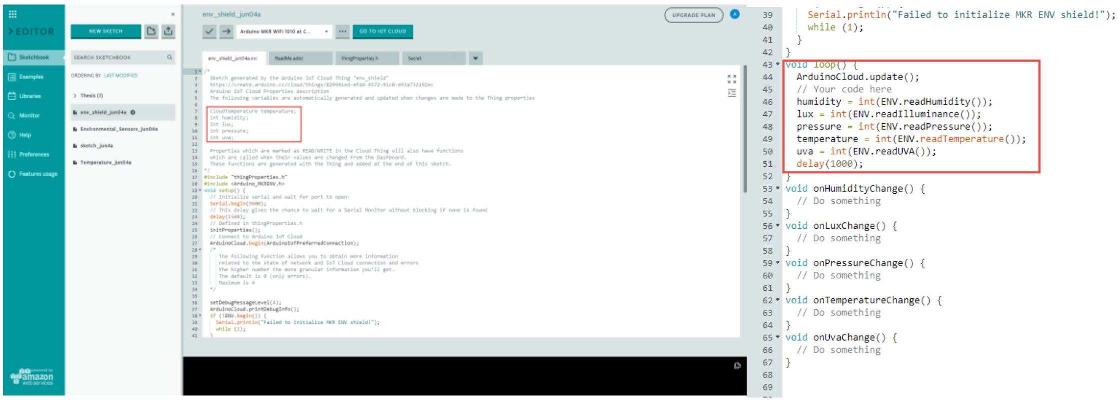


Figure 15 – Arduino IoT ‘Sketch Edition’ showing the environment properties to be exhibited (see Appendix 3)

6.3. Create properties entry and dashboard

To display the values of data collected by the sensor in the MKR ENV SHIELD it is necessary to define each property individually (figure 16).

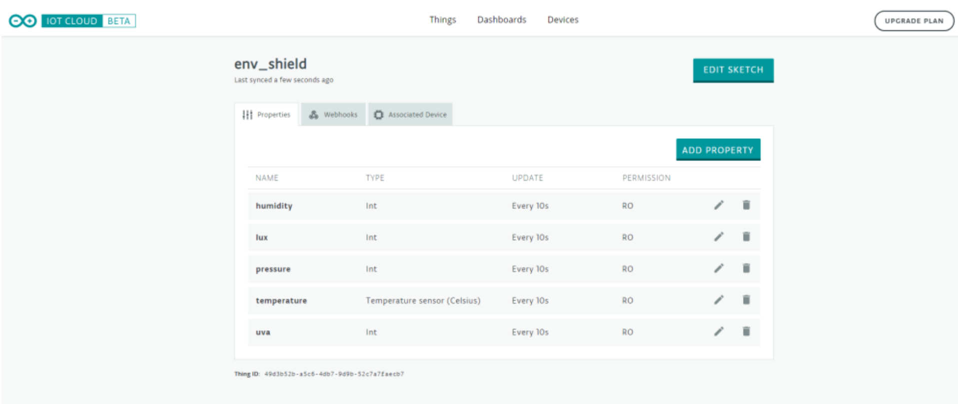


Figure 16 – Adding properties in Arduino IoT Cloud

For each property it is necessary to define some characteristics including the type, that in this case was defined by an integer (Int) and the interval of time as10 seconds, meaning that every 10seconds the value for the property will be updated (figure 17).

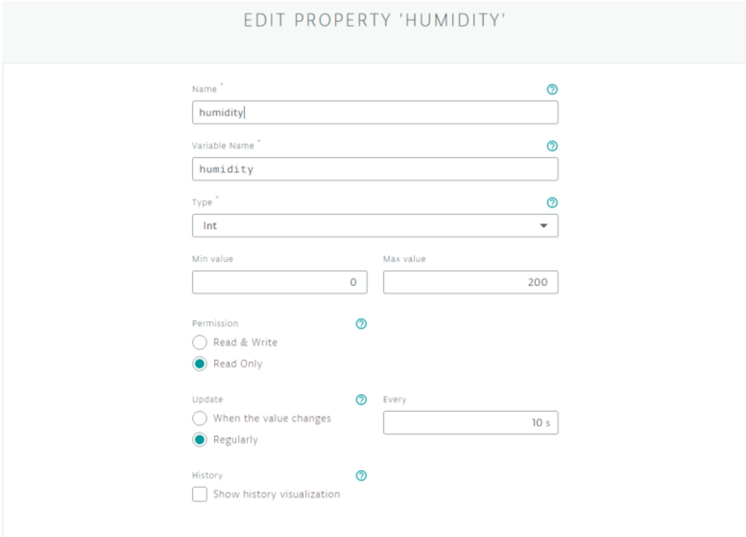


Figure 17 – Adding properties in Arduino IoT Cloud

After the properties are defined it is possible to create a dashboard to see the real data values being updated at the defined time interval (figure 18). This dashboard only works as a graphical representation of data and does not have a format that can be used to import data into our digital twin inside our BIM model. For that reason, it is necessary to push this data into a more suitable system like Google Sheets.

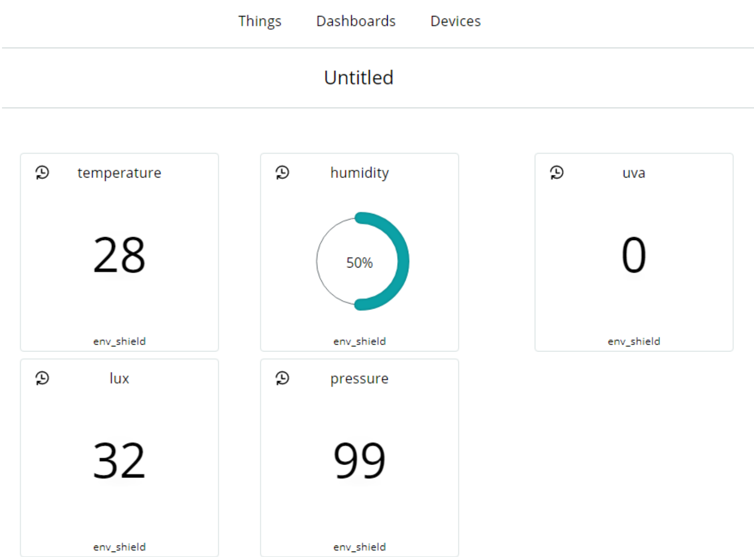


Figure 18 – Properties dashboard

6.4. Arduino IoT and Google Sheet connection through JavaScript

At this point it was necessary to use javascript to establish the connection between Arduino IoT and Google Sheets. A new sheet was created and saved with the 'RawData' name. After that, using the

'Script Editor' under the 'Tools' menu, it was possible to access the code for the Google Sheet. The code belongs to Marco Passarello and was available online. However, line 27 was changed to contain the URL of the 'RawData' sheet to publish the data into that specific sheet (figure 19).

> 'varid="19h9mFmg2Ks0WE6l_bl9t-Bil1K_Iz1CQC887B2rhdkw";//docs.google.com/spreadsheetURL/d'

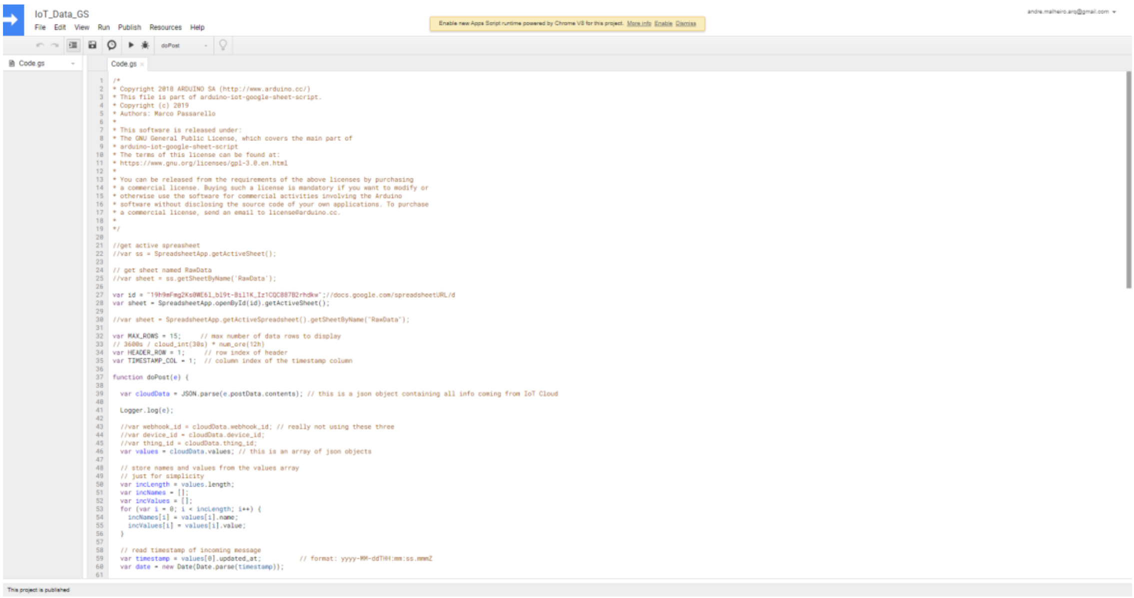


Figure 19 – Properties dashboard

Once the code was complete it was necessary to use the option 'Deploy as a web app' under 'Publish' menu which generated a 'web app URL' to input into the Arduino IoT to close the data circuit and communicate the values into the Google sheet (figure 20).

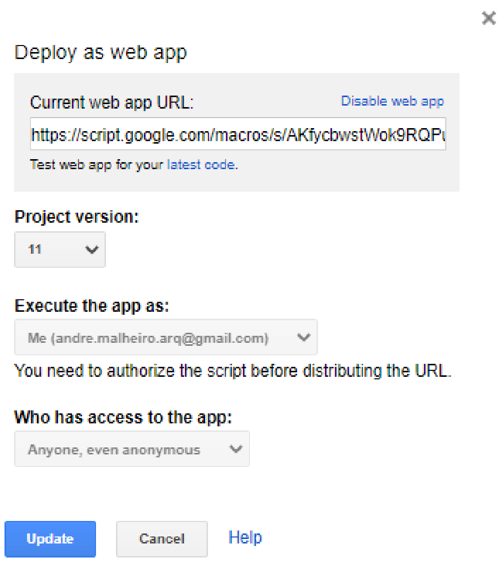


Figure 20 – Publish code

The 'web app URL' was then copy-pasted into the Arduino IoT Cloud webhook. The active webhook field was filled with the URL establishing the link between the two platforms (figure 21).

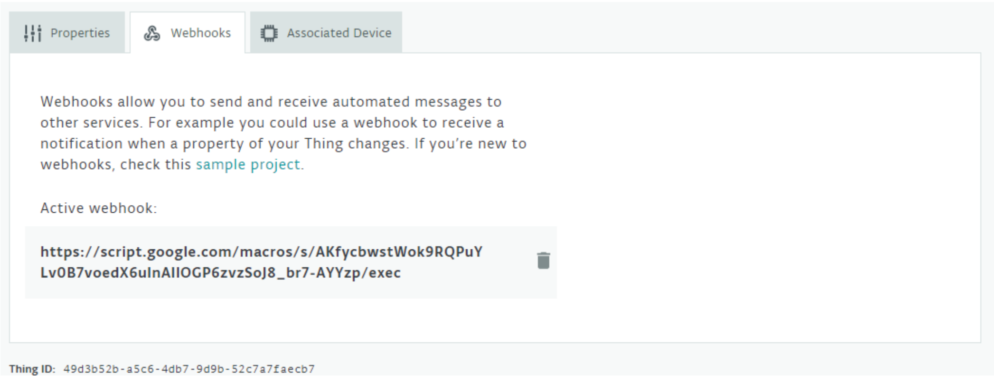
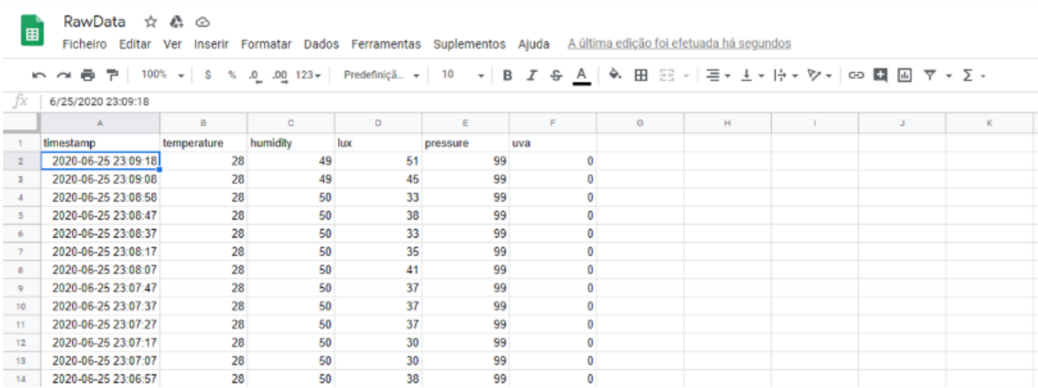


Figure 21 – Arduino Webhooks

In the end, the data gathered by the sensors was displayed both in the Arduino IoT Cloud dashboard and in the Google sheets (figure 22).



timestamp	temperature	humidity	lux	pressure	uva
2020-06-25 23:09:18	28	49	51	99	0
2020-06-25 23:09:08	28	49	45	99	0
2020-06-25 23:08:58	28	50	33	99	0
2020-06-25 23:08:47	28	50	38	99	0
2020-06-25 23:08:37	28	50	33	99	0
2020-06-25 23:08:17	28	50	35	99	0
2020-06-25 23:08:07	28	50	41	99	0
2020-06-25 23:07:47	28	50	37	99	0
2020-06-25 23:07:37	28	50	37	99	0
2020-06-25 23:07:27	28	50	37	99	0
2020-06-25 23:07:17	28	50	30	99	0
2020-06-25 23:07:07	28	50	30	99	0
2020-06-25 23:06:57	28	50	38	99	0

Figure 22 – Google sheet displaying real-time data

6.5. Access Google Drive and Google Sheets APIs

After pushing the environmental sensors data into Google Sheet, it was necessary to create a Dynamo script to read the data at a specific time interval and report it as a family parameter. The next steps followed the procedure described by BIMOne (2020) to use their visual script.

First, it was necessary to create Google API credentials:

1- Create a Google APIs console project to access https://console.cloud.google.com/

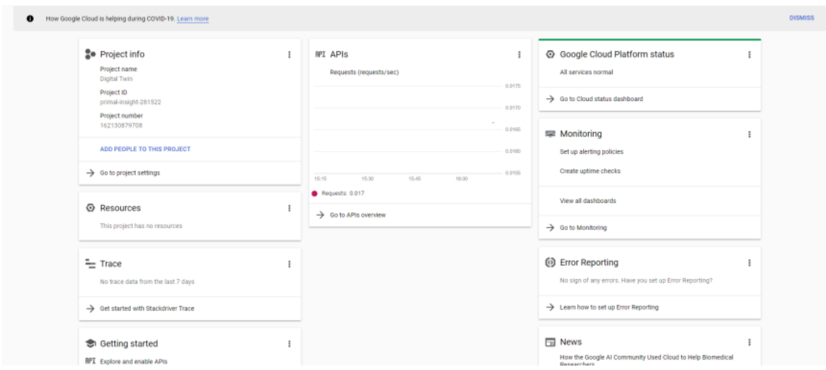


Figure 23 – Google API console project

2- Access the API console, selecting the project created on the previous point, navigate to ‘APIs & services’ and select ‘Library’. After enable Google Drive and Google Sheets API.

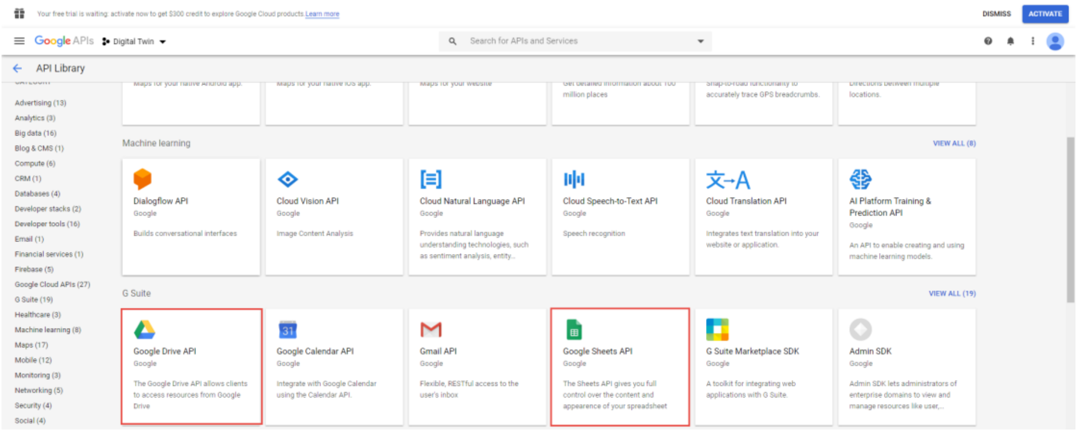


Figure 24 – API Library

3- Create OAuth Google API credentials for Google Drive and Google Sheets API.

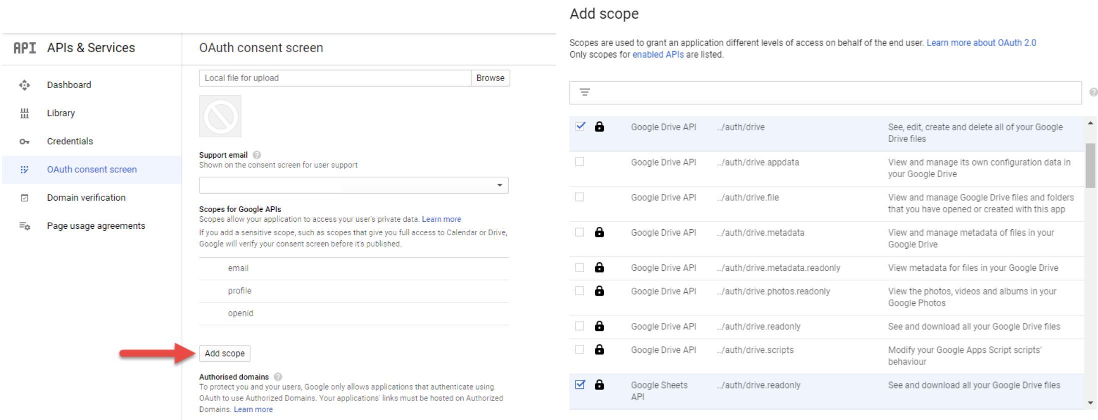


Figure 25 – OAuth add scope

4- Download the JSON file from the credentials tab. The file will then need to be copy-pasted to the folder of BIMOne package :

C:\Users\’user name’ \AppData\Roaming\Dynamo\DynamoRevit\2.3\packages\BIMOneGoogleAPI

6.6. Dynamo script

Once the access to the Google sheets was configured through API, it was necessary to create a dynamo script capable of reading the specific cell content and report it inside Revit (figure 26).

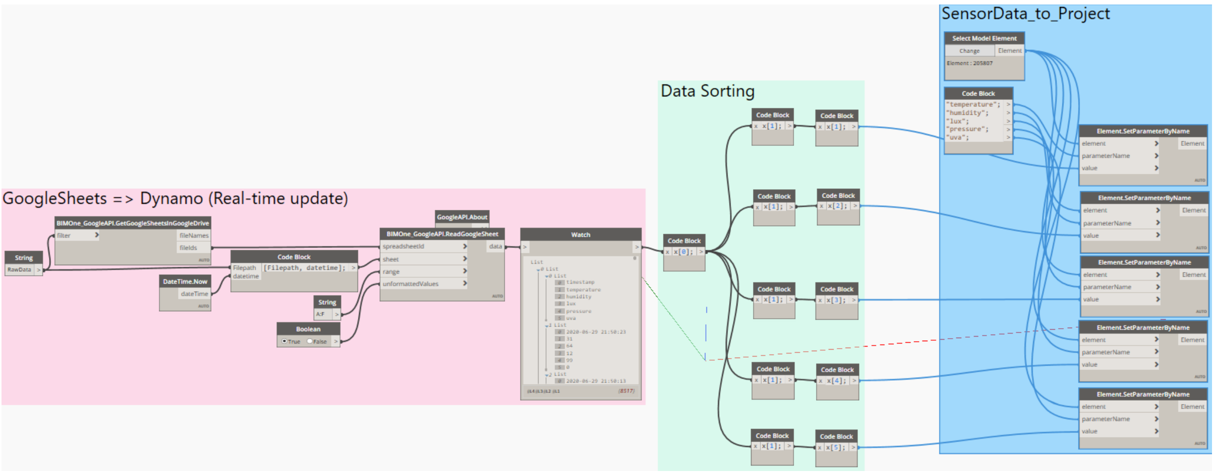


Figure 26 – Dynamo Script (complete)

The first part of the script (GoogleSheets to Dynamo) was responsible for selecting the specific sheets inside user Google Drive (figure 27). The node from BIMOne selected all the sheets on Google Drive while the string identified the name of the sheet 'RawData'. Using a Code Block, it added a timestamp to relate the data with the time it is extracted. The node watch showed the data gathered and the data structured.

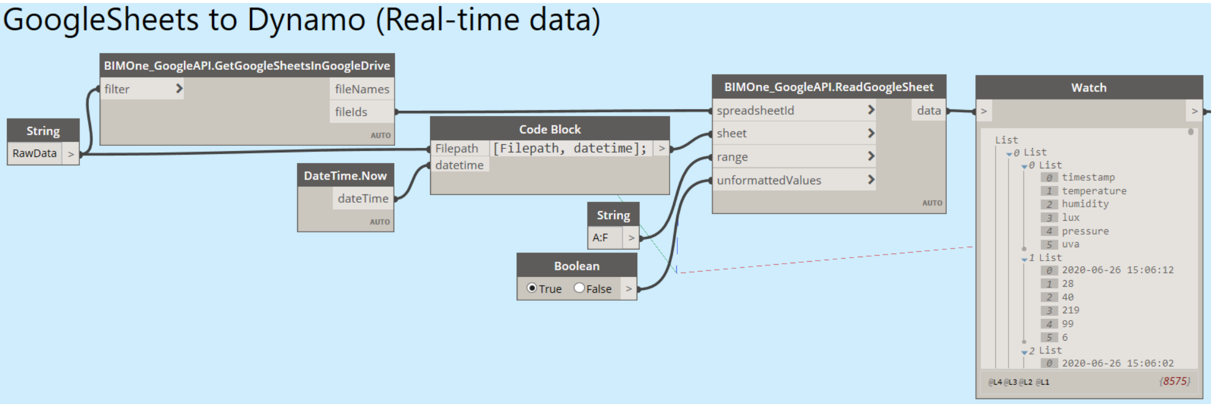


Figure 27 – Sheet selection

The second part of the script (Data Sorting) structured the information, extracting only the values of the list 1. The second group of nodes isolated each value of the multiple environment properties (figure 28).

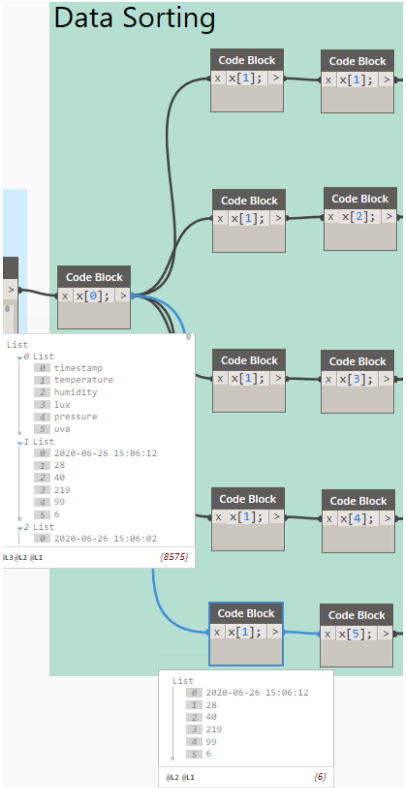


Figure 28 – Data structure

After, each value was associated with the corresponding parameter. To create this relation the ‘SensorData to Project’ script group was already using the family created with Revit (figure 29). The shared parameters created in the model and shared with the sensor family will assume the values reported in the Google Sheet and will be automatically updated every 10 seconds. This part of the script populates those parameters with real-time data from the Arduino MKV ENV SHIELD sensors.

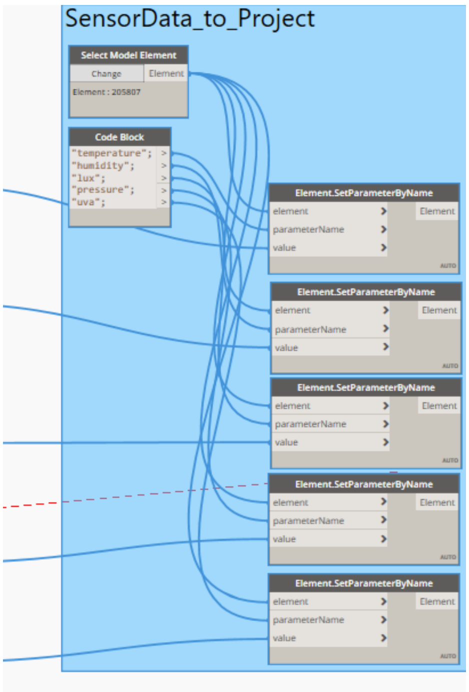


Figure 29 – Data connection to the family

To finish the dynamo script, instead of using manual or automatic run, it was defined that the script would be running ‘Periodically’ every 1000 milliseconds, being this the timeframe used to collect the data into the Arduino IoT platform. That ensured that the data indicated by the digital twin would be the most recent data.

6.7. Data into Digital Twin

As mentioned before, a family called ‘sensor’ was created in order to represent the real Arduino MRK WIFI 1010. The model works as a digital twin of the boards and communicates the data collected by the real asset (figure 30).

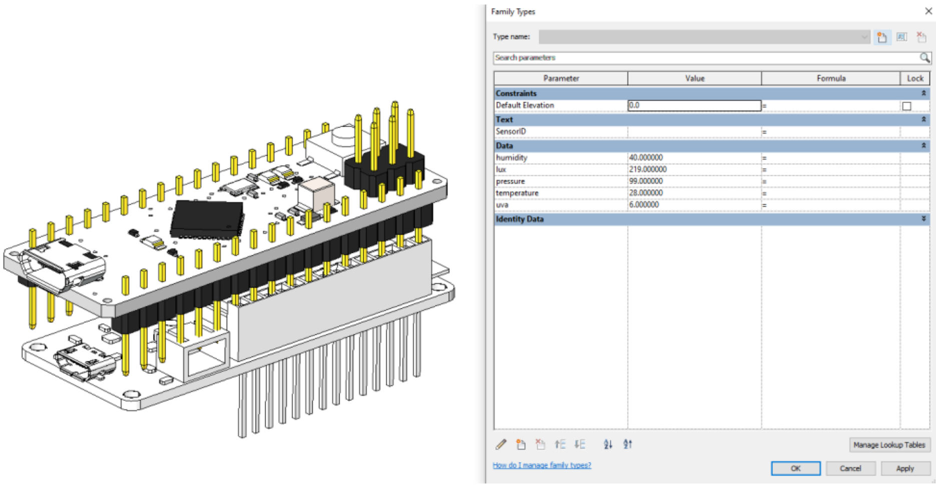


Figure 30 – Arduino digital twin

This is the element that is integrated into the model created through the use of photogrammetry technology. It allows a more accurate representation of the room’s conditions and the real position of the asset in the world which can be accessed by the real-world position of the room (figure 31).

Figure 31 – Room model using photogrammetry to generate point cloud

Once the real and the digital world were connected it was possible to access, both by the instance parameters and by generic tag, the real environmental data of the room according to the digital twin position in the model (figure 32).

Figure 32 – Digital Twin positioned on the BIM model (overlapping the point cloud) and displaying real-time data from the Arduino sensors

6.8. Data interpretation

Certainly, with a bigger scope and more development time, the current use case could have been improved showing the potential of data analytics. At this time, it was possible to expose the connection between DT and BIM through the connection between real-time environmental data associated with a real position of a BIM model in the world that determines the right conditions according to the city and country. Also, it is important to mention that the sensor board was positioned at a height of 1.00m above the ground floor.

Nevertheless, it was possible to demonstrate with this experience that for a certain space or building it is possible to determine if the required conditions and specifications demanded by the client and legislation are achieved. In this specific case, the objective was to verify if the room and the work area of the user were considered comfortable to promote better working conditions.

According to the Portuguese legislation for the thermal comfort, the interior temperature should be between 20ºC to 25ºC which means the temperature verified on the 26th of June 2020 at 15h 06minutes 12 seconds was 3ºC above the maximum temperature recommended for the interior spaces (Matias and Santos, 2013).

As for the air humidity values, the value (40%) is below 50% defined as the maximum limit for comfortable air conditions, according to the Portuguese legislation.

Regarding the luminosity, it is variable according to the spaces, however, for an office, it should be around 750 LUX which means the value of 219 LUX is not enough for the room purpose (Exporlux, N.D.). Also, the UVA radiation values are low, what may be related to both the position of the sensor in a corner of the room and to the window orientation to the north.

The atmospheric pressure value is constantly near 100hPa and has no relevant impact on the present use case.

To be able to test the complete condition of the room it would be necessary to analyze the data from the sensors positioned in multiple places of the room. However, for this academic work, the real-data obtained from the sensors showed that the place where the sensor was positioned in the real world does not offer good user conditions.

The data gathered would be useful to check the EIR and specifications defined in the BEP and/or other documentation important for the project, to verify the built conditions and adapt future projects, to promote space optimization with a redesign of the space layout, to introduce alternative mechanical systems increasing the light and air control, etc.

In conclusion, this use case exposes a simplified version of a system where digital twins are integrated into a BIM model communicating the real-time data. Further development could result in the automation of the processes that will automatically verify spatial conditions and transmit information to other DTs that will activate real-world objects to control the environmental conditions.

A good example that could be used in further exploration of this academic work will be related to the action activated by the sensors' value. For example, the decrease of luminosity could trigger an alarm on the user model that will provide information about the environmental statistics and he/she will be presented with alternatives to control the comfort levels, like opening a curtain shade or turning on the light. Other possibilities will not require user action and will benefit from ML/AI integration. In this case, the computer will decide on its own and will consider other inputs like the hour, the day an month of the event. This will mean that depending on the day time and even on the day and month of the event, the system will decide either to open the curtain shade if it is still day time and it is not too hot, or it will simply turn on the light and control its intensity to reach the recommended levels of luminosity for the defined space.

If we were able to integrate a vast amount of real-time data into a DT in an automated environment, we will contribute to a more dense and consistent ecosystem where decisions are a test and pattern-based instead of a creative or single project experience knowledge.

7. CONCLUSIONS

The full potential of data utilization in the AECO context remains to be seen. Data is generated without a clear state of necessity which results in a waste of time, money, and information. From the strategic definition to the use stage, and even before any design starts, data is being composed and a branch of complex interconnections is taking place inside the BIM core processes. From the point where projects began to formalize the concepts, data starts to increase in complexity, volume, velocity, and variety in a dense CDE. A single project can generate a vast amount of data that most of the time is not even considered as relevant for present or future projects.

Only by evolving BIM data will allow the merge with other fields of knowledge, such as computer science, and the AECO industry will gain business value. As seen in other industries, like manufacturing with PLM processes, there is a value in the adoption of Big Data engineering and Big Data analytics. The potential of machine learning, deep learning, and artificial intelligence mechanisms are yet to be discovered.

Certainly small, and medium projects, and even some projects considered nowadays as bigger, may not bring together the full benefit of these mechanisms and systems. However, considering the research done, there are at least three ways to add the prefix 'big' to data: multiple projects database to prevent design or construction mistakes and decrease construction time and waste; densified projects to expose patterns and relations; and digital Twins with real-world properties put in perspective with the amount of data automatically generated by sensors and other devices.

In all the previous cases, the data properties will fully benefit from machine learning and artificial intelligence algorithms. Data can be understood by computers as long as it is in the right format. If data is managed and processed by a mechanism, it will mean that we will be advised by the computer about how to better implement requirements and align them with the user and environmental needs.

AECO industry needs to understand the data value instead of using it in an autistic and uncontrolled way. Only after defining the needs and importance of data, all the stakeholders will be able to consider user experience and business benefits in the same equation. The entire lifecycle of a building is responsible for generating data and the capacity to track the information flow is one of the key elements to promote the industry's technological and operation maturity. Big Data generated by the physical and digital world has the power to align the AECO to other industries.

The interconnection of BIM and the DTs ecosystem may increase the richness of the digital world with a more precise and accurate relation to the future built reality. Despite the uncertainty about the boundaries of BIM and DT, it is believed that the main focus should be the integration of both, to reach a more complete and capable system to manage the information through planning, building, operation, and integration (PBO-I).

Furthermore, the interconnection of DT's in an ecosystem that benefits from a connected data environment, another kind of CDE, will promote automatic optimization of the design and construction as they will benefit from external factors that will influence the performance of each DT. For instance, an HVAC DT will be affected by environmental data, such as sun orientation, but also by the temperature generated, and communicated to its DT, by another equipment and users that will increase the interior's temperature and humidity. The data gathered will not be a static communication of the values but rather an integrated systematization of decisions that will collect, structure, and analyse data to automatically propose solutions or identify issues.

The use case presented in the last chapter is a test to reveal that the systems are becoming easily accessed and can improve all projects despite their scale. It shows the increasing capacity to verify requirements and implement solutions to minimize or eliminate project issues. The test tries to expose the necessity of a decision-making process based on real data instead of options based on instinct. It is believed that the increasing complexity of the current projects may result in the increasing of issues related to design and construction that will later influence the operation and use of buildings and assets. Data can support the decisions in a semi or completely automatized way to prevent mistakes as a result of undiscovered patterns that are not human-readable.

It is not believed that everything exposed on this academic work will be achieved without mistakes and maintaining traditional processes. There are also those who do not believe that BIM generates enough information to be considered as Big Data and that artificial intelligence will not be useful in the AECO context. However, working with DiRoots company and seeing the projects they are involved in, exposed the integration systems' advantages and how they reduce time and data waste.

In conclusion, the research done revealed that some players in the industry are making improvements regarding the subjects mentioned before. Also, technology companies are investing in systems that will integrate data-driven design and data-supported decision-making. Some identify DT as an evolution of the BIM process, bringing more maturity and the integration of operation stages. The data flow identified in this process exposes the amount and format of data produced during the building's lifecycle and the introduction of DT works as an expansion of BIM methodology into a connected and integrated data system.

Due to the extent of topics mentioned in this academic work and the peculiar conditions that affect the world at this moment, deeper knowledge may remain to be developed and explored. The need for a framework that integrates some of the raised subjects could be the first major step for future research.

This page is intentionally left blank

REFERENCES

Journal article

Benner, J. and McArthur, J. J. (2019) 'Data-driven design as a vehicle for BIM and sustainability education', *Buildings*, 9(5). doi: 10.3390/buildings9050103.

Bi, Z. and Cochran, D. (2014) 'Big Data analytics with applications', *Journal of Management Analytics*, 1(4), pp. 249–265. doi: 10.1080/23270012.2014.992985.

Bilal, M., Oyedele, L. O., Akinade, O. O., *et al.* (2016) 'Big Data architecture for construction waste analytics (CWA): A conceptual framework', *Journal of Building Engineering*, 6, pp. 144–156. doi: 10.1016/j.jobe.2016.03.002.

Bilal, M., Oyedele, L. O., Qadir, J., *et al.* (2016) 'Big Data in the construction industry: A review of present status, opportunities, and future trends', *Advanced Engineering Informatics*. Elsevier Ltd, 30(3), pp. 500–521. doi: 10.1016/j.aei.2016.07.001.

Binnekamp, R. (2019) 'Preference-Based Design in Architecture', *ResearchGate*. IOS Press Nieuwe Hemweg 6b 1013 BG Amsterdam The Netherlands, (May), pp. 1–19.

Boire, R. (2013) 'Predictive analytics: The power to predict who will click, buy, lie, or die', *Journal of Marketing Analytics*, 1(3), pp. 184–185. doi: 10.1057/jma.2013.14.

Caragliu, A., del Bo, C. and Nijkamp, P. (2011) 'Smart cities in Europe', *Journal of Urban Technology*, 18(2), pp. 65–82. doi: 10.1080/10630732.2011.601117.

Carrillo, P., Harding, J. and Choudhary, A. (2011) 'Knowledge discovery from post-project reviews', *Construction Management and Economics*. doi: 10.1080/01446193.2011.588953.

Chen, H. *et al.* (2012) 'Business Intelligence Research Business Intelligence and Analytics: From Big Data To Big Impact', *MIS Quarterly*, 36(4), pp. 1165–1188. Available at: www.freakonomics.com/2008/02/25/hal-varian-answers-your-questions/.

Comiskey, D. *et al.* (2017) 'An analysis of data sharing platforms in multidisciplinary education', *Architectural Engineering and Design Management*. Taylor & Francis, 13(4), pp. 244–261. doi: 10.1080/17452007.2017.1306483.

Correa, F. R. (2015) 'Is BIM big enough to take advantage of Big Data analytics?', *32nd International Symposium on Automation and Robotics in Construction and Mining: Connected to the Future, Proceedings*, (June 2015). doi: 10.22260/isarc2015/0019.

Ekanayake, L. L. and Ofori, G. (2004) 'Building waste assessment score: Design-based tool', *Building and Environment*, 39(7), pp. 851–861. doi: 10.1016/j.buildenv.2004.01.007.

Grieves, M. (2014) 'Digital Twin : Manufacturing Excellence through Virtual Factory Replication This paper introduces the concept of a A Whitepaper by Dr . Michael Grieves', *White Paper*, (March). Available at: https://www.researchgate.net/publication/275211047_Digital_Twin_Manufacturing_Excellence_through_Virtual_Factory_Replication.

Holzwarth, V. *et al.* (2019) 'Data driven value creation in AEC along the building lifecycle', *Journal of Physics: Conference Series*, 1343(1). doi: 10.1088/1742-6596/1343/1/012046.

Howard, R. and Björk, B. C. (2008) 'Building information modelling - Experts' views on standardisation and industry deployment', *Advanced Engineering Informatics*, 22(2), pp. 271–280. doi: 10.1016/j.aei.2007.03.001.

Hu, H. *et al.* (2014) 'Toward scalable systems for Big Data analytics: A technology tutorial', *IEEE Access*. IEEE, 2, pp. 652–687. doi: 10.1109/ACCESS.2014.2332453.

Kim, H., Soibelman, L. and Grobler, F. (2008) 'Factor selection for delay analysis using Knowledge

Discovery in Databases', *Automation in Construction*. doi: 10.1016/j.autcon.2007.10.001.

Loyola, M. (2018) 'Big Data in building design: A review', *Journal of Information Technology in Construction*, 23(November), pp. 259–284.

Meier, H., Roy, R. and Seliger, G. (2010) 'Industrial Product-Service systems-IPS2', *CIRP Annals - Manufacturing Technology*, 59(2), pp. 607–627. doi: 10.1016/j.cirp.2010.05.004.

Parrott, A. and Warshaw, L. (2017) 'Industry 4.0 and the digital twin', *Deloitte University Press*, pp. 1–17. Available at: https://dupress.deloitte.com/dup-us-en/focus/industry-4-0/digital-twin-technology-smart-factory.html.

Provost, F. and Fawcett, T. (2013) 'Data Science and its Relationship to Big Data and Data-Driven Decision Making', *Big Data*, 1(1), pp. 51–59. doi: 10.1089/big.2013.1508.

Tao, F. *et al.* (2019) 'Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison', *Engineering*. Chinese Academy of Engineering, 5(4), pp. 653–661. doi: 10.1016/j.eng.2019.01.014.

Tsai, C. W. *et al.* (2015) 'Big Data analytics: a survey', *Journal of Big Data*. Springer International Publishing, 2(1), pp. 1–32. doi: 10.1186/s40537-015-0030-3.

Vanlande, R., Nicolle, C. and Cruz, C. (2008) 'IFC and building lifecycle management', *Automation in Construction*, 18(1), pp. 70–78. doi: 10.1016/j.autcon.2008.05.001.

Book chapter

Arayici, Y. *et al.* (2017) '*Heritage building information modelling', Heritage Building Information Modelling*, pp. 40-41 doi: 10.4324/9781315628011.

Eastman C.; Teicholz P.; Sacks, R. . L. K. B. (1886) '*Handbook – a guide to Building Information Modeling for owners, managers, designers, engineers, and contractors'*, *John Wiley & Sons, Inc.,* pp. 266-281. doi: 10.1093/nq/s7-II.32.110-e.

Farsi, M. *et al.* (2020) '*Internet of Things Digital Twin Technologies and Smart Cities'*. Pp. 6-7 https://doi.org/10.1007/978-3-030-18732-3

Han, J. and Kamber, M. (2001) *Data Mining: Concepts and Techniques*, *Data Mining: Concepts and Techniques*. Academic Press, Morgan Kaufmarm Publishers, 2001. pp. 66-68 doi: 10.1016/C2009-0-61819-5.

Krijnen, T. and Tamke, M. (2015) 'Assessing Implicit Knowledge in BIM Models with Machine Learning', in *Modelling Behaviour*. Springer International Publishing, pp. 397–406. doi: 10.1007/978-3-319-24208-8_33.

Lu, W. and Lu, W. (2018) *Big Data for construction cost management*, *BIM and Big Data for Construction Cost Management*. doi: 10.1201/9781351172325-6.

Conference proceedings

Bertoni, A. (2018) 'Role and Challenges of Data-Driven Design in the Product Innovation Process', *IFAC-PapersOnLine*. Elsevier B.V., 51(11), pp. 1107–1112. doi: 10.1016/j.ifacol.2018.08.455.

Bertozzi, C. (2009) 'Compositions and Methods for Modification of Biomolecules', *IFIP Advances in Information and Communication Technology*, 517(19), pp. 324–334. doi: 10.1007/978-3-319-72905-3.

Bolpagni, M. and Ciribini, A. L. C. (2016) 'The Information Modeling and the Progression of Data-Driven Projects', *Proceedings of the CIB World Building Congress 2016. Volume III. Building up business operations and their logic. Shaping materials and technologies*, (May), pp. 296–307. Available at: http://www.bimthinkspace.com/2016/07/the-many-faces-of-lod.html.

Deutsch, R. (2015) 'Leveraging data Across the Building Lifecycle', *Procedia Engineering*. Elsevier B.V., 118, pp. 260–267. doi: 10.1016/j.proeng.2015.08.425.

Kargah-Ostadi, N. (1993) 'Computing in Civil and Building Engineering', in *Computing in Civil and*

Building Engineering, pp. 1222–1229.

De Mauro, A., Greco, M. and Grimaldi, M. (2014) ‘What is Big Data? A Consensual Definition and a Review of Key Research Topics Call for Papers: In (Big) Data we trust: Value Creation in Knowledge Organizations View project COST ACTION “OPen innovation Excellence Network (OPEN)” View project’, in. doi: 10.13140/2.1.2341.5048.

Reid, J. B. and Rhodes, D. H. (2016) ‘Digital System Models: An investigation of the non-technical challenges and research needs’, in *Conference on Systems Engineering Research*, pp. 1–10. Available at: http://seari.mit.edu/documents/preprints/REID_CSER16.pdf.

Standards

British Standards Institution (BSI) (2019) ‘BSI Standards Publication Transition guidance to BS EN ISO 19650’, pp. 1–40.

ISO 19650-1 (2018) ‘Organization and digitization of information about buildings and civil engineering works , including building information modelling (BIM) - Information management using building information modelling’, Part 1: Co, pp. 1–46.

ISO 19650-2 (2018) ‘Organization and digitization of information about buildings and civil engineering works , including building information modelling (BIM) - Information management using building information modelling’, pp. 1–46.

NIST (2014) *NIST Big Data Interoperability Framework: Volume 1, Definitions*.

RIBA (2020) ‘RIBA Plan of Work 2020’, *Health and Safety*, pp. 8–9. doi: 10.4324/9780429346637-2.

Richards, M. (2013) ‘Building information management: A standard framework and guide to BS 1192 BSI Standards’, *BSI Standards Publication*, (1), pp. 1–68. doi: Published by the British Standard Institute. British Standard Limited. ISSN9780580781360. /BIM TASK GROUP.

UK BIM Alliance (2019a) ‘Information Management according to BS EN ISO 19650 - Guidance Part 1: Concepts’, *UK BIM Alliance*, (july), p. 42. Available at: https://www.ukbimalliance.org/stories/information-management-according-to-bs-en-iso-19650/.

UK BIM Alliance (2019b) ‘Information Management according to BS EN ISO 19650 - Guidance Part 1: Concepts’, *UK BIM Alliance*, (April), pp. 1–42.

Non-Print Material

Autodesk, 2019 (2019) ‘Dialects of Digital Twins’, p. 1. Available at: https://www.autodesk.com/campaigns/digital-twin (Accessed: 22th July 2020)

BSI Group (2019) ‘Digital twins for the built environment’, *BIMToday*, p. 24. Available at: https://www.pbctoday.co.uk/news/bim-news/digital-twin-4-0/64519/. (Accessed: 22th July 2020)

Robins, B. (2019) ‘Advancing BIM: Digital Twins’. 2019 Bentley Systems.

Trimble Inc. (2019) ‘Why BIM Needs Digital Twins’.

Report/Article

Bolton, A. *et al.* (2018) ‘The Gemini Principles’, p. 15. doi: 10.17863/CAM.32260. Available at: https://www.cdbb.cam.ac.uk/Resources/ResoucePublications/TheGeminiPrinciples.pdf (Accessed:22th July 2020)

buildingSMART (2020) ‘*Enabling an Ecosystem of Digital Twins*’.

Deloitte (2017) ‘*New Technologies Case Study: Data Sharing in Infrastructure*’.

Laney, D. (2004) ‘Evidence of two effects in the size segregation process in dry granular media’, *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 70(5), p. 16. doi: 10.1103/PhysRevE.70.051307.

Mussomeli, A., Gish, D. and Laaper, S. (2015) ‘The Rise of the Digital Supply network’, *Deloitte*, 45(3), pp. 20–21. Available at: http://ezproxy.lib.monash.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=110198124&site=ehost-live&scope=site.

Mohammadi, N. and Taylor, J. (2020) 'Knowledge Discovery in Smart City Digital Twins', in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, pp. 1656–1664. doi: 10.24251/hicss.2020.204.

UK BIM Alliance (2018) 'A Fresh Way Forward for Product Data - State of the Nation', pp. 1–37.

Website

Accruent (2020). 'Ensures Global and Immediate Access to Engineering Documentation' [Online].

https://info.accruent.com/rs/167-BOY-362/images/48373_MeridianManufacturing.pdf?_ga=2.240375043.1299790203.1593979664-1692842576.1593979664 (Accessed: 24 July 2020)

Allplan (2020). 'BIM and the Digital Twin Model' [Online]. Available at:

https://blog.allplan.com/en/bim-and-the-digital-twin-model (Accessed: 24 July 2020)

Bolpagni M. (2016). 'The Many Faces of "LOD"' [Online]. Available at:

https://www.bimthinkspace.com/2016/07/the-many-faces-of-lod.html

Borne K. (2014). ' Top 10 Big Data Challenges – A Serious Look at 10 Big Data V's' [Online]. Available at:

https://mapr.com/blog/top-10-big-data-challenges-serious-look-10-big-data-vs/ (Accessed: 24 July 2020)

buildingSmartt Internation (N.D.). 'Model View Definitions (MVD)' [Online]. Available at:

https://www.buildingsmart.org/standards/bsi-standards/model-view-definitions-mvd/(Accessed: 24 July 2020)

Burger R. (2019). 'How the construction Industry is Using Big Data' [Online]. Available at:

https://www.thebalancesmb.com/how-the-construction-industry-is-using-big-data-845322 (Accessed: 24 July 2020)

Chastain A. (2017). 'Five big benefits of data-driven design' [Online]. Available at:

https://builtworlds.com/news/five-big-benefits-of-data-driven-design/ (Accessed: 24 July 2020)

Cross A. (2018). ' The Importance of Scalability in Big Data Processing' [Online]. Available at: https://www.ngdata.com/the-importance-of-scalability-in-big-data-processing/ (Accessed: 24 July 2020)

Dassault Systemes (N.D.). '3DEXPERIENCE connector for REVIT' [Online]. Available at: https://www.3ds.com/partners/products/200000000086802_PDT00000001922_3DEXPERIENCE_Connector_For_Revit_/(Accessed: 24 July 2020)

Designing Buildings (2020). 'Data drops for BIM ' [Online]. Available at:

https://www.designingbuildings.co.uk/wiki/Data_drops_for_BIM (Accessed: 24 July 2020)

Designing Buildings (2019). 'Level of detail for BIM' [Online]. Available at:

https://www.designingbuildings.co.uk/wiki/Level_of_detail_for_BIM) (Accessed: 24 July 2020)

Designing Buildings (2019). 'PAS 1192-3' [Online]. Available at:

https://www.designingbuildings.co.uk/wiki/PAS_1192-3 (Accessed: 24 July 2020)

Designing Buildings (2019). 'PAS 1192-5:20115' [Online]. Available at:

https://www.designingbuildings.co.uk/wiki/PAS_1192-5:2015 (Accessed: 24 July 2020)

Google (N.D.) 'Understand how structured data works' [Online]. Availabe at: https://developers.google.com/search/docs/guides/intro-structured-data (Accessed: 24 July 2020)

Interantion Organization for Standardization (2001). 'IEC 82045-1:2001(en) Document management — Part 1: Principles and methods' [Online]. Available at:

https://www.iso.org/obp/ui/#iso:std:iec:82045:-1:ed-1:v1:en) (Accessed: 24 July 2020)

Meridian Enterprise (2020). 'Accruent Meridian Enterprise 2020 Supported Software' [Online].Available at:

https://documentation.bluecieloecm.com/BCWebHelp/en/meridian/2020/ss/Content/EasySync/Authoring%20applications.htm (Accessed: 24 July 2020)

Meridian Enterprise (N.D.). 'Introducing Accruent Meridian Entreprises' [Online]. Available at:

https://documentation.bluecieloecm.com/BCWebHelp/en/meridian/2020/sr/Content/Meridian%20AG/Introducing%20InnoCielo.htm (Accessed: 24 July 2020)

NBS (N.D.). 'What is the PAS 1192 framework?' [Online]. Available at:

https://www.thenbs.com/knowledge/what-is-the-pas-1192-framework (Accessed: 24 July 2020)

Ofnisytems (N.D.). 'User Requirements Specifications' [Online]. Available at:

http://www.ofnisystems.com/services/validation/user-requirement-specifications/ (Accessed: 24 July 2020)

Stoica I. (N.D). 'Conquering Big Data with Spark and BDAS' [Online]. Available at: https://www.usenix.org/conference/icac14/technical-sessions/presentation/conquering-big-data-spark-and-bdas (Accessed: 24 July 2020)

Taylor C. (2018) 'Structured vs. Unstructured Data' [Online]. Available at: https://www.datamation.com/big-data/structured-vs-unstructured-data.html (Accessed: 24 July 2020)

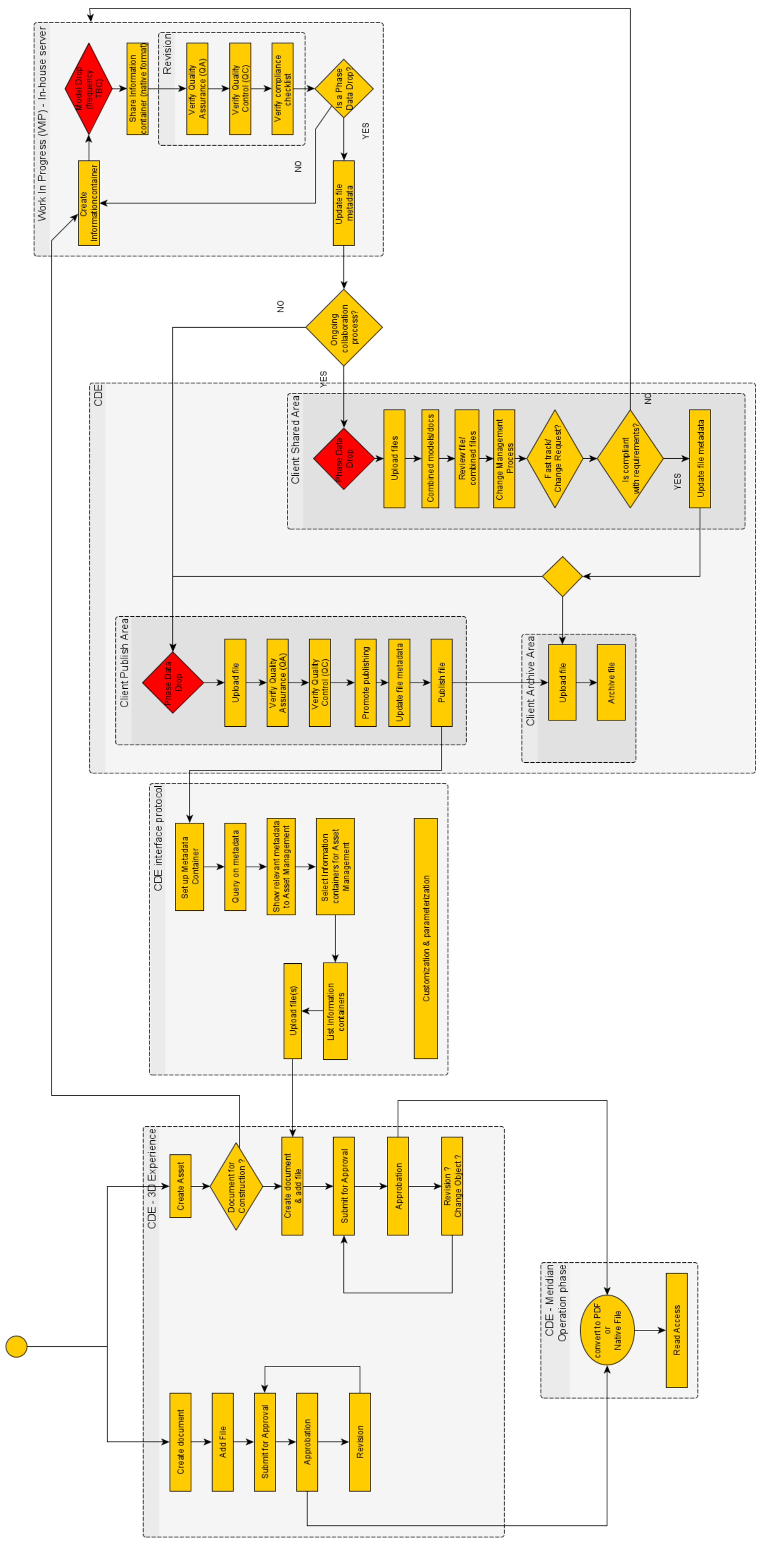
LIST OF ACRONYMS AND ABBREVIATIONS

API	Application Programming Interface
ACE	Architects Council of Europe
AECO	Architecture, Engineering, Construction, and Operation
AI	Artificial Intelligence
AIM	Asset Information Model
ANN	Artificial Neural Networks
AIR	Asset Information Requirements
AR	Augmented Reality
BC	Blockchain
BDAS	Berkeley Data Analytics Stack
BEP	BIM Execution Plan
BDA	Big Data Analytics
BDE	Big Data Engineering
BIM	Building Information Modelling/Management
BoM	Bill of Materials
bSI	buildingSmart International
DAG	Directed Acylic Graphs
CAD	Computer Aided Design
CDE	Common Data Environment
CDE	Connected Data Environment
COBie	Construction Operations Building information exchange
CWA	Construction Waste Analytics
DT	Digital Twins
EBD	Evidence-based design
EIR	Exchange Information Requirements
FDA	Food and Drug Administration
FM	Facility Management
GA	Genetic Algorithm
GIS	Geographical Information Systems
HDFS	Hadoop Distributed File System
HSP	Horizontal Scaling Platforms
IDM	Information Delivery Plan
IFC	Industry Foundation Classes
IoT	Internet of Things
IT	Information Technology
ISO	International Organization for Standardization
LOD	Level of Detail/Level of Definition
LOI	Level of Information
MIDP	Master Information Delivery Plan
ML	Machine Learning
MR	MapReduce

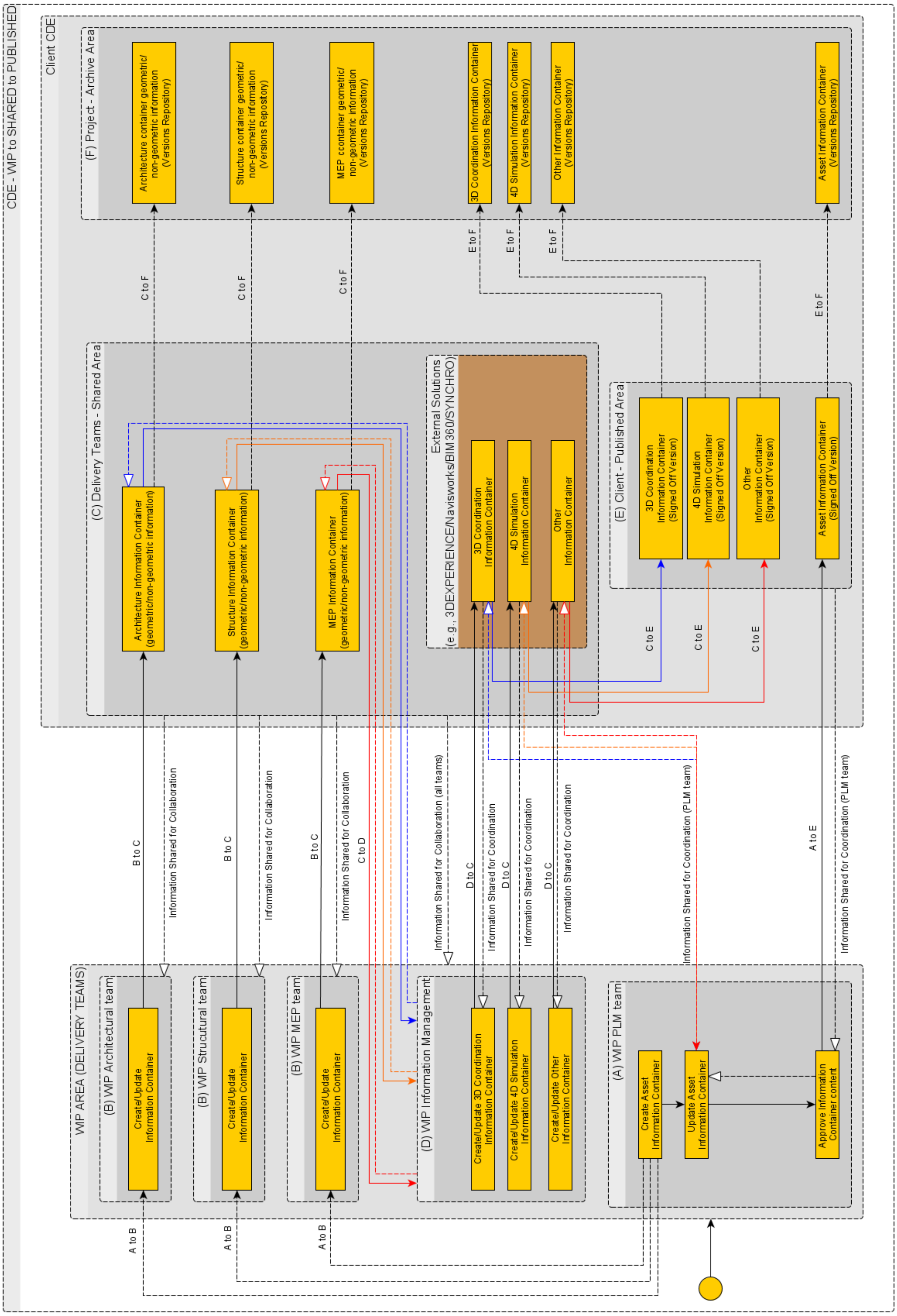
MVD	Model View Definition
NDT	UK National Digital Twin
OIR	Organization Information Requirements
OT	Operation Technology
PBO-D	Plan-Build-Operate-Decommission
PBO-I	Plan-Build-Operate-Integrate
PDM	Project Delivery Manager
PIP	Project Implementation Plan
PIR	Project Information Requirements
PLM	Product Lifecycle Management
QTO	Quantity Take-off
RFID	Radio Frequency Identifications
RIBA	Royal Institute of British Architects
TIDP	Task Information Delivery Plan
URS	User Requirements Specifications
VR	Virtual Reality
VSP	Vertical Scaling Platforms
SVM	Support Vector Machines

APPENDICES

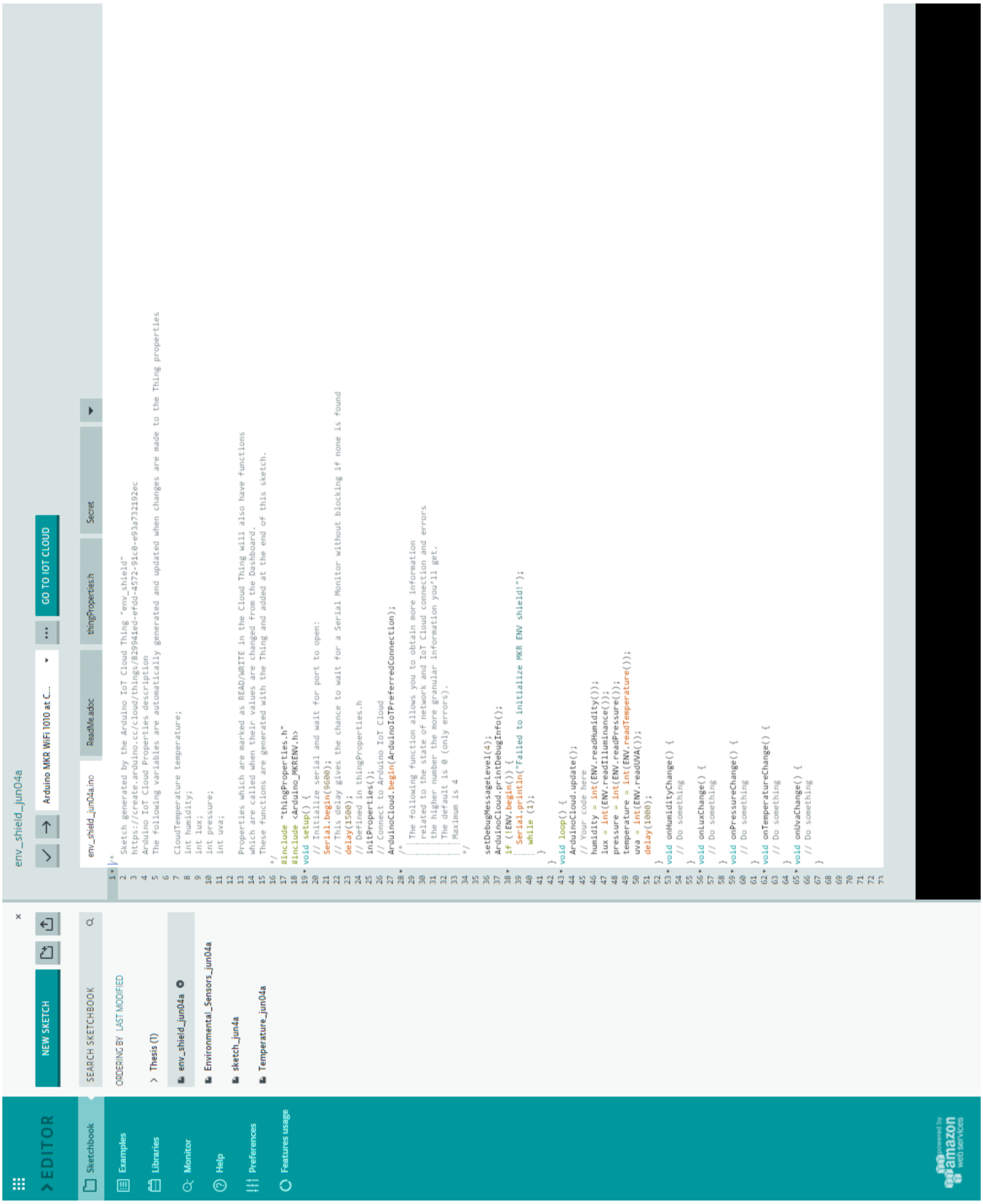
APPENDIX 1: FLOW DIAGRAM – MULTIPLE CDE'S



APPENDIX 2: FLOW DIAGRAM – SINGLE CDE



APPENDIX 3: ARDUINO IOT MKR WIFI 1010 + ENV SHIEL CODE



APPENDIX 4: DYNAMO VISUAL SCRIPT