POLITECNICO DI MILANO

Master in

Building Information Modelling

European Master in
Building Information Modelling

Systematic BIM Models Quality Check

Supervisor:

Claudio Mirarchi

Supervisor:

Miguel Azenha

Author:

Mohamed Safwat

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

a.a. 2020/2021

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

ACKNOWLEDGEMENTS

I would like to express my gratitude to My GOD, my belief in him always gives me the support to continue my path without any hesitations. After that, I am grateful to my parents and friends, who give me unconditional love and support.

My acknowledgments and gratitude are expressed in appreciation to the European Commission, which supported me during the whole BIM A+ Master Program with an Erasmus Mundus scholarship. Without it, this work would not be possible.

I would like to thank Prof. Dr-Miguel Azenha and Prof. Dr Claudio Mirarchi for their exceptionally helpful comments and guidance during the review process. Although the endless requests, revisions, and modification that sucked the life out of a poor master student. But it was really helpful and push my work further and beyond my limits.

No words can describe my gratitude to the unknown soldier who appeared from nowhere when I was seeking help and guidance in programming. Mohamed Rashad, your tutoring and support are more than appreciated, and you proved to me that good people still exist in our world. Also, I want to thank Sara Amiri Rigi for her support and her help.

Lastly, I want to thank Mr Google, StackOverflow members, and Lynda website for their unlimited sources and for answering my doubts and questions during my work process.

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 modelli di Building Information sono sviluppati secondo l'aggregazione di una serie di oggetti caratterizzati da informazioni grafiche e non grafiche. Oggi, questo insieme di informazioni è definito secondo il principio del livello di fabbisogno informativo (ISO 19650-1) che richiede un'identificazione a priori di tutte le informazioni necessarie per soddisfare gli obiettivi e il modello informativo specifico. Anche se i modelli informativi si basano su una semantica condivisa e su una struttura dati che facilita la trasmissione delle informazioni tra macchine (interoperabilità) e dalla macchina all'uomo, le peculiarità del settore delle costruzioni richiedono spesso l'integrazione di informazioni personalizzate. Inoltre, non tutte le informazioni integrabili in un oggetto sono correlate e/o parametrizzate. Ciò comporta la possibilità di problematiche legate alla qualità del modello che possono ostacolare sviluppi futuri.

Lo scopo di questo lavoro è quello di esplorare le principali classi di problemi di qualità che possono essere trovati in un modello di building information e fornire soluzioni sia in termini di processi e regole che possono essere integrati nei documenti di Exchange Information Requirements (EIR), sia di sviluppare sistemi automatici o semiautomatici che possono aiutare i progettisti, le imprese edili, il committente, ecc. nel controllare che tutte le informazioni siano compilate secondo le specifiche esigenze, ovvero che il modello sia di buona qualità.

Come attività complementare, la ricerca dovrebbe guardare ai mezzi per comunicare i problemi di qualità in modo efficace anche a coloro che non hanno esperienza nella modellazione e/o lavorano con modelli informativi. In questa direzione il lavoro riguarderà lo sviluppo di una dashboard strutturata che dovrebbe essere automaticamente estratta dall'analisi della qualità sopra menzionata.

Parole chiave: qualità dei dati, automazione, analisi dei dati, gestione delle informazioni, controllo qualità, BMC.

ABSTRACT

Building information models are developed according to the aggregation of a series of objects characterised by graphical and non-graphical information. Today, this set of information is defined according to the level of information need principle (ISO 19650-1) that requires an a priori identification of all the information needed to satisfy the objectives and the specific information model. Even if information models are based on a shared semantics and a data structure that facilitates the transmission of information between machines (interoperability) and from the machine to the humans, the peculiarities of the construction sector often require the integration of personalised information. Moreover, not all the information that can be integrated into an object are related and/or parametrised. This brings to the possibility of issues related to the quality of the model that can hinder future developments.

The aim of this work is that of exploring the main classes of quality issue that can be found in a building information model and provide solutions both in terms of processes and rules that can be integrated in Exchange Information Requirements (EIR) documents, and to develop automatic or semi-automatic systems that can help the designers, the construction companies, the client, etc. in checking that all the information are compiled according to the specific requirements, i.e. that the model is of a good quality.

As a complementary activity, the research should look at means to communicate the quality issues in an effective way also to those with no experience in modelling and or working with information models. On this directing the work will look at the development of structured dashboard that should be automatically extracted from the quality analysis above mentioned.

Keywords: Data quality, Automate, Data analytics, Information Management, Quality Control, BMC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

BIM was invented in the construction industry in recent decades and its significant potentials led to fast development and application around the globe (World Economic Forum, 2018). BIM has left a considerable influence on the management and delivery of construction projects. It still has enormous potential to develop further. Taking advantage of the digitalization side that focuses on efficiency, BIM has added to the productivity of construction projects over time. Such management provides an interconnected set of information that is more accessible to all stakeholders (Mihindu and Arayici, 2008). Thus, all the participating parties in a project, like , owners, designers, users and manufacturers, would better have a clearer image of the project phases, related costs, and workflow. (Waterhouse, 2017)

BIM models are defined as geometrical and non-geometrical data and information(Wang, 2014). The main backbone of BIM is information that is collected either according to the needs of the process defined in the project context or the needs of the company context. But different features such as technical characteristics can limit the extent of information (Mirarchi and Pavan, 2019). Here interoperability among different systems becomes one of the most common issues that designers will face.

An essential challenge in the construction industry is still replacing manual processes that check the architectural data with digital ones. In this way, Not only effort and time will be saved, but also the possibility of human errors will reach a minimum level (Solihin, Eastman and Lee, 2015). A primary factor in the use of information management properly is validation and formal verification, called Digital Model Checking (DMC)(Eastman *et al.*, 2009). This process provides interconnected and collaborative information during the whole life cycle of the project and makes the data a more valuable asset to the BIM field.

Therefore, the quality of the model that will be exchanged and the process of reaching this model are of core significance (Törmä, 2013). There is a concept in computer science that says “garbage-in garbage-out (GIGO)” and believes nonsense input leads to faulty output. In BIM, the input model’s quality should be checked to guarantee the quality of exchanged ones. Several studies in BIM indicate that the quality of data contributes to the quality of the model, which is why the quality of the data model is critically significant. An example study carried out (de Farias, Roxin and Nicolle, 2018) shows that faulty results in the output of an algorithm are rooted in issues of data quality of input models.

The concept of data quality has always been familiar in the field of database management. So, it is not uncommon to be crucial in digitalized Architecture, Engineering, Construction and Operation (AECO) industry. Some researchers have analyzed the quality of models exchanged through IFC (Solihin and Eastman, 2015). However, only limited studies tried to understand issues that are possible to arise in the native models. Moreover, errors are inherent to data quality, and studies in Information Technology shows that it is tremendously difficult to reach an error-free database. Therefore, evaluation of data quality in BIM contributes to the development of automated processes (like rule checking processes) and proposes an automated structured process to guarantee the model`s quality.

1.1. Motivations

According to a study about BIM adoption globally (Succar and Kassem, 2015) there are three stages in BIM, named modelling, collaboration and integration. Once someone takes control of one stage entirely and goes to the next, the efficiency and opportunity for innovation start to grow. The first stage of digital modelling has developed much faster than the second and third ones. That is true while collaboration and integration turn to be of higher importance regarding BIM uses such as authoring design, specification, and quantity-take off.

The motivation here can fall under one of the following two categories. First, the study focuses on quality issues in the real world, including accuracy, consistency, and completeness. Second, the studies dealing with the quantity of data that is created in building an information model. The first studies try to identify quality issues in the practical field and figure out how factors have a significant effect on BIM model quality. Proposed solutions also try to measure the data quality of the information model based on testing and checking semantically the model information.

1.2. Objectives

Regarding this outcome, the purpose is to focus on quality issues in the real world, including accuracy, consistency, and completeness. These objectives are also gathered within a model quality checking to take significant steps toward an optimum model quality.

From motivations and to reach the above objectives, a review of literature and questionnaire is carried out to reply to the following questions:

What is the knowledge of BIM model quality currently?

How is this knowledge associated with AEC industry requirements?

What is to be done to improve the current modelling quality process?

What can be added or modified to the current state to create a quality BIM model?

1.3. Methodology

The methodologies to reach those objectives are thought as shown in Figure 1 ; a) comprehensive review about BIM model quality. b) Data collection that has been done using a questionnaire and interviews. c) A data analysis to cluster the different topics and identify the areas of interest. d) A development stage considering different techniques. e) A conclusions stage with the analysis results and final product.

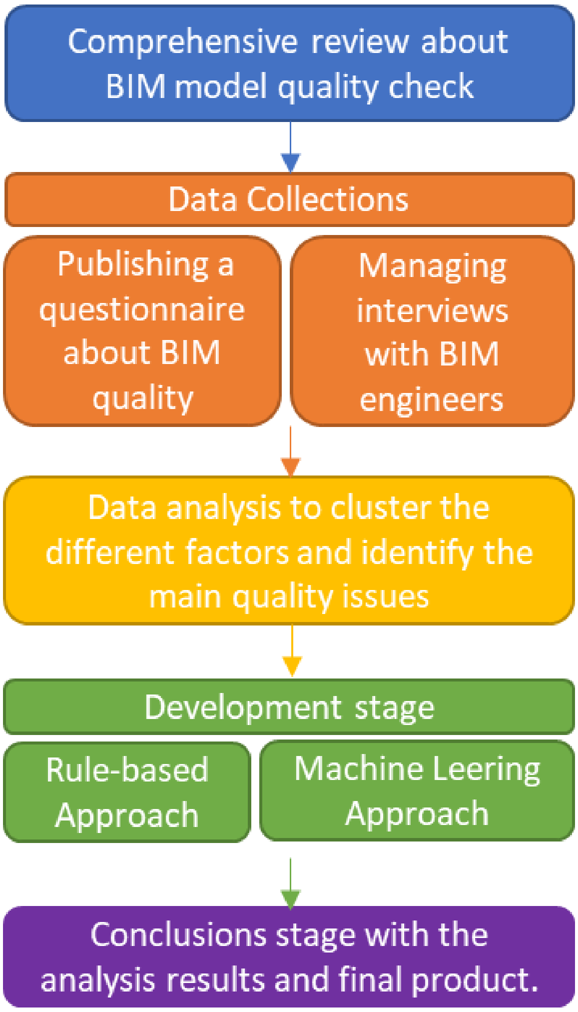


Figure 1 - Research workflow and methodology

1.4. Thesis structure

Apart from the ***"Introduction"***, the four following chapters shape the main structure of this dissertation. Named ***"BIM model quality challenges",*** chapter two provides an overview of the meaning of model checking in main concepts for quality checking, code compliance, and detecting clashes between models or even in the same model. These chapters are subdivided into sections, A) Overview of BIM model quality check; B) Review of Questionnaires in BIM; C) BIM and Machine Learning intersections.

The following chapter, ***"A Quantitative survey in BIM quality."*** Is a highlighted area of how the different companies and firms are trying to reach a well and significant model with an acceptable quality according to the different BIM uses?. For reaching such an objective, interviews have been conducted with practitioners of different BIM roles in different companies, and a survey has been conducted. This data is to be analyzed to reach the results in this chapter. This chapter is divided into three subsections. The first section reviews the questionnaire and interview questions and structures. The second illustrates the results and the summary of it from different perspectives. Moreover, the final section indicates the review of interviews` outcomes and discussions which stated the main and common issues and problems which have a high impact on the BIM model quality; those outcomes were discussed and come up from the interviews discussions and questionnaire results. Moreover, it has been illustrated those issues with their impacts and the proposed solutions.

Chapter Foure, with a ***"Consistency check for BIM models"*** target, has been divided into several subsections. The first section indicates the selected issue from the previous chapter issues . Section two discusses the working frame and road map to solve the selected proposal. Section three illustrates the coding structures, algorithms and tries to code and check if it works according to the plan. Finally, the test section is about testing the add-in that is created in a real case scenario and discusses the future development possibilities.

Lastly, in Chapter Five, ***"The conclusion"*** states the results of the developed work. It indicates further developments for the BIM models Quality checking to reach a higher level of quality.

2. BIM MODELS QUALITY CHALLENGES

2.1. Overview of BIM model quality check

According to the McGraw-Hill Construction, Research and Analytics survey (2007) named "The business value of BIM for construction in major global markets , that Architects employs about an average of 50 hours per project for code checking, and above 100 hours have been spending from 6% of engineers. In the meantime, Approximately 85% of architects are enthusiastic about using model checking tools to aid in code review. On the other hand, engineers appear to be driving efforts to improve the flow of data utilized for checking. In addition, the report indicates that engineers have attempted Modeling checks about twice as often as architects and owners. Even though contractors are the least interested, a Modeling check appears to be an excellent low-cost option for contractors.

The majority of the participants in the AEC sector anticipate that the checking of the Model process, where economic value, efficiency, and time are key factors. Value refers to the advantages an organization may get from the model checking, which is typically reliant on the capabilities the company aims at defining and establishing platforms, formats, content, the data structure, and the outputs that fulfil the demands of owners modellers, and clients. These main results are vital to improving the company's profit, business, and competitive edge.

2.1.1. BIM models quality review

Quality evaluation of the BIM process can be divided into two main areas, as shown in Figure 2 : quality of the process and quality of the model (Donato, Lo Turco and Bocconcino, 2018). Besides, the quality of the model depends on the quality of the entire model (like what is found in clash detection or model coordination) and the objects that compose that model. Hence, data contained in the objects and their coherence with the project (and information) requirements are considered.

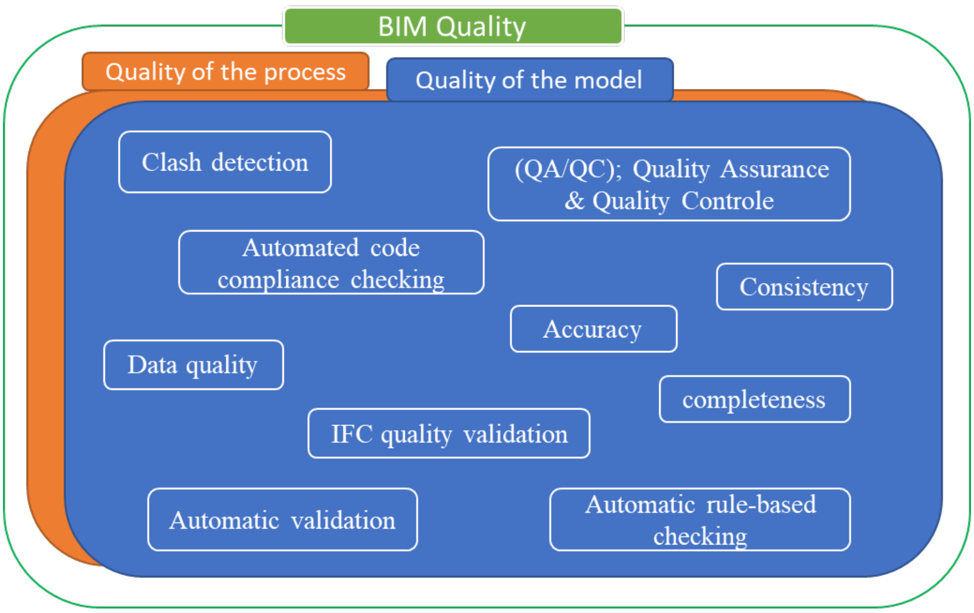


Figure 2 - Quality evaluation of the BIM process areas

According to (Mirarchi and Pavan, 2019), Issues faced in real-world models can be considered under three categories: accuracy, consistency, and completeness. This research is focused on possible issues that can influence the data quality of models. Based on background researches, as shown in Figure 3, the data can be investigated and evaluated from the most used objective dimensions of accuracy, consistency, completeness and timeliness (Wang and Strong (1996), Naumann and Rolker (2000), Delone and McLean (2003)

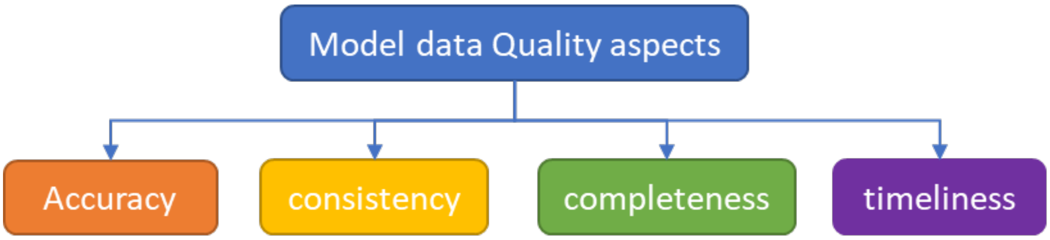


Figure 3 - BIM model data quality dimensions

2.1.1.1. Accuracy

As far as concerning the accuracy, three different types of issues turned up in BIM models:

- Inconsistency between information fields. (for example, inconsistency between an object`s name and its informative contents.
- Precompiled information fields: whether the information is not known or is known but not used by designer
- Other issues related to human performance (like typos)

One example of not applying data despite knowing it is leaving the function field as “exterior” while, in reality, it is an “interior” element.

2.1.1.2. Consistency

Two main issues that happen as a result of poor consistency are

- Faulty categorization of objects in authoring tools
- Provision of information in the wrong section.
- Not coherent use of the same information field.

The first issue happens because authoring tools use defined semantic structure in order to limit interoperability. As a result, it can be said that defined semantic structure deviates from the BIM authoring tools function. Another issue related to consistency is the possibility of filling text-free fields in authoring tools. Such a possibility is an area of mistake for users. An empty field that can be filled with thermal transmittance can also be filled with generic comments or even dimensions of the element (Mirarchi and Pavan, 2019).

2.1.1.3. Completeness

As far as concerned completeness, the model must meet requirements defined in EIR on one side and requirements to create the structure of information on the other side. It is worth mentioning that even meet the exact defined requirements in a project; every design team defines a different information structure.

According to (Mirarchi and Pavan, 2019) study showed that different modelling teams could define their different structure of information for the same target LOD. This attitude will lead to severe issues in data accessibility and interoperability by not using a standard structure for each mode. Also, it indicates that in the BIM authoring tool, similar levels of information on the same object can be led with a varied number of information fields.

An issue in database evaluation is regarded as null values that are difficult to understand and interpret. One reason is that we do not know if empty cells are unknown or nonexistent or a result of a mistake in model development. This fact makes evaluation more complicated (Zaniolo, 1984).

2.1.1.4. Timeliness

For the timeliness dimension, it is meant by it as the extent to which data are sufficiently up-to-date for a task.

2.1.2. Quality check Clusters

After reviewing the literature and articles related to the BIM model quality check, it can be summarised and cluster the BIM model quality check into four main classes; Data integrity, Ifc Validation, Code compliance, and Clash detections. As shown in Figure 4

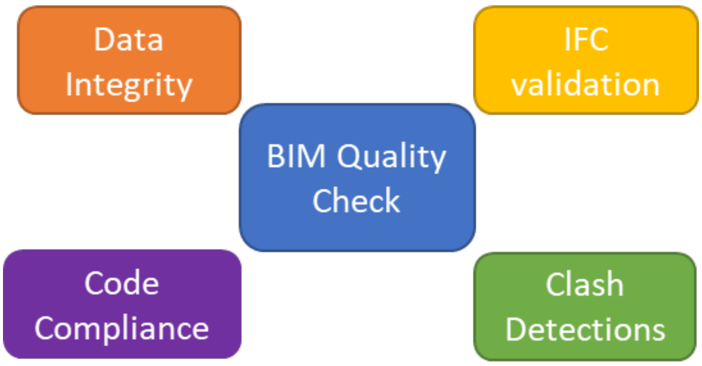


Figure 4 - BIM model data quality Clusters

2.1.3. Short Review of BIM models quality

2.1.3.1. Semantic Enrichment of BIM Models for Code Compliance Checking

The main obstacle in the BIM field is proved to be insufficient, deficient and inaccurate data. Different code checking platforms ask users to complete input data of modelling before checking. However, since this is problematic and takes more energy, the process can be automatically done by semantic enrichment. Identification, specification and classification of the information are required through such a semantic enrichment. To code compliance checking, there are ten classified tasks.

Nevertheless, it is needed to research and work further on this classification system and create a more affluent base. This research provides a layout of building code concepts and their properties. It also brings forward a classification to develop semantic enrichment (Bloch and Sacks, 2020).

Another problem in information exchange between BIM tools is that the IFC exchange schema is not explicitly designed for each BIM tool and, consequently, cannot capture whole semantic meaning according to the requirements of a specific BIM tool. Though BIM standards that direct MVDs for domain-specific exchanges are developing, there is a gap of semantic definition in exchanges that prevents faultless interoperability in BIM. In this regard, Michel Belskey and Rafael Sacks (Belsky, Sacks and Brilakis, 2016) propose an innovative approach that equips IFC exchange with semantically valid concepts deduced from data contained in the building model. They afterwards tested the applicability of their approach by the employment of prototype software. This research proved that integration of inference adaptable rule sets to different types of domains is possible(Belsky, Sacks and Brilakis, 2016).

2.1.3.2. BIM-QA/QC in the architectural design process.

This research is devoted to exploring the importance of quality evaluation of the BIM models in decision making. Evaluation of the BIM processes and BIM models here are based on BIM Quality Assurance (BIM-QA) and BIM Quality Control (BIM-QC). For having usable data for fair quality evaluation, customized checklists and queries were used within a database management system that uses model-checking software. This method was applied on three projects in Italy and proved to bring forward an excellent possibility for quality evaluation during the design process (Donato, Lo Turco and Bocconcino, 2018).

2.1.3.3. BIM verification at the lifecycle stage of construction

This study investigates a fair system that verifies the BIM model, especially during the construction phase. It determines particular design parameters to control the objects based on a set of rules. It claims that objects in BIM can be evaluated through compliance with such a verification tool, and such a system can assure the successful completion of a project. (Galkina and Kuzina, 2018)

2.1.3.4. Model Checking of Precast Concrete Structures

Prefabricated structures are not considered in the early phases of construction projects, and this negligence leads to a considerable loss in terms of time, energy and quality. Considering the requirements of precast concrete structures enables architects to improve projects efficiency. By Providing a checklist, this study ensures an optimum consideration of such requirements on a small scale. Furthermore, since this checklist is used in the early stages of BIM modelling, changes in the projects become easy to implement and, accordingly, less costly and time-consuming. (Mekawy and Petzold, 2018)

A knowledge exists that shows what the deficiencies of precast concrete structures in finished projects are. This knowledge has been gathered and inserted into the computer system as an extracted set of values. This computerized checklist can be applicable in BIM modeling. This study examines three reasoning examples for such a purpose, while three of them were reviewed from automated constructability. (Mekawy and Petzold, 2018).

2.1.3.5. Semantic BIM Reasoner for the verification of IFC Models

Recently, different technics and tools have been developed to code compliance with the IFC model. However, there is a requirement for full automatic code compliance. For developing such a tool, research is carried out by M. Fahad et al. that concentrates on developing one specific tool named Semantic BIM Reasoner (SBIM Reasoner). This mechanism first takes the input of the IFC model via several preprocessors (IFC to RDF converter, Geometry Extractor) and, based on that, creates a semantic repository. Then Stardog builds a graph based on created knowledge for semantic verification. The resulted comprehensive inference and reasoning mechanism assures meeting one of the fullest possible verification. This research employs several IFC models to examine SBIM Reasoner. It finally proves that semantic technology provides the richest mechanism that answers vast queries to verify the IFC model. (Fahad *et al.*, 2019).

2.1.3.6. IFC Quality validation

Undoubtedly ifc format has become increasingly common in BIM, providing features of interoperability. However, it has its restrictions when it comes to the quality of the product model. The problem is on one side the quality level of ifc models and, on the other side absence of a reliable set of criteria to evaluate this quality. These two problems made researchers think to find a reliable evaluation system. A study by Wawan Solihin et al. (2015) tries to address this issue in two steps: first, it is defined the level of good quality for the ifc model. Second, it suggests rules that automatically measure the quality of the ifc model in terms of completeness and accuracy. The offered measurement is considered comprehensive in that it examines all aspects of the project, such as geometrical properties. Another advantage of this approach is the ability to be quantifiable. (Solihin, Eastman and Lee, 2015)

To guarantee whether the ifc model is of good quality or not is difficult because a comprehensive evaluation should consider the variety of factors such as multiple characteristics of data, including synthetic well-formedness, consistency among multiple redundant representations, the integrity of translation from the sources, the accuracy of the data, etc. The mentioned research by Solihin et al. aims to create a groundwork to achieve well-defined, clear, and accurate criteria to measure the quality of the IFC model. This quality was also previously a concern of Kako and Kiviniemi (2012), who overviewed the development of ifc and its various issues. They clearly and for the first time proposed a tool that can perform geometry-based comparison used in the STEP-based manufacturing domain.. (Solihin, Eastman and Lee, 2015).

2.1.3.7. SmartCodes and BIM

The design and construction of buildings process are highly vast and complicated to be in line with code provisions. To address this challenge, we need to work on the relationship between smart codes and BIM to make it more enhanced but straightforward. This time, Nowari O. Nowari (2013) decides to study Smart Codes implementation in building design along with BIM model necessities and their relationship with automated code checking systems.(Nawari, 2013)

Since the human brain has a different mechanism than computers, rules behind the IFC schema are complicated for regulators to understand. So, maintaining these rules will become increasingly complicated over time. There is another issue here that relates to the nature of creating models by designers. This creating is being done over different authoring BIM software, not a single one. Therefore, automated code-checking must be employable inside different applications. (Nawari, 2012)

Smart code checking is determined based on two main factors. First, building codes should be transformed into a computerized set of rules, named SmartCode. Second, a necessary level of detail should be identified for complete model checking in BIM. (Nawari, 2013)

2.1.3.8. Automatic rule-based checking of building designs

Some surveys investigated the rules of checking systems and the criteria they are based on it. One of the by C. Eastman et al. investigates five primary industrial efforts used in IFC building models and compares them. They categorize the stages of checking systems into four functional ones to have a more transparent research framework and then evaluate each side's technology and structure side. Since the rule checking system is a newly emerging field, only limited related experienced data is presented. The output of such research is a framework for evaluation of these systems (Eastman *et al.*, 2009)

2.2. Review of Questionnaires in BIM

The main ingredient of research is the collected data. So it is not strange that the quality of data has to determine the effect on the quality of research. There are three main techniques for data collection: observation, questionnaire, and interview, and based on the nature of research and its circumstances, any of these three ways are selected. For the current research, the best data collection method is recognized to be a questionnaire. So, it is crucial to set a professional questionnaire that enables the researcher to collect data best.

A successful questionnaire is firstly goal-oriented, providing sufficient and accurate data to the interviewer and extracts the information with target quality. Besides, it needs to indicate what interviewees think well.

2.2.1. Questionnaires as a research tools

An inadequate questionnaire reaches a poor result, and to produce a good questionnaire, the researcher needs to realize the research problem well. Therefore, to plan an appropriate questionnaire below factors should be considered (Jenkins, 1940):

1. Comprehension investigation.
2. The main subject of questions must be essential and intensive and exciting, and possible to answer.
3. Designing the right questionnaire

2.2.1.1. Questionnaire design

Questions are planned to figure out about four mains below facts:

1. The nature of people
2. Their actions and feelings
3. Their impressions about the above actions and feelings
4. What they think to be true

It is important to explore not only the features of such believes but also where they come from. Selected words in the questionnaire play a significant role in the extraction of pure data. Poor and vague words and phrases result in misunderstanding and confusion and contribute to inadequate responses. Besides, such questions make the respondents feel bored and try to answer in an irrelevant way not to be embarrassed. So, respondents must not be left to personal interpretations through wrong phrasings(Jenkins, 1940).

A hugely significant matter here is people's tendency to social acceptance. Thus, the questionnaire has to be designed in a way that minimizes the tendency to social desirability. For example, it can be noticed on the top of the questionnaire that low-prestige responses are equally good and just as acceptable. Generally, extracting biased responses must be avoided with any possible technique(Jenkins, 1940).

All questions can be categorized under two groups of "closed" or "open". A closed question is replied to by a specific number of choices like "yes'/'no". But open-ended ones are answered with individually selected words and descriptions. Answers to open questions provide rich data and are associated with feelings and beliefs. These questions follow the closed questions to reveal more about the first answers. Open questions are easy to ask but difficult to reply to because they allow various answers. This is while close questions are easy to answer and bring forward more reliable data that is easy to code and reach. There also exist semi-closed questions which offer several alternatives extended with additional information(Jenkins, 1940).

In comparing a study with previous work, the same questions in previous research with identical words need to be used. Nevertheless, the whole questionnaire must become adjusted to the purpose of new research. The appearance of the questionnaire is also of high importance. For example, questions are presented in order and followed by plenty of space for expanded replies. The most relevant and direct questions should come at first, and the more get to the end, the more detailed and specific become questions. The extent to which the results of an investigation can be generalized to other samples is named external validity. The two types of external validity are population validity (whether you can generalize to other groups of people or not) and ecological validity (whether you can generalize to other situations and settings or not). These two terms need to be regarded and balanced based on the circumstances of the research. Sampling right respondents. (Jenkins, 1940)

Since sharing a questionnaire with the entire population of the globe is impossible, many acts as representatives to draw information from. The primary approaches in scientific and clinical studies are incidental and random sampling. Whatever the approach is, there exists a possibility of biased data collection. Nevertheless, we should bear in mind that a response rate of 70% eliminates the chance of bias. Two main factors determine the size of sampling: first, how serious is the problem under study.

Second, how big is the involved population. Anyhow, sample size determines the level of certainty and reliability of results.(Jenkins, 1940)

2.2.1.2. Piloting

Despite the valuable efforts of researchers, studies are always subject to errors, and questionnaires must be tested on a small sample of respondents to assist the researcher in identifying both potential problems and possible solutions. For pilot research, respondents are singled out from the population under study. Most pilot work is oriented around the most severe problem of study. In some research, the population is not divided evenly, and some groups are discriminated against over others. Though such discriminations are necessary in some cases, it lowers the internal consistency.(Jenkins, 1940)

2.2.1.3. Validity and reliability

For the assessment of the validity of research, two concepts of validity and reliability are evaluated. Validity contributes to the accuracy of the measure, while reliability is about the consistency of a measure. Validity is not difficult to prove, and there are accepted ways to evaluate the validity of the research.(Jenkins, 1940)

2.2.2. BIM questionnaires

Since the BIM technology is a new and the implementation or adoption faced some obstacles, numerous questionnaires have been published to review those obstacles and manifest the transformation process between the traditional method and the BIM process. The following sections mention some of the questionnaires that have been done and published in the BIM context.

2.2.2.1. BIM for Sustainability Analysis

The percentage of buildings consumption of the energy used in the United States is about 40% of the total energy, and 30% of the greenhouse gas emissions are from its share. With the development of the environment and the increase in energy, sustainable buildings with a minimum impact on the environment are increasing. Effective decisions related to sustainable building design can be small in the initial design and pre-construction stages. Building information formation can assist in complex building performance analysis procedures ensure an optimized, sustainable building design. The objectives achieved by this research and the survey in the feasibility of sustainability analysis on BIM; 1) Evaluation of various building performance analysis programs, 2) Develop a conceptual framework to demonstrate the use of BIM for sustainability analytics throughout the project lifecycle.(Azhar and Brown, 2009)

2.2.2.2. Building information modelling: An international survey

Several surveys have been carried out at the local level, but little is known about the international scale. This research is a preliminary report of a large-scale electronic survey of BIM implementation and the impact on AEC project delivery and project stakeholders in Australia and on the international direction. Local and regional patterns of BIM use will be identified. These patterns will include disciplinary users, project growth stages, technology integration, software compatibility, and organizational issues such as human resources and interoperability. The current status of including BIM in higher-level curricula and the possibility of creating a new system are also considered. (Gray *et al.*, 2013)

2.2.2.3. Contribution and obstacle analysis of applying BIM in promoting green buildings

Based on literary research, A questionnaire consisting of 27 questions has been designed. In the next step, questionnaires are distributed to 300 concerned service owners in BIM and green buildings. The interviewees mainly cover building design engineers, construction company managers, and employees of a green building certification company. Then SPSS and AMOS software were used to test the reliability of the questionnaire to ensure the accuracy of the results. According to the survey notes, BIM technology has made critical green buildings' design and construction phases. Finally, the barriers are being summarized from four aspects of technology, management, economics, and the social environment, besides distinguishing barriers. also, rank the incision of the barriers for different stakeholders to suggest more beneficial countermeasures.(Huang *et al.*, 2021)

2.2.2.4. BIM Investment, Returns, and Risks in China's AEC Industries

Building Information Modelling (BIM), emerging digital technology, is subject to many applications in developing countries, including China. Both government policy and industry patterns indicated that BIM had become the standard method of construction in China. However, one of the main concerns is the uncertainty of BIM investment for architecture, engineering, and construction firms. AEC should be aware of the areas in which BIM investment can focus (e.g., BIM software); 1) Expected returns from a BIM investment. 2) ways to enhance revenue from the use of BIM. 3) Risks in implementing BIM.(Jin *et al.*, 2017)

This study adopted a questionnaire-based model to address these issues related to BIM implementation and risks in China. BIM engineers from several AEC fields and different levels of experience were recruited as the survey sample. It was found from the survey that both internal and external agreement should be a priority for BIM investment, along with interoperability between multiple BIM software tools. Multilateral communication and understanding were the highest recognized return on a BIM investment. Survey respondents had high expectations for the application of BIM in green building projects. Architects tend to hear more conservative opinions about the impact of BIM on the marketing of their work, project planning, and staffing. Findings from this pilot study provide an overview of BIM investment, implementation risks, and implementation risks for AEC professionals. As an extension of existing BIM implementation studies in developed countries, this study provides a broad view of BIM practical experience and associated risks in China and other countries by adopting a comprehensive approach and summarizing perceptions of AEC professionals across disciplines and levels of expertise. Furthermore, the knowledge gained from this study can be further applied in other developing countries where information technology is increasingly being applied in prefab and green building projects.(Jin *et al.*, 2017).

2.3. Before ML (Rule based approaches)

BIM implementations have a considerable effect and benefits in the AEC industry. However, it has significant issues or challenges in the interoperability matters between different domains. Also, code compliance, regulation checking, and model qualities are sophisticated processes. In implementations and working in those aspects, it needs to define rules and set procedures to justify it using IFC formats. IFC files are the way to exchange data and information through different domain authoring tools. However, the IFC files checking themselves and ensuring it been exported entirely with all required information.

Solihin (Solihin and Eastman, 2015) showed in several kinds of research the proposed solutions for checking the IFC files and implementing the code compliance checking by creating an external software called “FORNAX”. It is mainly based on a rule-based approach, that the rules being defined and checking the attributes or model objects according to that rule.

Most checking platforms such as Solibri, Autodesk Navisworks, and Bexel Manager are based on setting and defining rules for checking the information and model data. Those rules are extracted from regulations or according to the requirements of the project. Setting the rules is limited with the defining and specific requirements, and once the requirements or changes are changed, they need to back to the rules and re-define them to match the requirements.

2.3.1. Artificial intelligence and building information modelling

From (Galle, 1995) definition of intelligent modes, it has to be with an “ability to maintain semantic integrity”. The transformation from CAD formates or drawings dealing with the objects as ignorant lines to the BIM models reaches a level of computer integration and compliance. However, the BIM models are still not integrated with the definition of “Intelligence”. That even the objects contain its information but still can be considered “ignorant”.

It can be considered that the word” intelligence “exists in BIM platforms in parametric modelling and behaviour design, energy consumptions, and maintenance. In those aspects, as the human being, the common sense knowledge of the real world are maintained on it, not only knowledge of the world objects physically. (Bloch and Sacks, 2018). This common-sense knowledge of the world is not only physical is missing in BIM platforms, which is why BIM is still at the beginning of the ML implementation.

For example, if a modeller neglects to place a railing on a balcony, no objection or warning will be raised unless a purpose-built rule-checking routine is invoked (Zhang *et al.*, 2015). However, to a human expert, the error is immediately apparent. Likewise, if the width of the opening to the laundry room is not large enough for a washing machine to be carried through it, we understand that this is a design fault, though the BIM platform will raise no red flags. As object-oriented systems, BIM tools include representation of objects, their properties, and the relationships between objects, but we are still far from referring to BIM tools as being intelligent..

2.3.2. Short history about BIM and ML intersections

2.3.2.1. BIM and machine learning integration framework

The Big data applications in the real estate domain and computer valuation systems are growing up very fast; for example, AI-enhanced automated valuation models (AVMs) which can address those issues, and plethora can increase the performance of AVMC at the same time while using a little information to be required valued. Also, the value relevant for design information is decreased for property valuation because of the increasing data of BIM. This research shows the difference between AEC projects and property and automates workflow for their information exchange by BIM and machine learning integration. It has been used comprehensive research to clarify the qualities and quantities of data analyses, and BIM is proposed to contain fundamental database interpretation. IFC also proposed for extracting the valuation, which is already based on a genetic algorithm for optimizations. The important finding indicates; 1) Partial information requirements can be extracted from BIM models. 2) Property information can be performed more efficiently. That research could help manage the exchange of information between BIM and ML, and many other digital technologies might add value to the property industry.(Su, Li and An, 2021)

2.3.2.2. Comparing machine learning and rule-based inferencing in BIM checking.

The complicated process in data checking, code compliance, regulations checking in BIM models and sophisticated processes makes it challenging to manage and be considered obstacles to implement or apply. This research using the semantic enrichment concept by implementing ML and compare e it with the rule interface in checking and validate the room types in a residential building. So, comparing it to the rules of applications, the semantic enrichment offers manual pre-processing. So, the results prove that the Machin learning is directed applicable to classification the problems that could lead to the BIM required a different approach to be observed. Furthermore, it has been mentioned that both methods have been used for semantic enrichment are depending on the nature of the problem. Also, the future of research would be a good guideline for the ML implementation approach.(Bloch and Sacks, 2018)

2.3.2.3. BIM and machine learning and computer vision techniques

Launching technology ways for the constructions industry has been tested and experienced by engineering in the AEC industry. Machin learning could own their approaches or applications to be more advanced in the research field besides BIM and digital techniques. Urban underground infrastructures developments need planning and maintenance and are considered valuable assets. While it is challenging to visualizing and reliability the process of constructing infrastructure systems, machine learning and computer vision techniques allow setting those assets and overcome those challenges. Furthermore, this research was seeking to present the current status and future perspective for BIM and machine learning and digital technique as underground constructions as following; 1) Present the global demand of adopting those technologies. 2) Introduce those terminologies, fundamentals of it. 3) Review the BIM in the traditional mechanisms. 4) Examine the critical applications at different stages. 5) Discuss the recent challenges. 6) Summarize the current situation and figure out the gaps and future directions. (Huang, Ninić and Zhang, 2021)

This page is intentionally left blank

3. A QUANTITATIVE SURVEY IN BIM QUALITY

3.1. Information gathering

This section targeted creating data collections and analysed them to highlight the main and common issues that affect BIM model quality. For achieving that objective, it was considering two ways to collecting data and information. Firstly, publishing a questionnaire focuses on the different aspects of BIM quality and how different companies and engineers consider the quality issues during different BIM phases. Secondly, by managing interviews with interested participants in BIM such as CEO, BIM managers, BIM coordinators, and BIM modellers.

3.2. Questionnaire

3.2.1. Questionnaire design

After reviewing research in BIM quality as referred to in chapter two section (2.1.3) and reviewing questionnaires in the BIM context as referred to in chapter two section(2.2.2), it has been drawn an overview of the questionnaire sections' design and its questions. The questionnare consostye of six pages or sections as shown in Appendix 1.

Section one was just introductions and explaining the objectives and the estimated time required for answering them. Section two has been designed to know the background of the participants, and it was expected and evident that this background affects the answers to the followed questions in the other sections as shown Appendix 1.

Section three considered the BIM modelling manual as the first step in conserving or defining the model quality procedure to control the BIM quality. Therefore, that section has been designed to be an interactive question asking a Yes/No question. If the participant answered No, it would lead him directly to the next section. Moreover, if he answered yes, it would open an interactive new window containing more detailed questions about that BIM modelling manual.

Section four considered the BIM modelling quality check procedures mentioned in the BIM modelling manual or just as structured rules. These sections were focusing more on answering the question of how would check the model quality and. Also, that section has been designed to be an interactive question asking a Yes/No question. If the participant answered No, it would lead him directly to the next section. Moreover, it would open an interactive new window containing more detailed questions about that BIM modelling quality check if it been answered yes.

Section five was about BIM modelling and checking Authoring tools. It was created to collect analytical data for reviewing the authoring tool used for BIM modelling and either the companies using native authorities tools capability or additional software for checking the model quality.

Finally, in section six, before submitting the answers, the participant asks if he consent accepts to identify his identity and the region or country he is based. That would help review the BIM competence level based on the region.

3.2.2. Publishing and entities

In the first place, after finishing and completing the design of the questionnaire, it been sent directly to about 10 participants to test it and receive feedback about the questions and expecting answers. After that, it has been published on all professional and social platforms. Furthermore, it was setting time to receive entries, and then it closed for not received more entries. Finally, it considered publishing and sharing it in various countries and different interest group in the BIM context. The overall entities were 76 during ten days of publishing and sharing. The following section will show how it was shared diversity and received an answer from different partners.

3.2.3. Results analysis and discussion

3.2.3.1. BIM Background

This partition is about six questions related to the personal background and the organizations' specialisation. It helps evaluate the participants' responses and summarise the BIM competence level either at the personal or organisational level. By analysing the participants' entries, the majority of them are holding master studies, as shown in Figure 5. It was expected that the main domain involved in BIM is architectures as shown in Figure 6, it is up to 41 % of the participant who specialises in Architectures.

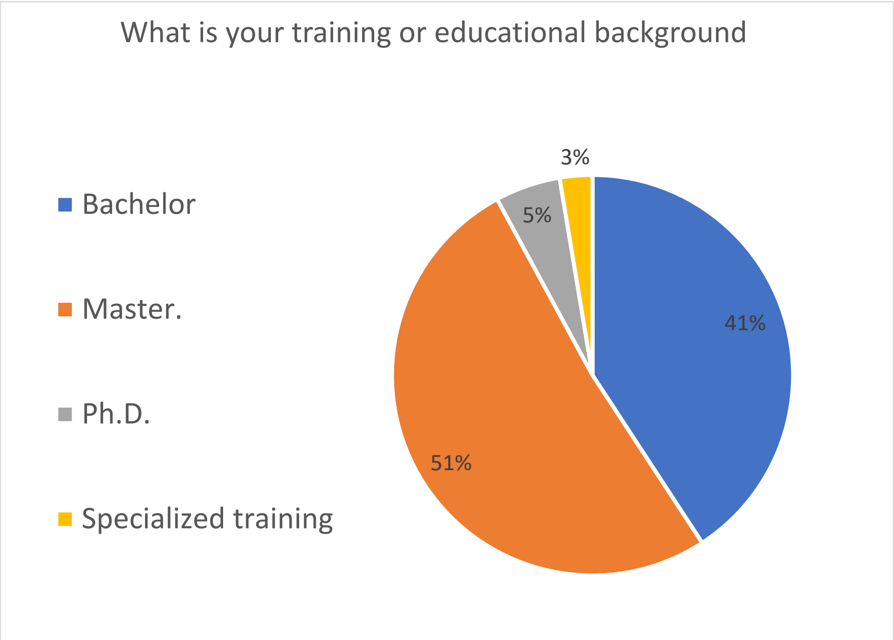


Figure 5 - Personal Education

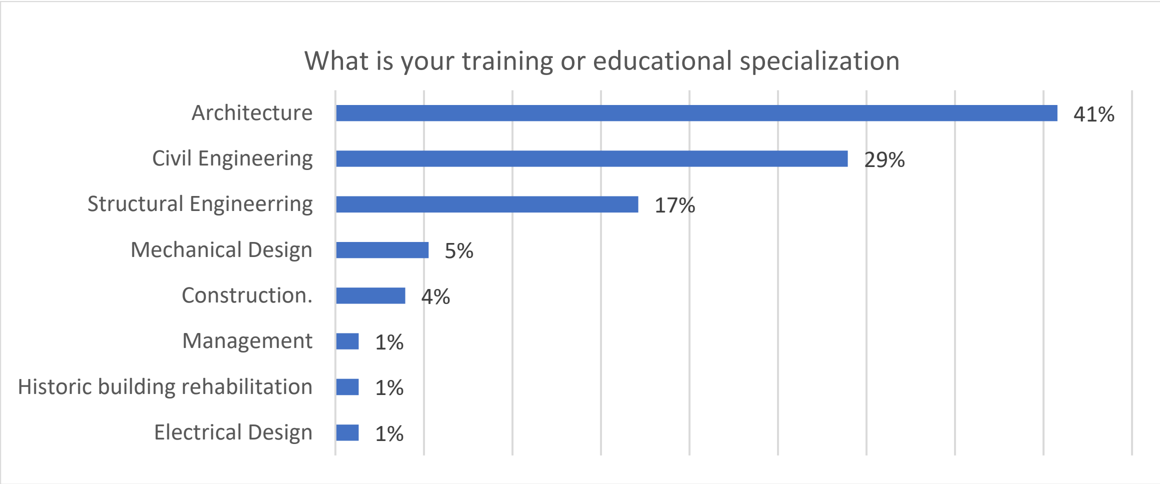


Figure 6 - Personal specialization

It was noticeable that there was an acceptable balance in the questionnaire answers that the main domains of the AEC field (Architectures, Engineering, Constructions) companies were participating in that survey, that add confidence and trusted point in the variety of the survey results and the final outcomes. As shown in Figure 7, the main field in AEC specialization were more than 40 % of the participants. Meanwhile, the participation of Electrical and management was not too far from the participation..

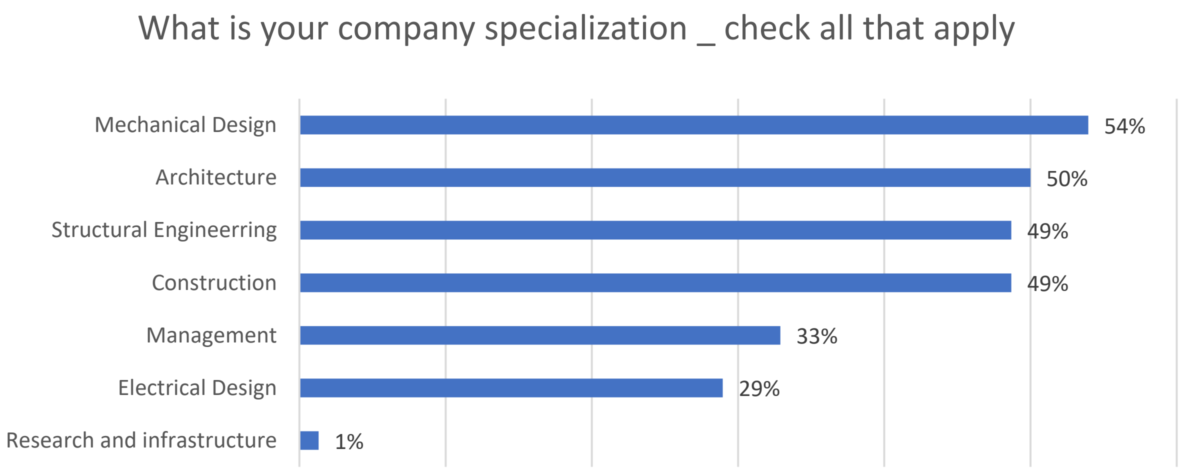


Figure 7 - Companies specialization

In term of personal experience either in general or through BIM context, it was apparent that the participants whose have (5 and 10) and more than ten years of experience were the majority by more than 50% of the answers as shown in Figure 8 and Figure 9. However, the variety of answers in the experience through the BIM context was balanced. Although most participants have more than ten years of experience, it is limited in the BIM context. That was expected because the BIM technology itself is recent to the field.

Lastly, it was referencing the definition of each BIM role (CEO, BIM manager, BIM coordinator, and BIM modeler). So, it was required to describe their roles to define the effects of those roles in answers, especially in terms of BIM model quality. As shown in Figure 10, it was clear that the three majorities of the participants' answers were BIM roles by 75 %. Furthermore, that add a more trusted point in the following question answers for the survey.

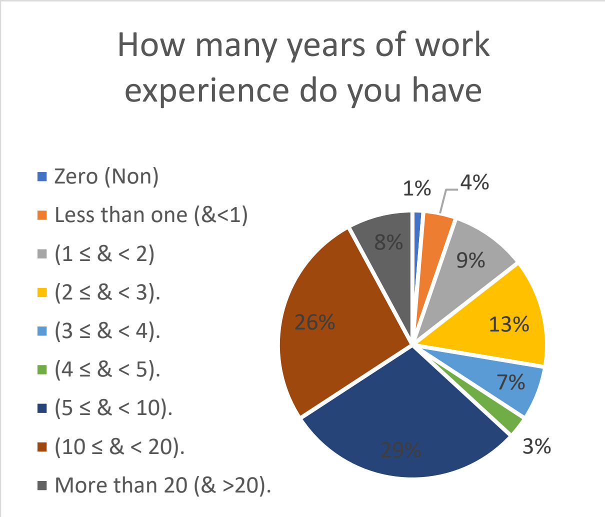


Figure 9 - General Work Experience answers distribution

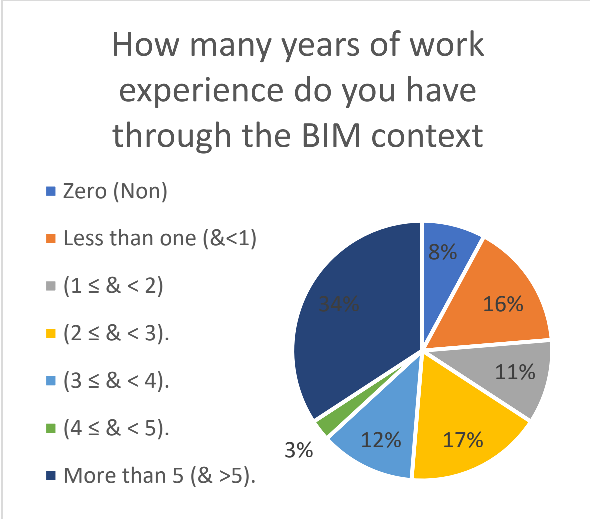


Figure 8 - BIM Work Experience answers distribution

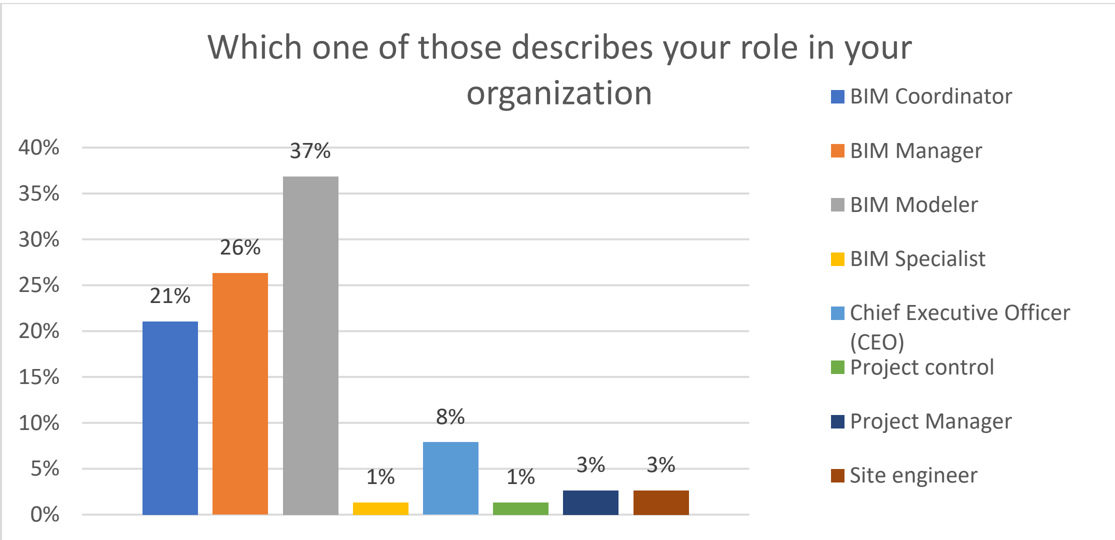


Figure 10 - Roles distributions in the organizations

3.2.3.2. BIM Modelling Manual

In this part of the questionnaire, the first and only appeared question was a Yes/No question about whether the organization has an internal manual for modelling containing modelling rules?. That question was considered a first step to know if the organizations have structured steps and ways to be following in modelling objects in the model to reduce the number of issues and errors. Moreover, it has to lead to producing an accepted quality model. The result of that question was that about half of the participants did not have an internal manual for modelling to follow. As shown in Figure 11, the first question for the organization with a manual was 45 %.

Meanwhile, 45% when been asked if this manual was introduced to every new employee in the company, the answers were about 75 % positive. Thus, it can be illustrated that the modelling steps and procedures have been controlled by the company and not let it be according to the individual experience or background. That is a positive sign that the quality of the model can be controlled in the early phase by the internal modelling manual..

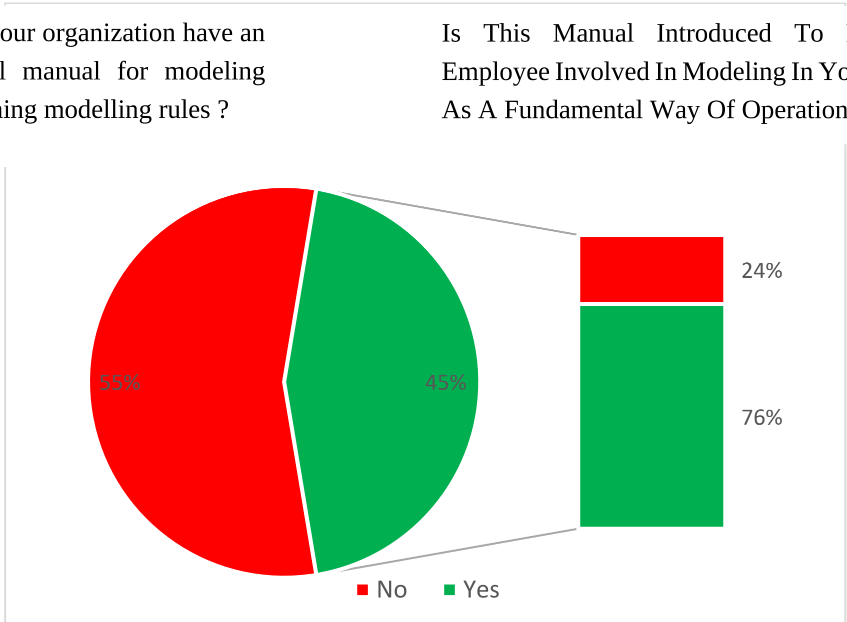


Figure 11 - Internal BIM modeling Manual questions

The participants who answered yes have been asked which bim uses are included in this internal manual. Furthermore, the answers are illustrated as shown in Figure 12. The Coordinating, Design and Cost estimation were the majority. Finally, before moving to the next part, it was essential to know if those manuals are being applied entirely and easy to followed or not, and if not wholly applied, what were the reasons from each participant perspective as shown in Figure 13 and Table 1.

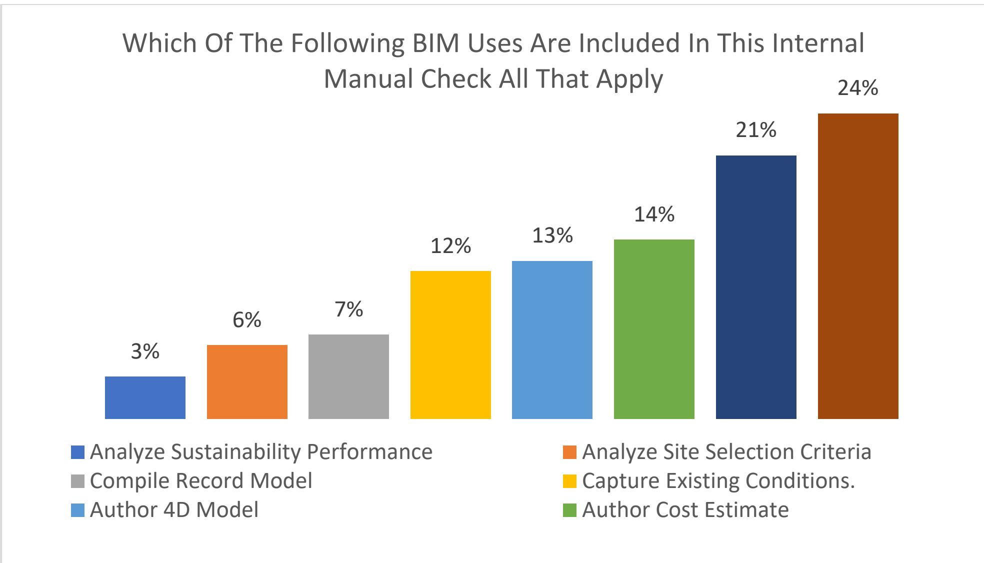


Figure 12 - BIM uses included in the internal modelling manual

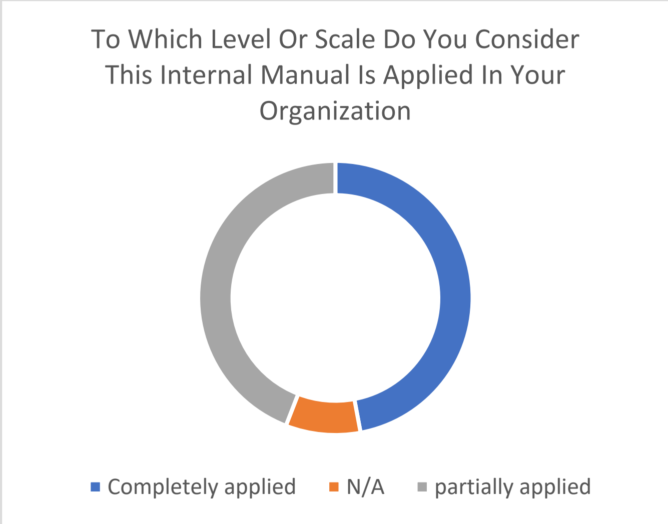


Figure 13 - The level of applied the modelling internal manual

Table 1 -Participants answers the question about the application of BIM manuals

Question	**Participants answers**
Why Is Not It Completely Applied?	Because it's not a written manual that has a checklist and is shared on our server. Its rather rules communicated verbaliser
	Delivery dates are too tight, most of the times BIM models must only comply minimum requirements
	No strong technical support is provided
	Lack of update frequency makes it somewhat obsolete as the processes evolve.
	Incompetence of people
	not to clear easy to follow, did not have the right disability to follow the information inside it

3.2.3.3. BIM Modelling Quality Checking rules

As following the same approach of interactive questions, this part started with only one Yes/No question about *"If the organization have structured rules for controlling and checking the quality of the BIM models?"*. The answer to that question will define if it continues to detail questions about those structured rules or continue to the next chapter. AS shown in Table 2, the percentage of answers by Yes was 47%, and that was expected from the previous chapter answers, and that means the quality check need more concern in companies' manuals and working flow.

Table 2 – The Answers count of the question about structural rules in the organization

Question	Participants answers	count	Percentage
Does your organization have structured rules for controlling and checking the quality of the BIM models	NO	40	53%
	YES	36	47%
Grand Total		76	100%

The participants whose answers were yes were asked, "*How are these rules integrated into your company?*". It was multi checkboxes questions with an opening at the end if needed to add other answers—the answers to that question as illustrated in Figure 14 - Integrated structured rules. The checklist is the most popular tool been used by the coordinators and the BIM modellers. It helps them more for control, but it can be an obstacle because of the time need to check the criteria, and it is manual checking, so the second question was about in which scale they apply those rules and easy to follow. Figure 15 - The scale of applying and following the rules for controlling and checking the quality BIM modelsand Table 3 -The participants answers for the question why it is hard to apply the integrated structured rulesshow the result question on which scale they apply and the reasons if those rules have not been entirely applying ?.

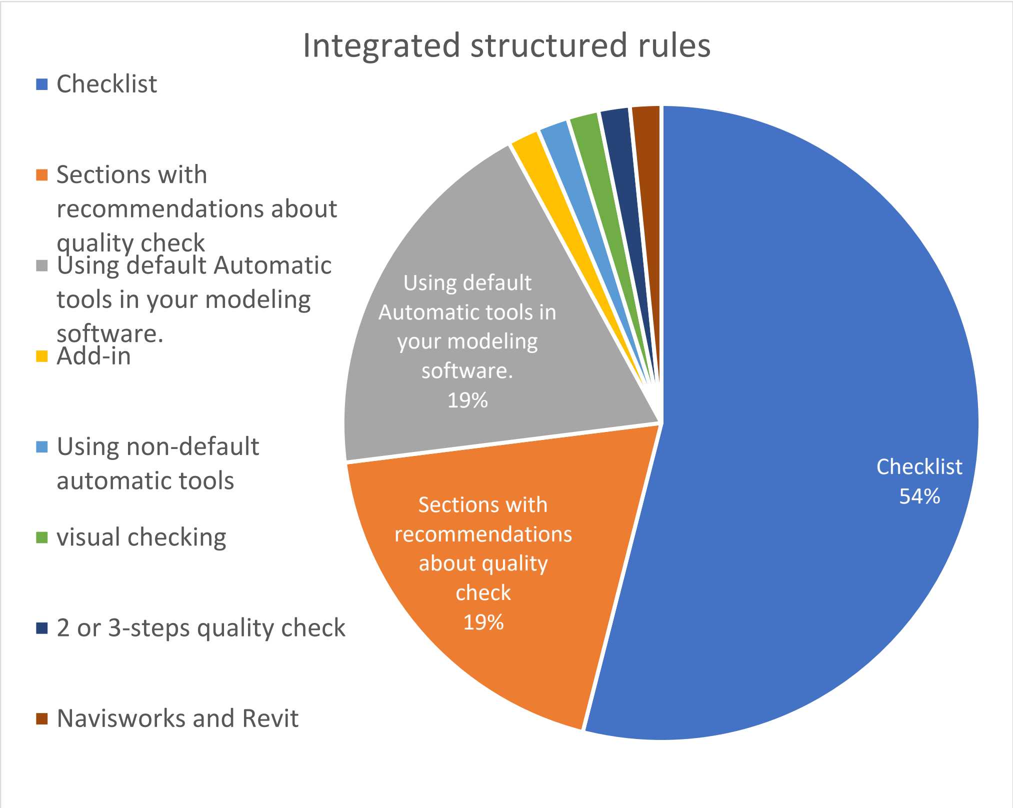


Figure 14 - Integrated structured rules

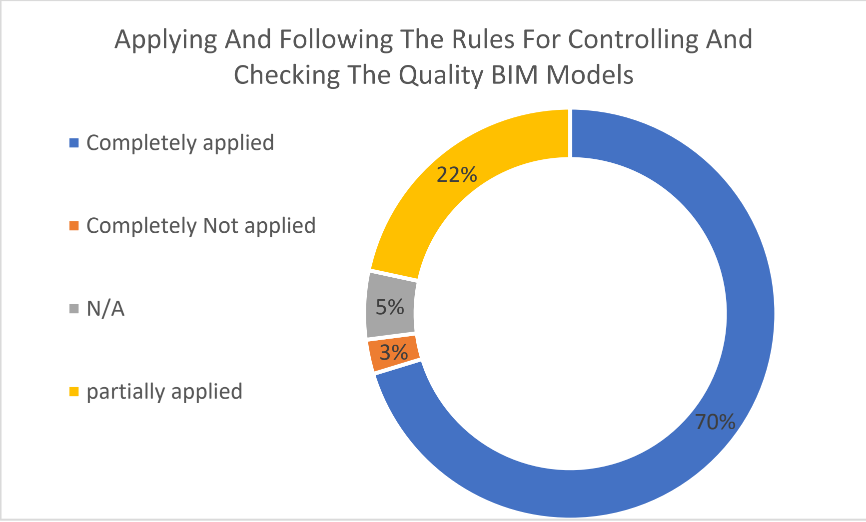


Figure 15 - The scale of applying and following the structured rules

Table 3 -The participants answers for the question why it is hard to apply the integrated structured rules

Question	Participants answers
Why Is Not It Completely Applied?	Even though I am responsible for the model authoring and take some time to check the overall model health, there is a BIM coordinator in charge of the model check. Most of the time again, since the delivery dates are getting closer most of the modellers deliver the minimum (quality-wise) in order to produce more significant amounts of information
	Due to tight deadlines
	Some issues/comments about it.
	because its time consuming and usually we are working on rush projects, so the market seeks automation
	people resistance

Finally, at this part, and by reviewing many structured rules and checklists for BIM model quality control, it came up with some standard criteria and asked the participants if those criteria exist in their structured rules and if it is in which scale it affect controlling the model quality. Table 4 - criteria that can be exist in the structured rules and its impact on model quality shows the participants answers and the percentage of each criteria importance or effects. It is recommended to consider those criteria and each effect in the organization structured rule because, as shown, most of them got a response of high impact in model quality.

Table 4 - criteria that can be exist in the structured rules and its impact on model quality

Which of the following criteria are existing in your structured rules and to which extent do they significantly affect the quality of the BIM model?	Not exist	Low Impact	Medium Impact	Hight Impact
Following the correct naming convention for views and sheets	53%	3%	6%	8%
Project location accurately identified	53%	3%	3%	28%
Building model checked and clean of modelling errors	53%	0%	6%	25%
Components modelled using correct objects	53%	0%	6%	39%
Excess and unused elements removed and purged from model	53%	0%	3%	19%
All floor levels defined and uniquely named	53%	8%	14%	67%
Model objects on the correct level	53%	3%	3%	25%
All tags should be parametric	53%	0%	8%	19%
All drawings and exported sheets must be extracted from the model	53%	8%	17%	39%
All cad imported files must be linked not imported	53%	0%	0%	31%
The building sections are completely dimensioned and annotated	53%	8%	6%	39%
Check that all interior elements are coordinated with other disciplines	53%	6%	14%	47%
All modelled elements shall be floor by floor	53%	0%	8%	31%
All BIM extracted schedules shall be parametric	53%	11%	3%	47%
Checking about empty cells in the model	53%	3%	6%	36%
Classification	53%	17%	28%	42%
Information checks when applicable	53%	3%	17%	28%

3.2.3.4. BIM modelling and checking authoring tools

The last section in the survey was about the authoring tools and software been used in different companies. Also, it concerns which tool or way the companies implement their structured rules for model quality checking. The following Figure 16, Figure 17 and Figure 18 illustrate the majority and percentage of the popularity of different authoring tools used by the companies to create the BIM models.

It has been added a question to the participants about "if some of that additional software was not included in their company tools, what are the reasons behind that?". As shown in Figure 19, It can be considered those answers from the companies` CEOs and software developer companies.

3.2.3.5. Identifications and submission

Before submitting the answers, it asked the participant if they agree to show their identity and their country or region to show the diversity of the publishing this questionnaire. The answers for that part were 63 out of 76 entries. Thus, it is shown that the distribution of the participant through the world map was varied, As shown in Figure 20.

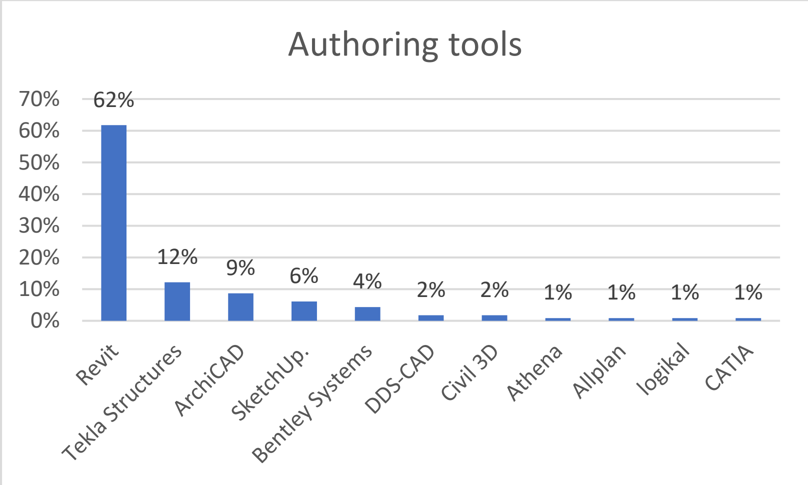


Figure 16 - BIM modeling authoring tools

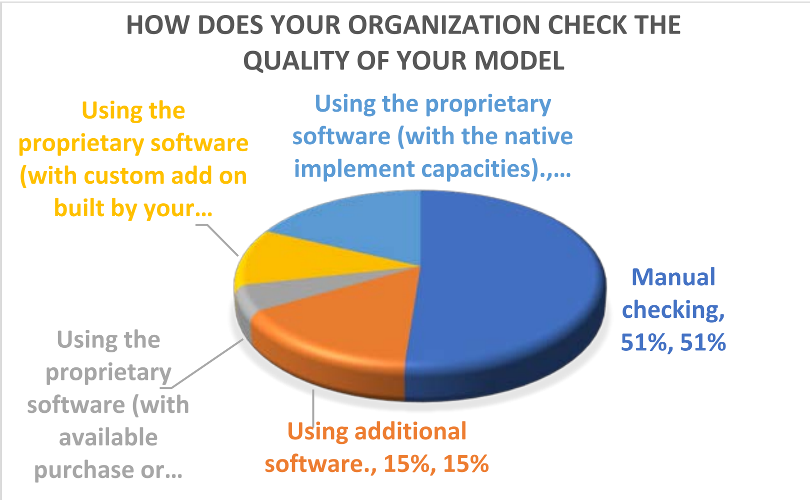


Figure 17 - How does your organization check the quality of your model

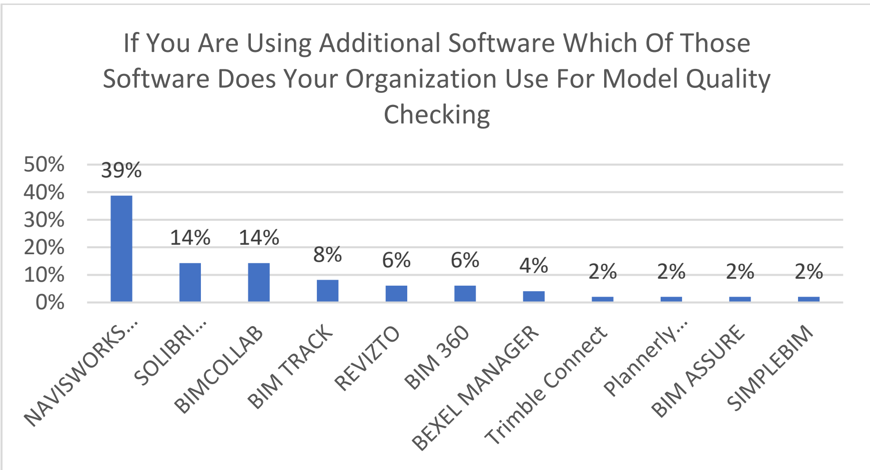


Figure 18 - The additional software been used for model quality checking

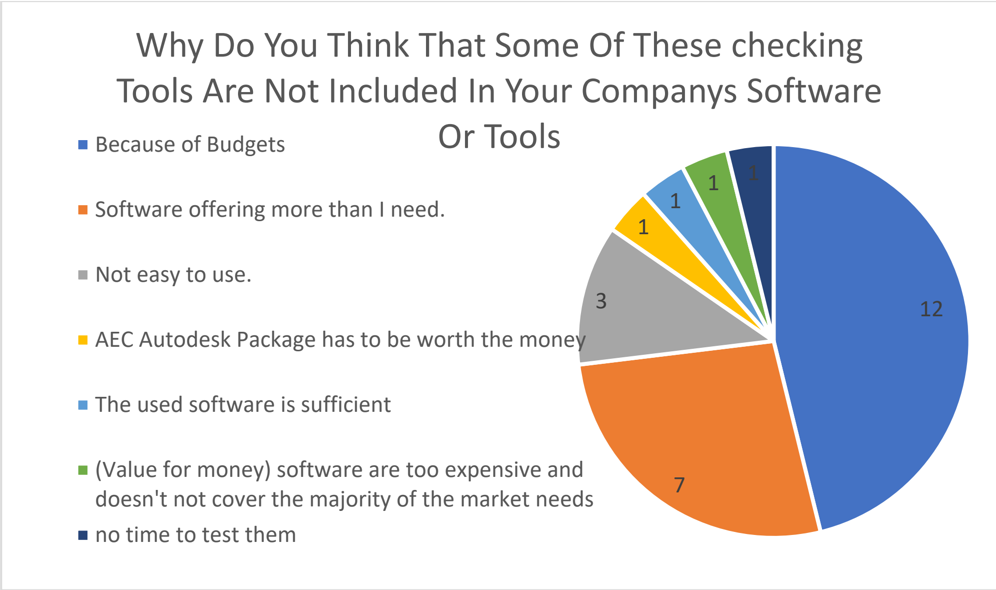


Figure 19 - The participants answers about the main reason their companies not included some useful additional software in their tools.

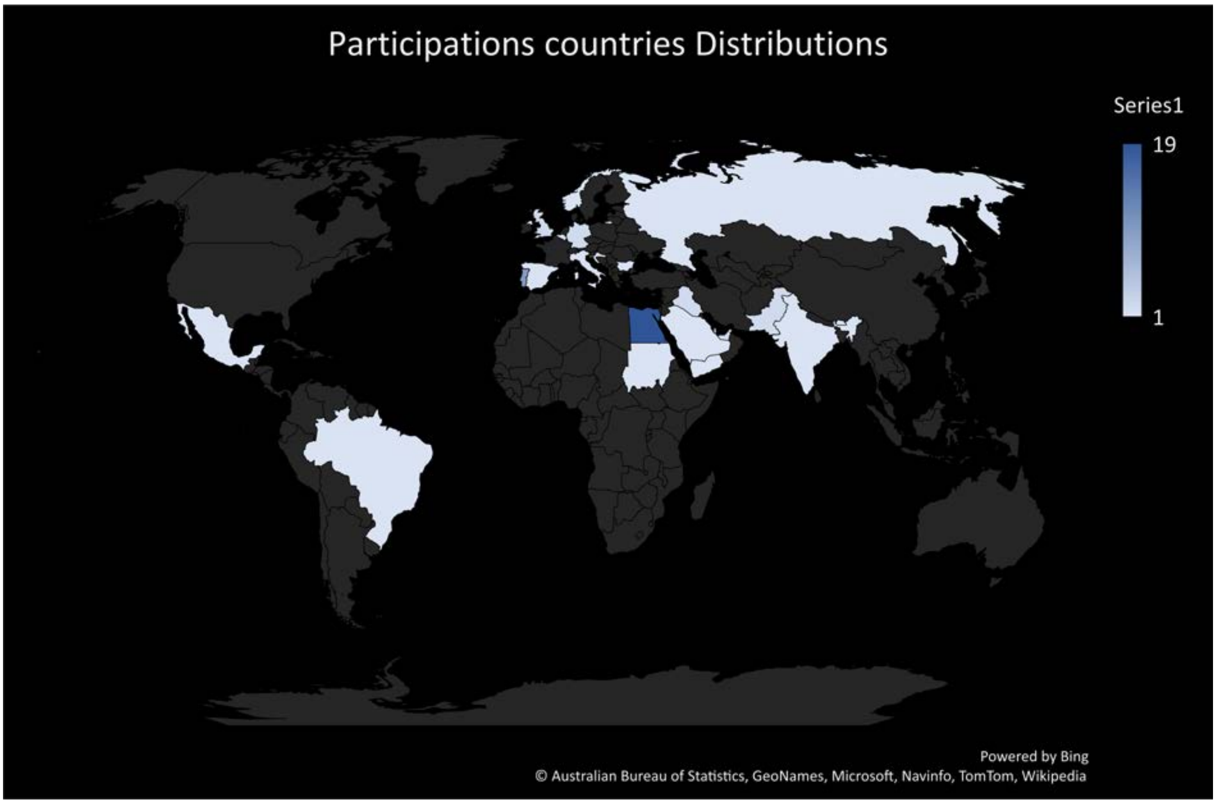


Figure 20 - Participations countries Distributions

3.3. Interviews

An interviews and meeting been managed with BIM Engineer holding different roles such as BIM Interviews and meetings have been managed with BIM Engineer holding different roles: BIM managers, BIM Coordinators, and BIM modellers. Those interviews were with engineers from Europe, Asia, and the Middle East, considering the diversity in regions. It was also considered to have the opportunity to discuss the BIM model quality from different perspectives, especially in disciplines such as Architectures, Structures, MEP, and coordination. In addition, the diversity in region or country was a better chance to have an overview of the working process and workflow. The countries that participated in that interview were Portugal, Germany, Egypt, Belgium, Qatar, United Arab Emirates, Russia, and Mexico.

The ninth interviews with different BIM positions discuss two main parts; First, the questionnaire questions but with comments and sincerely practice and details. It takes time to discuss the questions themselves and their point of view in the questions and the answers from their point of view. The second part was about the development that they can see will impact the enhancement of the BIM models quality. The following points have been focused on in the discussions:

- Time factor during the modelling and delivery is the main Issue. Furthermore, try to find what consuming more time in modelling and reduce it by automate.

- The work in progress with the internal coordination affects checking the quality of the models

- Early clashes detection can make it easier for the (work in progress) phase

- Available tools for model quality check, such as the Autodesk Interoperability tool, were good in semantic check but not easy to use if needed to set specific checklists according to certain BIM uses.

Those interviews and the questionnaire responses manifest some crucial points that significantly affect BIM models quality. It came up with five issues been highlighted and mentioned several times during the interview and from questionnaire participant answers. The following sections illustrate those issues and their proposals for improving productivity in modelling and enhancing the model quality. The Solutions to all those issues have not been addressed deeply in this thesis work; only the last issue has been selected for developing a practical solution.

3.4. Quality Issues in BIM Models

BIM model checking methods can be divided into two ways as shown in Figure 21 . The first method checks the native file such as (Revit model.rvt), (ArchiCAD model.pln). The other type checks the model by exporting the native file to the IFC file then using an external checking tool or software. The presented issues deal with the first method of checking (the native file), especially the Revit model. Selecting the Revit platform depended on the questionnaire result; the Revit platform is the most popular tool most companies use. Refer to Figure 16 in the previous section.

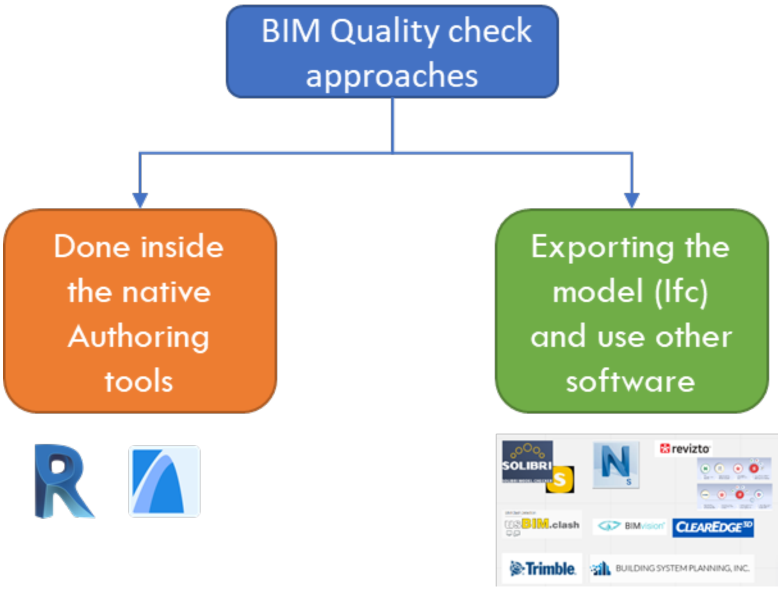


Figure 21 - BIM Quality check approaches

3.4.1. Model quality checking tools

The Autodesk Revit users are using many ways to check their models in term of quality. But none of them are simple in term of easy use and time needed to finish the task. One of the most popular add ion been used in model quality checking is "*Autodesk BIM Interoperability Tool*". It is an add ion been developed by Autodesk developers. The tools provide more than just model checker for Revit as shown in Figure 21.(Autodesk, 2021)

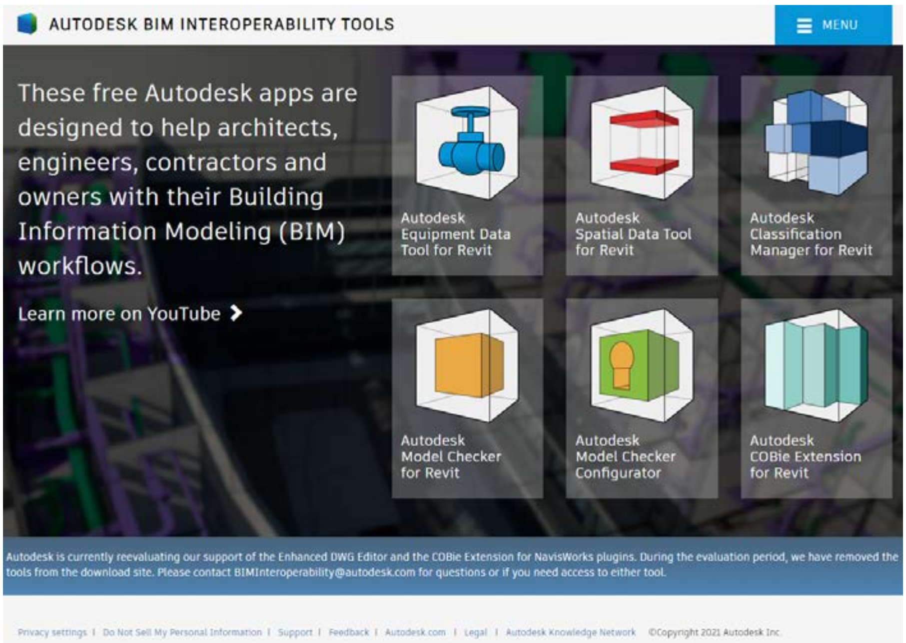


Figure 22- BIM interoperability tools features (Autodesk, 2021)

This tool has many features such as classification manager, Cobie extensions, and Model checker. However, when the BIM engineers been asked why they are not using these tools in model quality checking, the answer was that it is not easy to use. Although this tool provides flexibility to set and define internal checking criteria or rules, it is so hard to deal with, and time-consuming for developing those criteria in the first place and modifying it according to the different BIM uses or according to various project requirements checking.

The proposed solution for those issues come from interviewing the BIM coordinators. Since the ready libraries and check files in those tools are limited and cannot apply for various projects, they suggest that creating ready libraries be used directly according to their checking needs. So, the idea is to create a tool or additional add-ion just choosing which BIM uses need or which checking criteria and then this tools export XML file that can be used for that tool to apply it. So, it will reduce the time consuming from the reviewers to create or even update the checklist (XML file) and help overcome the complication of creating this file using the BIM interoperability tool from scratch.

3.4.2. Real-time clash detection

From interviewing the MEP modelers, they proposed issues have effect in their working progress and consuming more time. This issue is that Revit tool cannot detect the interfaces or clashes between Cable tray and the ducts. But this clash can be detected only at the end of modeling by using interface checking button inside the Revit as shown in Figure 22. That is the point, the modelers cannot detect the clashes during the modeling, they need to wait at the end of the modeling to detect and modify it, unless they use this check each time after drawing every single tray or duct. And from that it come the time consuming if using that interface checking button regularly in addition the using it at the end of modeling and then back to modify the model again.

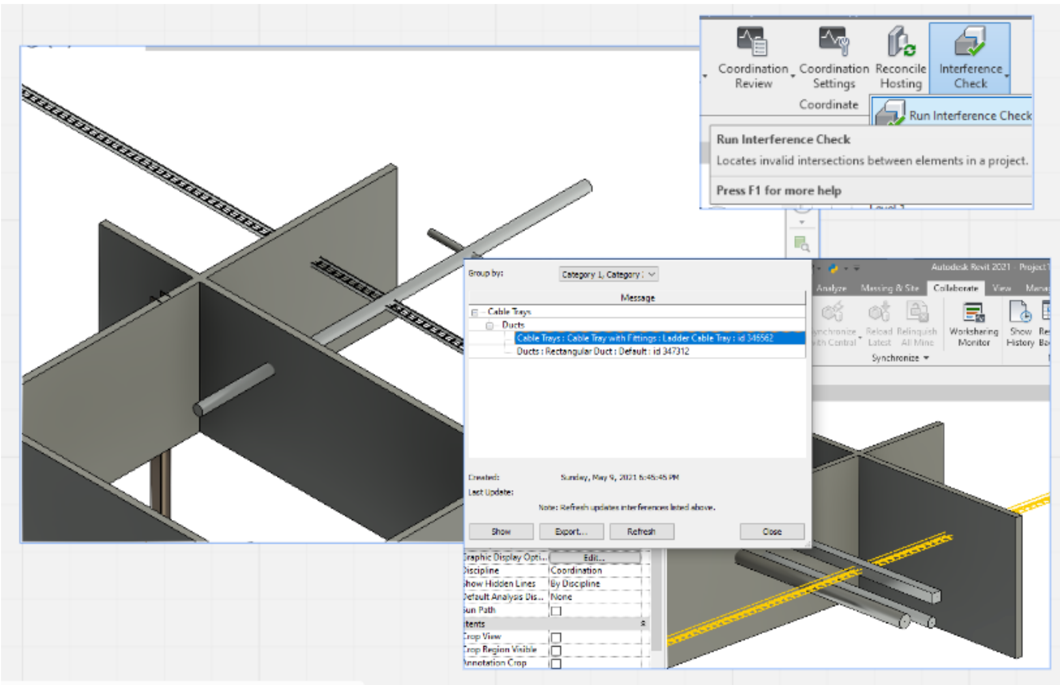


Figure 23 - The interface checking inside Revit

The proposed solution for that issue can be creating real-time clashes detection add-ion. It can happen by creating a tool that can be activated and drawing the cable ray and ducts. Then, in the background of the software, the tool will check the interferences after every single element has been drawing if it clashes or interferes with other elements or not. Furthermore, if it crashes or interferes, it will pop up an alarm message for the user that it has a clash with other elements. That solution will help the modeller to reduce the time of modifications and remodel the elements more times.

3.4.3. Structural Analytical model adjustment

The analytical model is being creating automatically once the model active the box of "*Enable analytical Model*". But the issues or the errors appear in some elements when it is creating their analytical objects. For instance, the analytical members of the columns and walls are being crated at the centre of the objects. Although both elements are attached to each other physically, the analytical members will not be connected to each other as sown in Figure 23. The issues are that in any structural analysis software, it is required all elements to be connected to each other, otherwise the analysis cannot be performed in the right way. So, the modeller or the structural engineer need to manually check these errors and solve it. And that is a repetitive task and consuming more time.

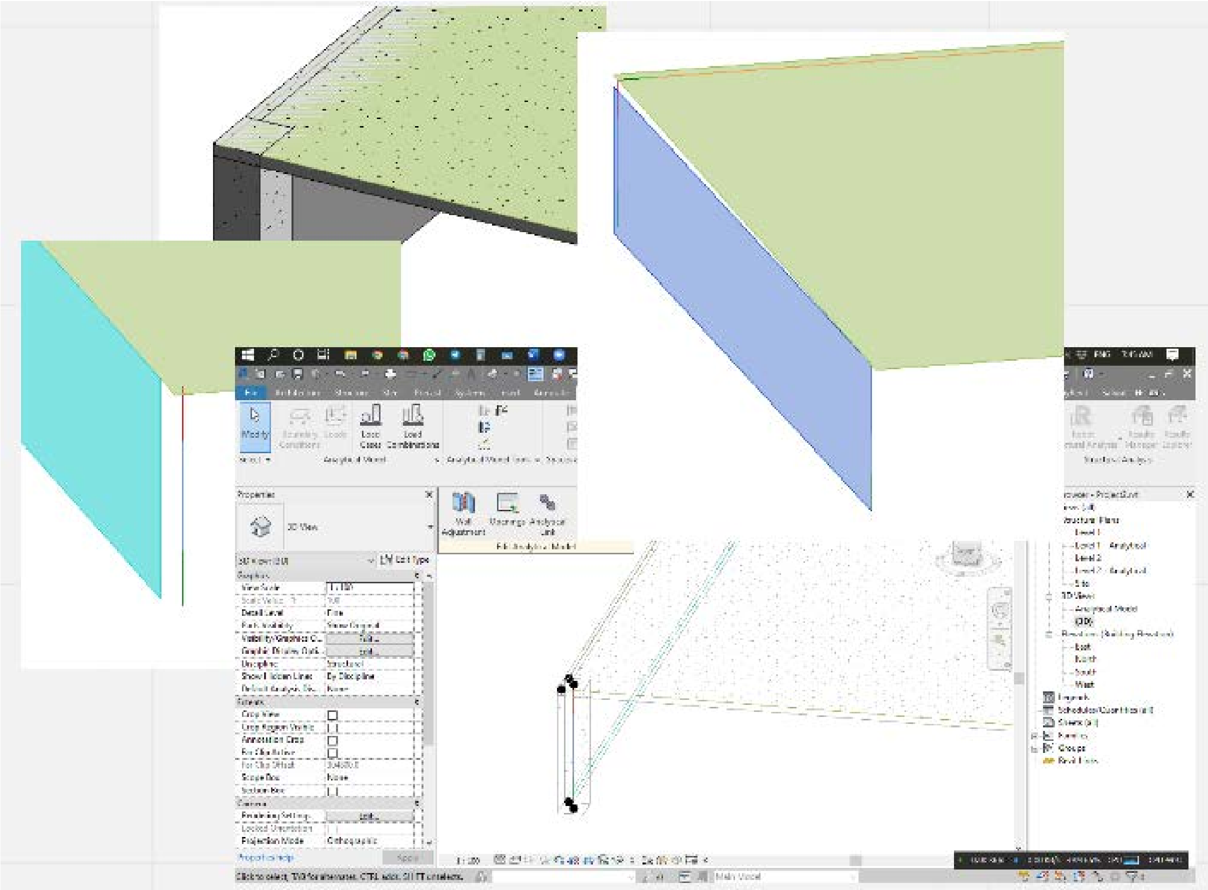


Figure 24 - Physical and analytical model connection in Revit

The proposed solution for this issue is by creating an add-ion reviewing the nearest two analytical elements and check the connection between them. It is not recommended to make it utterly automated since sometimes it is required those closed analytical elements not be connected, so the solution will appear a message to the modeller to review those unconnected elements. However, this solution will help the user to review the analytical models in a semi-automated way instead of reviewing them manually by vision. Furthermore, it is a time reducing solution.

3.4.4. Defining and setting rules during the modelling

From interviewing MEP modeller, one of the issues that have the most significate effect on the BIM model quality is setting rules and real-time checking it. It is needed to check it during the modelling, not just wait to finish the model and export it as an Ifc file and set those rules in additional software such as Solibri or Autodesk Navisworks to check the model quality. For instance, there is a rule in MEP modelling element that the clearance between two ducts or pipes has not been less than 10 or 15cm for a serviceability and construction matter, as shown in Figure 24. Modelling those elements and considering these rules is complicated and not easy to follow or check during the modelling phase. It is required to export the model and check it in additional software. That issue is time-consuming, and an error or issue need a later modification.

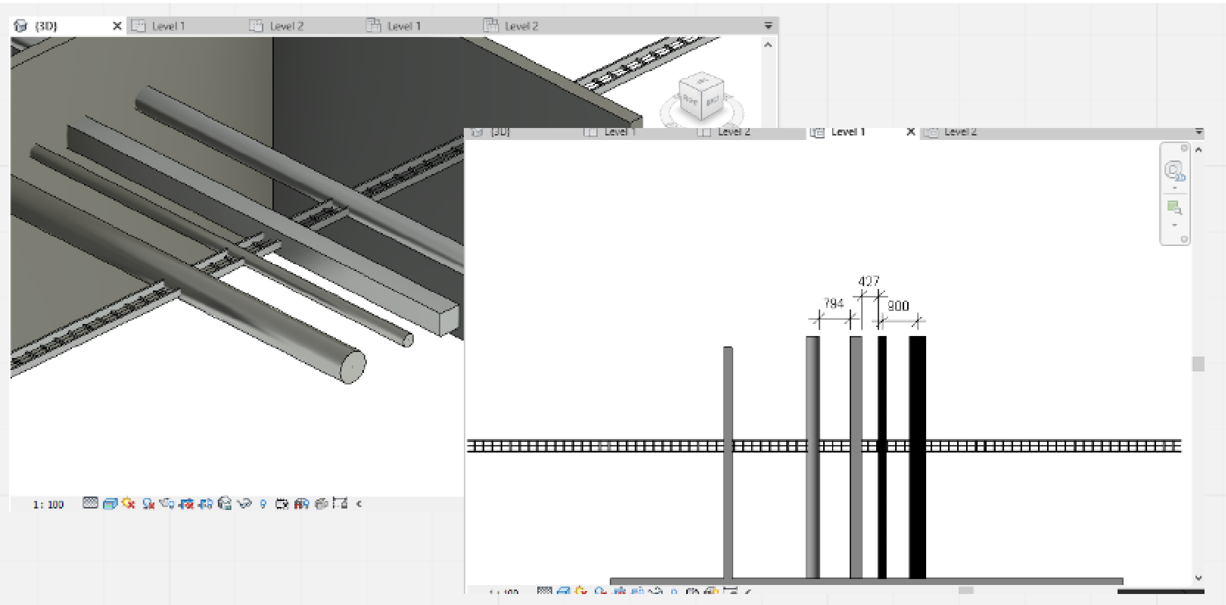


Figure 25 - Clearance between two MEP elements

The proposed solution for that issue is to create an interfaced add-ion allowing the user to select the elements of modelling and setting or defining the rules, such as the clearance between different elements. After that, and during the modelling of those elements, if there are any clashes or elements not following the defined rule, it will pop up alarm message mentioning that the last modelled element is not following the set rules.

3.4.5. Checking information Semantically

3.4.5.1. The issue

This issue was mentioned during the BIM engineers meeting and mentioned as an aspect for the model quality check in a published article by (Mirarchi & Pavan, 2019). The issue is defined as the consistency between the attributes, parameters of the objects and the semantically meaning of that information. For instance, that issue is like defining a concrete column family according to the company naming conventions and renaming the column family type with the dimension of the column. However, the error and inconsistency in that column can exist when those dimensions of the column parameter are inconsistent with the column type name, which contains distinction dimensions. As shown in Figure 25, the family name of the column contains or is defined as a concrete column. However, the assigned parameter for the material inside the model object is copper. That can consider as conflicting information inside the objects. The same error can happen while replicating the column type and renaming it with its new dimensions and forgetting to change those dimensions inside the column family parameters. This error will lead to an error in the quantities, the drawing details, and the column schedules for construction drawings.

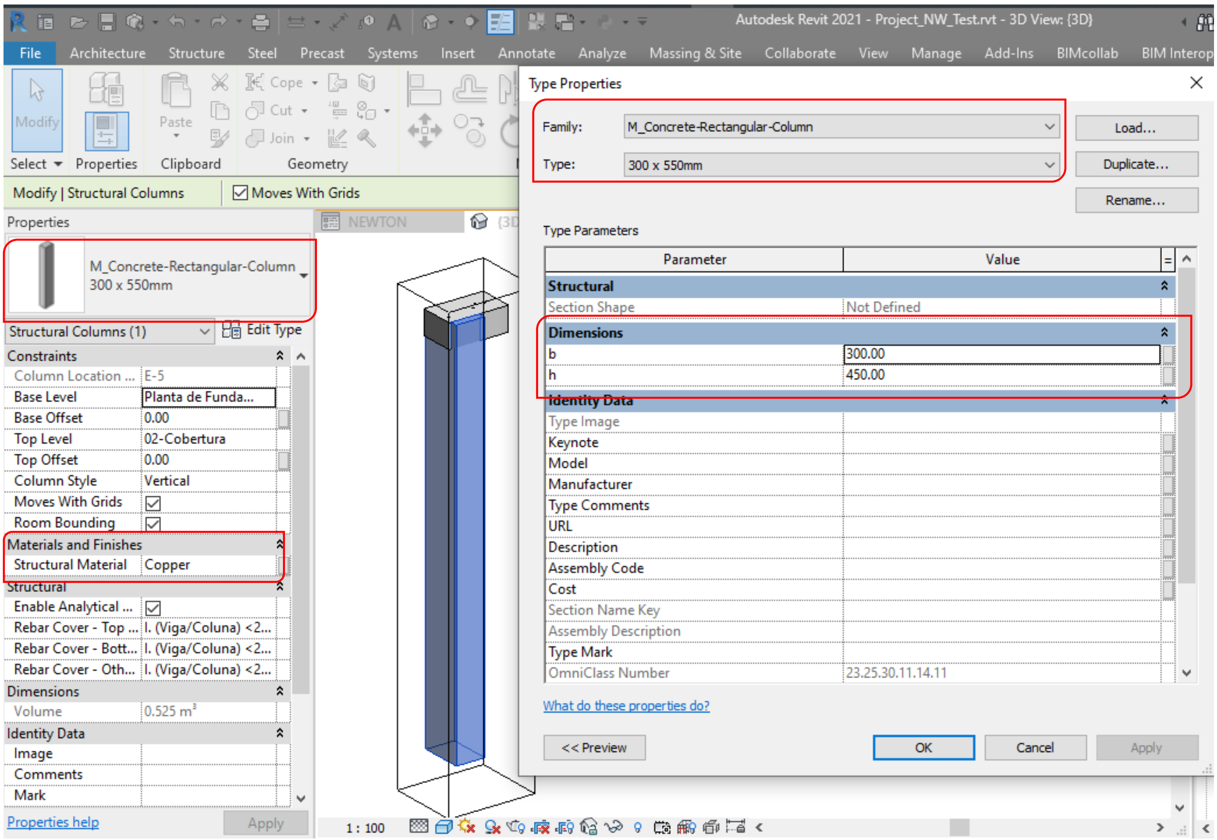


Figure 26 - Example of inconsistence object`s information

3.4.5.2. The source of the error

This error is considered a human error. When the modeller tries to duplicate the family type to add a new type with new parameters or properties, sometimes because of the time factor or not paying more attention, the modeller forgets to change the new parameters according to new names of the new type. If the error is in the family type, all objects with that family type will have the same error and wrong or conflicting information. For example, sometimes, the modeller needs to add a new column with a different dimension. For adding a new column, as shown in Figure31, just press “Duplicate”. Then it will create a typical copy of that column family, and a new screen pup up to the user to rename this newly added column with its new dimension.

Another source of that error is that changing a parameter in the instance type or the object instance; for example, the modeller can change the material of the object instance at the same time; the family name refers to another kind of material different from the assigned material.

3.4.5.3. The important and effect of error

Looking at those semantically inconstant errors leads to cumulative errors and issues in several phases in the model. Furthermore, it is significantly a significant factor effect in the BIM model quality. It needs human being revision, consuming time and not such tool to automate a solution for this error. The impact of this kind of error has a significant effect on the quantities generated from the model. For example, in a model have several types of columns. When the modeller forgets to change the column dimension according to the new name (200x100 mm), the column parameter's actual dimensions are (300X1000). As shown in the Figure 26 and the columns schedule, the quantities have been calculated for a 3.0 m floor height; this column volume is 0.9 m3. Simultaneously, according to the column's name, the drawing and schedule (built in the construction site) is 0.6 m3. The dimension that has been drawn around the column is to clarify the error of the actual dimension; in the real case, the drawing is exported with the column names, not drawing each column dimension. So, the construction engineer will use the name has been referred to in the drawings for the construction, and in the meantime, that name gives wrong information about the actual dimensions and quantity.

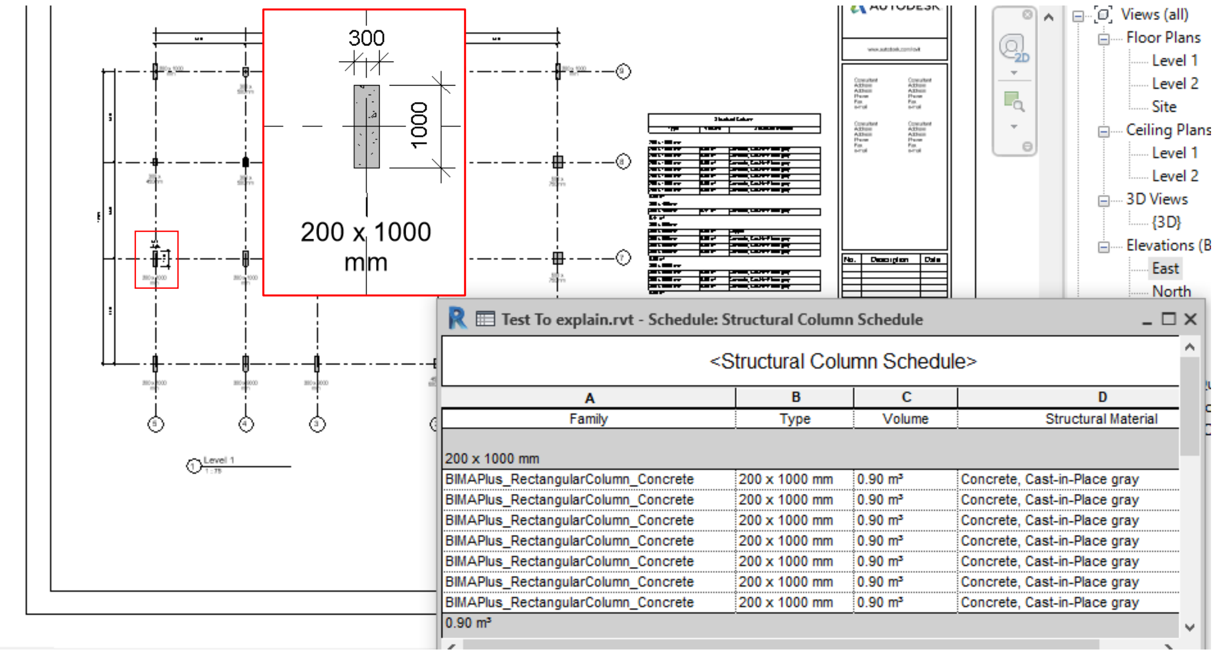


Figure 27 - Column inconsistent data example

4. CONSISTENCY CHECK FOR BIM MODELS

This chapter will discuss the selected issue for the developing phase, its impact factor, the proposed solutions, and the approach been followed to develop the proposed solution. At the end of this chapter, it will discuss the result and outcome from the approaches been implemented and the future development work.

4.1. The Selected Issues for a development phase.

The five mentioned issues have a significant effect on the BIM model quality in several ways. Most of those issues and their proposed solution in the last sections have an alternative solution. However, it takes more time. Some issues have a redundant task to solve, some time-consuming modifications, and the last one has a high challenge in solving it in an automated solution. The proposed solution has been designed to reduce the time consumed and enhance the model quality in the early phase except for the last issues about the BIM model consistency data check. This issue has a challenging approach that can solve it. First, it is needed to create something that can have human intelligence. Then, when reading the object's Family Name and Type Name, recognize the correct information or data needed to check inside that object's parameters. For example, in the case of naming conventions of the family name of the column, the following is the family name example *(Nw_Column_RC_Rectangular)*. From that name, the human being, once looks at it, will recognize that the first a few letters might refer to the company or the organization, the second section in the name *"Column"* is referred to the type of the object. Any Engineer will recognize the third section of the family name "*RC*" as it refers to the family material, which is "Reinforced Concrete". Finally, the last part of the name is referring to the type of column. It is the same in the case of the type naming, from it can recognize the dimensions of the column and will have to check those inside the column parameters in the model. The challenging part is how to transfer that human intelligence to a coded tool and do it automatically?. So the issue been selected to bee find a solution in the developing phase of this thesis is issue number five (BIM model consistency check).

4.2. The proposed solutions.

The proposed solution from the first step any developer can think about is a Rule-based Code to detect words and numbers using a regular expression library (Regex). A regular expression (shortened as regex or regexp (Goyvaerts, 2016); also referred to as rational expression (Mitkov, 2005) is a sequence of characters that specifies a search pattern. Usually, such patterns are used by string-searching algorithms for "find" or "find and replace" operations on strings or for input validation. It is a technique developed in theoretical computer science and formal language theory. The proposed workflow of the tool can be summarized as shown in Figure 27.

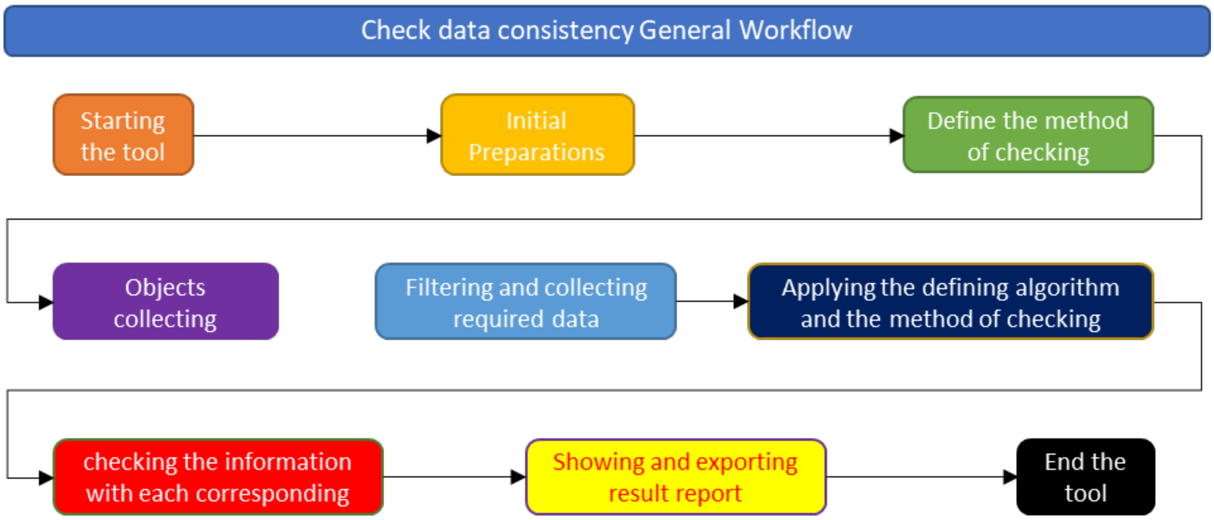


Figure 28 - Check data consistency general workflow

The first step is starting the tool. The second step is initial preparations, such as importing needed libraries, setting active documents in the Revit database. Step three in the script and the workflow is setting or defining the methodology applied for information processing. The fourth step is object data collection (Family and type names) and (object instance material and dimensions). The fifth step is filtering and collection the required data from objects. The sixth step is applying the defining mechanism of data and information processing to get meaning from the names of objects and set a list of the recognizing information and corresponding data from object parameters. The sevinth step is to check the information corresponding in each list, such as; the recognized material from the family name and the material parameters from the objects. The sevinth step also checks the recognized dimension from the type name and its corresponding object parameters. The eight step is showing and exporting a report with both entintes from the user and checking result if the object passed or failed in the check in term of information consistance. Last step is ending the tool.

4.3. Rule based approach.

4.3.1. Regular expression approach

As mentioned, the rule-based approach transfers human intelligence and most of the probabilities into programming code and algorithms. For example, the consistency check of naming an object and its parameters will be solved by defining a dictionary inside the code. This dictionary defines most words, acronyms, and words refer to such meaning or word. As shown in Figure 28, in the defining dictionary, the possibilities to write the concrete material in the family name can be like ['concrete' , 'conc' , 'con' , 'reinforcedconcrete' , 'rc', 'rcconcrete' , 'renconc']. The same concept has been followed for other materials and types.

```
All_elements={
    'material': {
    'concrete': ['concrete','conc','con','reinforcedconcrete','rc','rcconcrete','renconc'],
    'steel': ['steel','st','stel','metal','metallayer','S275'],
    'wood': ['wood','wo','woood','wd',],
    'masonry':['masonry', 'brick', 'stone','ConcreteMasonry']
    },
    'type': {
      'rectangular': ['rectangular', 'rect','rec'],
      'square': ['square', 'sq','squ'],
      'round': ['round', 'circular'],
      'with drop': ['withdrop', 'with_drop', 'with drop'],
      'floor': ['floor', 'flor', 'slab ','roof'],
    }
}
```

Figure 29 - Defining the dictionary with possible words in term of material and object type

The next step after defining the dictionary is to apply it to detect the material type inside the family name. For example, the following Figure 29 shows the searching in the family name to find any words or patterns like the definition in the dictionary. It then will return the key to that word, such as the material is concrete or steel.

```
Splitters = ["_", "-"]
def text_cleaning(text, splits):
    for splitter in splits:
        text = text.replace(splitter, " ")
    text = text.lower() + " "
    return text
def get_Material(text):
    result = {x:"" for x in All_elements.keys()}
    text = text_cleaning(text, Splitters)
    #print(text)
    for key, values in All_elements.items():
        for k, v in values.items():
            for name in v:
                if name in text:
                    result[key] = (result[key] + " " + k)
                    break
            """
            try:
                # <Prefix> (Number) (Unit)
                match = re.search(f"({'|'.join(v)})[column]*\s+", text).group(1)
                result[key] = (result[key] + " " + k)
            except:
                continue
            """
    #return list(result.values())
    return result["material"]
```

Figure 30 - Recognizing Material function in the rule-based approach code

The second part of this check is defining the method to detects the numbers inside the object type name and define them as a dimension; after that, check those recognized dimensions with the actual object dimensions parameters. Figure 36 explains the implementations of the regular expression represented by “re” in the code to find and search in the name to detect the numbers patterns according to several probabilities of writing dimensions.

There are two different functions been defined in the code shown in Figure 30. The first trial, “get_dimension”, detected the dimensions numbers from the name according to the different way of writing it. For example, the pattern of writing the dimension as (300x200mm) the idea was searching in the name for that pattern (Number) <Splitter> (Number) (Unit) and detect the numbers inside that pattern and returned them as a recognized dimension. However, applying this trial of defining the expected patterns of writing the dimensions is not flexible and limited. Each company has its naming convections and way of representing dimensions in several ways. Therefore, it needed to solve these limitations in defining the patterns by making them general. The second trial “get_number” was by searching in the type of name and detect all numbers inside this name and put it in a list then use those numbers whatsoever are and check them with the list of object dimensions parameters.

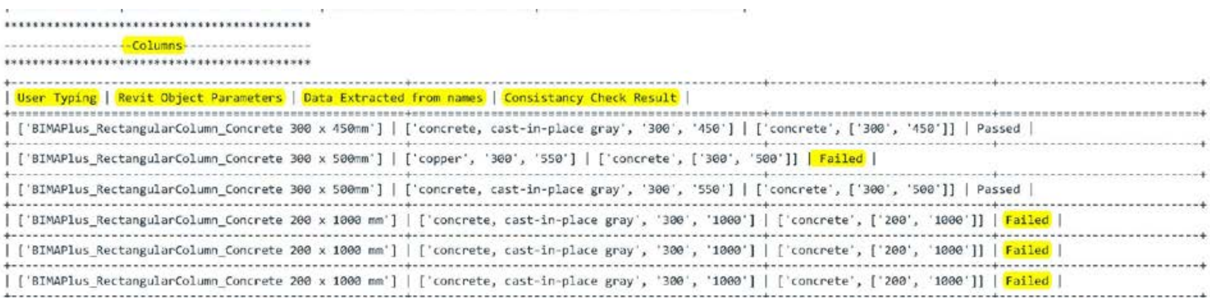
```
def get_dimension(inp):
    #inp = text_cleaning(inp, splitters)
    #print(inp)
    try:
        try:
            # (Number) <Splitter> (Number) (Unit)
            match = re.search(f"(\d*\.\d+|\d+)[{splitters_rx}]+(\d*\.\d+|\d+)[\s]*({units_rx})", inp)
            return match.group(1,2,3)
        except:
            # (Number) <Splitter> (Number)
            match = re.search(f"(\d*\.\d+|\d+)[{splitters_rx}]+(\d*\.\d+|\d+)", inp)
            return match.group(1,2)
    except:
        try:
            try:
                # <Prefix> (Number) (Unit)
                match = re.search(f"[{prefixes_rx}]+(\d*\.\d+|\d+)[\s]*({units_rx})", inp)
                return match.group(1,2)
            except:
                # (Number) (Unit)
                match = re.search(f"(\d*\.\d+|\d+)[\s]*({units_rx})", inp)
                return match.group(1,2)
        except:
            try:
                # <Prefix> (Number)
                match = re.search(f"[{prefixes_rx}]+(\d*\.\d+|\d+)", inp)
                return match.group(1)
            except:
                # <space> (Number)
                match = re.search(f"[_]+(\d*\.\d+|\d+)", inp)
                return match.group(1)
    return get_dimension

def get_number(text):
    text = text_cleaning(text, splitters)
    return re.findall(r'\d+(?:\.\d+)?', text)
```

Figure 31 - Recognizing dimensions function in the rule-based approach code

4.3.2. The result and the limitations

By implementing and testing the rule-based approach (Regular expression), the result was as expected. Detection the material and numbers from the names and compare it with the list of the actual object parameters. As shown in Figure 31, the check result report is divided into four columns. The first column is the user typing (company families naming and type names); the second column is the actual parameters extracted from the Revit database; The third column is the result of data processing by the defining functions. Finally, the last column shows the consistency check result, either Passed or Failed.



--Columns--

User Typing	Revit Object Parameters	Data Extracted from names	Consistancy Check Result
['BIMAPlus_RectangularColumn_Concrete 300 x 450mm']	['concrete, cast-in-place gray', '300', '450']	['concrete', ['300', '450']]	Passed
['BIMAPlus_RectangularColumn_Concrete 300 x 500mm']	['copper', '300', '550']	['concrete', ['300', '500']]	Failed
['BIMAPlus_RectangularColumn_Concrete 300 x 500mm']	['concrete, cast-in-place gray', '300', '550']	['concrete', ['300', '500']]	Passed
['BIMAPlus_RectangularColumn_Concrete 200 x 1000 mm']	['concrete, cast-in-place gray', '300', '1000']	['concrete', ['200', '1000']]	Failed
['BIMAPlus_RectangularColumn_Concrete 200 x 1000 mm']	['concrete, cast-in-place gray', '300', '1000']	['concrete', ['200', '1000']]	Failed
['BIMAPlus_RectangularColumn_Concrete 200 x 1000 mm']	['concrete, cast-in-place gray', '300', '1000']	['concrete', ['200', '1000']]	Failed

Figure 32 - The check result report

There are some limitations in this approach and make it difficult to generalize the solution. The first limitation is the defined dictionary. Every time the company has a new term or word to represent material or types, it will be required to open the background code and add the new word. Moreover, even it can be solved by the user interface box to make it easier, but it is still limited to fixing any issues in the dictionary in terms of code editing. The second limitation is in the dimensions detecting methods. The first method depended on the patterns and the possible ways of writing the dimensions, and once the company or other users have a different pattern, the code will not be able to detect it. Therefore, it will be required to edit the code and add or modify the patterns or define new ones. Also, in the same limitation of detecting the dimensions by implementing the second method, which blindly detects all numbers from the name and puts them in one list, this method will not be accurate if the name has several numbers, not just the dimension. Some companies define the grad of the concrete and the column number in the column name, for example, “Con 25_ 200 x 300 cm_C04”. This name contains the grade of concrete, which is 25 MPa, and column numbers in the schedule, which is 04, in addition to the column dimension 200x300. By applying the second method of the dimensions detector, the result will be a list of several numbers [‘25’, ‘200’, ‘300’, ‘04’]. In that case, the tool will not be accurate in applying this method because sometimes there will be conflicts between the column material numbers or the dimensions. From that conclusion and limitation, it was necessary to develop a new approach that can easily generalize better than a rule-based approach.

4.4. Machine Learning approach

Thinking for a flexible solution, easy to generalize, and overcomes rule-based solutions’ limitations lead to the Machine Learning (ML) approach. It is more reasonable to implement Artificial Intelligence (AI) through machine learning models to simulate or transfer human intelligence better than fixed coding. The main reason for implementing ML models is that it is easy to add new data to the dataset and re-train the model to learn the new terms and ways of naming objects without modifying, editing, or manipulating the background code.

4.4.1. Explaining the framework

This framework is typically the same mentioned in **section 4.2.1,** except the change is in processing the collecting information. Instead of applying rule-based codes, using ML model to process the information and classify or predict the results based on the feed dataset. Also, the differences and variations of naming conventions between different companies make it hard to create a one code cover all those varieties. As shown in Figure 32, the naming conventions examples between companies in term of ordering or inter of words used to refer to specific information.

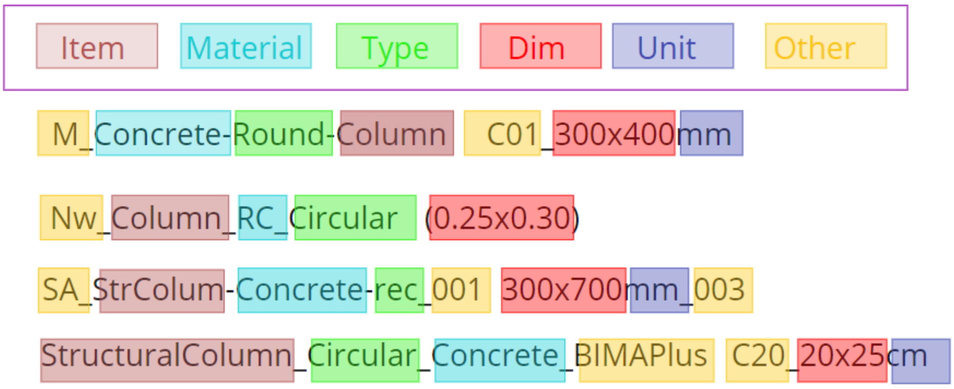


Figure 33 - The varieties in the naming conventions

4.4.2. Datasets

The backbone of any successful ML model is the dataset. Begin with several examples to create an available dataset for columns examples with different dimensions and material. The total number of samples been created reached up to 2700 samples for only concrete and steel material. An excel sheet has been created with seven columns; the first two present the features that contain the expected Family names and the columns type names. The other five columns represent the corresponding labels which have to be the predicted result from the ML models, as shown in Figure 33.

	A	B	C	D	E	F	G
1	Features		Labels				
2	Family Name	Type Name	Type	Material	Dim1	Dim2	unit
3	ectangular-Column w	M_200 x 250	ngular With	Concrete	200	250	mm
4	ectangular-Column w	M_200 x 300	ngular With	Concrete	200	300	mm
5	ectangular-Column w	M_200 x 350	ngular With	Concrete	200	350	mm
6	ectangular-Column w	M_200 x 400	ngular With	Concrete	200	400	mm
7	ectangular-Column w	M_200 x 450	ngular With	Concrete	200	450	mm
8	ectangular-Column w	M_200 x 500	ngular With	Concrete	200	500	mm
9	ectangular-Column w	M_200 x 550	ngular With	Concrete	200	550	mm
10	ectangular-Column w	M_200 x 600	ngular With	Concrete	200	600	mm
11	ectangular-Column w	M_250 x 300	ngular With	Concrete	250	300	mm
2693	BIMAPLUS_StColumn_HEM	HEM_300	hem	steel	310	340	mm
2694	BIMAPLUS_StColumn_HEM	HEM_320	hem	steel	309	359	mm
2695	BIMAPLUS_StColumn_HEM	HEM_340	hem	steel	309	377	mm
2696	BIMAPLUS_StColumn_HEM	HEM_360	hem	steel	308	395	mm
2697	BIMAPLUS_StColumn_HEM	HEM_400	hem	steel	307	432	mm
2698	BIMAPLUS_StColumn_HEM	HEM_450	hem	steel	307	478	mm
2699	BIMAPLUS_StColumn_HEM	HEM_500	hem	steel	306	524	mm
2700	BIMAPLUS_StColumn_HEM	HEM_550	hem	steel	306	572	mm
2701							

Figure 34 - The created Dataset for ML training

4.4.3. ML Libraries

4.4.3.1. Panda Library

“Pandas” – short for “Panel Data” (A panel is a 3D container of data) – is a library in python which contains in-built functions to clean, transform, manipulate, visualize, and analyse data. It is an econometrics term for multidimensional structured data sets.” Pandas is an essential library in analysing data with Python, and it is one of the most preferred and widely used tools in data munging/wrangling, if not the most used library. Pandas is an open source written by Wes McKinney, free to use (under a BSD license).

The exciting point of using "Panda" is that it allows importing or exporting data formats such as (CSV, TSV files, or SQL database). Furthermore, the developers create a Python object with rows and columns called a data frame that looks very similar to the table in statistical software like Microsoft Excel.

4.4.3.2. Scikit-learn(sklearn)

David Cournapeau initially developed Scikit-learn as a Google Summer of Code project in 2007. Later, Matthieu Brucher joined the project and started to use it as a part of his thesis work. Finally, in 2010 INRIA got involved, and the first public release (v0.1 beta) was published in late January 2010. The project now has more than 30 active contributors.

Scikit-learn is the most helpful library for machine learning in Python. It provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. In addition, the sklearn library contains many efficient machines learning and statistical modelling tools, including classification, regression, clustering, and dimensionality reduction.

4.4.4. Dataset preparations

Before training, testing, and using ML models, It is been required to clean and process the dataset used in the models. As shown in Figure 34, Starting by importing the Pandas library and reading the dataset's Excel sheet to process and clean it. Then, by using some features in Panda libraries like defining the headers of the data features and labels, also dropping any null data or duplicated data from the dataset.

Figure 35 shows a preview of the data after reading, defining the headers and dropping distracting data. As shown, the rows (samples) total numbers after dropping out of 2700 samples became 2504, and the numbers of the features and labels are 7. Those samples for only columns in two types of material (steel and concrete) but with various naming conventions as mentioned in the dataset creating (4.2.3.1.1. Datasets).

Also, to remove or to clean any distracting items inside the names or data, it was required to define a function to clean all splitters between words or characters such as “_”, “-“, and “x”, As shown in Figure 36 . Furthermore, to train this kind of data to ML models, it has been required to transform those strings or text into numbers. “ *TfidfVectorizer()* “ was the method been applied to transform all text to unique values as numbers to start training the modes.

Preprocess Dataset

Importing, Reading data and make some cleaning on it for further processing

```
[1] # Read in, clean, and vectorize data
    import pandas as pd
    import re
    from sklearn.feature_extraction.text import TfidfVectorizer

    # Read Data
    df = pd.read_excel('dataset_clean.xlsx')
    header = [x.strip() for x in df.iloc[0]]
    df = df[1:]
    df.columns = header
    df.dropna(inplace=True)
    df.drop_duplicates(inplace=True)
    df
```

Figure 35 - Dataset processing and cleaning

	Family Name	Type Name	Type	Material	Dim1	Dim2	unit
1	M_Concrete-Rectangular-Column with Drop Caps	M_200 x 250	Rectangular With Drop	Concrete	200	250	mm
2	M_Concrete-Rectangular-Column with Drop Caps	M_200 x 300	Rectangular With Drop	Concrete	200	300	mm
3	M_Concrete-Rectangular-Column with Drop Caps	M_200 x 350	Rectangular With Drop	Concrete	200	350	mm
4	M_Concrete-Rectangular-Column with Drop Caps	M_200 x 400	Rectangular With Drop	Concrete	200	400	mm
5	M_Concrete-Rectangular-Column with Drop Caps	M_200 x 450	Rectangular With Drop	Concrete	200	450	mm
...	...	...	...	...	...	...	...
2694	BIMAPLUS_StColumn_HEM	HEM_360	hem	steel	308	395	mm
2695	BIMAPLUS_StColumn_HEM	HEM_400	hem	steel	307	432	mm
2696	BIMAPLUS_StColumn_HEM	HEM_450	hem	steel	307	478	mm
2697	BIMAPLUS_StColumn_HEM	HEM_500	hem	steel	306	524	mm
2698	BIMAPLUS_StColumn_HEM	HEM_550	hem	steel	306	572	mm

2504 rows × 7 columns

Figure 36 - Dataset preview

```
input_df = df["Family Name"] + " " + df["Type Name"]
df["Dim1"] = df["Dim1"].fillna(value="None").apply(str)
df["Dim2"] = df["Dim2"].fillna(value="None").apply(str)

def clean_input(text):
    splitters = ["_", "-", "x"]
    for spl in splitters:
        text = text.replace(spl, " ")

    text = [x for x in text.split(" ") if x != ""]
    text = " ".join(text)
    return text

for i, row in enumerate(input_df):
    input_df.iloc[i] = clean_input(row)

tfidf = TfidfVectorizer()
Tf = tfidf.fit(input_df)
X_tfidf = Tf.transform(input_df)

input_df
```

```
1       M Concrete Rectangular Column with Drop Caps M...
2       M Concrete Rectangular Column with Drop Caps M...
3       M Concrete Rectangular Column with Drop Caps M...
```

Figure 37 - Combine the features and clean texts

one more step required before going further to the ML training and testing by different algorithms is Printing unique values (classes) of every column to get a sense of the complexity of the problem. As shown in Figure 37

```
# print(f"Data Frame Column names: {df.columns.to_list()}\n")
print(f"Raw number of unique values: {len(input_df.unique())}\n")
print(f"Family Name number of unique values: {len(df['Family Name'].unique())}\n")
print(f"Type Name number of unique values: {len(df['Type Name'].unique())}\n")
print(f"Types number of unique values: {len(df['Type'].unique())}\n")
print(f"Material number of unique values: {len(df['Material'].unique())}\n")
print(f"Dim1 number of unique values: {len(df['Dim1'].unique())}\n")
print(f"Dim2 number of unique values: {len(df['Dim2'].unique())}\n")
print(f"unit Unique names: {df['unit'].unique()}\n")
```

```
Raw number of unique values: 2504

Family Name number of unique values: 48

Type Name number of unique values: 2276

Types number of unique values: 10

Material number of unique values: 2

Dim1 number of unique values: 121

Dim2 number of unique values: 155

unit Unique names: ['mm' 'm ' 'cm']
```

Figure 38 - Showing the unique values inside the dataset

4.4.5. Selecting algorithms and Model training

In this problem solving, there are two ways or algorithms that can be applied to solve it. First, as shown in Figure 36, the classification algorithms have been tried to train the model with that dataset and check the outcomes or the model's accuracy after training. Furthermore, regression algorithms will be checked to develop accuracy and determine which one will be most suitable for solving the problem.

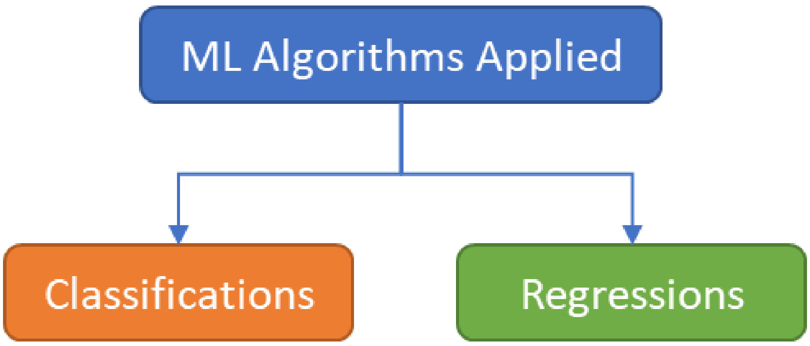


Figure 39 - ML applied Algorithms

4.4.5.1. Classification Trials

Random Forest for classification

From sklearn importing RandomForestClassifier, GradientBoostingClassifier, and import MLPClassifer. And then, the dataset has been divided into two parts: training and testing part. After running the different models with different classifiers, those models' results, and accuracy in predictions for the different classes [Type, Material, Dim1, Dim2, unit]. As shown in the Figure 39, the model predictions` accuracy was 100 % for Material classes, however in the case of Dim1 was about 90%, and dim2 was about 77%. Those results are expected from ML models. However, this accuracy cannot be acceptable in the case of checking the consistency of data. It is not acceptable to use data predicted with not 100% accuracy.

```python
# Import Random Forest for classification from sklearn
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import precision_recall_fscore_support, max_error
from sklearn.model_selection import train_test_split

models = []
classes = ['Type', 'Material', 'Dim1', 'Dim2', 'unit']
for cls in classes:
    X_train, X_test, y_train, y_test = train_test_split(X_tfidf, df[cls], test_size=0.2)
    # rf = RandomForestClassifier()
    rf = GradientBoostingClassifier()
    # rf = MLPClassifier()
    rf_model = rf.fit(X_train, y_train)
    y_pred = rf_model.predict(X_test)
    P, R, F = precision_recall_fscore_support(y_test, y_pred, average='micro')[:-1]
    print(f"{cls} => Precision: {P}, Recall: {R}, F1 Score: {F}")
    models.append(rf_model)
```

```
Type => Precision: 1.0, Recall: 1.0, F1 Score: 1.0
Material => Precision: 1.0, Recall: 1.0, F1 Score: 1.0
Dim1 => Precision: 0.8962075848303394, Recall: 0.8962075848303394, F1 Score: 0.8962075848303394
Dim2 => Precision: 0.7764471057884231, Recall: 0.7764471057884231, F1 Score: 0.7764471057884231
unit => Precision: 1.0, Recall: 1.0, F1 Score: 1.0
```

Figure 40 - Try classifications models

Try All Classifiers in sklearn

Since the accuracy from trying the most popular classifier for this kind of problem, “RandomForestClassifier”, was not close to 100% accuracy, it was needed to try all available classifiers in the sklearn library. Figure 40 and Table 5 shows the coding to import and try all classifiers and the results of the trials.

It was expected that not all available classifiers would work accurately, and some have been unable to import and run. However, from all those results, it was apparent that the best and most accurate result is by importing and using “*BaggingClassifier*”. Although it is not obtaining 100% accuracy as needed, it will be developed in the tool code. In addition, it will check the possibility to develop or improve the dataset for more accurate results.

```
from sklearn.utils import all_estimators
from sklearn.metrics import precision_recall_fscore_support, max_error, mean_absolute_error
from sklearn.model_selection import train_test_split

estimators = all_estimators(type_filter='classifier')

all_clfs = []

classes = ['Type', 'Material', 'Dim1', 'Dim2', 'unit']
for name, ClassifierClass in estimators:
    print('Trying', name)
    for cls in classes:
        X_train, X_test, y_train, y_test = train_test_split(X_tfidf, df[cls], test_size=0.2)
        try:
            clf = ClassifierClass()
            clf.fit(X_train, y_train)
            y_pred = clf.predict(X_test)
            P, R, F = precision_recall_fscore_support(y_test, y_pred, average='micro')[:-1]
            print(f"{cls} => Precision: {P}, Recall: {R}, F1 Score: {F}")
            all_clfs.append(clf)
        except Exception as e:
            print('Unable to import', name)
            break
    print("--------------------------------")
```

Figure 41 - Trying all Sklearn classifiers

Table 5 - Some results from different trails of classifiers

	Precision	Recall	Score
	Trying AdaBoostClassifier		
Type	0.764	0.764	0.764
Material	1.000	1.000	1.000
Dim1	0.157	0.157	0.157
Dim2	0.047	0.047	0.047
unit	1.000	1.000	1.000
	Trying BaggingClassifier		
Type	1.000	1.000	1.000
Material	1.000	1.000	1.000
Dim1	0.900	0.900	0.900
Dim2	0.786	0.786	0.786
unit	1.000	1.000	1.000
	Trying BernoulliNB		
Type	0.956	0.956	0.956
Material	1.000	1.000	1.000
Dim1	0.479	0.479	0.479
Dim2	0.359	0.359	0.359
unit	1.000	1.000	1.000
	Trying CalibratedClassifierCV		
Type	1.000	1.000	1.000
Material	1.000	1.000	1.000
Dim1	0.896	0.896	0.896
Dim2	0.788	0.788	0.788
unit	1.000	1.000	1.000

4.4.5.2. Regression Trails

Since the results obtained from classifications trials were more accurate up to 100 % in material, units, and type, dimensions were not accurate, specifically the dim2. So, it was worth trying regression algorithms for dimensions. Before training and testing the data by regression algorithms, the data must be proceeded and modified to fit in those trails, such as changing the numbers from strings to numeric to be compatible with the regression's algorithms. As shown in Figure 41, the code of transform numbers from strings to numbers and dropping the nun cells.

```
# Make a copy of the dataframe for regression trials
df_regression = df.copy()

for name in ["Dim1", "Dim2"]:
    df_regression[name] = pd.to_numeric(df_regression[name].fillna(value=0))
df_regression.dropna(inplace=True)
df_regression.drop_duplicates(inplace=True)
```

Figure 42 - Dataset processing before traying Regression

The regression trails been applied only on Dim1 and Dim2 as shown in Figure 42. on those trails, it was testing or trying all available regression estimators in the Sklearn library. And their results of all trials are mentioned in the Appendix 2, Some of selected results from regression trial been shown in Table 6.

```
from sklearn.utils import all_estimators
from sklearn.metrics import max_error, mean_absolute_error
from sklearn.model_selection import train_test_split

estimators = all_estimators(type_filter='regressor')

models = []
classes = ['Dim1', 'Dim2']
for name, estimator in estimators:
    print(name)
    for cls in classes:
        X_train, X_test, y_train, y_test = train_test_split(X_tfidf, df_regression[cls], test_size=0.2)
        try:
            rf = estimator()
            rf_model = rf.fit(X_train, y_train)
            y_pred = rf_model.predict(X_test)
            max_err = max_error(y_test, y_pred)
            mean_error = mean_absolute_error(y_test, y_pred)
            print(f"{cls} => Max Error: {max_err}, Mean Absolute Error: {mean_error}")
            # models.append(rf_model)
        except:
            print("Failed!")
            break
    print("----------------")
```

```
ARDRegression
Failed!
----------------
AdaBoostRegressor
Dim1 => Max Error: 664.5818077144509, Mean Absolute Error: 162.43431158637222
Dim2 => Max Error: 485.322628458498, Mean Absolute Error: 189.1233120527161
----------------
BaggingRegressor
Dim1 => Max Error: 174.89999999999998, Mean Absolute Error: 5.198424579412605
Dim2 => Max Error: 645.9, Mean Absolute Error: 28.168962566929633
----------------
```

Figure 43 - Applying all available regression estimators on Dim1 and Dim2

Table 6 - Some results from regression estimators trials

	Max Error	Mean Absolute Error
	Trying AdaBoostRegresso	
Dim1	664.581	162.434
Dim2	485.322	189.123
	Trying BaggingRegressor	
Dim1	174.899	05.198
Dim2	645.900	28.168
	Trying DecisionTreeRegressor	
Dim1	201.0	4.485
Dim2	950.0	25.01
	Trying ExtraTreeRegressor	
Dim1	300.0	6.73
Dim2	700.0	28.15

4.4.6. The result and the limitations

The results were optimistic from the previous trial in terms of material, type, and units predictions. It was up to 100% accuracy. This also might be the cause of model Overfitting, which means the model tailored the training data and learned only from the example and deals with any new sample precisely like the training samples, and otherwise will fail in prediction. If it exists, dealing with this overfitting will be acceptable in the data set created, and the model will be re-trained for any new data.

Implementing the ML approach was compatible with generalizing the problem and easily adapting to any new data. Thus, there is no need to customize the code and the tool for any new information like the Rule-based approach. Furthermore, ML models accurately predicted material, type, and units because the number of classes was limited and small. However, the machine learning trial, either the classifications or the regression, failed to predict the complete accuracy specifically in dimensions because dealing with the dimensions as a class makes them too many classes for predictions. Moreover, dealing with the dimensions as numbers for the regression estimation cannot befitting the regression number to predict 100% accuracy. Thus, the dataset's limitations are the main obstacle in developing an accurate machine learning model used in the consistency check tool.

4.5. Testing and Reporting

Obviously, using python shell with IronPython 2.7.7 and importing the regular expression library was not working, so it was needed to use CPython 378 engine to import the regular expression library into the code and run it. Nevertheless, it was not allowable to import the Pyrevit library, which contains the functions and the tool for presenting the results and reporting the result in presenting way when using CPython instead of IronPython.

The same problem was noticed in implementing the ML approach during import Sklearn, Panda and other machine learning libraries. It was required to use CPython, not Iron python. Moreover, in this case, it was not clear to create reports or presenting the outcome results. So, it was thinking of creating an exe tool outside of Revit but connected to the exported file containing the required data from Revit, then implementing both approaches and reporting the result.

The current testing was testing a model containing different types of columns with various naming conventions. Some of them have errors in naming and dimensions corresponding to the object's parameters. The tool has been designed to import a CSV file which has been exported from the Revit database. The sequence of work in that checking tool can be illustrated as shown in Figure 44, the pushbutton created to collect column required information [Family name, Type name] as user input and [object material, dimension] as an object parameter and create lists with both variable. After active the exporting button, it reported that those data that has been exported to the same folder have the background code of the tool, As shown in Figure 45 ,. It can be reached to the file location by pressing (Alt + Right-click) to the pushbutton "Export Data".

In Figure 46 it shows the interface when the user opening the tool and import the CSV file. First, it shows the user inputs and objects parameters from Revit in a tab. The next step is to choose the method of the check, either rule-based or machine learning. For instance, Figure 47 once pressed the Predictions button, shows the predictions results from the user inputs names. Then compares it with the list of objects parameters; therefore, if the data are consistent with each other, it presents that this object passed in green colour; otherwise, it shows that it is failed and showing in red font colour.

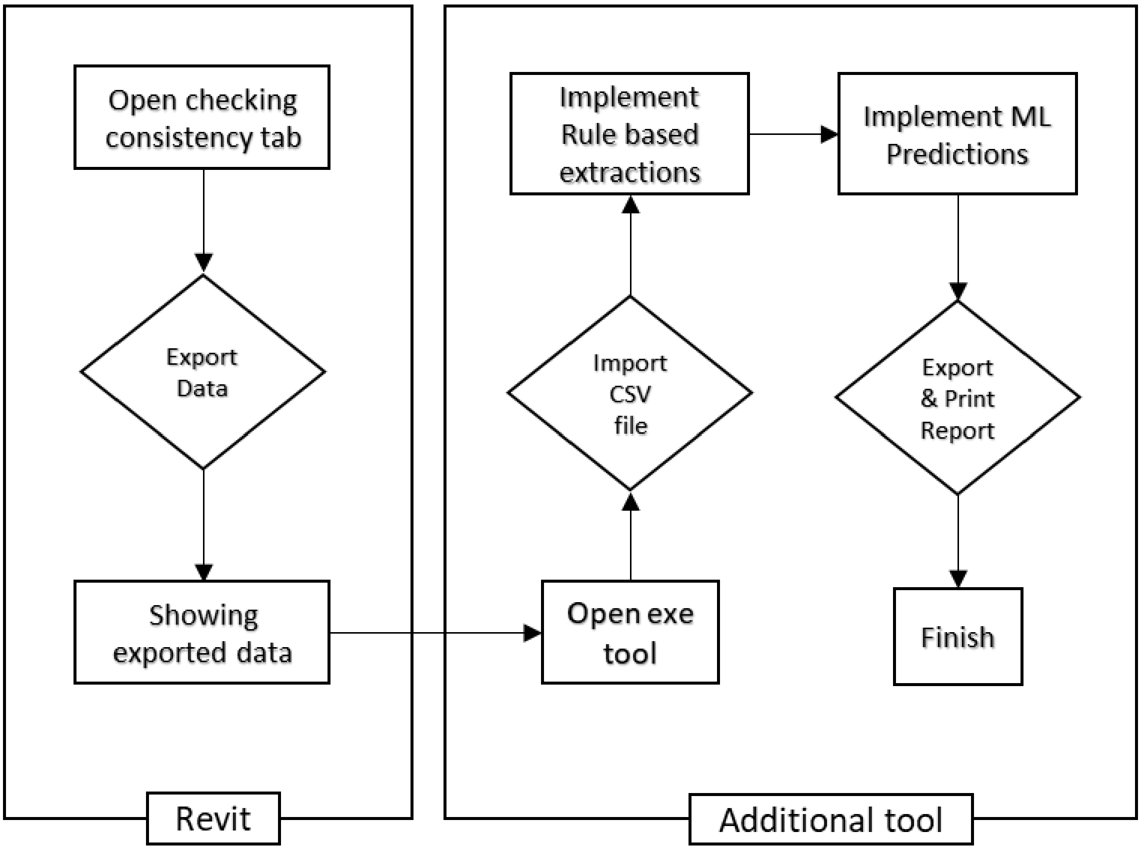


Figure 44 - Checking tool working sequence

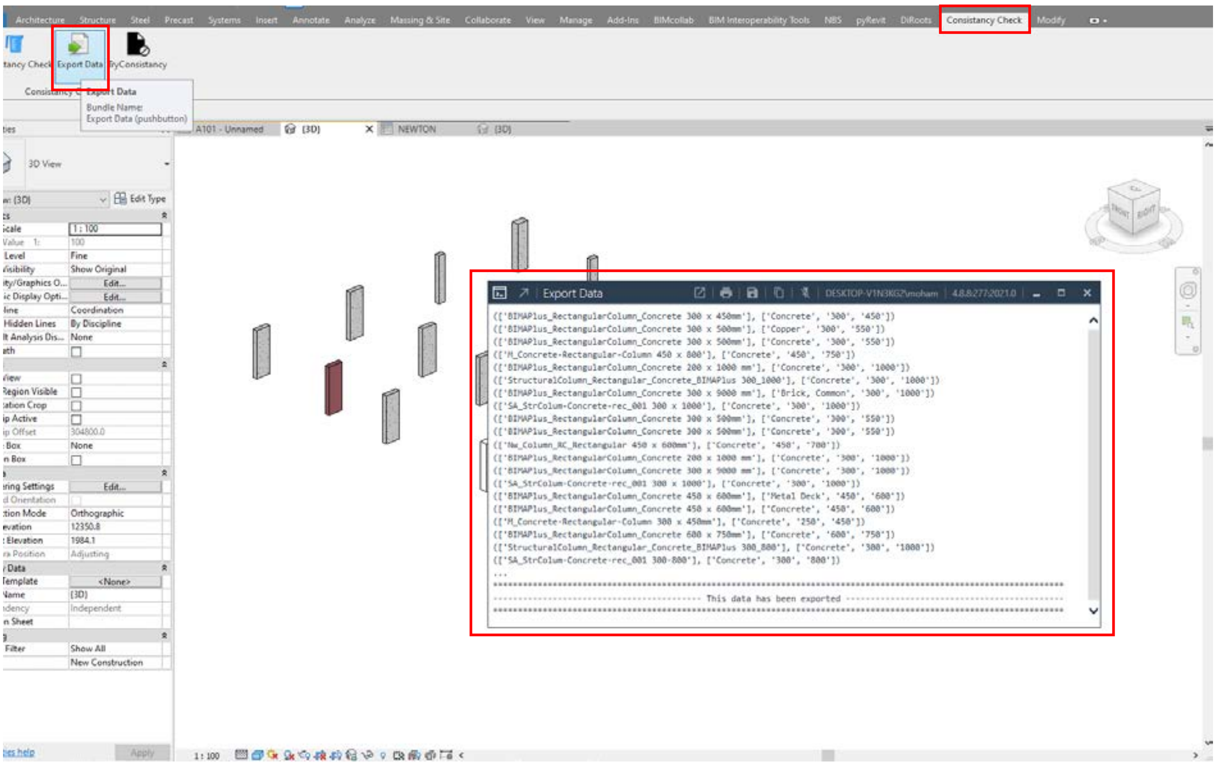


Figure 45 - Exporting data from Revit step

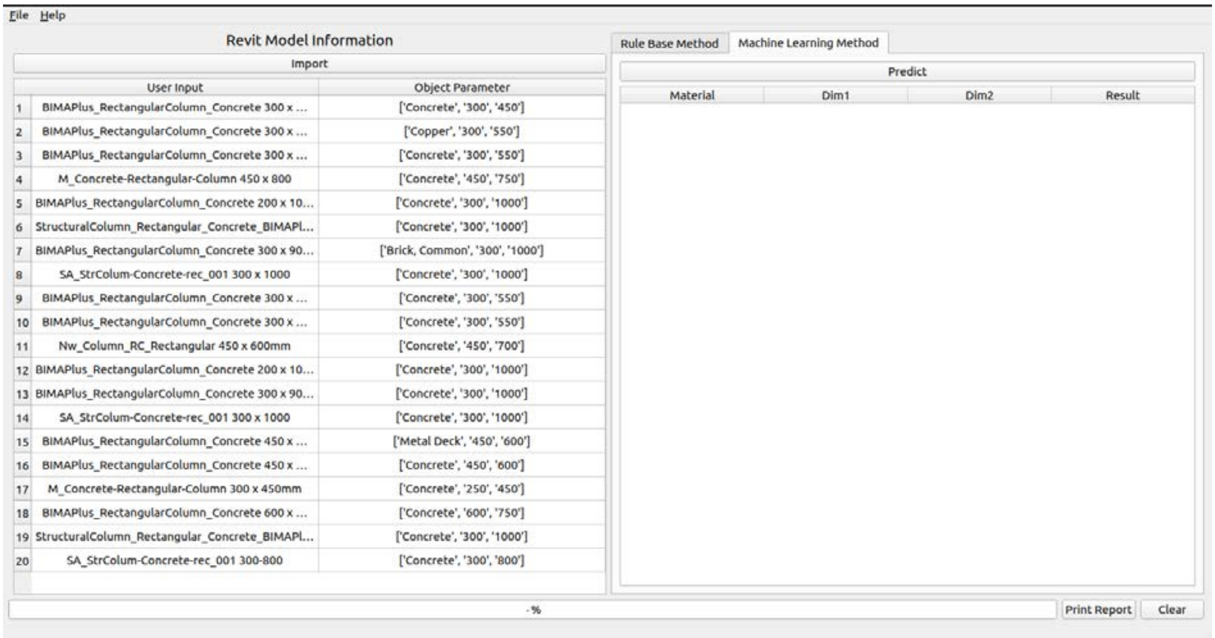


Figure 46 - Tool user interface and importing data

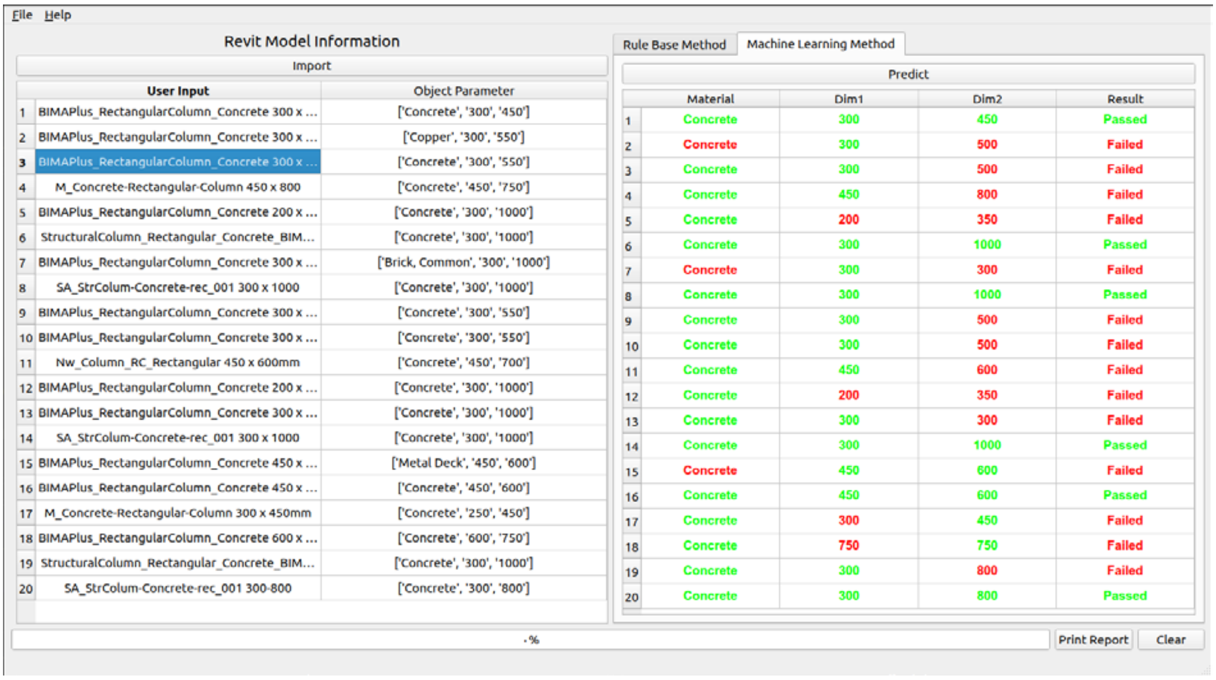


Figure 47 - Result reporting

4.6. Overall Discussion and Future Developing

It was apparent the impact factor and importance of the selected issue. The consistent information inside BIM models can be considered a high impact effect, specifically in quantity take-off, detailed drawings, and cost estimation. Following the proposed solution for solving that problem through either a Rule-based approach or machine learning approach both have limitations and pros and cons. However, both approaches reach an acceptable point of accuracy or dealing with issues in an automated way. Moreover, in terms of the availability to generalise the solution to cover the varieties in objects and companies' standards, machine learning is ideal than rule based.

The most straightforward approach for solving that issue was the rule-based approach. Nevertheless, the main cons of this approach are that it cannot be generalised; also, it needs to be customised for each case and with the diversity of companies naming conventions will be not effective. For instance, each company has its naming conventions, types or writhing and defining the materials or dimension; for that case, it is needed to customise the code and dictionary every time to check the model with the tool by implementing the rule-based approach. Moreover, creating an interface window to add the new terms and naming conventions will not be sure that the code will work perfectly with new data and mishandle it.

On the other hand, checking the model consistency by implementing a machine learning approach will allow the user to add the new naming conventions if it does not exist in the database and re-training the models again. Thus, it will get a more accurate result without customising the code or checking blindly like the rule-based model. Still, the core idea of implementing machine learning approaches needs more development to overcome the limitations of accuracy in the dimensions explained in the previous section. The following Figure 48 explains the core idea and develops it generalised and more accurate for use in the BIM Model consistency check.

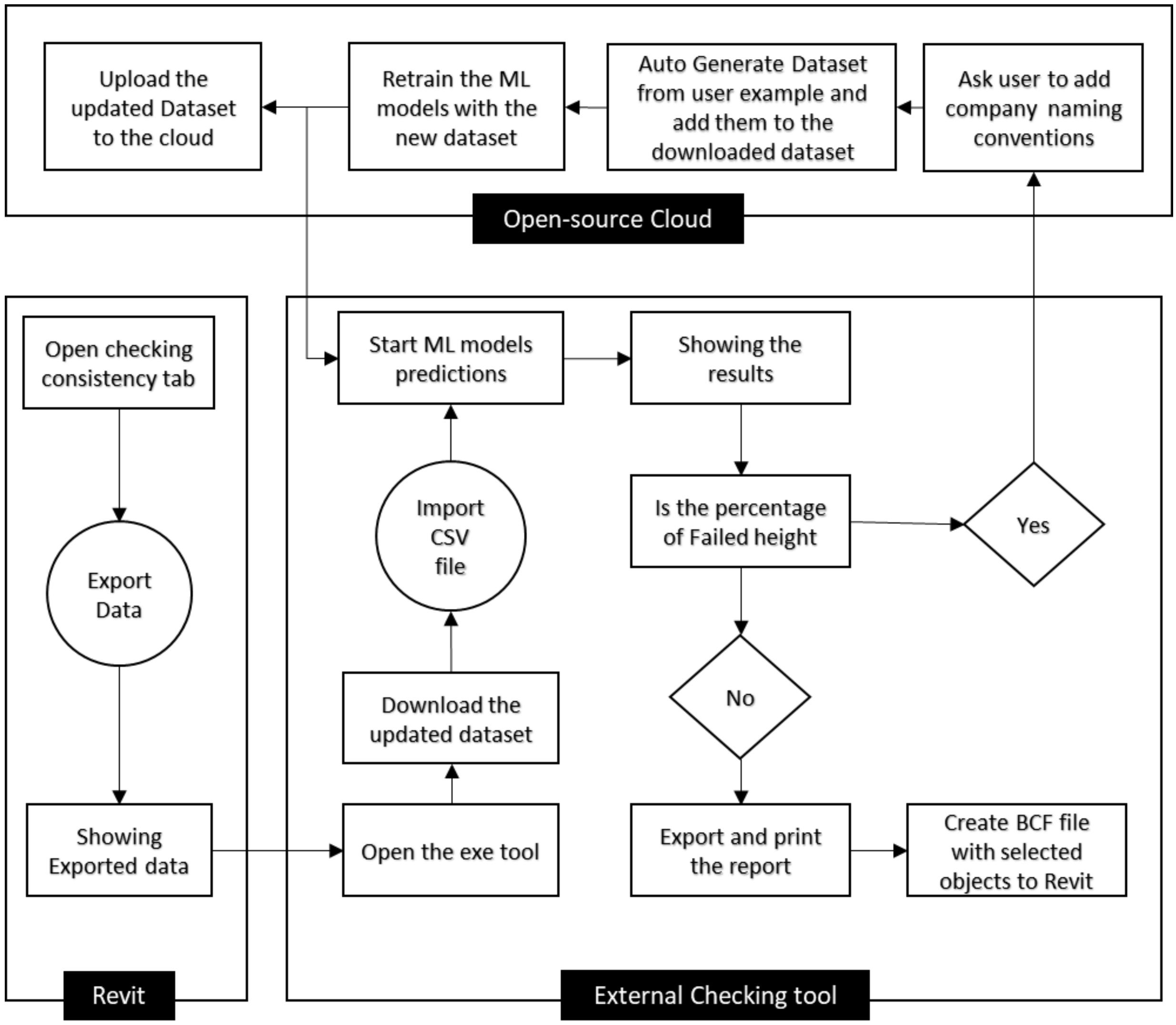


Figure 48 - Process map for checking tool future development

5. CONCLUSIONS

Building information modelling (BIM) models quality in checking, quality assurance, and quality control is a comprehensive expression and cannot summarise in one aspect. This thesis has been developed to answer several questions related to the BIM model quality. The first question is what is the clusters and definition of BIM model quality have been done?. The second question was How it has been controlled and defined the BIM models from the practical field. The third question was what are the factor, issues, errors, and challenges influence BIM model quality. The last question was how it could be solved and overcome those factors or issues.

After reviewing several published articles in terms of BIM model quality, it was considered that the model quality has different phases and levels. BIM model quality can be checked in the native file format level or the exported model in another format such as ifc. Both approaches have different limitations and ways of checking. The clusters of BIM model quality in either checking the federated model in the authoring tool or exporting the model to additional checking software can be classes to several classes. Code compliance checking is one of BIM model quality checking. Code compliance is known as checking the object's geometry or spaces following the code regulation. Code compliance is done using additional software and cannot be performed by the authoring tool, which is one of the issues illustrated in this thesis. The coordinator or checker only performs the code compliance by setting the rules in the coordination phase, and the modeller cannot have guidance or rules guide him during the modelling phase. Furthermore, that is a time-consuming factor and affects the BIM process and the workflow.

Another term within the BIM model quality context is clash detection. The clashes or interference between model objects have several levels: soft clashes and brutal clashes. Also, clashes can exist between the same model objects or between different objects from different models such as the architecture, structural, and MEP models. Clash detections are the most popular term in the context of BIM model quality. Issues like early clashes detecting during the modelling phase, clashes or connecting elements in structural analytical models are illustrated in this thesis and discussed their effect and proposed solutions.

Answering the second thesis question about How it has been controlled and defined the BIM models from the practical field was needed to figure out how to control and create quality BIM models. Therefore, A questionnaire has been designed with several parts to detect and ask about the companies, organizations, and individual practises in the context of BIM model quality. The participants for this questionnaire had suitable varieties in BIM competence level and diversity in the domains such as Architecture, Engineering, and constructions.

The designing sections of the questionnaire were divided into main three parts. The first part was about the internal BIM modelling manual. About half of the participants had not specified a modelling manual to follow according to the level of information needed. That also illustrated one of the crucial factors that affect the produced BIM model quality. A second important section of the questionnaire asked how it has been checking the quality of the model. the result summarized from participants answers that about 50% of individuals or companies had not specified structured rules to follow for controlling and checking the model quality. Besides, this part of the questionnaire mentioned some critical criteria that affect model quality. The participants were asked to set the scale of the importance and add other criteria to affect if needed. The criteria and factors have a high impact on the BIM model, such as the objects classifications, coordinating the model objects with other disciplines model, and the All-floor levels defined and uniquely named. The last part of the questionnaire was about the tools and software used in modelling and quality checking and the obstacles or reasons for not using specific tools. Most companies have obstacles to using specific useful software because of budges. For example, some software offers more than the client needs, and others see it because it is not easy to use.

In addition to the questionnaire as a tool for gathering information related to BIM model quality, interviews had been managed with several BIM engineers. It has been considered diversity in the regions, countries, and BIM positions such as BIM managers, coordinators, and modellers. Those interviews were more helpful for completing the whole image and frame under the BIM model quality control and checking context. The outcomes of those interviews in terms of factors effects on model quality were many. Some of those factors were time-consuming in repetitive modelling tasks, the time needed to check the model quality manually by checklist. Other factors mentioned were that the workflow and the BIM process map do not have specifics for model quality control and software or tools barriers in terms of cost or use. Furthermore, the results and proposal ideas that might be helpful to overcome those issues have been illustrated in part of the thesis. Those issues and solutions ideas were heightened by the engineers working daily and facing those issues on a daily basis.

Considering the time and working limits during the thesis development, and after reviewing and illustrating the outcomes issues from the interviews. One issue or error that significantly influences BIM model quality has been selected for the developing phase and creating an automated tool for solving it. This selected issue is the consistency of BIM objects check. It is the way to check the consistency of the information of the objects with the object assigned parameters. It focuses on objects, family names, and type names that contain information inside those names semantically, such as the material and the object's dimensions.

Furthermore, this consistency check compares the extracted information semantically from the name and the actual object parameters (dimensions and material of the object type instance). This error comes from the modeller when duplicating the object type and renaming it with the new parameters. But, because of time, they forget to modify the object`s parameter consistency with the new names. The selected error's importance to solve relies on its crucial effect on the quantities, schedules and drawings produced from the model that contains such error. Another significant factor is that this kind of error needs human intelligence and manually checking; no tool or add ion solved it in terms of semantically checking, not only naming conventions or existence.

The straightforward proposed approach to solve this issue was by creating rule-based code depending on a defining dictionary. This dictionary has all possibilities for writing and defining materials, types, objects. This approach uses Regular Expression to search on the object`s family and types of names, detect any words or acronyms that match with ones in the dictionary, and return the dictionary key with the corresponding material or type. The same concept has been applied in dimensions extracted from the objects type names. Using Regex extracts all numbers from the texts (names) and put them in a list. By testing this code, it works properly. Nevertheless, there are many limitations, and this approach cannot be generalised in terms of the diversity of companies naming conventions and standards.

The limitations of the first approach (Rule-based) code can be concluded as; 1) Every time the company has a new term or word to represent material or types, it will be required to open the background code and add the new word. Moreover, even it can be solved by the user interface box to make it easier, but it is still limited to fixing any issues in the dictionary in terms of code editing. 2) The dimensions detecting methods. The first method depended on the patterns and the possible ways of writing the dimensions, and once the company or other users have a different pattern, the code will not be able to detect it. Therefore, it will be required to edit the code and add or modify the patterns or define new ones. Also, in the same limitation of detecting the dimensions by implementing the second method, which blindly detects all numbers from the name and puts them in one list, this method will not be accurate if the name has several numbers, not just the dimension.

The second and more generalised approach for solving that issue was to apply the machine learning approach. The core idea of this approach was to create a dataset for the names of the objects and train the model to use artificial intelligence to predict the information from the objects name and compare it with its corresponding information in the model. Implementing the ML approach was compatible with generalizing the problem and easily adapting to any new data. Thus, there is no need to customize the code and the tool for any new information like the Rule-based approach. Furthermore, ML models accurately predicted material, type, and units because the number of classes was limited and small. However, the machine learning trial, either the classifications or the regression, failed to predict the complete accuracy specifically in dimensions because dealing with the dimensions as a class makes them too many classes for predictions. Moreover, dealing with the dimensions as numbers for the regression estimation cannot befitting the regression number to predict 100% accuracy. Thus, the dataset's limitations are the main obstacle in developing an accurate machine learning model used in the consistency check tool.

The ML approach will need more developments, such as connecting it to an open-source cloud with updated models and datasets after each company check. For a while, the dataset and the machine learning models will cover the most used naming conventions and make the tool accuracy almost 100%. that is, a generalised approach can be developed in the machine learning approach for BIM model consistency check.

The selected issue to propose and find a proper and practical solution for it is only one of the possible causes of data integrity and consistency of BIM models. Furthermore, the solutions could be extended to the more generic issues: checking non-parametric information fields with other fields in the model and other data from different sources. Those issues are more crucial for model quality because they need to be checked manually and be coordinator vision or review. Although, the proposed and tested solution still needs more developments; however, it allows the coordinator or checker to check those kinds of errors with the help of ML and in an automated way.

This page is intentionally left blank

REFERENCES

Autodesk (2021) *AUTODESK BIM INTEROPERABILITY TOOLS*. Available at: https://interoperability.autodesk.com/.

Azhar, S. and Brown, J. (2009) 'Bim for sustainability analyses', *International Journal of Construction Education and Research*, 5(4), pp. 276–292. doi: 10.1080/15578770903355657.

Belsky, M., Sacks, R. and Brilakis, I. (2016) 'Semantic Enrichment for Building Information Modeling', *Computer-Aided Civil and Infrastructure Engineering*, 31(4), pp. 261–274. doi: 10.1111/mice.12128.

Bloch, T. and Sacks, R. (2018) 'Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models', *Automation in Construction*, 91(July 2017), pp. 256–272. doi: 10.1016/j.autcon.2018.03.018.

Bloch, T. and Sacks, R. (2020) 'Clustering Information Types for Semantic Enrichment of Building Information Models to Support Automated Code Compliance Checking', *Journal of Computing in Civil Engineering*, 34(6), p. 04020040. doi: 10.1061/(asce)cp.1943-5487.0000922.

Donato, V., Lo Turco, M. and Bocconcino, M. M. (2018) 'BIM-QA/QC in the architectural design process', *Architectural Engineering and Design Management*, 14(3), pp. 239–254. doi: 10.1080/17452007.2017.1370995.

Eastman, C. *et al.* (2009) 'Automatic rule-based checking of building designs', *Automation in Construction*, 18(8), pp. 1011–1033. doi: 10.1016/j.autcon.2009.07.002.

Fahad, M. *et al.* (2019) 'Semantic BIM Reasoner for the verification of IFC Models To cite this version : HAL Id : hal-02270827 Semantic BIM Reasoner for the verification of IFC Models', *Proceedings of the 12th European conference on product and process modelling (ECPPM 2018), Copenhagen, Denmark, September 12-14, 2018, eWork and eBusiness in Architecture, Engineering and Construction - Jan Karlshoj, Raimar Scherer (Eds) Taylor & Francis* .

de Farias, T., Roxin, A. and Nicolle, C. (2018) 'A rule-based methodology to extract building model views', *Automation in Construction*, 92. doi: 10.1016/j.autcon.2018.03.035.

Galkina, E. and Kuzina, O. (2018) 'Building information model verification at the lifecycle stage of construction', *IOP Conference Series: Materials Science and Engineering*, 365(6). doi: 10.1088/1757-899X/365/6/062031.

Galle, P. (1995) 'Towards integrated, "intelligent", and compliant computer modeling of buildings', *Automation in Construction*, 4(3), pp. 189–211. doi: 10.1016/0926-5805(95)00004-K.

Goyvaerts, J. (2016) *Regular Expression Tutorial - Learn How to Use Regular Expressions*. Available at: www.regular-expressions.info.

Gray, M. *et al.* (2013) 'Building Information Modelling: an international survey', *World Building Congress 2013, 5-9 May 2013, Brisbane, QLD*, (May), pp. 5–9. Available at: http://www.researchgate.net/publication/258642116_A_tentative_comparison_of_the_performance_of_Strategic_Alliance_and_DesignBidBuild_project_delivery_methods/file/3deec528c6b941e678.pdf%5Cnhttp://eprints.qut.edu.au/56516/.

Huang, B. *et al.* (2021) 'Contribution and obstacle analysis of applying BIM in promoting green buildings', *Journal of Cleaner Production*, 278, p. 123946. doi: 10.1016/j.jclepro.2020.123946.

Huang, M. Q., Ninić, J. and Zhang, Q. B. (2021) ‘BIM, machine learning and computer vision techniques in underground construction: Current status and future perspectives’, *Tunnelling and Underground Space Technology*, 108(February). doi: 10.1016/j.tust.2020.103677.

Jenkins, J. G. (1940) ‘The Questionnaire as a Research Instrument*’, *Transactions of the New York Academy of Sciences*, 2(5 Series II), pp. 118–140. doi: 10.1111/j.2164-0947.1940.tb00075.x.

Jin, R. *et al.* (2017) ‘BIM Investment, Returns, and Risks in China’s AEC Industries’, *Journal of Construction Engineering and Management*, 143(12), p. 04017089. doi: 10.1061/(asce)co.1943-7862.0001408.

Mekawy, M. and Petzold, F. (2018) ‘BIM-based model checking in the early design phases of precast concrete structures’, *CAADRIA 2018 - 23rd International Conference on Computer-Aided Architectural Design Research in Asia: Learning, Prototyping and Adapting*, 2(May), pp. 71–80.

Mihindu, S. and Arayici, Y. (2008) ‘Digital Construction through BIM Systems will Drive the Re-engineering of Construction Business Practices’, in *2008 International Conference Visualisation*, pp. 29–34. doi: 10.1109/VIS.2008.22.

Mirarchi, C. and Pavan, A. (2019) ‘Building information models are dirty’, *Proceedings of the 2019 European Conference on Computing in Construction*, 1, pp. 131–140. doi: 10.35490/ec3.2019.180.

Mitkov, R. (2005) *The Oxford Handbook of Computational Linguistics (Oxford Handbooks)*. USA: Oxford University Press, Inc.

Nawari, N. O. (2012) ‘Automating Codes Conformance’, *Journal of Architectural Engineering*, 18(4), pp. 315–323. doi: 10.1061/(asce)ae.1943-5568.0000049.

Nawari, N. O. (2013) ‘Smartcodes and BIM’, *Structures Congress 2013: Bridging Your Passion with Your Profession - Proceedings of the 2013 Structures Congress*, pp. 928–937. doi: 10.1061/9780784412848.082.

Solihin, W. and Eastman, C. (2015) ‘Classification of rules for automated BIM rule checking development’, *Automation in Construction*, 53, pp. 69–82. doi: 10.1016/j.autcon.2015.03.003.

Solihin, W., Eastman, C. and Lee, Y. C. (2015) ‘Toward robust and quantifiable automated IFC quality validation’, *Advanced Engineering Informatics*, 29(3), pp. 739–756. doi: 10.1016/j.aei.2015.07.006.

Su, T., Li, H. and An, Y. (2021) ‘A BIM and machine learning integration framework for automated property valuation’, *Journal of Building Engineering*, 44. doi: 10.1016/j.jobe.2021.102636.

Succar, B. and Kassem, M. (2015) ‘Macro-BIM adoption: Conceptual structures’, *Automation in Construction*, 57, pp. 64–79. doi: https://doi.org/10.1016/j.autcon.2015.04.018.

Törmä, S. (2013) ‘Semantic Linking of Building Information Models’, *2013 IEEE Seventh International Conference on Semantic Computing*, pp. 412–419.

Wang, X. (2014) ‘BIM Handbook: A guide to Building Information Modeling for owners, managers, designers, engineers and contractors’, *Construction Economics and Building*. doi: 10.5130/ajceb.v12i3.2749.

Waterhouse, R. (2017) ‘Specification is at the heart of construction information’. Available at: www.linkedin.com/in/.

World Economic Forum (2018) ‘‘An Action Plan to Accelerate Building Information Modeling Adoption’.

Zaniolo, C. (1984) ‘Database relations with null values’, *Journal of Computer and System Sciences*, 28(1), pp. 142–166. doi: 10.1016/0022-0000(84)90080-1.

Zhang, S. *et al.* (2015) ‘BIM-based fall hazard identification and prevention in construction safety planning’, *Safety Science*, 72, pp. 31–45. doi: 10.1016/j.ssci.2014.08.001.

This page is intentionally left blank

LIST OF ACRONYMS AND ABBREVIATIONS

AECO	Architecture, Engineering, Construction, and Operation
AEC	Architecture, Engineering, Construction
AI	Artifical Inteeligance
AVMs	Automated Valuatoin Models
BDS	Berkely Source Distribution
BIM	Building Information Modeling
CEO	Chief Excutive Officer
CSV	Comma Spwerated Values
DMC	Digital Model Checking
EIR	Employee Infroamtion Requirments
ifc	Industry Foundation Classes
Lod	Level Of Details
ML	Machine Learning
MEP	Mechamical, Electrical, and Plumbing
MVD	Model View Defintions
Pndas	Panal Data
QA	Quality Assurance
QC	Quality Controle
SQL	Structures Query Lanaguage
STEP	Standared for Exchange of Product data
TSV	Tab Seperated Values

This page is intentionally left blank

APPENDICES

APPENDIX 1: QUALITY CHECK QUESTIONNAIRE

BIM model Quality Check Questionnaire

① Page 1 ② Page 2 ③ Page 3 ④ Page 4 ⑤ Page 5 ⑥ Page 6

In a part of the research for a master thesis about " **Systematic Quality check of Building Information Models,"** we have the target of surveying and defining the main issues, repetitive works, and the problem of the workflow for creating complete and quality BIM models during the different phases.

So the part of this study is contacting with the leading companies in BIM field, justify where is the area of the lack of information can be, and the need to develop it in order to develop and deliver BIM models as complete as possible with low effort and less time-consuming.

Please think carefully and answer truthfully as your anonymity is completely assured.

it wouldn't take more than 15 minutes maximum

Next

1 / 6

BIM model Quality Check Questionnaire

(1) Page 1 (2) Page 2 (3) Page 3 (4) Page 4 (5) Page 5 (6) Page 6

Personal and Organization Backgrounds:

1.1. What is your training or educational background? *

- ◯ Bachelor
- ◯ Master.
- ◯ Ph.D.
- ◯ Specialized training
- ◯ Other

1.2. What is your training or educational Specialization? *

- ◯ Architecture
- ◯ Civil Engineering
- ◯ Structural Engineerring
- ◯ Construction.
- ◯ Mechanical Design
- ◯ Electrical Design
- ◯ Plumbing design
- ◯ Management
- ◯ MBA
- ◯ Other

1.3. What is your company specialization? (Check all that apply.) *

- ☐ Architecture
- ☐ Structural Engineerring
- ☐ Mechanical Design
- ☐ Electrical Design
- ☐ Plumbing design
- ☐ Construction.
- ☐ Management
- ☐ Other

1.4. How many years of Work experience do you have? *

- ◯ Zero (Non)
- ◯ Less than one (<1)
- ◯ (1 ≤ & < 2)
- ◯ (2 ≤ & < 3).
- ◯ (3 ≤ & < 4).
- ◯ (4 ≤ & < 5).
- ◯ (5 ≤ & < 10).
- ◯ (10 ≤ & < 20).
- ◯ More than 20 (& >20).

1.5. How many years of work experience do you have through the BIM context? *

- ◯ Zero (Non)
- ◯ Less than one (<1)

- (1 ≤ & < 2)
- (2 ≤ & < 3).
- (3 ≤ & < 4).
- (4 ≤ & < 5).
- More than 5 (& >5).

Which one of those describes your role in your organization? (for reference, see below) *

- A chief executive officer (CEO)
- BIM Manager
- BIM Coordinator
- BIM Modeler
- Other

The definitions of BIM roles According to the BIMe Initiative Dictionary

- **BIM Manager:**

A BIM Role played by an individual or an organization on behalf of the whole Project Team. The Project BIM Manager has many responsibilities (typically defined with the BIM Management Plan) which include: BIM Facilitation, coordinating data-exchange activities, fulfilling pre-defined Design Specifications and Delivery Specifications, and overall Model Quality control

- **BIM Coordinator**

A BIM Coordinator is BIM Role combining Model Management, Project Information Management and process management activities. The definition of a BIM Coordinator varies between companies/markets and is sometimes used interchangeably with BIM Manager or Project BIM Manager

- **Model Element Author (BIM modeler)**

The party originally responsible for generating a Model Component or set of components. The term is typically used in conjunction with Model Progression Specifications or - in the UK - Model Production and Delivery Table. A Model Element Author (MEA) is not always the same as the ultimate Model Element Owner

Back **Next**

2 / 6

BIM model Quality Check Questionnaire

(1) Page 1 (2) Page 2 (3) Page 3 (4) Page 4 (5) Page 5 (6) Page 6

BIM modeling manual

Does your organization have an internal manual for modeling containing modelling rules? *

- ◯ Yes
- ◉ No

Back | Next

3 / 6

BIM model Quality Check Questionnaire

① Page 1 ② Page 2 ③ **Page 3** ④ Page 4 ⑤ Page 5 ⑥ Page 6

BIM modeling manual

Does your organization have an internal manual for modeling containing modelling rules? *

- (●) Yes
- ( ) No

Is this manual introduced to every new employee involved in modeling in your company as a fundamental way of operation? *

- ( ) Yes
- ( ) No

Which of the following BIM uses are included in this internal manual? (Check all that apply.) *

- ☐ Capture Existing Conditions.
- ☐ Analyze Site Selection Criteria
- ☐ Coordinate Design Model(s)
- ☐ Author Cost Estimate
- ☐ Author Design
- ☐ Compile Record Model
- ☐ Author 4D Model
- ☐ Analyze Sustainability Performance
- ☐ Other

To which level or scale do you consider this internal manual is applied in your organization? *

	Completely Not applied	partially applied	Completely applied	N/A
Applying or observation of those internal manuals? *	○	○	○	○

To which level or scale do you consider this internal manual is clear, and it can be followed? *

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
internal manual is clear, and it can be followed *	○	○	○	○	○

Back **Next**

BIM model Quality Check Questionnaire

① Page 1 ② Page 2 ③ Page 3 ④ **Page 4** ⑤ Page 5 ⑥ Page 6

BIM modeling quality checking

Does your organization have structured rules for controlling and checking the quality of the BIM models? *

○ Yes

◉ No

Structured rules mean some formalized or non-formalized methods that you apply to check the model. (e.g. checklist, recommendations section, or automatic interference tools)

Back **Next**

4 / 6

BIM model Quality Check Questionnaire

① Page 1 ② Page 2 ③ Page 3 ④ Page 4 ⑤ Page 5 ⑥ Page 6

BIM modeling quality checking

Does your organization have structured rules for controlling and checking the quality of the BIM models? *

- (●) Yes
- ( ) No

Structured rules mean some formalized or non-formalized methods that you apply to check the model. (e.g. checklist, recommendations section, or automatic interference tools)

How are these rules integrated into your company? (Check all that apply.)

- ☐ Checklist
- ☐ Sections with recommendations about quality check
- ☐ Using default Automatic tools in your modeling software.
- ☐ Other

Are you updating those model quality checking procedures according to sequential experiences from different projects? *

- ( ) Yes
- ( ) No

To which level or scale, do you apply and follow those (Rules, instructions,...., model quality checking procedures)? *

	Completely Not applied	partially applied	Completely applied	N/A
Applying and following the rules for controlling and checking the quality BIM models? *	(●)	( )	( )	( )

Why is not it completely applied? *

To which level do you agree that those procedures are clear, and can be followed to control the BIM model quality? *

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
model quality checking procedures are clear, and can be followed *	○	○	○	○	○

Which of the following criteria are existing in your structured rules and to which extent do they significantly affect the quality of the BIM model? *

	Not Existing	Low Impact	Medium Impact	High Impact
Following the correct naming convention for views and sheets. *	○	○	○	○
Project location accurately identified *	○	○	○	○
Building Model checked and clean of modelling errors. *	○	○	○	○
Components modelled using correct objects. *	○	○	○	○
Excess and unused elements removed and purged from model. *	○	○	○	○
All floor levels defined and uniquely named. *	○	○	○	○
Model objects on the correct level. *	○	○	○	○
All tags should be parametric. *	○	○	○	○
All Drawings and Exported Sheets must be extracted from the model. *	○	○	○	○
All CAD imported files must be linked not imported. *	○	○	○	○
The building sections are completely dimensioned and annotated. *	○	○	○	○
Check that all interior elements are coordinated with other disciplines. *	○	○	○	○
All modelled elements shall be floor by floor, except indicated by Arch`s requirements *	○	○	○	○
All BIM extracted schedules shall be parametric. *	○	○	○	○
Checking about empty cells in the model *	○	○	○	○
Classification *	○	○	○	○
Information checks when applicable *	○	○	○	○

Is the quality of the models checked through the early stages of the BEP?

○ Yes

○ No

How are the quality model rules coordinated when working with external companies?

Back **Next**

4 / 6

BIM A+ European Master in Building Information Modelling

POLITECNICO MILANO 1863

Erasmus+

BIM model Quality Check Questionnaire

(1) Page 1 (2) Page 2 (3) Page 3 (4) Page 4 (5) Page 5 (6) Page 6

BIM modeling and checking Authoring tools

Which of the following software does your organization use as a BIM Authoring tool for creating the proprietary model? (Check all that apply.) *

- [ ] Allplan
- [ ] ArchiCAD
- [ ] Bentley Systems
- [ ] DDS-CAD
- [ ] Edificius
- [ ] Revit
- [ ] SketchUp.
- [ ] Tekla Structures
- [ ] Vectorworks Architecz
- [ ] Other

How does your organization check the quality of your model? (Check all that apply.) *

- [ ] Manual checking
- [ ] Using the proprietary software (with the native implement capacities).
- [ ] Using the proprietary software (with custom add on built by your developers)
- [ ] Using the proprietary software (with available purchase or free add on)
- [ ] Using additional software.

Is there anything you would like to add that might be useful for our survey ?

Back Next

5 / 6

BIM model Quality Check Questionnaire

① Page 1 ② Page 2 ③ Page 3 ④ Page 4 ⑤ **Page 5** ⑥ Page 6

BIM modeling and checking Authoring tools

Which of the following software does your organization use as a BIM Authoring tool for creating the proprietary model? (Check all that apply.) *

- [ ] Allplan
- [ ] ArchiCAD
- [ ] Bentley Systems
- [ ] DDS-CAD
- [ ] Edificius
- [ ] Revit
- [ ] SketchUp.
- [ ] Tekla Structures
- [ ] Vectorworks Architecz
- [ ] Other

How does your organization check the quality of your model? (Check all that apply.) *

- [ ] Manual checking
- [ ] Using the proprietary software (with the native implement capacities).
- [ ] Using the proprietary software (with custom add on built by your developers)
- [x] Using the proprietary software (with available purchase or free add on)
- [x] Using additional software.

If you are using the proprietary software (with a purchase or free add-on) what is the name of this add-on? *

If you are using additional software, which of those software does your organization use for model quality checking? (Check all that apply.) *

- [ ] BEXEL MANAGER
- [ ] BIM ASSURE
- [ ] BIMCOLLAB
- [ ] BIM TRACK
- [ ] NAVISWORKS MANAGE
- [ ] SIMPLEBIM
- [ ] SOLIBRI MODEL CHECKER
- [ ] TRIMBLE QUADRI
- [ ] REVIZTO
- [ ] Zbuilder
- [ ] Trimble Connect
- [ ] Other

Do you think that additional sofware would help in the quality checking activity? *

- ○ Yes
- ○ No

why do you think that some of these tools are not included in your company`s software or tools? *

- ☐ Because of Budgets
- ☐ Software offering more than I need.
- ☐ Not easy to use.
- ☐ Other

Can you please mention the software you think will be useful and have not been included in your company tools?

What are the customizations you need to be added for those tools to make it more efficient for checking the quality of your model?

Is there anything you would like to add that might be useful for our survey ?

Back | Next

5 / 6

BIM A+ European Master in Building Information Modelling

POLITECNICO MILANO 1863

Erasmus+

BIM model Quality Check Questionnaire

① Page 1 ② Page 2 ③ Page 3 ④ Page 4 ⑤ Page 5 ⑥ **Page 6**

The questionnaire will be **anonymous** by default. However, to collect information from different stakeholders in the same office or the same company and relate them, also relate country-wise responses, and for a potential follow-up interview, we would like you to provide your identification. It enables us to get back to you later and keep in touch with you.

That means **your identification will not be revealed** in any way in the public disclosure of the questionnaire results. **Any identification disclosure would be preceded by your express authorization.**

According to the mentioned conditions above, you may require to fill the following cells. Anyhow, if you are reluctant to fill the cells, feel free to leave them blank.

Name

First | Last

Email

Organization or company Name:

Location (Country)

Back **Submit**

6 / 6