CHALMERS
UNIVERSITY OF TECHNOLOGY

The application of Big Data technology to improve the information sharing and enhance the decision-making process in construction projects

Master's Thesis in the master's Program Design and Construction Project Management

ISSA ALALI
SAAD ALMERI

Department of Architecture and Civil Engineering
Division of Construction Management
CHALMERS UNIVERSITY OF TECHNOLOGY
Thesis XXXXXXX
Gothenburg, Sweden 2020
www.chalmers.se

MASTER'S THESIS, 2020 **XXXXXXX**

The application of Big Data technology to improve the information sharing and enhance the decision-making process in construction projects

Master's Thesis in the master's Program Design and Construction Project Management

ISSA ALALI
SAAD ALMERI

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Construction Management
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020
www.chalmers.se

The application of Big Data technology to improve the information sharing and enhance the decision-making process in construction projects

Master's Thesis in the master's Program Design and Construction Project Management

ISSA ALALI
SAAD ALMERI

© ISSA ALALI AND SAAD ALMERI, 2020.

Supervisor: Dimosthenis Kifokeris, Department of Architecture and Civil Engineering
Examiner: Christian Koch, Department of Architecture and Civil Engineering

Examensarbete xxxxxxx

Institutionen för arkitektur och samhällsbyggnadsteknik

Chalmers tekniska högskola, 2020

Department of Architecture and Civil Engineering
Division of Construction Management
Chalmers University of Technology
E-412 96 Göteborg
Sweden
Telephone: + 46 (0)31-772 1000
Department of Architecture and Civil Engineering Göteborg, Sweden, 2020

The application of Big Data technology to improve the information sharing and enhance the decision-making process in construction projects

Master's thesis in the master's Program Design and Construction Project Management

ISSA ALALI
SAAD ALMERI

Division of Construction Management
Department of Architecture and Civil Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY

Abstract

Construction joint-ventures projects suffer from difficulties in information sharing due to data fragmentation, the usage of different information sharing systems, and the temporary nature of projects, which hinders long-term investments in better systems, and the cultural and organizational differences between joint-ventures entities. These difficulties have a negative effect on the decision-making process. The aim of this study is to investigate if the application of Big Data technology can enhance information sharing and improve decision making by studying a construction joint-venture project in Sweden. To achieve this aim, a literature review was conducted to investigate the difficulties in the theoretical part, and a model based on Big Data software was created and used for data preparation, data linking, data analysis, and data visualization. The model was linked to different data sources on the cloud, and the results are published online to the concerned stakeholders to solve four major issues related to information sharing in the project. These issues were: providing and visualize construction costs details to the construction team, calculate and visualize cost performance index CPI to executive management, show details of payments certificates and budget to design team, and lastly matching list of invoices in the ERP system with a list that is sent to the client. The four issues were solved using the Big Data software, and four online dashboards were presented to the concerned stakeholders that were interviewed later. The interviews' results showed enhancement in information sharing in which most interviewees agreed that the presented dashboards helped to understand the status of the project/package and reduced the effect of the organizational and cultural differences in the joint venture. The results also showed improvement in decision-making, but the middle-level decision-makers that make tactical decisions benefited more from the presented dashboards than the high-level decision-makers that make more strategic decisions.

Keywords: Big Data, Decision-making, information sharing, Visualization, Dashboards,

Preface

Table of Contents

Table of Figures

Table of Tables

1. Introduction

1.1 Data-driven decision-making

From the tendering phase of a construction project to the project closing process, decisions are made daily. Deciding wither to take the project or not, what kind of technical solution to be used, how much risk reserve will be considered regarding cost and schedule, in-house construction or subcontracting, make or buy, lease or rent, which supplier to buy from, which person to hire, and hundreds of other decisions are an example of what decision-makers are dealing within the construction industry.

When dealing with a change-averse industry like construction where the cost of changes becomes very high later during the project (PMBOK, 2017), it is crucial to make as many right decisions as possible not only from the early phases of a project but also in every major activity.

To make the right decision, all information should be available to the decision-maker, including the facts, consequences, and forecasts of the decision. This is known as an informed decision (Amendola, 2002). However, this does not guarantee that the outcome of the decision is going to be as expected, but the informed decision rises the chances of getting positive outcomes.\

The decision making of these informed decisions is called data-driven decision-making DDDM, It means that the decision has been taken based on the analysis of data and not based on intuition only (Provost and Fawcett, 2013). The project managers who need to decide between two technical solutions, for example, will make a decision either based on their experience and intuition or will consider time, cost, and scope data analysis. However, a combination of the two approaches is usually used in reality (Provost and Fawcett, 2013). Data-driven decision making proved to influence the organization's productivity and financial measures (Brynjolfsson et al., 2011).

1.2 Big Data trends

For simple decisions, not a lot of data is needed. However, for complicated decisions, more data is required not only from one source but also from different sources. If we add the advancing technologies like the Internet of Things IoT for example, we get more data to be analyzed, and thus we need more advancing technologies to do that.

Hence the term Big Data emerges. It means that analyzing datasets with the traditional methods is not sufficient and new technologies are needed for data processing (Provost and Fawcett, 2013). Generally, special software is needed to process Big Data (Provost and Fawcett, 2013). Big Data Analysis aims to assist decision making by uncovering the latent trends inside the data (Bilal et al., 2016).

1.3 Data needs in construction

Nevertheless, the question that arises here does the construction industry have Big Data comparing to other industries like sales and social media for example? The answer is yes. Lue et al. (2018) advocate for this answer by giving two examples from the construction industry. The first example is about a Quantity Surveying firm that needs to manage datasets from several sources like design data, BIM models data, Bill of Quantities, Bidders information and offers, schedule data, and

construction reports. The second example is about a construction company that keeps a record for each offer it gets from suppliers and uses that as a database when bidding for new projects.

1.5 JVs

Getting data in construction projects requires information and knowledge sharing. Moreover, one of the top cases that suffer from information sharing is construction joint ventures due to their unique features. Kivrak et al. (2014) found that cultural differences, language barriers, communications difficulties, and trust are hindering information and knowledge sharing in joint ventures in construction projects.

Knowledge sharing in construction is handled partially by using Enterprise Resource Planning ERP systems, but ERP systems do not usually handle all types of information. Building information modeling BIM is found to handle this issue, but it is more dominant in the design stage and the facilities management stage with less use in the construction stage. That leaves a gap regarding knowledge and information sharing.

1.6 Initial problem

The case in this research is a joint venture between three companies working in an infrastructure project in Sweden. Both authors are working at this project in the Cost Control and Contracts departments, respectively.

In addition to the joint venture common issues, there are more factors in this case that affects information sharing, which in turn affects the decision-making process. Firstly, The ERP system that is used as a basis for some project operations belongs to one of the JV partners, which makes the other partners have a lack of knowledge on how to use it. Moreover, the used ERP system has issues with user experience and user interface. Secondly, as the ERP system was not able to link all data together, several sources of data emerged and were used by the teams. That created more problems as those sources cannot "talk" to each other. Briefly, the problem can be explained as follows:

- There are big raw data stored in deferent places (servers, systems, and files) and there is no system or method to gather it in one place where everyone can get access to it.
- This raw data is continuously subject to change and there is no system in which gives the ability to consider these changes or new inputs and visualize it easily and continuously in real-time.
- This raw data are also not analyzed efficiently to obtain useful information and knowledge in which can support the decision- making process by giving a clear vision about what is going on in the project and what are the risks/opportunities now and in the future.
- Finally, there is no common platform where the generated information is visualized and shared by all players to take advantage of the time factor (the right person receives at the right time the right information/analyzed data to make the right decision)

1.7 Research question

The aim of the study is to test and investigate if Big Data technology can solve the information sharing related issue and enhance the decision-making process in construction projects.

1.8 Plan to solve the problem

In order to reach the study's aim, the following steps will be performed:

1. Study the current information-sharing methods.
2. Analyze the needs and prioritize them based on the importance of decisions.
3. Prepare data sources and datasets
4. Build an information-sharing model by using a Big Data application to store, connect, and analyze datasets.
5. Visualize the integrated information in dashboards.
6. Share the results with decision-makers and work on their feedbacks.
7. Conduct a final survey and interviews to assess the benefits of the model regarding information sharing and decision making.

1.9 What the research will contribute

If the results of this model were positive, it is expected to offer a potential solution to a common issue in the construction industry, information sharing. Moreover, the ERP systems usually maintained by an IT specialist, but the model can be built and maintained by anyone with advanced MS Excel skills. That opens the way for engineers and industry practitioners to dive into data analysis without being specialized in it.

The aim of the model is to provide more information and enhance decision making. This will consequently lead to better cost-efficiency in the project and probably will increase the productivity of workers, which leads to better time efficiency. Those two aspects are vital for the economic sustainability of the joint venture company.

Moreover, the contract of the project is partially cost plus, meaning each cost-saving for the contractor will lead to the same cost-saving for the client, which is a public organization. That means a benefit to the society as the overall owner funds this project and can use the cost-saving for other projects.

2. Background

As mentioned before in the introduction, the project that authors are working in is managed by a joint venture. This joint venture consists of three companies from different countries. More than eleven different nationalities are working in the project and it has a mix between local and expatriate staff. The project organization gives the project manager an ultimate responsibility for project success objectives. A board consists of partners' representatives meet regularly to follow up on the project.

3. Theoretical framework

3.1. Big Data

3.1.1 What is Big Data

By 2001, broadband wires, portable phones, and B2C commerce were spreading widely over the world (Cheng, 2017). In those days, computer networks were released to open societies, and people could search for information on the Web and send messages to any place at any time. Around 2005, individuals and companies wanted to get access to information whenever or wherever they are (Cheng, 2017). This demand was fulfilled by cloud infrastructures, which provide free services and have become quite common. At this stage, a challenge was perceived that the stored information generated by information communication technologies could become too big to be treated. Later, around 2010, the term Big Data has shown up (Cheng, 2017).

Big Data is data that is large in quantity, traded high speed, and formed in many styles. These three features are called volume, velocity, and variety of data. Together they make Big Data impossible to be processed by traditional technologies (Cheng, 2017). Big Data technology consists of many computer nodes with specific hardware and software technologies (Cheng, 2017). This new technology can analyze different types of a massive amount of data to come out by hidden patterns, correlation, and other valuable information that could support decision- making process and help to obtain better business planning(Lu et al., 2018). Figure 1 by Cheng (2017) shows the different phases of Big Data processing.

Figure 1 - Phases of Big Data processing (Cheng, 2017)

Most of Big Data is generated by users in an unstructured way like graph data, voice, images, etc. However, in general, there are four Big Data categories, first, external structured data such as credit history, second, internal structured data such as inventory data, third, unstructured external data such as Facebook posts, and finally, unstructured internal data such as sensors data(Cheng, 2017). Big Data could also be categorized according to its nature to structured, which are well defined and organized in a standard way within databases, unstructured data that do not have a clear structure and semi-structured, which have kind of light organization(Ngo et al., 2020). Lu et al (2018) show that Big Data in the construction industry could be in different types, for example in case cost management data, it exists as drawings, bill of quantities, bidding document, schedule of rates, process Excel sheets, accounting records, suppliers and quotations' data, labor data and maintenance tables. Furthermore, it could be sorted by variety as graphics, texts, voice, pictures, and videos or by its status to paper-based or digital format created by software like Office Automation and AutoCAD (Lu et al., 2018).

3.1.2 Why is Big Data important for business?

The idea of analyzing data to understand the current situation of businesses and obtain a clear vision is not new. Since at least 1954, the activities of analyzing data and create useful information out of it have been called decision support (Davenport, 2014). Later, it took different names like executive support, online analytical processing, business intelligence, and now Big Data (Davenport, 2014).

While traditional information management approaches focus on supporting internal decisions, Big Data is somewhat different in which it is also used to analyze a pool of data and come up with a model that advises the decision-maker about the criteria and timeframes for which decisions need to be made (Davenport, 2014). Big Data analytics has been used to realize and manage the process, keeping high management aware of opportunities and inform them of irregularities (Davenport, 2014). `

Davenport (2014) states that by analyzing unused data, it could impact the whole business by shaping or transforming its strategy. This means that there are values of using Big Data technologies such as cost cut, decision improvement, and enhancing of products and services. Moreover, using Big Data analyses in the construction industry could allow the adoption of sustainable practices through adjusting policies that lead to better performance, decrease wastes, enhance energy efficiency and develop human assets (Omran and Chen, 2016).

In other words, the main aim of Big Data's digital strategy is to transform raw data into information and knowledge. This is not only about transforming the infrastructure of the information system but also about transforming the company's business (Cheng, 2017). Modern strategy theory states that competitive advantage is not something that companies won; the position game-changer innovation is the key to business success. Both the business strategy and business performance management are driven by decisions based on statistical models that optimize the data using analytic/visualization technologies, as shown in figure (2) (Cheng, 2017).

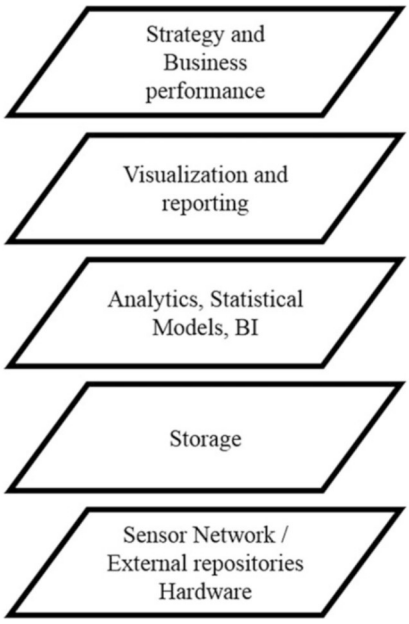

Figure 2 - Big Data layers (Cheng, 2017)

3.1.3 Visualizing Big Date

Visual representation is one of the most useful and valuable instruments to perceive Big Data. A good visualization can expose features, trends, and patterns, allowing creating actionable analysis and obtaining more in-depth insight (Cheng, 2017). There is a need for new techniques to overcome the challenges of visualizing data due to its speed, size, and diversity. Graphs and charts in which decision-makers are familiar to deal with whether it is possible to achieve the goal efficiently, but a new way to look at the data is required (Cheng, 2017). Using visualization to explore data and support the decision-making process is not new. It has been for a long time an essential part of data analysis and model building activities (Cheng, 2017). However, the challenges facing the visualization of Big Data have to do with the extent of the representation of the details and the relationships between variables (Cheng, 2017). In general, the main purposes of data visualization are the exploration of data, highlighting uncovered patterns and relationships, supporting decision-making processes, and finally, for communication (Cheng, 2017).

Cheng (2017) presents three levels of Big Data visualization. First, elementary -level visualization, which is useful to give fast and focused insights, the focus here is on specific, too narrow data, and one summary is produced per category. The second level is the intermediate -level visualization where a more significant amount of detail about the data is shown and graph semiotics are used to get a considerable amount of information in a single graphic; however, this level of visualization does not capture the whole database. The third level is overall-level visualization, where the focus is to produce responses that cover the entire data with an emphasis on general trends (Cheng, 2017).

3.2 Data needs and decision - making in construction projects

3.2.1 Decision making

Companies usually depend on their managers' intuition and experience to make decisions. Especially in case of a lack of digital form data, this approach has proven not to be efficient (Cheng, 2017). Organizations could lose market competitiveness by not obtaining higher maturity by adopting a new strategy to process their unused fragmented data systematically and create useful information (Cheng, 2017). Lu et al (2018) state that management should strive to pursue informed and rational decisions instead of entirely correct decisions. This means that even if the outcomes later are lower than what was forecasted, decision-makers should be aware of the facts, effects, and consequences of their decisions. Furthermore, Lu et al (2018) argue that decisions are made under three limitations according to bounded rationality theory, as shown in figure 3. First, the limited information about possible alternatives and their consequences, second, limitation of capacity to analyze the provided information, and finally, limitation of time to make decisions.

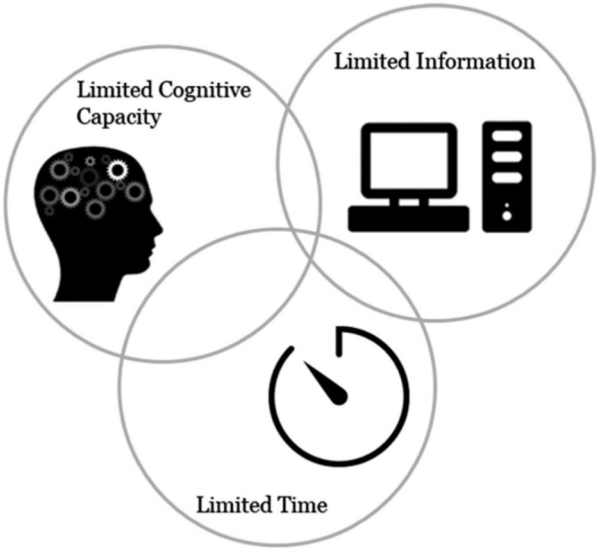

Figure 3 - Bounded rationality theory - (Lu et al., 2018)

Bounded rationality theory explains logically how individuals make decisions in the real world, taking the three limitations into consideration (Lorkowski and Kreinovich, 2018). According to Herbert Simon's bounded rationality theory from 1955, the three limits of rationality are classified as external and internal limitations (Lorkowski and Kreinovich, 2018). The external limitation is that our knowledge of the external world is not perfect, or what is called the uncertainty. This means that individuals do not spend their time, in case they have it, to gather information on all possible variables that could affect their decisions and its consequences. Instead, they get likely what is considered the most relevant information to make the decision within a specific time (Lorkowski and Kreinovich, 2018). The internal limitations, on the other hand, are the limits to compute and process a large amount of data. That means even if we have all possible data and inputs related to a specific given question, we cannot handle all of them because of the limited time and the internal cognitive bounds which characterize human rationality (Lorkowski and Kreinovich, 2018). Furthermore, these internal and external boundaries shape the procedure to be

most suitable to make a decision (Lorkowski and Kreinovich, 2018). Organizations are forced under this condition of unperfect information, limited resources, and restricted information processing capacity to make suboptimal and satisficing decisions based on what is available (Walker et al., 2014).

Marwala (2014) defines rational decision making as "*a process of making decisions based on relevant information, in a logical, timely and optimized manner*." The process starts with studying the decision-making space and then identifying the relevant necessary information and finally presenting this information to the decision-maker logically and consistently (Marwala, 2014). However, since a fully rational decision is impossible practically because of the three limits, according to Simon (1991), these limits can be **partially** corrected by implementing advanced information analysis methods (Marwala, 2014). In this case, the bounds of rationality can be expanded, and thus bounded rationality theory is transformed into the theory of flexibly bounded rationality (Marwala, 2014).

3.2.2 Decision making under Big Data

When it comes to making business decisions, companies usually depend on their high paid managers who gained good experience through their professional life, especially in case of data is not available or costs a lot to be obtained in digital form (Cheng, 2017). This approach of decision management, which relies on leaders' intuition, has proven that it is not sufficient (Cheng, 2017). Shifting toward facts-driven decisions is needed to enhance the quality and speed of decisions; as a result, the overall performance will be developed in the middle and long-term periods (Cheng, 2017).

Available and easy to read data can make it more possible for construction leaders to discover patterns, relations, and facts which could enhance the progress steadily (Safa and Hill, 2019). By gathering, linking, and analyzing raw data, management can increasingly know more about their business status and act on what they know (Davenport, 2014). Big Data analytics help to create useful information and knowledge, which in turn can support the human decision (Lu et al., 2018). By adopting specific and automated processes, organizations can covert the result of processing data into practical information, knowledge, and finally wisdom (Lu et al., 2018). Figure 4 by Lu et al (2018) shows that raw data could be a number formula and facts. Analyzing this raw data in structured way produces information. Knowledge, on the other hand, is created as justified information.

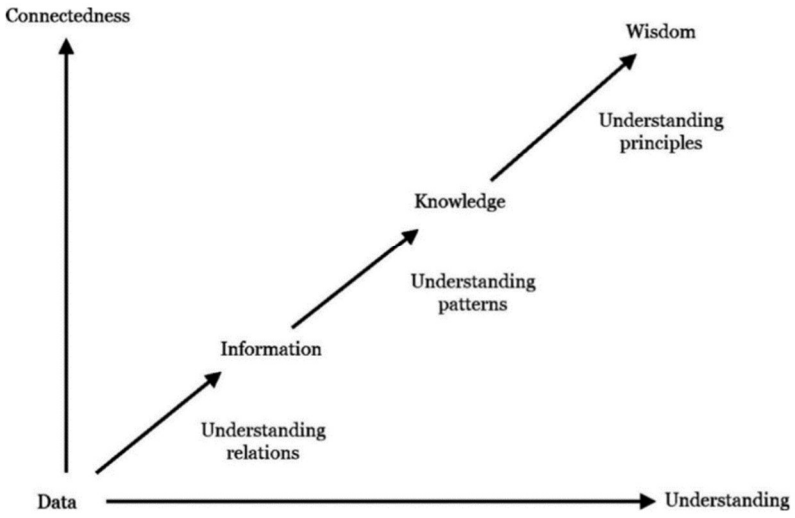

Figure 4 - Data, information, knowledge, and wisdom (Lu et al., 2018)

Accurate and on-time information is vital for good and realistic decisions. More and more firms are changing their strategic intelligence from an intuitive approach to a more detailed and systematic one (Cheng, 2017). Digital facts-based systems allow linking different and fragmented functions together in one Big Data platform (Cheng, 2017). Managers, in this context, receive combined reports that give the completed and real-time situation of the work. This makes them more capable of making rational decisions (Cheng, 2017). In addition, Cheng (2017) states that faster decisions could be a critical factor for organizations to take advantage of market competition. Faster decisions could be achieved by providing the right information at the right time to the right decision-maker in the right way, and that is what Big Data stands for (Cheng, 2017).

3.2.3 The impact of temporary status of construction projects and the need for data management system

As the construction projects become more complex and larger, the generated information and data through the lifecycle of projects become bigger in volumes and types (Chassiakos and Sakellaropoulos, 2008). This makes it challenging to manage all this information in an effective and reliable way (Chassiakos and Sakellaropoulos, 2008). In this context, Databases and internet information management technology could be a useful and powerful way to handle this challenge (Chassiakos and Sakellaropoulos, 2008). Moreover, project leaders' decisions in all project's phases and especially in the pre-construction phase, are considered as a fundamental factor for the success of the project(Nasir et al., 2016). Managers should make sure that all necessary information during these phases is valid and can provide the correct inputs to the decision-making process in order to avoid possible risks and dilemmas (Nasir et al., 2016).

The temporary nature of construction projects or what it is called one-off projects makes it difficult or impossible in most cases to establish long-term unified information and knowledge management system. (Pirzadeh and Lingard, 2017). As a result of this temporary nature and the lack of a tool to control and ensure efficient information sharing within project's departments and the project as a whole, individuals and teams are forced to depend on and use the existing temporary networks to interact and exchange needed information, knowledge and ideas among each other (Pirzadeh and Lingard, 2017). Consequently, these temporary networks between the different players can not ensure the efficient and permanent flow of information and facts, which constitutes an obstacle to obtain satisfying performance (Pirzadeh and Lingard, 2017).

Furthermore, Pirzadeh and Lingard (2017) explain how efficient information sharing through a well-centralized communication network could support decision making. It is known that the decision-making process in the construction industry is a complex process in which different decision-making situations need different and specific information and knowledge (Pirzadeh and Lingard, 2017). Clear channels and interaction networks can provide this kind of suitable sources of knowledge and viewpoints collected from many specialists to achieve improved and informed decisions based on facts (Pirzadeh and Lingard, 2017).

Chassiakos and Sakellaropoulos (2008) criticize the document-based approach used in the construction industry, which manages information by dividing and storing documents into separated information entities. In this approach, the documents themselves are managed rather than the information included in them. Here, the information uniformity and flow can not be guaranteed since, first, the data contained in the documents is not structured and it takes too much time and effort to reach it, second, this kind of information management systems are rigid in which and allow limited interactivity with the end-user (Chassiakos and Sakellaropoulos, 2008). On the other hand, Chassiakos and Sakellaropoulos (2008) propose an alternative data-based approach that has the ability to overcome the previously stated limitations of document-based systems and improve the efficiency of information management. A Data-based system is based on deconstructing information to a chain of smaller pieces and then saving it in a structured, systematic, and organized way (Chassiakos and Sakellaropoulos, 2008). This system gives the ability to track, search, and analyze these data easily, which in turn guarantee to address the information needs consistently (Chassiakos and Sakellaropoulos, 2008).

3.3 Cultural and organizational differences' impact on construction joint venture projects

A joint venture in the construction industry, as well as in other industries, is recognized as a strategic choice for two or more parent companies to share risks and opportunities (Ozorhon et al., 2007). Diversity of cultures, behaviors, languages, and procedures often led to difficulties in cooperating, communicating, and sharing information and ideas between partners (Ozorhon et al., 2007). Besides, it was estimated that in case the commercial, financial and technical information is managed effectively, the cost of joint venture projects could be reduced up to 25% and this can be achieved through new data transfer technology (Baldwin et al., 1999).

Adnan and Morledge (2004) found in their research that communication and information sharing is the 8^{TH} ranking factor for the joint venture project's success. The partners have to reach constructive structure and systems built on sharing information and resources to ensure long-term

and well-established activities (Adnan and Morledge, 2004). Without appropriate communication allowing for good mutual understanding, problems can take place because of differences between partners' cultures (Adnan and Morledge, 2004). The different organizational structure and management practices of the JV partners, which are reflected in their different organizational processes, may damage the JV relationship (Ozorhon et al., 2007). It can cause weak and unstable communication, cooperation, and behaviors; as a result of that, the problem of interaction and knowledge exchange can expand to a level in which they create kind of uncertainties that led to possible risks and failures (Ozorhon et al., 2007).

The construction project nature forces firms to share information quickly internally and with business partners, and this is achievable by applying new technologies like electronic Data interchange (Baldwin et al., 1999). The most attractive features of these technologies are the structured nature of information sharing and the ability to integrate the transferred data into the receiver's internal database system and present it in a readable format (Baldwin et al., 1999). The shared database could be an effective solution for different cultural communication styles of joint venture projects. It can ensure the free flow of information in which all project participants follow the most updated instructions and be informed permanently about the latest status of the project (Köster, 2009).

3.4 The paradox of logic and intuition

De Wit (2017) states that managers face opposites of logic and intuition when it comes to making decisions. The intuition consists of tacit knowledge and experience gained through years of work (De Wit, 2017). On the other hand, logical thinking is based on actual facts and using information from other stakeholders during the process (De Wit, 2017). De Wit (2017) argues that in case strategic managers build their decisions based only on their intuition, that could lead to a risk of biased and unreliable decisions. Leaders should use both thinking styles together, even if they are opposites. This can be achieved by mixing their intuitive thinking with facts-based rational analysis, especially for a strategic decision where much more time and energy should be spent to avoid intuitive bias (De Wit, 2017).

Strategists and managers find themselves in an uncomfortable situation of needing to adopt paradoxical ways of thinking together. It is not easy to answer the question of which approach should be dominant, logical thinking, or more intuitive process (De Wit, 2017). To manage this paradox, De Wit (2017) shows different options, one of them is to apply parallel processing in which management define and differentiate the processes and activities of the organization in separate classes based on which level of logic and intuition they need. For example, production and operational processes mainly benefit from analytic and facts- based thinking, while business development decisions tend toward more intuitive thinking (De Wit, 2017).

4. Methodology

This section shall explain the method used in this study regarding the research strategy, adopted design, data collection, data analysis, delimitations, and ethical considerations.

4.1 A qualitative study

As this study aims to investigate whether Big Data technology can solve the information sharing related issue and enhance the decision - making process in construction projects, a wide range of variables such as personal cognition and experiences could manipulate the results. This forced authors to adopt a qualitative approach based on interviews (Bell et al., 2018). Qualitative studies are criticized that it is too subjective in which qualitative findings depend on the researcher's unsystematic opinion about what is essential. However, the nature of this study is more interpretive and concern with the words rather than numbers, moreover, Qualitative study gives the ability to the researcher to change his/her direction easily, all that makes it the most suitable choice to proceed with (Bell et al., 2018). A qualitative language-based method was applied in which participant observation is a part of a principal amount of qualitative interviewing (Bell et al., 2018). The first part of the qualitative approach starts with an observation to understand the current situation of information sharing inside the project and evaluate the factors that are most famous for the decision-making process. This observation was made with short interviews and a survey with selected decision-makers and managers. The second step was to build an information-sharing model by using one of Big Data application to prepare, connect, analyze, and finally visualize information to the same decision-makers. The final step was to complete the empirical data by conducting interviews and compare, analyze, and evaluate how the model affected the information sharing and decision-making process.

In order to link the empirical research and theory, an abductive approach took place through waving back and forth between empirical results and theory (Bell et al., 2018). This abductive strategy with qualitative research is not in one direction, which means it is not about testing theory by using collected data as in deductive strategy and is not to generate theory from collected data as an inductive strategy. However, it uses theory as a background to qualitative investigations (Bell et al., 2018). A mixed abductive approach is more suitable for the nature of this report. The aim here is not to generate theory nor test the validity of it, but it is to answer a specific question of how Big Data can enhance information sharing and decision-making process. This requested a mixed approach to use a concrete theoretical framework as a base to analyze the empirical results and come up with a rich and in-depth discussion, which led in turn to a cohesive conclusion.

4.1.1 The literature review

The theory research in this study was made using online resources and libraries like Google scholar and Chalmers Library. As the first step of a systematic review to narrow the search results, keywords of this report like "Big Data in construction," "Decision-making process," and "Joint Venture" were used during the search (MacLure, 2005). Sometimes the search was carried out by first, dividing the research question to parts, then putting it on the online browser, and then checking the titles of the results to choose which one worth going with. Moreover, as a second step, the scanning process of the abstract and conclusion of each chosen article and book has been made to evaluate and assess if the resource is valuable and can support the discussion later on

(MacLure, 2005). The last step in the adopted systematic literature review is to read through the selected sources, scan and filter the source text, and finally extract relevant data (MacLure, 2005). Even though search strategies like checking the references list of an article to find something interesting or what is called snowballing strategy might seem as not systematic, it could lead to useful and relevant literature(Greenhalgh and Peacock, 2005). The snowballing strategy was used during the theory research in this study as a try to find relevant articles that may enrich the theoretical framework by adding new theories and arguments that serve the purpose of this study (Greenhalgh and Peacock, 2005).

It is worth mentioning here that Big Data is an interdisciplinary and compound topic that is applied in many other fields like healthcare, marketing Etc. For this reason and because Big Data technology in the construction industry is still a new topic and there are limited sources about it, literature from outside the construction field generally exploring the topic was screened to review and develop a theory (Webster and Watson, 2002).

4.1.2 Data collection method

After introducing the model to selected departments' managers and other decision-makers like the Chairman of the Board and Board members, they started using and interacting with the model and its dashboards. These dashboards were used as a tool in the Board meetings and work teams' meetings to explain the current situation and plan activities. In addition, dashboards were used individually to extract, compare, and check needed information related to a specific topic if. In order to measure the model performance, evaluate its results, and obtain a good overview of whether the model helped those decision-makers to enhance their decisions and improved the information sharing among each other, semi-structured interviews with the same selected managers and decision-makers were conducted as shown in Table 1 below. The interviews aimed to gather data in order to understand the effect of the model on four different directions. First, if the model succeeded in providing enough information about the scope, cost, and budget so the quality of decisions will be improved. Second, to know if the model helped managers to understand the current situation of the project better and be able to compare it with what was planned before and know where the project is going. Third, to know if the model provided the needed information and facts to plan for the upcoming operational activities. Finally, to understand the way the model affected the information sharing process between players and departments.

The interviews were open and based on pre-prepared questions to cover and understand specific aspects (Bell et al., 2018). All interviews took place in a face-to-face setting, in Gothenburg, Sweden, in English, and each lasted between 30 to 45 minutes. Questions, in turn, were divided into three phases, first, to introduce the interviewees and their positions and backgrounds. Second, to cover and evaluate the decision-making process using the model. Lastly, to learn their assessment of the model's performance regarding information sharing between different departments. A short survey was conducted during the interviews to get an overall evaluation of the model from a different point of views. The interviews were audio-recorded and then transcribed carefully to make sure that all the information presented by the interviewees were taken into consideration and used later when a comparison with the theoretical framework happens.

Table 1 - interviewees' positions, and consumed time

interviewee	Role	Duration	Date	How
A	Chairman of Board	45 minutes	12.05.2020	Face to face
B	Board member and commercial manager	45 minutes	11.05.2020	Face to face
C	Design manager	40 minutes	13.05.2020	Face to face
D	Finance manager	30 minutes	13.05.2020	Face to face

4.1.3 Data analysis

A qualitative study with an abductive strategy based on semi-structured interviewing usually generates a massive amount of data where respondents have the ability to talk about areas not included in the interview schedule and not covered in theory (Bell et al., 2018). This makes it more difficult to define themes within the collected data, and even when themes that not covered in the theoretical framework are defined, a dilemma appears in which it is problematic to know to what extent should the authors go back and forth between data and theory to read more and make sense of data (Bell et al., 2018). Since there is a deadline to submit this study and because of a lack of time and capacity, the focus was to find balance in adopting just the themes that are backed in the theory and relevant to the research question. Analysis of collected data in this study includes an abductive process in which, after the collecting of data, coding, filtering, and cleaning processes were performed to find relevant and critical ideas. The final step was to face the empirical results with the theoretical framework to discuss and assess the presented model and come up with a conclusion (Bell et al., 2018)

4.2 The software: Microsoft Power BI

The main software that is going to be used in the empirical part to handle and present Big Data is Microsoft Power BI. As stated by Microsoft, it is a software used to "create a data-driven culture with business intelligence." Power BI was launched in 2016 and is used to visualize data interactively with some business intelligence tools that are used to create dashboards and reports (Microsoft, 2020). Some main features of the software as following:

- It uses different sources as data sources. Data are not stored in Power BI, but in the original data source, which can be Excel file, Databases (MS Access, SQL…), Webpage…
- It can handle a considerable amount of data comparing to MS Excel. Currently, it can import datasets that are 1 GB per dataset.
- It allows end-users to create their own filters, show data from different sources, and comment on graphs and dashboards in a shared way.
- It is designed for a shared experience. Users do not need to install the software and they can access dashboards and reports on a web version or mobile phones.

Figure 5 - Process of Data Processing and Sharing, Study authors' own elaboration

4.3 Delimitation

MS Power BI is a fairly new software that was introduced in the market only a few years back. Thus, many interviewees were not familiar with it and needed time to learn how to use it and interact with the data and graphs. Moreover, since the timeframe of the study is limited, and the interviewees did not have more than one month, their evaluation of the model maybe not complete, and the potential benefit of the model may not be fully utilized.

Another delimitation is that it is hard to assess how much the decision-maker relies on the model when he/she makes a decision. In other words, to what extent the decision-maker will use his own experience and the information from the model in the decision-making process.

Moreover, since the authors of this study are working in the studied project, this may make the interviewees respond to the interviews' questions in a biased way that gives the model a slightly higher estimation of benefit.

Lastly, the authors themselves are not entirely familiar with the used software (MS Power BI) in a professional way. This means that some functions and features of the program may not have been used and could have maybe enhanced the performance of the model and affect the final results.

4.4 Ethical considerations

One of the challenges during this study is how to manage the collected data ethically. In order to overcome this challenge, many actions have been adopted. First, all the individuals and organizations which are involved in this study were presented anonymously in a full and strictly confidential way. Secondly, the collected data from interviews, as well as the used data by the model, were hidden and treated carefully in which it is kept confidential and protected internally and externally.

Besides, many procedures have been followed to take the privacy and sensitivity of the parties into consideration. These procedures include performing separate interviews in a closed and suitable place, asking for permission to record the interviews, providing explanations at the beginning of

each meeting on how to deal and handle the personal data. Furthermore, the authors believe that presenting the data from interviews and references honestly is an important issue. For that reason, repetition and revision processes occurred to achieve a high level of data quality.

5. Empirical part

5.1 Introduction

5.1.1 The Project

The project is an infrastructure project in Sweden. Its goal is to build a new metro system under the city of Gothenburg. It started in 2017 and planned to be finished by 2026. The joint venture that we are studying is responsible for two parts of the project. The project is considered a complex project, not only due to the advanced construction technologies that are used in it but also due to the contracts' types of the project. The following figure shows the contracts types:

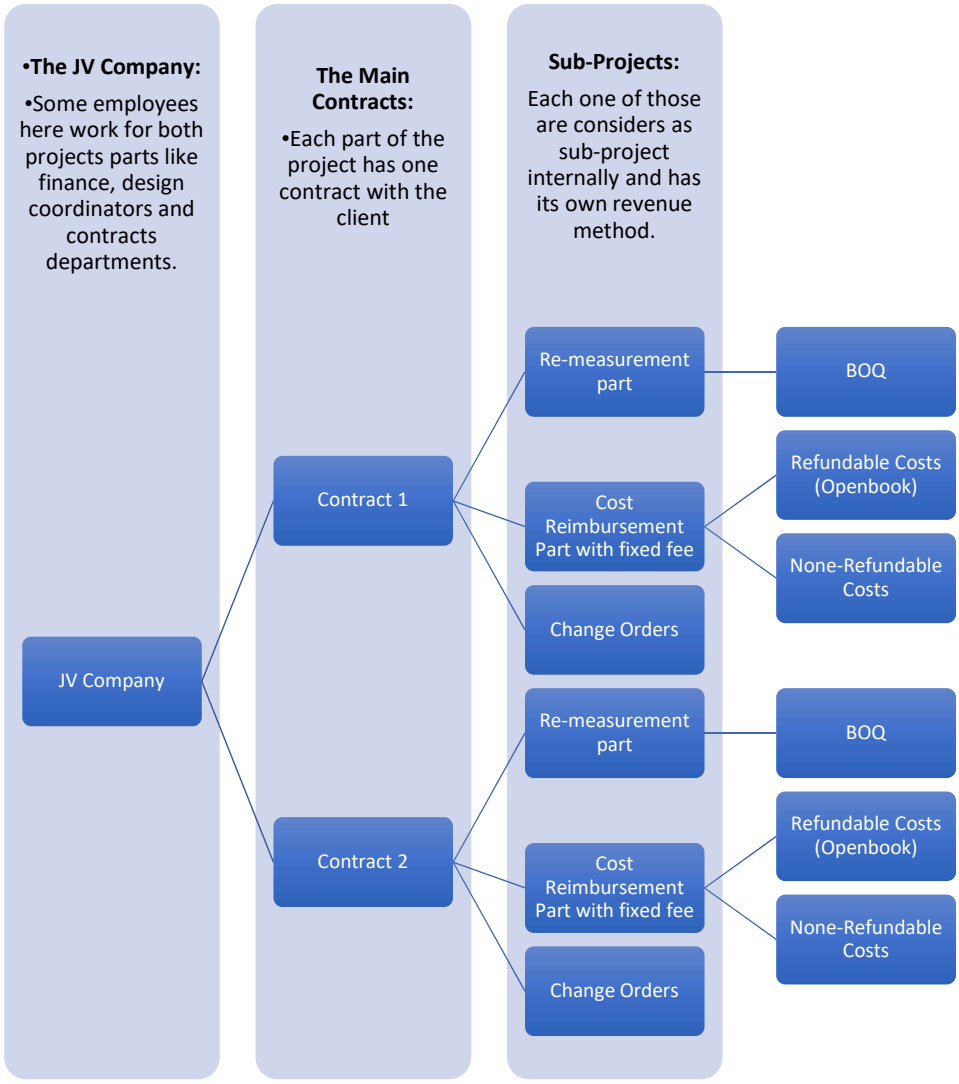

Figure 6 - Contracts Types in the project, Study authors' own elaboration

The complexity of the project comes partially from the different contracts' types that are under each contract. As we see in the figure, under main contract number 1, we have three contract types (which are internally named as sub-projects):

- Re-measurement part: The scope of work and the design here is known. It has a clear bill of quantities BOQ and the JV is paid for the quantities of accomplished works.
- Cost reimbursement part with a fixed fee: This is also known as the Cost-Plus part. The JV is reimbursed for all invoices plus a fixed fee, which represents a percentage of the invoices. However, not all costs are reimbursed. The refundable part is called Openbook part. Furthermore, some costs are not refundable like insurance, taxes, travel and some other costs. For the Openbook part, there is a cap cost that the client and the JV agreed to reach. The two parties share the costs above the cap and share the gains below it.
- Change Orders: If a change order was agreed between the client and the JV, all invoices related to the change order are reimbursed with a fixed extra percentage.

Most of the work in the project is subcontracted. Thus, most costs come from invoices from those subcontractors and suppliers. The most common issue that was noticed here is the mix between invoices, which means an invoice that belongs to Openbook can be registered under the re-measurement part by mistake. That is because the line that defines each scope of the three sub-project is fine, and also because costs are inputted from different persons that may leave the project and be replaced.

Finding anomalies in the cost allocation is done manually and a few days are required each month to check invoices and allocate them to the correct project.

5.1.3 Big Data in the empirical part

As mentioned in theory, Big Data has two aspects: Big Data engineering and Big Data analytics. In this study, only Big Data analysis will be considered for working on. Here we have four steps to work on as shown in this figure:

Figure 5 - The four steps of using Big Data in the empirical part

Data Preparation

Generally, data exist in several structures and organized differently. For this study, the authors will define three types of data based on how it is organized as follows:

1- Structured data: datasets in tabular forms containing unique identifiers that can be linked with other tables. For example:

Table 2 - Structured Data Example

ID	Date	Activity	Quantity
FA-390001	22-Mar-20	Excavation	20
FA-390002	23-Mar-20	Backfilling	30

2- Semi-structured data: tabular forms without unique identifiers that cannot be linked with other tables. This can also include tables that have unique identifiers, which, however, are not used those identifiers. For example:

Table 3 - Semi-structured Data Example

Date	Activity	Quantity
22-Mar-20	Excavation	20
23-Mar-20	Backfilling	30

3- Unstructured data: different forms that are not organized in tabular forms. For example:

Table 4 - Unstructured Data Example

Daily Report				Date:	23-Mar-20
Machinery			**Manpower**		
#	Description	Hours	#	Description	Hours
1	Loader 1	8	1	John	8

Data preparation means we need to convert data structure from unstructured to semi-structured or structured. Data preparation aims to have logs or tables of data, where each line contains all information about that line. If we took the last example, we need to make it in the following way:

Table 5 - Example of a proper data structure

Date	Type	Description	Hours
23-Mar-20	Machinery	Loader 1	8
23-Mar-20	Manpower	John	8

Data preparation takes much time, especially in an industry like construction where construction data like daily reports are usually unstructured.

Data Linking

After preparing data and having structured or semi-structured tables, we need to link those tables together. Linking tables require a column that contains an identifier of each row of data. The identifier column from table A will be linked to the identifier column from table B. This identifier

can be a date, cost center, name, number, etc. Each one of the two tables can be linked directly using MS Power BI. The best practice is that the identifier column in one of the tables to be unique. However, if the two tables contain non-unique identifiers, then we need to create an intermediary table to link them together.

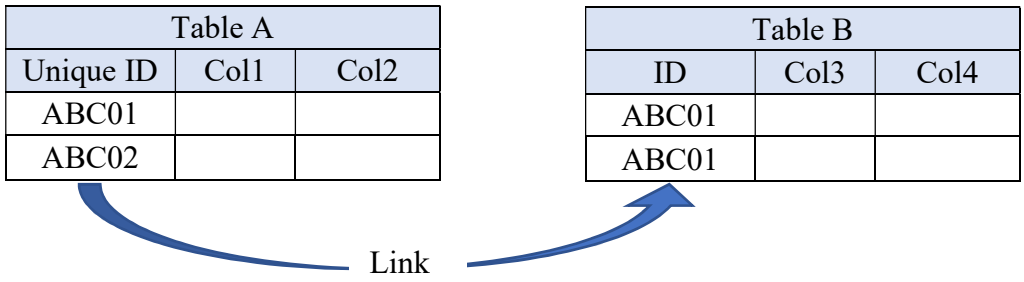

Figure 6 - Linking Tables A&B with one at least has Unique Identifier

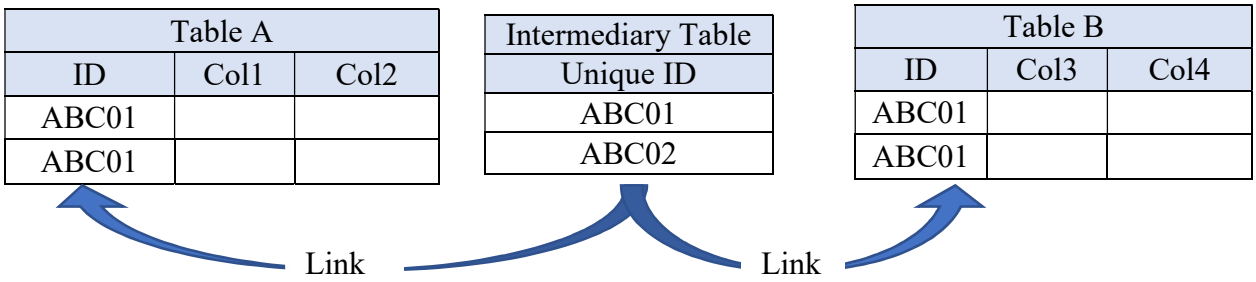

Figure 7 - Linking Tables A&B when the identifiers are not unique

Data Analysis

In this part, we will analyze the data after preparation and linking and convert it from raw data to useful information. For example, the cost table from the ERP system contains raw data about cost, and the budget table from cost control contains raw data about the budget. After linking both tables together, we can analyze the cost comparing to the budget, check the remaining budget, and, if needed, adjust the budget based on the actual cost data.

Earned value tools and techniques are used in data analysis to assess the performance of the project regarding cost and schedule. We can also check if there are abnormalities in the data.

Data Visualization

We mentioned in the theoretical part, the importance of data visualization in making a decision. Here we will show the results of data analysis through dashboards and reports that target different stakeholders. Data visualization usually depends on static graphs and charts like pdf or excel graphs, but with MS Power BI, we are able to create dynamic dashboards. It makes it possible for stakeholders to filter data depending on their needs.

5.2 Information Sharing Methods in the project

There are different methods for sharing information currently in the project between the different departments. As shown in the following figure:

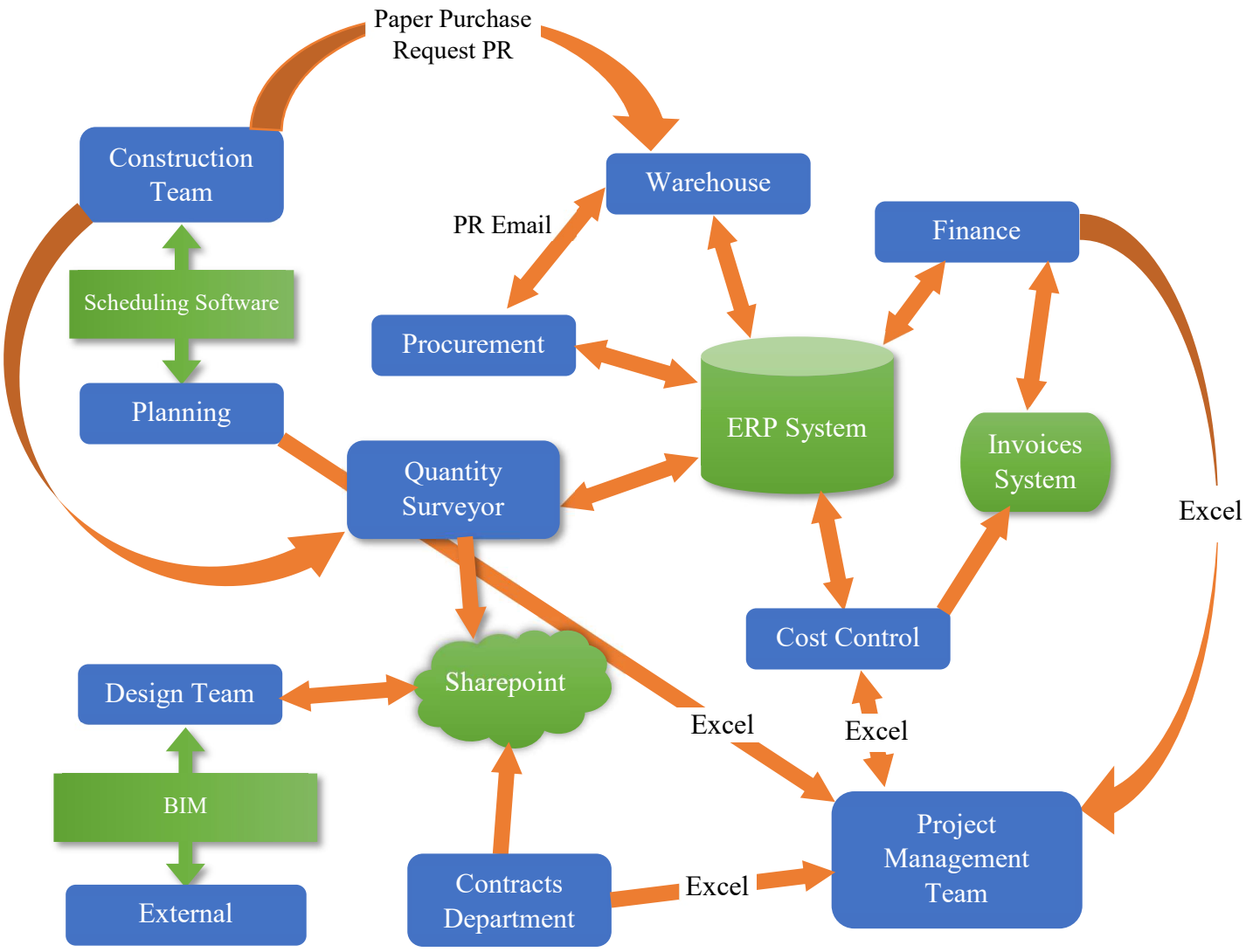

Figure 8 - Information Sharing Methods in the project, Study authors' own elaboration

Figure interpretation:

- Color Codes:
  - Green: Information management systems.
  - Blue: Departments and stakeholders.
  - Orange: Information flow
- Shapes and thicknesses have no specific meaning

- Sharepoint is used by all stakeholders; the links that are related to it means that it is used as a primary method of sharing regular information.
- Some links were not shown even though they exist, but they do not represent regular information flow like the link between construction and contracts, which is an on-demand information method.

Information management systems are not linked together. All links between systems are done externally using unique identifiers that are manually inputted.

5.2.1 Enterprise Resource Planning ERP system: BAAN

BAAN is the main ERP system used in the project. This system is used as the main ERP system also for the leader company of the Joint Venture.

BAAN was found in the early nineties. After 2003, BAAN Developer Company was acquired by SSA Global technologies, which changed the name of the system to SSA ERP. However, the leader company is still using the version of 2003 and the last major update to the system was in 2007 (Reference of Business, 2020), which still named BAAN. BAAN is used to manage financial accounting mainly. However, in the studied project, it is used by the accounting and finance departments, cost control, warehouse, quantity surveyors, and procurement department.

BAAN can store a massive amount of data, and it is known as a secure system. However, the system has some pitfalls, for example:

- It was described by many users as a rigid system.
- It has a poor user interface.
- The user experience is also not good compared to other ERP systems like SAP.
- There is no communication method inside BAAN. That means users cannot comment on inputs of other users or have some kind of check and control inside the system.
- The changes are not reflected immediately in the system and not seen directly by other departments.
- The way that BAAN displays information is very basic. That is why every user needs to export BAAN reports to excel or PDF to be able to see the information.
- There is no unique identifier for transactions in the system, and that is, in our opinion, the biggest disadvantages in the system. There is an identifier for financial transactions, but it has noticed that this identifier is not unique and is being used for multiple transactions.
- Reports are not generated from the system directly, but through excel and macro scripts.
- Most of the changes to the system need to be done by technical support from the leader company's main office.

The issues with the system are known to its users. However, the leader company may not have any problems with it in other projects as its employees know the system very well and can solve their issues directly with fast technical support. The problem in this project is that different companies use the system, which allows for more mistakes to happen. Moreover, each request from other companies needs to go through the proxy before it reaches technical support.

Data can be retrieved from BAAN by PDF or Excel formats.

5.2.2 Invoices management system: Billbox

Most of the works in the project were given to subcontractors. In two years, there were more than 600 subcontractors or as they called in the project: Business Partners. The invoices from those business partners are about 500 monthly. Each invoice has limited time to be approved, processed in BAAN, and paid. Managing invoices include registering them, send them to the stakeholder who will approve the content, get management approval, and track payment and accounting registration. BAAN was not used for managing invoices as it is not able to store documents. At the beginning of the project, Sharepoint was used to manage invoices. But then, after one and a half years, it was replaced with the invoices management system Billbox. Billbox can store, manage approvals, and export to excel with a strong filters feature. It needs some enhancements, but the developer company is local, and they are easy to contact regarding any feedback.

The finance department manages Billbox. It is used by cost control, warehouse, and all stakeholders that have the authority to approve invoices like site managers and departments' managers. Billbox is also used to share information between cost control and finance regarding cost centers of each invoice. Besides, it is used to share information between the entity that is responsible for the invoice and both finance and cost control regarding invoice location and related work.

5.2.3 Files Sharing System: Sharepoint

Some departments use Sharepoint as the primary method of sharing information. Sharepoint is a cloud storage system by Microsoft that is developed for collaboration between teams. In the project, it is used by the document controller to keep a log of all correspondences. It is also used by the quantity surveyor to share payment certificates with other departments. The contract department keeps a copy of each business partner contract on Sharepoint with a log of all contracts to be used by a quantity surveyor, finance, and cost control. Sharepoint is managed by the IT department that fully controls the accessibility.

5.2.4 Scheduling tool: Primavera

Primavera is a scheduling software owned by Oracle Corporation. In the project, it is used only by the planner to illustrate the construction and design schedules. Scheduling data is agreed with the construction team and then shared with concerned stakeholders in the type of bar charts and construction sequence drawings.

5.2.5 Building Information Modelling BIM tool: Revit

Like other works in the project, the design is subcontracted to several designers. Designers' models are aggregated by coordinators using Revit software. Information from design is not shared using Revit, but the coordinators generate reports like quantities take-off and share them with other departments.

5.3 Current Data Structures

Each department in the project generate data or collect data to share it with other departments. The following table shows the data in each department and how they are organized:

Table 6 - Data Tables in the project, Study authors' own elaboration

Department	Data table	Source	Data Structure	Identifier	Notes
Finance	1. Transactions	ERP System	Structured	Finance ID	
	2. Openbook Invoices	ERP System	Structured	Finance ID	This table represents the list of invoices that are sent to the client for the cost-plus refundable part.
	3. Revenue	Manual	Semi-structured	Project Code	Summary of revenue by sub-project
	4. Invoices	Invoices System	Semi-Structured	Invoice Number	List of invoices and their amounts.
	5. Assets	Manual	Structured	Finance ID	Assets that are subjected to depreciation.
Cost Control	6. Actual Cost Allocations	ERP System	Structured	Finance ID or Cost Center	Table with all cost details like cost centers, finance IDs, invoices numbers, order numbers, depreciation cost. But without warehouse stock and assets as they are not considered cost yet.
	7. Budget	Manual	Semi-Structured	Cost Center	
	8. Monthly Project Status	ERP System	Semi-structured	Project Code	Summary of cost versus revenue by sub-project
Warehouse	9. Warehouse Balance	ERP System	Semi-structured	Order Number	Materials enter and exist, and remaining balance with their values and orders numbers.
Quantity Surveyor	10. Payments Certificates Details	ERP System	Semi-structured	Certificate Number	
Procurement	11. Orders Details	ERP System	Structured	Order Number + Order Line	
Contracts	12. Contracts List	Manual	Structured	Contract ID	Details about contracts like amounts, business partners, duration, and status. But contract ID is not used by other departments' tables.
Planning	13. Schedule data	Scheduling Software	Semi-structured	Activity ID	Schedule activity ID is not used by other departments.
Construction	14. Construction Reports	Manual	Unstructured	-	Details about Labors and Equipment usage.
Design	15. Quantities	BIM Software	Semi-structured	Area Number	Quantities take-off by objects.

These above 14 data tables store all data types in the project. We will call them data sources. Currently, there are many issues when getting information from more than one data source.

5.4 Current Major Issues with Information Sharing

5.4.1 Issue #1: Construction Costs Details in the ERP System

The ERP system in the project represents a mine of data. However, these data are only visible for system users that are not many. Only a few figures from the system are shared with departments' managers and the project management team currently. Moreover, the construction team does not have much information about the details of costs for their areas, the budget, and the cost performance index CPI. This is a general issue; many stakeholders will benefit from easily reading ERP system data since we mentioned before the user interface and the user experience issues with the current ERP system. ERP system data can be organized, analyzed, and linked with other data sources to produce integrated reports or dashboards.

5.4.2 Issue #2: Cost Performance Index CPI

CPI comes from the famous Earned Value Management techniques. The aim of calculating it is to know how efficient we are comparing to our budget (PMBOK, 2017). The departments involved in this issue are planning and cost control.

At a given point in time, we have progressed in an activity of x%. The actual costs were y% of our budget. The CPI then equals x%/y%, which will be a decimal number of more than 0. Ideally, it should be 1. If it is more than 1 then we are getting more value for what we spent in that activity. If it is less than 1 then we are spending more and getting less value.

Currently, it is hard to calculate the CPI because there is no link between cost data and planning data. Even though both departments started from the same Work Breakdown Structure (WBS), but they had more development in their structure and combined some packages together for their internal benefits. That led after two years to the mismatch between work packages. Thus, it is not possible to calculate CPI for a level of more than level 2 in the WBS. Furthermore, even though it is possible, it has never been done.

5.4.3 Issue #3: Payments Certificates Details

This issue is between quantity surveyor (who is responsible for payment certificates to business partners) and design departments. It is a specific example of issue 5.4.1.

Design departments provide the quantity surveyor with the details of payment certificates. The quantity surveyor inputs them in the ERP system and issues the certificate in order to allow the business partner to invoice the JV for the required service. The design department keeps a record of the provided services and work hours but does not have many details about payments and the remaining amount in the contracts. Thus, they need details about payment certificates lines and their costs and business partners' contract values and remaining amounts, which is related to issue 5.4.3. The current cost report does not show payment certificates lines.

5.4.4 Issue #4: Matching Invoices in the ERP system with submitted invoices to the client.

The departments affected by this issue are Cost Control and Finance. In order to understand this issue, we will present the process of costs flow:

1. A stakeholder asks for material, for example.
2. If the material is not in the warehouse, it is requested to be procured by the procurement department.
3. After the material comes, it is followed by an invoice which linked on the system to the same order.
4. At the end of the month, the finance department export invoices the list from the ERP system and filter all invoices that are under the Openbook part of the project only. Then they are submitted to the client to be reimbursed.

Theoretically, costs that are under the Openbook part in the ERP system should match the submitted invoices to the client. However, we found that they are not matching. The reason behind the mismatch is that the cost allocation may change for costs in the past without reflecting the changes in the submitted invoices list. For example, if we have an invoice that is considered for the re-measured part of the project - which is not paid by invoices, but by quantities – then it will not be submitted to the client as an invoice. Moreover, if cost control found that this invoice should be in the Openbook part, then it will be changed in the ERP system, but most of the time, not submitted to the client for payment. In the beginning, it was possible to see the issue due to the total sum of what was submitted and what is in the ERP system, but it is not possible to find the specific invoices that cause the mismatch.

5.5 Solving Current Issues

5.5.1 Solving Issue #1: Construction Costs Details in the ERP System

The project consists of three major packages: Design, Construction, and Indirect costs. The construction package is the biggest one. Targeted stakeholders are the construction manager and three site managers. Moreover, the information that should be provided from the ERP system is the information about cost and materials quantities. Reports about budget, remaining budget, and cost performance index CPI also are provided.

Firstly, an investigation was conducted with the stakeholders to assess their current information and their needs. As a result, we found that the current information is:

Table 7 - Available information for construction team major stakeholders

Stakeholders	Budget details	Actual Cost Details	Materials	CPI
Stakeholder 1	Some	Some	Some	None
Stakeholder 2	Some	None	None	None
Stakeholder 3	All	None	Some	None
Stakeholder 4	None	None	Some	None

The aim here is to facilitate information sharing for those stakeholders. To do that, a web-based dashboard that contains all required information was built to achieve this goal.

Data Preparation

For this task, we need the following data tables; Budget, Actual Cost Allocations, and Schedule Data. Data preparation means making each row from those tables contain all needed information of the row. All data preparation steps are done on MS Power BI.

The first table, the budget table, is a semi-structured table that looks like the following:

Table 8 - Budget Table Template – Semi-structured

Activity	Cost Center	Quantity	Unit	Unit Cost	Total Cost
Sub-project					
Package (e.g. construction)					
Zone					
Work Package (Control Level)	**ABC01**				
Activity 1		100	m2	50	5,000
Activity 2		200	m3	100	20,000

In the budget table, we have the control level, which is the level used by finance and cost control for cost allocation in the ERP system. The identifier is the cost center, but it is not as unique as more than one activities hold the same cost center. The reason that we considered it semi-structured is that the activity line does not contain all required details (like work package, zone, package…) in the same line but different lines.

Preparing a budget table means converting it to a structured style. That means the following table style:

Table 9 - Budget Table Template – Structured

Sub-project	Package	Zone	Work Package	Activity	Cost Center	Quantity	Unit	Unit Cost	Total Cost
E1	Const.	Z1	WP1	Activity 1	ABC01	100	m2	50	5,000
E1	Const.	Z1	WP1	Activity 2	ABC01	200	m3	100	20,000

The actual cost allocations table is a structured table that contains all the required details about transactions. It is exported from the ERP system in excel format and it looks like the following:

Table 10 - Actual Costs Allocations Template - Structured

Sub-project	Cost Center	Type	Quantity	Unit	Unit Cost	Total Cost	Date	Business Partner	Order	Finance ID	Invoice Number
E1	ABC01	Material	100	M2	50	5,000	01/04/20	Partner 1	E10010	1FA00001	12
E2	ABC01	Rental	20	day	120	24,00	02/04/20	Partner 2	E20010	2FA00001	23

This table contains all types of costs like material, salaries, subcontractors, depreciation, and indirect costs. It does not need preparation as it is already structured.

For scheduling data, we mentioned before at issue 5.4.4 that the current schedule and budget are matching up to level 2 only, which is the Zone level. In order to have proper details, we need to match until level 3, which is the work package (control level), which is resolved later. The result will be a table like the following:

Table 11 - Schedule data template showing Cumulative Work Progress – Semi-structured

Sub-project	Cost Center	31-Jan-20	29-Feb-20	31-Mar-20	30-Apr-20	31-May-20
E1	ABC01	5%	6%	6.5%	8%	10%
E1	ABC02	0%	0%	0%	1%	2%
E2	ABC01	3%	3%	3%	3%	4%

The schedule data table is semi-structured, the columns are pivoted, and each row individually does not have all the needed information (data date). To prepare it, we will convert it to the following style:

Table 12 - Schedule data template showing Cumulative Work Progress – Structured

Sub-project	Cost Center	Date	Progress %
E1	ABC01	31-Jan-20	5%
E1	ABC01	29-Feb-20	6%
E1	ABC01	31-Mar-20	6.5
E1	ABC01	30-Apr-20	8%
E1	ABC01	31-May-20	10%
E1	ABC02	31-Jan-20	0%
E1	ABC02	29-Feb-20	0%
E1	ABC02	31-Mar-20	0%
E1	ABC02	30-Apr-20	1%
E1	ABC02	31-May-20	2%
E2	ABC01	31-Jan-20	3%
E2	ABC01	29-Feb-20	3%
E2	ABC01	31-Mar-20	3%
E2	ABC01	30-Apr-20	3%
E2	ABC01	31-May-20	4%

Data Linking

After data preparation and making tables in a structured style, we need to link tables (Table 9, Table 10, and Table 12) together. Linking tables in MS Power BI can be done by choosing a key column from each table. After looking at the three tables, we can notice that the Cost Center column is mutual between the three tables, but it does not contain unique values. As per the recommendation from MS Power BI we need at least one column from any table which contains unique values in order to get more accurate data. This issue can be solved by creating an intermediary table containing a unique column of all Cost Centers and linking the three tables to

the new one. We will call this intermediary table, "Cost Centers." In the Relationships Layout in MS Power BI we can see a view of each table and its columns like the following:

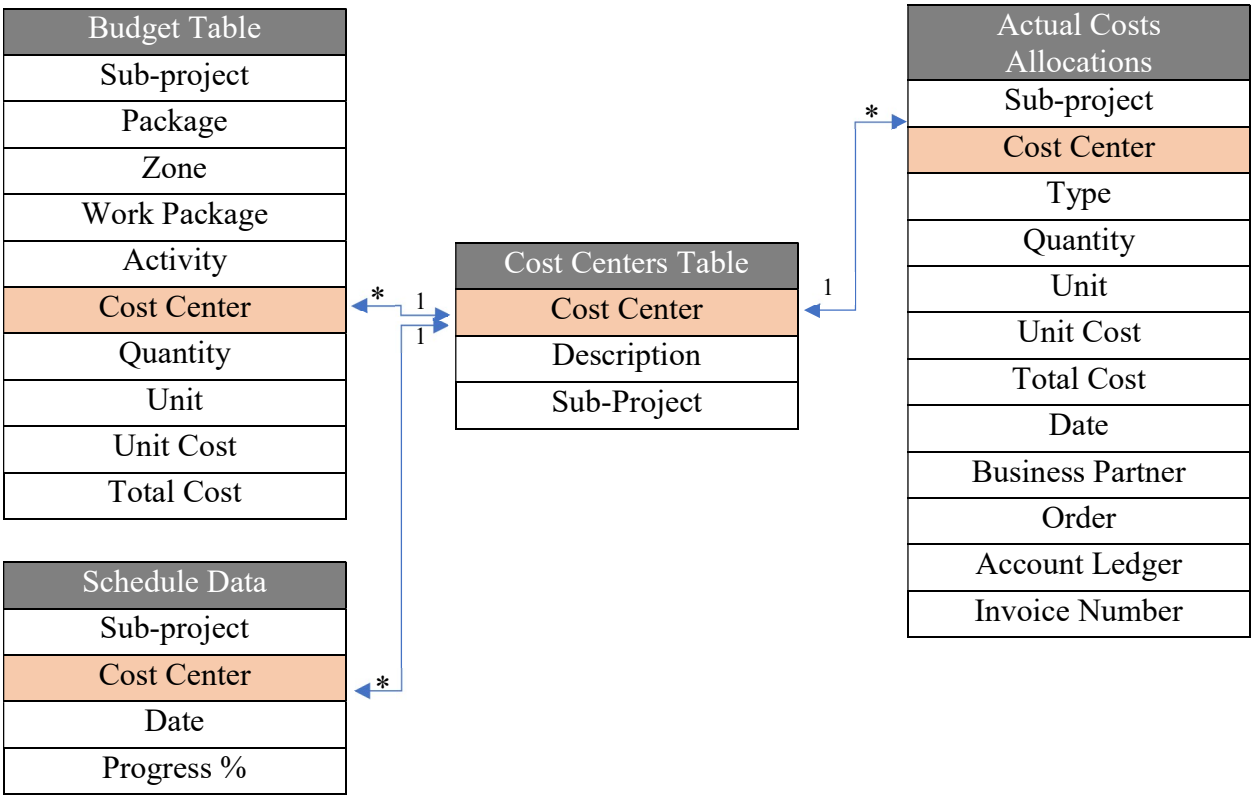

Figure 9 - Relationships Layout for Issue 1

The links between tables are marked with 1: * to illustrate relationship type, which is one to many. The side marked with 1 (cost centers table) means the values in the column are unique. The colors of columns titles are just to illustrate which columns are used.

Data Analysis

After data linking, we can now analyze data using values from different tables together. Basically, we have three tables: Cost, Budget, and Schedule.

Firstly, from Cost and Budget tables, we can compare the cost to the budget to illustrate how much was consumed from the budget and to know if there is a need to increase the budget. Thus, we need to create a "measure" as called in MS Power BI. This measure is: Actual % = Actual Cost / Budget Cost. That is applied to each cost center.

Comparing quantities between actual and budgeted is made by creating another measure: Quantity % = Actual Quantity / Budget Quantity.

The remaining cost measure is created to know how much more money there still is for each cost center. Remaining Cost = Budget Cost – Actual Cost.

Secondly, from Cost, budget, and Schedule, the cost performance index CPI can be calculated. In its simple equation, CPI = Progress % / Actual %. In this measure, we used a previous measure we created, which is Actual %.

The cost performance index is a technique from Earned Value Management, and it is an indicator of performance in terms of cost. It does not compare the cost of the planned cost. It compares it with earned value. If CPI is > 1 then the package is under budget, and there is more earning value than cost. If CPI is < 1 then the package is over budget, and I earned less than what I spent on.

Data Visualization

For this step, we will create a dashboard that contains tables, graphs, and charts to illustrate data. The dashboard can be accessed on the MS Power BI webpage only by specific stakeholders. There is no need to install the MS Power BI application to view it. The web-based dashboard has a dynamic view where the stakeholder can click on different filters to change the viewed information. The dashboard data will be stored on MS Power BI Service Cloud, where it will be accessible by stakeholders anytime using a web browser or MS Power BI mobile app. It is also possible to use the comment section on the dashboard to comment on it or to explain the information presented in some figures. It is always possible to print or export the dashboard as PDF or PowerPoint files.

Since we have different site managers, we need to restrict data to each one of them. Thus, we can use a function in MS Power BI called Row Level Security RLS. In RLS we can create profiles with limited access by cost centers, and we will apply RLS on the intermediary table (Cost Centers Table) that link the other three tables together. The result of the dashboard is the following:

Figure 10 - Data Visualization - Construction Dashboard

The upper part makes it possible to filter data by time, business partner (subcontractors, consultants, suppliers), or invoice number. While the rest of the dashboard contains different tiles, each tile serves a purpose as the following:

Table 13 - Construction Dashboard Tiles

Tile	Source Table	Purpose
1	Cost Centers	Filters the rest of data by cost center
2	Costs Allocations, Budget	Compares total cost to the total budget
3	Costs Allocations, Budget	Compares costs to budget by cost center
4	Costs Allocations	Shows cost by a business partner
5	Costs Allocations	Shows maximum details from the ERP system
6	Costs Allocations, Budget	Compares actual quantities to budgeted quantities
7	Costs Allocations, Budget, Progress	Shows costs performance by cost center
8	Costs Allocations	Shows cost by cost budget by descending order.
9	Costs Allocations, Budget	Shows how much percentage consumed from the budget by cost center
10	Budget	Maximum budget details for each cost center.

In the mid of April 2020, the construction dashboard was introduced to the relevant stakeholders. Furthermore, a small workshop was conducted to walk through the dashboard. Each user was able to view the costs centers that she/he is responsible for.

The aim from the dashboard as explained in the workshop, is the following:

- Give the ability to the construction team to see detailed costs of construction works that are registered in the ERP system and to see budget details for their part.
- Let them analyze the costs to calculate actual unit costs.
- Use cost analysis for enhancing budget calculations for remaining works.
- Show the quantities of materials that are registered in the ERP system
- Detect cost allocations mistakes.
- Compare costs to budget
- Monitor cost performance index CPI

5.5.2 Solving Issue #2: Cost Performance Index CPI

The original plan in this issue was to add cost centers to the schedule to get accurate data for progress by each cost center until level 6 of WBS. Unfortunately, we are not able to fulfill this plan during the study period due to out-of-control circumstances.

However, we still can use WBS Level 2 to connect the tables. Based on that, we will proceed with solving this issue and show high-level reports.

The tables needed in this task are costs allocations, budget, and schedule data. The aim here is to build a high-level report for executive management to show the overall status of the project.

Data Preparation

The three needed tables were already prepared in previous tasks. They are Table 9, Table 10, and Table 12 from 5.5.1 Solving Issue #1: Construction Costs Details in the ERP System.

Data Linking

The tables already linked in solving issue #1, as we can see in Figure 9 - Relationships Layout for Issue 1.

Data Analysis

The main indicator we are calculating here is the cost performance index CPI. In the earned value management technique:

$$CPI = \frac{Earned\ Value}{Actual\ Cost}$$

But Earned value = Actual Work Progress % * Budget. Thus,

$$CPI = \frac{Actual\ Work\ Progress\ \% * Budget}{Actual\ Cost}$$

The value of (Actual Cost / Budget) represents the percentage that is consumed from the budget. We will call it the Actual Cost %. Then:

$$CPI = \frac{Actual\ Work\ Progress\ \%}{Actual\ Cost\ \%}$$

If CPI >1, then the status is under budget. If CPI <1, it is over budget.

Data Visualization

In this task, we will create a report to view CPI. The targeted stakeholders are executive management. Thus, the report will not have too many details.

The report looks like the following:

Figure 11 - Earned Value Report

Each tile in the report represent the following:

Table 14 - Earned Value Report Tiles

Tile	Source Table	Purpose
1	Cost Centers, Budget	Filters the rest of the data by WBS Level 2; the size of each part correlates with its share of the budget.
2	Costs Allocations	Shows Actual Cost
3	Schedule Data, Budget	Shows Earned Value
4	Costs Allocations, Budget	Shows consumed percentage of the budget
5	Schedule Data	Shows Work Progress %

6	Costs Allocations, Budget, Schedule Data	Shows CPI
7	Costs Allocations, Budget, Schedule Data	Shows CPI, Actual Cost % and Work Progress % by WBS Level 2
8	Costs Allocations, Budget, Schedule Data	Shows line graphs of Work Progress % and Actual Cost % in the last six months.

Clicking on any part of tile 1 will filter the whole report for this part.

A preliminary dashboard of CPI was introduced to the project management team in early April. The project manager decided to include some more information. Later before the end of April, the final version was presented in the joint venture board meeting. After the board meeting, an executive manager asked to add one more chart to the dashboard to illustrate how CPI changes in the past few months. Thus, the CPI trend chart was added to the dashboard that was presented in the next board meeting in mid of May 2020. The dashboard aimed to illustrate cost performance in the project and the biggest work packages.

5.5.3 Solving Issue #3: Payments Certificates Details

Data tables needed for this issue is Payments Certificates Details and Contracts List. Also, we need a budget table.

The process of payments certificates currently looks like the following:

Figure 12 - Payments Certificates Process, Study authors' own elaboration

The payment certificate is issued using the ERP system. The system has full details about each payment certificate, but those details are not accessible easily by the design department, which does not have access to the ERP system. Currently, details of payment certificates are shared upon request.

The aim here is to build a regular and systematic way to show details about payment certificates, budget, and contracts in the same report.

Data Preparation

Payment certificates details is a table exported from the ERP system and have a structured style. The table looks like the following:

Table 15 - Payments Certificates Table – Structured

Sub-project	Cost Center	Business Partner	Certificate Number	Date	Activity Description	Unit	Unit Cost	Quantity	Total Cost
E1	ABC01	Partner 1	1	10/04/2020	Architect	Hours	500	150	75,000
E2	ABC02	Partner 2	1	10/04/2020	Technical Service	Lumpsum	1	25,000	25,000

The budget table is already transformed from semi-structured to structured table in 5.5.1 Solving Issue #1: Construction Costs Details in the ERP System. It looks like the following:

Table 16 - Budget Table – Structured

Sub-project	Package	Zone	Work Package	Activity	Cost Center	Quantity	Unit	Unit Cost	Total Cost
E1	Design	Z1	WP1	Activity 1	ABC01	1	LS	60,000	60,000
E1	Const.	Z1	WP1	Activity 2	ABC01	200	m3	100	20,000

The contracts list is also was built in structured style, and looks like the following:

Table 17 - Contracts List Table – Structured

Business Partner	Contract ID	Sub-project	Contract Amount
Partner 1	E1-A-01	E1	1,000,000

Data Linking

In this issue, we have three structured tables that need to be linked. The aim here is to show the Actual cost from Payment Certificates Details table versus budget and contract amount. The requirement from the design department is to show data for each business partner per sub-project. The first link is between Table 15 - Payments Certificates Table – Structured and Table 16 - Budget Table – Structured. It is needed to know how much we spent compared to the budget. Neither tables contain unique identifiers. Thus, we need an intermediary table that contains a unique identifier column (as recommended by MS Power BI). We can use the previously created Cost Center Table from 5.5.1 Solving Issue #1: as an intermediary table.

The second link is between Table 15 - Payments Certificates Table – Structured and Table 17 - Contracts List Table – Structured. It is needed to know how much we spent comparing to contract amounts. The mutual column between both tables is the Business Partner column. However, it is not unique, as, in the contracts list, we may have more than one contract with the same business partner. Thus, we need to create a new intermediary table to have unique values of business partners.

The final shape of relationships layout will be like the following:

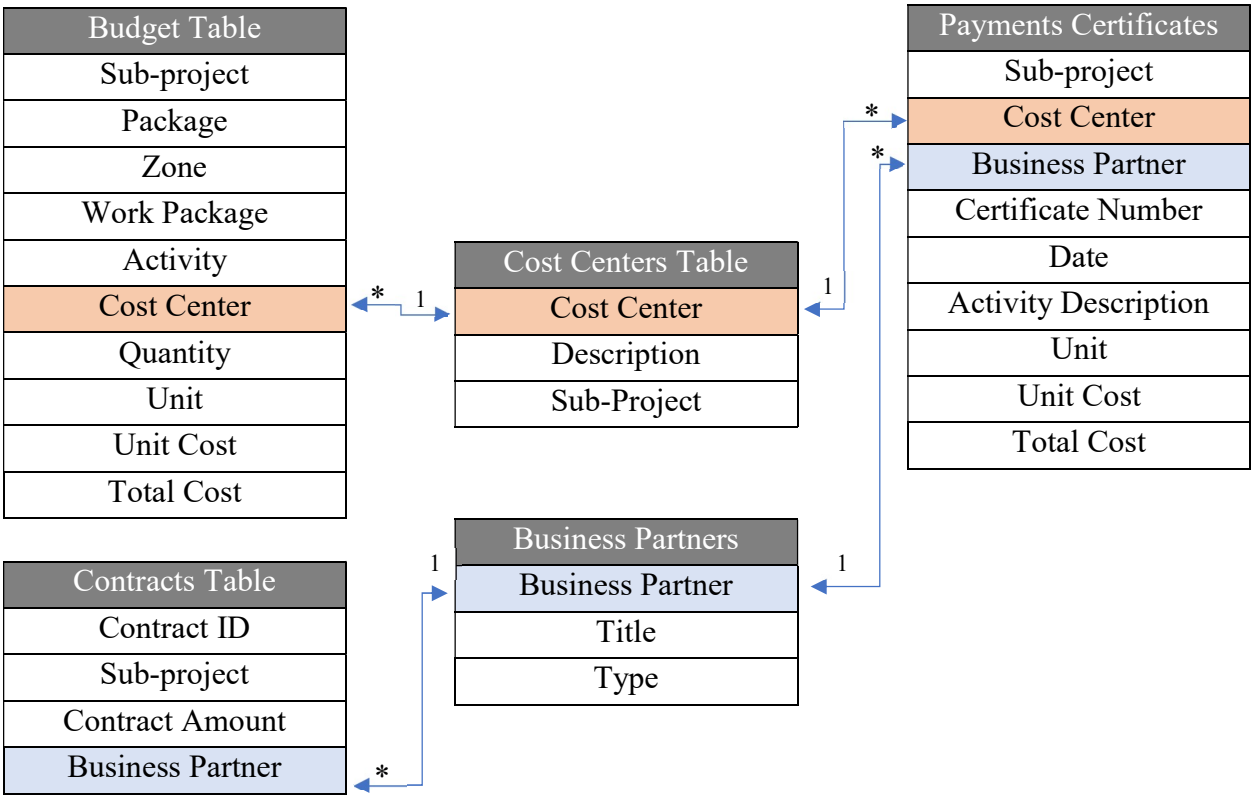

Figure 13 - Relationships Layout for Issue #3

This layout works well for the following combination of data:

- Total Cost per business partner
- Contracts Amounts per business partner and per sub-project

The required information from the design department is to show costs and contracts amounts per business partner and per sub-project in the same visual as the following:

Table 18 - Required Visual by Design Department

	Sub-project E1		Sub-project E2	
Business Partner	Contracts Amounts	Actual Costs	Contracts Amounts	Actual Costs
Partner 1				
Partner 2				

Nevertheless, if we tried to make the table using the links in Figure 13, it shows an error. The reason behind the error is that MS Power BI needs that all columns and headers of the visual (The blue cells of Table 18) should be from the intermediary tables. In this case, the sub-project is not from the intermediary table, but from the contract table. To fix this error, we can create a new intermediary table for the sub-project and link it with the contract table and payments certificate table. The final relationships layout now looks like the following:

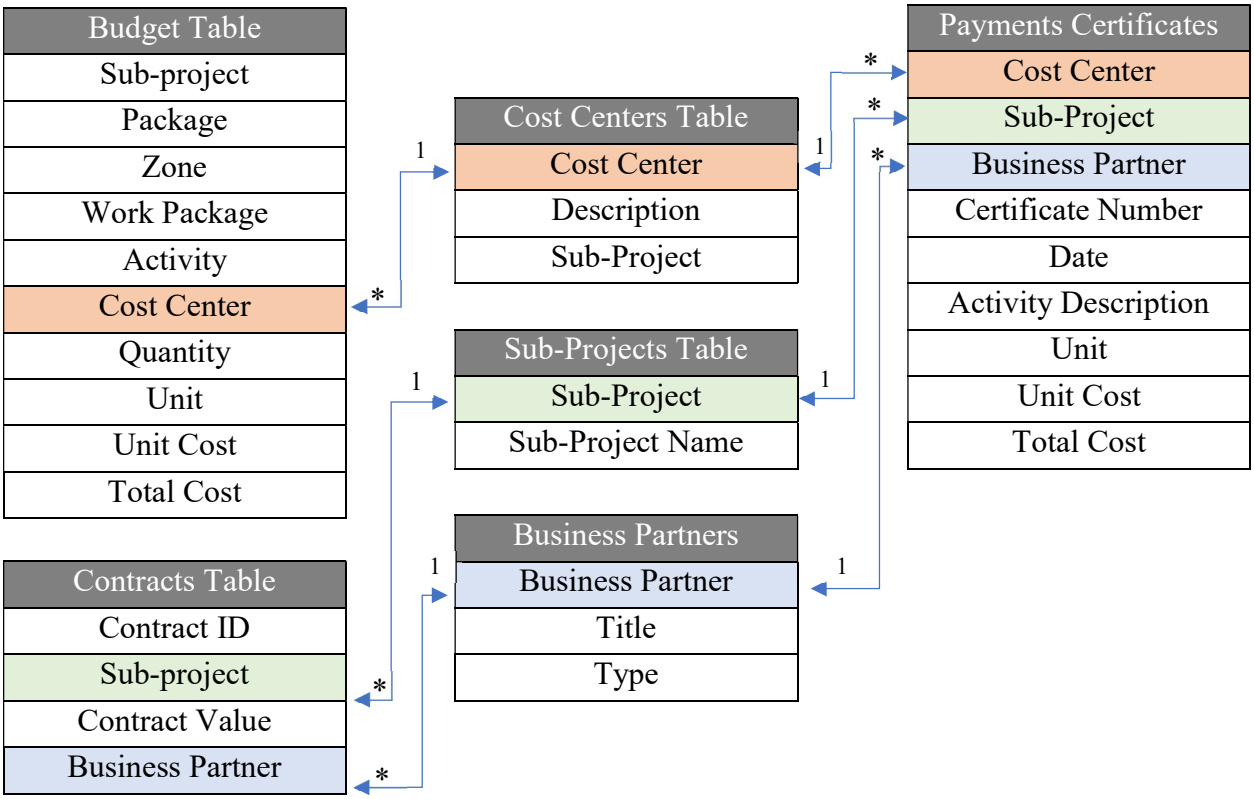

Figure 14 - Relationships Layout for Issue #3 after Modification

Colors of columns are just to illustrate which columns are used in the linking process.

Data Analysis

We have a few things to analyze in this task. Firstly, we can analyze cost performance by using Earned Value techniques. CPI is calculated by dividing the earned value by actual cost. Secondly, for the contracts, we can take the total sum of contracts versus the budget. By this, the remaining budget can be calculated in which it is easier to decide if this remaining budget is enough to make new agreements or there is a need to increase it and to what extent. The third indicator is to calculate how much we consumed from the contract amount by calculating costs/contracts. The result shows if we need to make a new contract with the business partner or increase the contract amount by using an addendum to the current contract.

Data Visualization:

The final dashboard for following up design costs looks like the following:

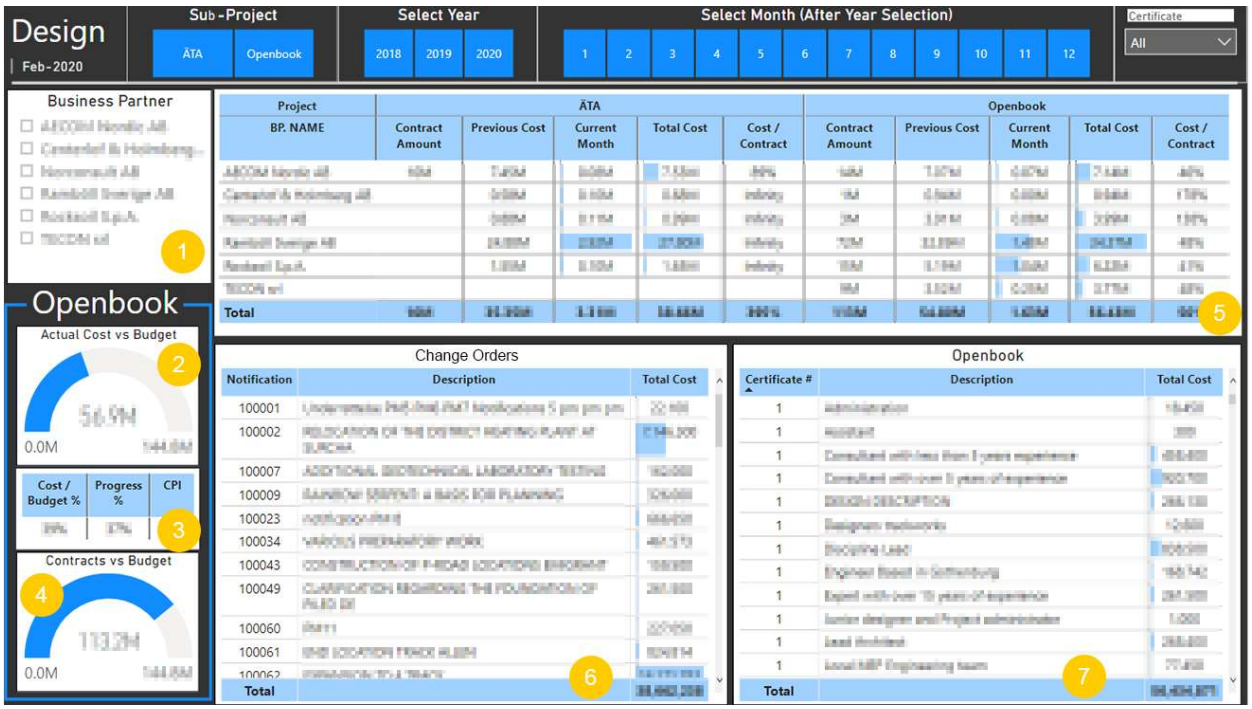

Figure 15 - Design Follow Up Dashboard

In a similar layout of the construction dashboard, we have the filters of time and sub-project in the upper part. The visualization tiles are:

Table 19 - Design Follow Up Dashboard Tiles

Tile	Source Table	Purpose
1	Business Partners	Filters report by a business partner
2	Payments Certificates, Budget	Shows actual cost compared to the budget
3	Payments Certificates, Budget, Schedule Data	Shows cost performance indicator CPI
4	Contracts, Budget	Shows total contracts amount comparing to the budget
5	Business Partners, Payments Certificates, Contracts	Shows Costs by business partner per project, and contracts amounts by business partner per project
6	Payments Certificates	Shows details of payments certificates for Change orders part by Change order number
7	Payments Certificates	Shows details of payments certificates for Openbook part by the certificate number

Tiles 6 and 7 are the details of payment certificates that the design department asked to have.

The dashboard of the Payments Certificate Details was introduced to the design team at the end of April 2020. This dashboard aims to:

- Show details of each payment certificate by the business partner
- Show contracts amounts
- Analyze remaining amounts from the contracts
- Show the cost performance index for design works

After the presentation, the design team asked to add extra information about the budget.

5.5.4 Solving Issue #4: Matching Invoices in the ERP system with submitted invoices to the client.

Data tables involved in this issue are costs allocations table from the ERP system (which contains an allocation of each invoice) and the list of invoices that are submitted to the client for reimbursement every month.

The following graph explains more about the issue:

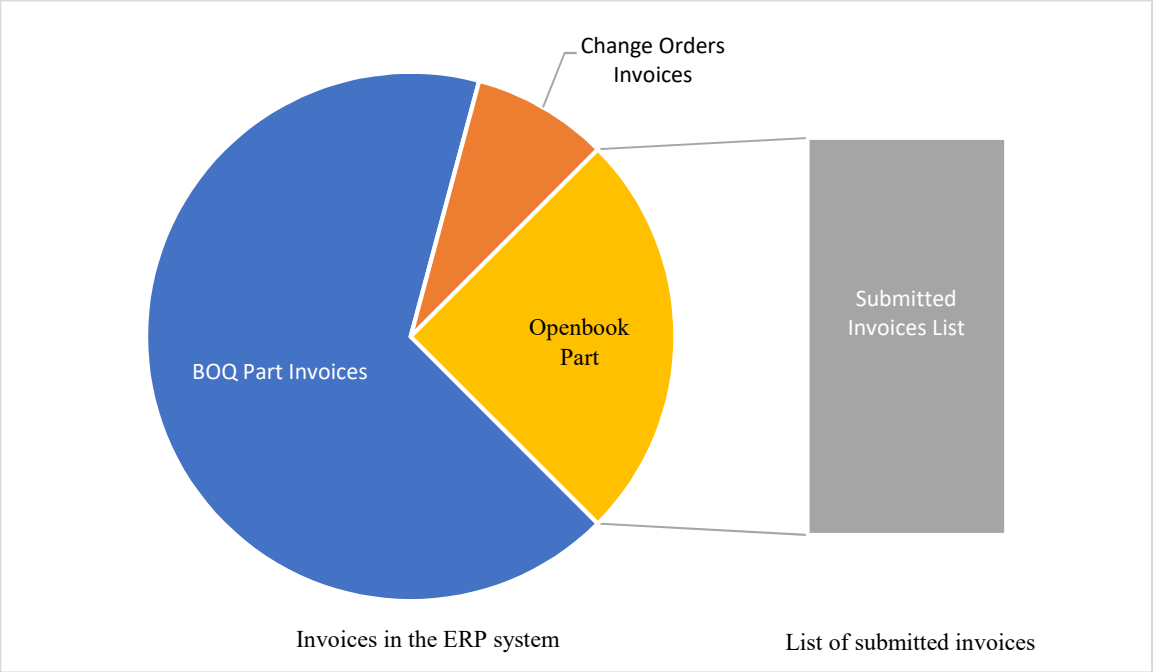

Figure 16 - Invoices in the project

The company gets each month hundreds of invoices. Those invoices are registered in the ERP system under different project parts. Every month, the administration clerk exports all invoices that are registered under the Openbook part, and submits them to the client for reimbursement, and keeps a record of the submitted invoices in an excel file.

Authors mentioned before that for costs that are in the past, the costs allocations between project parts might change, but the changes may not be reflected in the submitted invoices list. Thus, a mismatch between the two lists is found, i.e., the yellow part in Figure 16 and the gray part.

Data Preparation

The first table that can be exported from the ERP system contains details about invoices and their cost allocations. The table looks like the following:

Table 20 - Costs allocations table - Structured

Sub-project	Cost Center	Type	Quantity	Unit	Unit Cost	Total Cost	Date	Business Partner	Order	Finance ID	Invoice Number
E1	ABC01	Material	100	M2	50	5,000	01/04/20	Partner 1	E10010	1FA00001	12
E2	ABC01	Rental	20	day	120	24,00	02/04/20	Partner 2	E20010	2FA00001	23

The list of invoices that are submitted to the client looks like the following:

Table 21 - Submitted Invoices List - Structured

Total Cost	Date	Business Partner	Finance ID	Invoice Number
5,000	01/04/20	Partner 1	1FA00001	12
24,00	02/04/20	Partner 2	2FA00001	23

Both tables are structured and do not need more preparation.

Data Linking

Several trials to match the two data sources were failed, even though both lists hold an identifier, which is Finance ID. However, the problem with Finance ID is that it may change for a specific invoice if its cost allocation changed between sub-projects. Thus, the list that has static finance ID (Submitted Invoices) cannot be linked with the list of dynamic ID (ERP invoices list).

Linking with invoice number failed also. The reason behind that is that the invoice number is not unique since we may receive two invoices with number 12, for example.

For this task, to link data, we need to introduce a new identifier. This identifier is a combination of two values, which are Business Partner Number and Invoice Number. We will call the new identifier, Invoice ID. The invoice ID will be added as a new column to Table 20 and Table 21. The new tables will look like the following:

Table 22 - Costs allocations table with Invoice ID - Structured

Sub-project	Cost Center	Type				Business Partner	Order	Finance ID	Invoice Number	Invoice ID
E1	ABC01	Material	.	.	.	Partner 1	E10010	1FA00001	12	1-12
E2	ABC01	Rental	.	.	.	Partner 2	E20010	2FA00001	23	2-23

Table 23 - Submitted Invoices List with an Invoice ID

Total Cost	Date	Business Partner	Finance ID	Invoice Number	Invoice ID
5,000	01/04/20	Partner 1	1FA00001	12	1-12
24,00	02/04/20	Partner 2	2FA00001	23	2-23

The link can be established between Invoice ID columns in both tables. The relationships layout will be like the following:

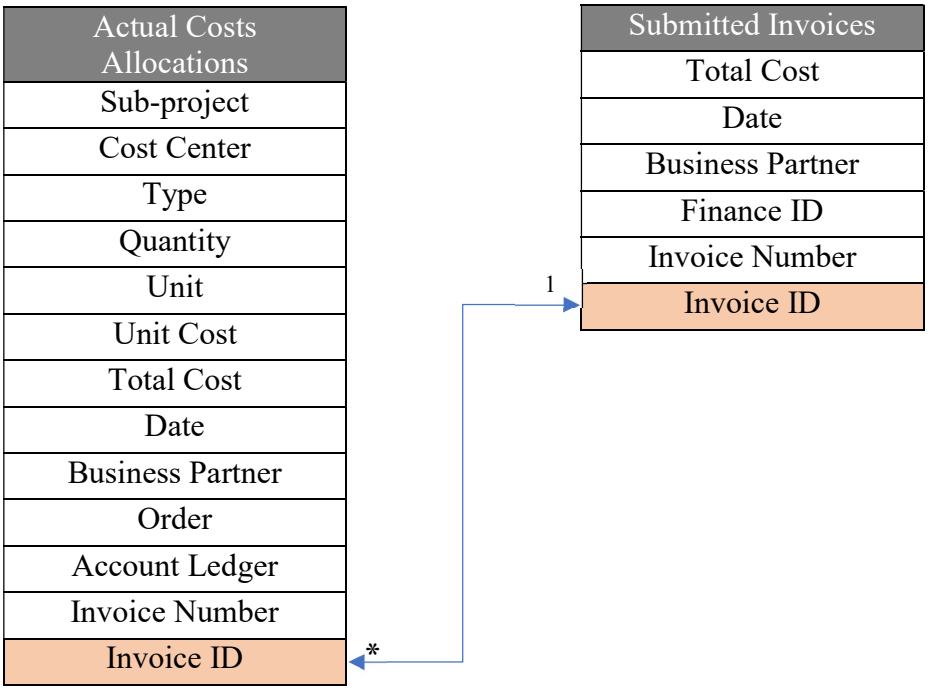

Figure 17 - Relationships Layout for Issue 4

The invoice ID in the submitted invoices list is unique as we submit the invoice once.

Data Analysis

In this part, we can make a variance analysis where variance measure equals the difference between the invoice amount in the ERP system and the invoice amount that is sent to the client.

Data Visualization

Since we do not have many data to visualize through graphs and charts, we will create a table like the following:

Table 24 - Variance between Amounts in ERP system and Submitted Amounts

Invoice ID	Business Partner	Amount in ERP System (A)	Submitted Amount (B)	Variance (A-B)	Comment
1-12	Partner 1	1,000	1,000	0	Ok
2-23	Partner 2	0	5,000	-5,000	The invoice is registered in the wrong place in the ERP system
5-1453	Partner 5	10,000	0	10,000	The invoice was not sent to the client

At the end of each month, we will check the variance column, which ideally should be 0. The anomalies need to be rectified either by changing cost allocation or by submitting the invoice to the client.

The final version of the dashboard was used by cost control since February 2020. The procedure here is that every month, before sending the list to the client, it was sent to the cost control to check if the submitted invoices are correct.

In the cost control, we used the dashboard to see if all invoices in the ERP system are included in the list. The comments on the list are sent to the finance department again, and the modifications are implemented to the list. Generally, every month around 5% of the invoices has issues to be added or deleted from the list.

6. Interviews and Findings

As mentioned before, in the methodology, four interviews were inducted with key decision-makers and professionals in the project. The interviewees are presented shortly as follows:

Interviewee A works as chairman of the board for the JV. He had worked first as a project manager for several years then, for the last 12 years, he worked as a country manager in different countries; Romania, Hungary, and Poland. His background in civil engineering with a master's degree in business administration.

Interviewee B works as a country manager of one of JV partners, commercial manager, and board member of the project. He graduated as a civil engineer, studied a master's degree in water resources and a master's in business administration MBA. His experience in the construction industry, in general, is almost 13 years, and he recognizes himself as a strategical decision-maker for 3.5 years since he starts work for the studied project.

Interviewee C works in the project as a design manager for 1.5 years. He graduated as a civil engineer, has 15 years of experience in the construction field, and worked on many different projects in Turkey, Qatar, and Sweden. He recognizes himself as a decision-maker for four years during his work in the last two projects.

Interviewee D works in the project as a finance manager. She obtained a degree in Economics in 2008 and a managerial position with decision making authority since 2017.

6.1 Information Sharing

6.1.1 General Information Sharing Findings

Interviewees A and B highlighted the importance of having **one centralized platform** in which all data is automatically linked and updated. Interviewee A argued that using interactive dashboards to present data and information gives the ability to cover different requests for data faster and more effectively compared with fragmented data in separated Excel reports.

"Even if you have numbers, not all board members were familiar with where to find it among too many numbers and reports." (Interviewee A)

Both interviewees A and B argued that in case this data is linked automatically to the model, it will be much easier to get the missing knowledge from a single place at any time. Interviewee B was more specific in this context. He said that in case the model is connected automatically to separate data resources, especially for the cost reporting, it will be much easier and effective to control the project from one centralized and always updated platform.

Both interviewees A and B believe that the model **makes it easier to read information** by providing rich and shorter reports. Interviewee A argued that the presented dashboards in the last board meeting (the first meeting to introduce the dashboard) made it easier for members - especially those who are not 100 percent involved in the project - to find, analyze, and understand things. However, Interviewee B thinks that they (the managers) need more time to interact with the dashboards, learn how to look, and read it. According to him, this process should be progressive. For Interviewee A, it is much desirable in the project management context to have one single short and rich report, including all the information in a summarized way rather than grouped subreports from different resources. Interviewee B also agreed that reports with numbers

and figures do not give insights like what graphs do. Interviewee B explained further that graphics is helping a lot during the meetings in which when they get lost between different figures, graphs can make it easier to discover problems, track it inside the figures, show it to others easily and discuss it. Interviewee A stated that currently, project management is not fully aware of sensitive and critical data regarding some costs. Interviewee B, on the other hand, highlighted this challenge by saying that before applying this model, it was difficult to compare because there is no record or data to compare with, and the monthly reports were very boring, full of numbers and figures, and so on.

Interviewee A, B, and C stated clearly that the model helped to **understand the current situation of the project** in general and the works by giving a visual summary of many data. Interviewee B stated that these presented dashboards are very much useful for reporting purposes (knowing the status of the project), and it gives a clear picture. Interviewee A added that in most cases, they had to wait for the cost controller to help them with finding and explaining submitted daily reports that were not providing the needed information to understand the status of the project in a simple and friendly way. However, interviewee B believes that what is important the data behind the figures to be correct, and internally there is still discussion about how to improve the authenticity of the original data

"If I compare, it is better with the presented model than what we had before. Those dashboards make it clearer what is the actual situation of the project, but still not complete; there is no time schedule" (Interviewee A)

Interviewee C dived deeper into practice, where he stated that control the design partners from a budget perspective is the primary responsibility for the design manager as well. It is true that the finance department and cost control department control the budget of the project designers. However, they focus only on the budget itself. There is a need from the design management side to have a general understanding of how things are going. For example, are the designers exceeding their budget, are they getting tasks that they should not. Design management is doing evaluation and revisions for the design contracts from time to time; in this context, it is essential before going to make renegotiations to know the real situation of each partner, where they are from a payment perspective. With the model, design management is more able to understand the progress of design partners; by connecting payments with contracts' amount in one dashboard, design management can control the situation better.

Interviewee A, B, and C argued that it is vital to ensure good and **efficient information flow between departments**. Interviewee A argued that it is more efficient to use this kind of approach to get information from project departments in which one does not need to ask and search to get what he/she wants. Interviewee B added that managing the information flow between different departments is vital for the project. Currently, there are difficulties in linking all departments together, and the project is suffering because of that. The solution, according to him, is to have an overall platform where anyone can easily monitor and control the costs flow; in other words, it can reach all the data from all the departments.

"I do not want to ask many people to get the information I need. If I can find it directly, it will be great" (interviewee A)

"From procurement to payment, all those steps/departments should talk to each other and sharing information" (Interviewee B).

Interviewee c also believes that it is crucial to get information from other departments. For design management and coordinating team, it is important to interact with the construction department, to know the reality, what is the real costs. Sometimes people are busy, and we need immediate cost information and estimations, that affect our decisions on which solution(design) will be efficient from the construction point of view in this context. The model is useful in which it is quick and easy to access this information.

"There are many many Excels; different departments keep different Excel files for the same information; it is actually a waste of time for everyone. The model is one common platform that everyone can rely on; it is useful in that sense" (Interviewee C)

6.1.2 Joint Ventures Related Findings

Interviewee A, B, and C agreed that the presented model helped to overcome the difficulties in information sharing and communication. From their point of view, these difficulties are because of the cultural and organizational differences of the partners. Interviewee A, for example, believes that different adopted procedures from the JV partners affect the communication inside the projects. Interviewee A, B, and C think that because the employees are from different backgrounds and used to work in one of the JV partners, they are not able to adapt to different procedures from another partner. Interviewee A and B stated that if we can make all the employees follow a standardized method, which is the model, in this case, that will surely improve and facilitate information sharing and communication.

"If you are alone, it is easier to make a decision, but when you are in JV like this one, it is always difficult to make a unanimous decision, means to reach agreement or consent about the decision. This dashboard helps to simplify the question, you can show it, and all understand that a decision should be taken. It makes it easier to convince people around you"(Interviewee B)

"Its additional load, sometimes people are sending emails that they are written in a way I cannot understand and in a different format, each time and I need to ask and search for what I want" (interviewee C)

Interviewee C took it further by saying that the model brings transparency to the JV. Sometimes partners tend to hide information on each other by not providing the requested information between professionals in deferent departments. This model is useful in this context; all information is available, automatically shared, and there is no need to ask, Interviewee C said.

6.2 Decision Making

6.2.1 General Decision-Making Findings

Interviewees A, B, and C linked the availability of information with the quality of the decision. All of them agreed that **the model provided more information, which in turn enhanced the decision** – making process. Interviewee A from his side argued that the amount of available information is higher now after using the model, and one can fairly say that the management is more capable of making better decisions. Interviewee B supported this argument by stating that information included in the presented dashboards during the board meeting is helping more to know how badly or well the project is. Consequently, top management can prioritize actions that are in one way or another kind of making decisions. Interviewee C was more specific than others;

he thinks that providing full information about design partners, costs, and budget, give the ability for the design management team to know the reality, to monitor the budget and make sense of the numbers. However, Interviewee A emphasized on a topic that data behind the model should be more complete in order to increase the reliability of the presented data. In diffident words, he believes that the right decision needs completed and correct information, interviewee A said.

According to Interviewee B and C, those **dashboards can be much more supportive and efficient for departments' managers** to make a specific decision. Interviewee B explained further that strategic decisions are not taken immediately in one meeting. Even if all the needed information is available, it takes more time in this case to monitor the information and discuss it between parties to, finally, make a decision, Interviewee B said. Interviewee C gave an example for a design partner who was included in the presented dashboard and he could see that the budget has exceeded the contract amount three times. In this case, Interviewee C decided not to give this partner any of the upcoming tasks since it is clear from the model that the costs increased, and there is a need to adopt a new payment strategy. This new strategy is to change the type of contract for this scope of tasks from Time & Material to Fixed Price. This is a direct decision that was made depending entirely on the information provided from the model.

6.2.2 Information vs. Experience Findings

Interviewees A, B, and C agreed that **both information and experience are needed** to make a decision. Interviewee A stated it clearly that it is vital to have enough and correct information in order to build on facts. However, still, the experience will be used to analyze alternatives. Interviewee C developed the argument by stating that data and experience are combined; if they do not match, data will be questioned by using the intuition and the experience. Data cannot be fully trusted, but it can give a warning that there is something wrong.

Interviewee A and B agreed that **each decision has its own circumstances**. Interviewee A explained further that each decision is unique and has its own conditions, sometimes the decision is made based only on numbers, but on the other hand, sometimes intuition and experience play the major role. Interviewee B argued that each decision has its own specificities; therefore, it is hard to define which is the most critical factor for a good decision, is it the numbers and information or the experience and intuition?

"*I think the experience verify in somehow the data; if there is a conflict between them, I would go for experience*" (Interviewee B).

6.3 Other findings

Both interviewees C and D stated that **the model helped to compare and check data,** which allowed them to discover meaningful patterns and issues. Interviewee C, for example, argued that presenting data in a customized visual way is always helpful to discover relations and patterns. Interviewee C added that the model is useful to compare the data in case the data is correct and reliable. Besides, by using a visual model, there is no need to deal with the ERP system, and that is very good since not all employees are familiar with it.

Furthermore, Interviewee D said also that the process of reviewing the submitted list of invoices by using the presented solution is very helpful for her, it was the first time in the project that her list is reviewed deeply by another party. Interviewee D said that the important thing is that the comments about the list arrive in a short time, as the time factor here is critical because there is due time for submitting this list to the client. Interviewee D also added that the list helped to discover many invoices that were not sent to the client. In March, we found around 4 million SEK, and in April, we found around 1.5 million SEK, besides a 1 million SEK of invoices for the past time of the project. Those amounts could probably get lost or at least take much time to be discovered if a similar technique were not used, interviewee D said.

6.4 Survey Results

During each interview, interviewees were asked four questions to evaluate the model from a different point of view. The results from the conducted survey are shown in table 25:

Table 25 - Interviews Survey Results

Question	Interviewee	Evaluation				
		Not at all	slightly	Moderately	Very much	Extremely
1. To what extend you think the model helped you to better understand the current situation of the project/Business Partners?	A				x	
	B				x	
	C			x		
	D				x	
2. To what extent you think the model succeeded in providing good and enough information so you can make a better decision?	A				x	
	B				x	
	C				x	
	D				x	
3. Does the model make it easier for you to get information from another department?	A					x
	B			x		
	C				x	
	D				x	
4. To what extend you think the model improved the information sharing between professionals inside the JV.	A				x	
	B				x	
	C					x
	D				x	

7. Discussion and critical results

This section is about discussing the results from the empirical study by using the theoretical frame as a tool to analyze the findings and reflect upon them.

7.1 Information Sharing

The first goal of the study is to make more information available for decision-makers and key professionals.

As stated by Pirzadeh and Lingard (2017) in the theoretical part in section 3.2.3 The impact of temporary status of construction projects and the need for data management system, the temporary nature of construction projects leads to inefficient information sharing systems. Consequently, the efficient flow of information is not guaranteed. Moreover, Chassiakos and Sakellaropoulos (2008) stated that adopting a document-based rigid system that manages documents separately rather than the data included in it allows limited interactivity with the end-user. This idea was clearly supported by interviewee C, who mentioned that it is a waste of time for everyone when keeping many different excel files in each department. Interviewee A also mentioned that the data in the project is fragmented and exists in separate excel reports, and not everybody in the board is familiar with how to get the information from among too many numbers and reports.

After introducing the model to the interviewees, it helped them to overcome information sharing issues. All Interviewees stated that they were able to get information faster and more efficiently than before. Interviewees A and C mentioned that getting the information from a single platform is very helpful for them. Moreover, Interviewee C mentioned that it is easier now to get information from different departments without the need to call or email them. Interviewee B emphasized that it is essential to allow for efficient information flow between departments from purchase requests to payment. Interviewee A stated that getting the most important information and present it in a single rich report is very helpful as it saves much time for executive management. Chassiakos and Sakellaropoulos (2008) clearly supported this argument in section 3.2.3 when they stated that a centralized data management system allows them to search, analyze, and present data easily and permanently.

Interviewees A and B mentioned that data visualization helped them to interpret data quickly. That is also supported by a study that was mentioned in the theoretical part in section 3.1.3 by Cheng (2017). Moreover, both theory (Section 3.1.3, Cheng (2017)) and interviewees agreed on the idea that visualizing data helps with finding trends and patterns in data quickly and easily. In the interview with interviewee A, when presenting him with a chart of CPI, interviewee A was able to get the trend in few seconds rather than getting the old fashion report, which contains many pages and takes much time to discover the trend.

Nevertheless, one can ask a question; could the same information be provided to the stakeholders with the old conventional method (i.e., excel, excel chart) without the need for Big Data technology or special software like MS Power Bi? The answer is probably yes. However, it will take a lot more time to generate the reports with excel rather than with Power BI. Technically, all Power BI modules that were provided in this study can be updated in 15 minutes, which is the time of exporting data from the ERP system only. The rest is automatically generated. While in the

conventional method, it may take days to generate the same reports. Besides that, the data sources that are growing will be harder and harder to be processed in the future conventionally. While in the Power BI, it takes the same time, it is taking today. This idea is very important as all of the interviewees mentioned the importance of the time factor.

Moreover, the conventional method needs data to be processed on the same device. That means we need first to collect all data source files and then analyze them and generate reports. While in Power BI, data files do not need to be on the same device and it can be linked from the data source owner device directly, and every time the original source of data changes, the changes are reflected directly to Power BI where it can be reviewed before publishing to the end-users.

Another point related to the information sharing is that even though the graphs that were presented in the dashboards visualize, the data were helpful to the managers. Interviewee B mentioned that they needed some time to learn how to look at the graphs for the first time, which was easier for the second time, and it is a successive process.

In addition, in section 3.3 Cultural and organizational differences' impact on construction joint venture projects, the cultural and organizational differences in joint ventures are hindering information sharing in joint ventures. Moreover, managing these differences in a good way is vital to the success of the project. This is supported by Interviewee A and B, who mentioned that the organizational differences and the different procedures have a negative effect on the information sharing in the project.

In theory, it is also mentioned by Köster (2009) that a shared database could be an effective solution for introducing a different communication style that minimizes cultural differences effects. This is also supported by Interviewee C, who stated that the model made it easier to obtain data from a colleague who came from a different working culture, and mentioned that it brings more transparency in which the information is available for all partners. Interviewee B stated clearly that the model represents a standardized tool to present information for people from different backgrounds, which facilitated the communication between the joint venture partners.

As interviewees belong to different levels of the organization's hierarchy, each one is interested in a different level of data visualization. Interviewee A, for example, is a high-level manager and more interested as he said to get a short and rich summary instead of too much information. Interviewee C, on the other hand, stated that he would like to have one report linking all payments with contracts' amount, in other words, all the details of payments and contracts. This is totally and clearly corresponding with what Cheng (2017), in theory, section 3.1.3, that there are different levels of visualization depending on which amount and depth of data are needed by the end-user. Based on what both interviewees mentioned above stated that the model provided the requested level of data and, It is fair to say that Big Data visualization can control the amount and the level of visualizing data to each end-user separately and efficiently.

7.2 Decision Making

The second aim of this study is to enhance decision making. If we reflect the bounded rationality in theory in section 3.2.1 Decision making by Lorkowski and Kreinovich (2018), the three limits that affect decision-making are: Limited Information, Limited Time, and Limited Cognitive Capacity. While Marwala (2014) argued that those limits could be partially corrected by adopting an advanced information analysis method. This argument suggests that if the presented model in this study works properly, then it will partially enhance the three pillars of bounded rationality. This argument is supported in the empirical part, as in the following:

Firstly, Limited Information. Interviewee A, B, C agreed that they received more information than before. Interviewee C was the stakeholder who benefited most from the higher availability of information from the model **in which the design team actually made a decision based on that**. They decided to change the contract type with a business partner, which was originally agreed based on uncertain information. However, interviewee A mentioned that even though they have access to more information, it is still not complete.

"If *I compare, it is better with the presented model that what we had before. Those dashboards make it clearer what is the actual situation of the project, but still not complete*" (Interviewee A)

Secondly, Limited Time. As a result of the solution, the required information was provided, and the reports were generated in a very short time. This allows the decision-makers to obtain information and discover patterns and trends from data faster which was clearly stated by interviewee A in which they needed to wait for the data owner to explain to the board how to read the numbers from the old reports, while in the new dashboard they were able to obtain the information more quickly. Interviewee C also mentioned that the design team became able to access information from the construction department quickly by using the model. Interviewee A and B emphasized the importance of graphs and charts as a tool for getting information in a fast way. Interviewee D also highlighted the importance of time and assured that getting the information in a quick way helped to submit the correct invoices to the client on time, where there is a due date for that.

Thirdly, Limited Cognitive Capacity: In the model, the graphs and charts provide the most essential analyzed data to the stakeholders. For example, the Cost Performance Indicator CPI analysis was a result of an in-depth analysis of data from different sources, which are cost, budget, and planning. Interviewee A clearly mentions this when stated that the model helped board members to find, analyze, and understand the information in a "fast and friendly way" compared to what they had before.

There is another point in the discussion of decision-making, the type of a decision, is it a strategic decision or tactical decision. Interviewee B stated that using the dashboards is more helpful in the tactical decision comparing to the strategic decision. The reason for this, as Interviewee B argues, is that the strategic decisions are not taken in one day, it takes more time while the tactical decisions are more frequent and needed on a daily basis. This is also mentioned in theory (Section 3.4 by De Wit (2017)) where it is argued that the strategic decisions like business development are based more on the experience (intuition) rather than the information. Interviewee B statement is

confirmed by Interviewee C – who is a middle-level manager that takes more tactical decisions - who mentioned that the model was beneficial for the design team, and they actually took a tactical decision based on it.

"*Information is pretty thin stuff unless mixed with experience*" (Clarence Day). Interviewee A, B, and C stated clearly that they make decisions based on both their experience and the available information and facts. The three Interviewees also confirmed that it is essential to have correct data and information. However, each decision has its own conditions and circumstances, so it is difficult to define which factor is the most important. Also, Interviewee B and C mentioned that experience is used to verify and question the data, interview A added that experience is important to analyze alternatives. This argument was fully supported by De Wit (2017) in the theory when he stated that decision-makers should use both intuitive thinking and facts-based rational analysis to achieve more realistic and unbiased decisions. Interviewee B choose to go for the experience in case of conflict with the presented data. This could be understandable since interviewee B is board member and more strategist than middle- level manager. A question could be asked here, is a quick and intuitive decision always risky, the answer maybe is no, as interviewees stated each decision has it won circumstances. Sometimes the quick decision is needed and cannot wait for the information to be provided. For example, the Corona pandemic case in 2020, where governments had to make quick decisions since time was a dominant factor. The same could be in case of too much information, mangers, in this case, can be lost, and that can affect the quality of the decision.

It is worth mentioning here that the model in this study is to enhance the part related to the logic thinking approach, which depends on information and fact to make decisions regardless of how much the decision-maker relies on it or the intuitive thinking.

7.3 Other Benefits

Even though the original aim for this solution is to enhance information sharing and decision-making in the project, other benefits were found from the solution that was not in the original plan. Firstly, the model was used to prepare many filters and to check the inputs in the ERP system and find anomalies of data; this was also mentioned by interviewee D, who stated that the model helped to find many invoices that were not sent to the client.

Also, other than the dashboards that already created, the model was used to answer quick questions in a work meeting and a "quick report" is generated to access the information, for example, how much we spent for this business partner, how consultants' costs are distributed through time, what are the high costs in the indirect costs.

8. Conclusion

This study aims to investigate if Big Data technology can improve information sharing and enhance the decision-making process in construction projects. The study was made in an ongoing construction project where the contractor was a joint venture company.

The main problem which was analyzed is the information sharing difficulties in the project that affect the decision-making process. Since any problem cannot be solved if it is not defined first, the initial problem of this study had to be described and analyzed deeply by dividing it into three specific components. First, data fragmentation problem in construction projects, the current situation of the project data was studied to understand how it is divided and stored in separated units and systems. Second, analyzing the negative effect of the temporary nature of construction projects on information sharing and knowledge flows between different departments and professionals. Lastly, studying how the organizational and cultural differences of the joint venture affect the information exchange inside the project. Furthermore, the decision-making process was studied to understand which factors are most important to generate reasonable and high-quality decisions, and how providing and presenting enough and well-analyzed data easily and simply could affect the process. As a result of the above-mentioned analysis, four major practical issues were selected to be solved. These issues were: provide and visualize construction costs details to the construction team, calculate and visualize cost performance index CPI to executive management, show details of payments certificates and budget to design team, and lastly matching list of invoices in the ERP system with a list that is sent to the client

To solve these four issues, a model that is based on a Big Data technology software (Microsoft Power BI) was created. The model went through many phases before being presented to a group of decision-makers and professionals. Firstly, the data preparation process of the raw data to re-shape and organize the datasets where it can be processed by the software. Secondly, the data linking process took place to connect the fragmented datasets together in one centralized platform. Thirdly, the data analysis process was conducted to calculate indicators and analyze data to generate useful information and reports out of it. Finally, the data visualization process through reports and dashboards and presenting those reports to the managers and professionals that are related to one of the four selected issues (This link for the final presented reports and dashboards with dummy data: https://www.datahouse.se/dashboard/). To study and evaluate the performance of the model, interviews, and a survey with the same managers and professionals were conducted. Interviews and survey results data were gathered and analyzed later by using a theoretical framework of relevant literature and studies to answer the research question.

Did Big Data technology improve the information sharing in the studied project?

From the interviews and survey results, we found that most interviewees mentioned that the Big Data model, which was presented in this study, helped them very much with improving information sharing between professionals and departments in the project, especially in a joint venture project. It also helped most of the interviewees very much to know the status of the project or the work package which they manage better than before. We can see that the model enhanced information sharing in the project. However, some interviewees mentioned that the information still not complete, and more data sources should be added.

Did Big Data technology enhance the decision-making process in the studied project?

All interviewees agreed that the information provided in the model could help them to make better decisions. Also, one of the interviewees stated that a tactical decision was made based on the information provided in the model. The results from the interviews showed that the model helped more for the tactical decision rather than the strategic decisions. This is because the timeframe of this study is not long enough to see the effect on strategic decisions.

9. Suggestions for future research

This study covered Big Data applications to enhance information sharing and decision-making in a construction joint venture project. However, through the process, we found a lot of other applications for Big Data like finding anomalies and make internal analysis inside the departments. This can be part of future research.

References

ADNAN, H. B. & MORLEDGE, R. 2004. *Joint venture projects in Malaysian construction industry factors critical to success.* Nottingham Trent University.

AMENDOLA, A. 2002. Recent paradigms for risk informed decision making. *Safety Science,* 40**,** 17-30.

BALDWIN, A., THORPE, A. & CARTER, C. J. A. I. C. 1999. The use of electronic information exchange on construction alliance projects. 8**,** 651-662.

BELL, E., BRYMAN, A. & HARLEY, B. 2018. *Business research methods*, Oxford university press.

BILAL, M., OYEDELE, L. O., QADIR, J., MUNIR, K., AJAYI, S. O., AKINADE, O. O., OWOLABI, H. A., ALAKA, H. A. & PASHA, M. 2016. Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced engineering informatics,* 30**,** 500-521.

BRYNJOLFSSON, E., HITT, L. M. & KIM, H. H. 2011. Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN 1819486*.

BUSINESS, R. O. 2020. Baan Company - Company Profile, Information, Business Description, History, Background Information on Baan Company. *https://www.referenceforbusiness.com/history2/40/Baan-Company.html*

CHASSIAKOS, A. & SAKELLAROPOULOS, S. J. A. I. E. S. 2008. A web-based system for managing construction information. 39**,** 865-876.

CHENG, M. 2017. "Big Data Management", Fausto Pedro García Márquez and Benjamin Lev. Springer.

DAVENPORT, T. 2014. *Big Data at work: dispelling the myths, uncovering the opportunities*, Harvard Business Review Press.

GREENHALGH, T. & PEACOCK, R. J. B. 2005. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. 331**,** 1064-1065.

KÖSTER, K. 2009. *International project management*, Sage.

LORKOWSKI, J. & KREINOVICH, V. 2018. *Bounded rationality in decision making under uncertainty: towards optimal granularity*, Springer.

LU, W., LAI, C. C. & TSE, T. 2018. *BIM and Big Data for Construction Cost Management*, Routledge.

MACLURE*, M. J. J. O. E. P. 2005. 'Clarity bordering on stupidity': where's the quality in systematic review? 20**,** 393-416.

MARWALA, T. 2014. *Artificial intelligence techniques for rational decision making*, Springer.

NASIR, N., NAWI, M. N. M. & TAPA, A. Dilemma in decision making at pre-construction phase-A state of art review. AIP Conference Proceedings, 2016. AIP Publishing LLC, 020080.

NGO, J., HWANG, B.-G. & ZHANG, C. J. A. I. C. 2020. Factor-based Big Data and predictive analytics capability assessment tool for the construction industry. 110**,** 103042.

OMRAN, B. A. & CHEN, Q. Trend on the implementation of analytical techniques for Big Data in construction research (2000–2014). Construction Research Congress 2016, 2016. 990-999.

OZORHON, B., DIKMEN, I. & BIRGONUL, M. T. J. J. O. M. I. E. 2007. Using analytic network process to predict the performance of international construction joint ventures. 23**,** 156-163.

PIRZADEH, P. & LINGARD, H. J. J. O. M. I. E. 2017. Understanding the dynamics of construction decision making and the impact on work health and safety. 33**,** 05017003.

PROVOST, F. & FAWCETT, T. 2013. Data science and its relationship to Big Data and data-driven decision making. *Big Data,* 1**,** 51-59.

SAFA, M. & HILL, L. J. S. D. 2019. Necessity of Big Data analysis in construction management.

WALKER, H., SEURING, S., SARKIS, J., KLASSEN, R., ROEHRICH, J. K., GROSVOLD, J., HOEJMOSE, S. U. J. I. J. O. O. & MANAGEMENT, P. 2014. Reputational risks and sustainable supply chain management.

WEBSTER, J. & WATSON, R. T. J. M. Q. 2002. Analyzing the past to prepare for the future: Writing a literature review. xiii-xxiii.

Lu, W., Lai, C.C. and Tse, T., 2018. *BIM and Big Data for Construction Cost Management*. Routledge.

Kivrak, S., Arslan, G., Tuncan, M. and Birgonul, M.T., 2014. Impact of national culture on knowledge sharing in international construction projects. *Canadian Journal of Civil Engineering*, *41*(7), pp.642-649.

Guide, P. M. B. O. K. (2017). A guide to the project management body of knowledge. Sixth Edit. *Project Management Institute, Inc*.

De Wit, B., 2017. Strategy-An International Perspective. 6th ed. Boston: Cengage Learning.

Microsoft Power BI https://powerbi.microsoft.com/en-us/

link to the final presented reports and dashboards with dummy data:
https://www.datahouse.se/dashboard/

Department of Architecture and Civil Engineering

Master's Thesis ACEX30-XXX

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden

www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY