EINDHOVEN UNIVERSITY OF TECHNOLOGY

Master Thesis

The utilization of BIM-based as-planned and as-built scheduling data

Exploratory research into the capture, storage, analysis and (re)use of BIM-based as-planned and as-built scheduling data for the estimation of task durations for future construction projects

2019 – 2020
Construction Management and Engineering
Max Wessel

(page intentionally left blank)

Colophon

General

Title	The utilization of BIM-based as-planned and as-built scheduling data
Subtitle	Exploratory research into the capture, storage, analysis and (re)use of BIM-based as-planned and as-built scheduling data for the estimation of task durations for future construction projects
Date	24-08-2020

Author's information

Name	Max Wessel
ID	1278649
E-mail	wesselmax@hotmail.com
Master track	Construction Management & Engineering
Faculty	Faculty of the Built Environment
University	Eindhoven University of Technology

Graduation committee

Chairman	Prof.dr.ir. B. (Bauke) de Vries
First supervisor	Dr. ir.-arch. P. (Pieter) Pauwels
Second supervisor	Ir. B. (Bob) van Thiel

Graduation company

Company	BAM Infra

External supervisors

External supervisor	Ing. J.I. (Brigt) Rykkje
External supervisor	Drs.ir. P. (Paul) Hendrikx

“Prior
Proper
Planning
Prevents
Poor
Performance
(The 6 P’s)”

Preface

Imagine a world wherein all information is digitally stored. Everything is there, but there is one problem. There is no way to access the information. All the information is put away, unstructured, locked, behind imaginary bars, unreliable and so on. Now, imagine that information not being digitally stored, but instead, it is stored in people's minds. The information is there, you just need to find the right person, ask, and hopefully, that person can retrieve the information and provide you with a valuable answer. As a matter of fact, today, those minds form the foundation for the discipline of construction scheduling. But the question is, are those minds enough?

This thesis is the result of my intrinsic pursuit to improve for the better. And, the final step as part of the master Construction Management and Engineering (CME) at the Eindhoven University of Technology (TU/e). The graduation research is conducted in collaboration with BAM Infra.

I would like to express my gratitude to the people that provided me with advice, support, feedback and joy throughout my studies. First, I want to thank Pieter Pauwels and Bob van Thiel for their guidance and feedback to bring this thesis to a higher level. Secondly, I want to thank the supervisors from BAM Infra, Brigt Rykkje and Paul Hendrikx, for their continuous support, the knowledge they have shared and the opportunity they provided to conduct this research. Also, I would like to thank all the other colleagues that helped me.

Lastly, I want to thank my fellow students, family and friends.

Max Wessel,

Eindhoven, 2020

(page intentionally left blank)

Summary

The factors that define the success of a project are commonly described by the golden triangle of time, cost, and quality. Hence, the assurance of the construction schedule is important since delays can undermine that success and tend to result in consequent cost overruns and other problems. Strikingly, it was found that nine out of ten large-scale infrastructure projects face significant time delays and cost overruns. Therefore, change is needed.

Within the AEC (Architecture, Engineering and Construction) industry, the ability to learn from experiences and projects is considered to be difficult. As part of that problem, a significant number of construction planners is unable to use scheduling data of past projects. As a result, the overall scheduling process is primarily based on intuition and personal experiences, rather than well-founded figures. Furthermore, often used on-site progress tracking methods depend on manual interaction and paper-based methods that are inaccurate and time-consuming. Also, there is a lack of systematic feedback towards the planner which prevents the continuous improvement of construction schedules.

The use of BIM (Building Information Modelling) yields an enormous number of benefits as its data can become the basis of information and knowledge distribution. 3D BIM models can utilize tools to create so-called 4D BIM models that link the BIM model to the construction schedule. Furthermore, the use of these models in combination with BIM-based monitoring techniques is becoming more automated lately. Based on the identified problems and potentials, the aim of this study is to explore the use of BIM-based scheduling in combination with progress tracking methods to capture, store, analyse and reuse as-planned and as-built data in a systematic manner. More specifically, the aim is to reuse the data to estimate task durations. The objective of this research is to develop a system that is reliable, accurate, accessible and effectively usable for the estimation of task durations for future scheduling processes. This resulted in the following research question: "How can BIM-based as-planned and as-built scheduling data be utilized to estimate task duration for future construction projects in a pre-construction phase?".

A literature study is conducted on the process and state of the art concepts related to BIM-based scheduling, progress tracking and delay analysis, and the estimation of task durations. The fundamental idea of BIM-based scheduling is that the elements in the BIM model form the basis for which the construction schedule is made. The current state of the art techniques regarding on-site progress tracking are deemed promising. However, it was found that based on different criteria such as time efficiency, accuracy, level of automation, preparation, costs of the technique, none of the current techniques perform ideally. For that reason, the choice of technique depends on the situation at hand. These progress tracking techniques can facilitate in finding discrepancies, their extent and potentially discover causes that allow for delay analysis to take place.

On-site task durations are influenced by several factors that can be categorized as input factors, factors related to the internal environment, and exogenous factors. The tangible character of elements within a BIM model makes parametric estimation a suitable method. This method relies on a (statistical) relationship between variables. Since the knowledge and

quantification of these influential factors can potentially allow task duration to be modelled more accurately, this study proposes to enrich the as-planned and as-built schedules with information regarding these factors. Additionally, the proposal is to enrich tasks that face either a start or duration variation with cause information. This information allows root cause analysis to take place and subsequently utilize this data to identify risks.

The V-shaped software development model was used consisting of a development part and a testing part to develop the system. Requirements and specifications were collected for different use cases (planners, project managers and portfolio managers) to estimate task durations and potentially other uses of data. This study proposes the use of a Plan, Do and Measure, Analyse and Improve, and (Re)use cycle that allows for continuous improvement. In the first phase, a BIM-based as-planned schedule is created. In the second phase, the construction progress is tracked resulting in BIM-based as-built schedules. For the storage of the data, SQL and NoSQL databases were studied. SQL databases are generally suited for structured data, whereas NoSQL is generally preferred for unstructured data. For this proof of concept, a relational database model is developed in SQL. The relationship between tasks and BIM elements (i.e. many-to-one) is found to be fundamental for the ability to reuse task duration data, as well as, the use of classifications. BIM-based scheduling data which includes BIM element property data can be exported as an Excel file using Synchro 4D (BIM 4D tool). Subsequently, a Python script, that is developed in this study, can be used to extract, transform and load the data from Excel into the SQL database.

The data can be retrieved using queries once stored. In general, data in databases can be analysed using queries, dashboards and data mining techniques. In this study, the SQL database is connected to PowerBI which serves as a ‘temporary’ graphical user interface. In PowerBI, the data can be searched and visualized in various ways. Subsequently, task duration estimates can be derived in three ways: 1) the duration can be directly retrieved with the search/query of a specific task part for a certain element, 2) a productivity measure can be derived over a certain set of tasks related to elements, and 3) predictive data analysis can be performed on a dataset with certain tasks, elements and influential factors to derive a model which can potentially be reused for the estimation.

The system as part of the proposed continuous improvement cycle is validated by means of a fictitious case study. The case study provides insight into the different ways data can be retrieved, analysed and (re)used. The case study showed that the system is relatively accessible and useable; however, separate graphical user interfaces are needed to fully utilize the system. The reliability and accuracy of the task duration estimates could not be validated properly since no real project data was used. In particular, the usability of predictive models related to tasks, elements and influential factors should be further investigated.

Overall, this research succeeded in the development of a system as a proof of concept that allows a productivity database to be built, fed by BIM-based as-planned and as-built project scheduling data to estimate task durations. This study recommends the use of BIM-based scheduling techniques, classifications and standards, as well as, the uniform definition and collection of inflectional factors upon task duration during the execution of a project. Furthermore, it is recommended to store BIM-based as-planned and as-built scheduling data as part of a database whereby tasks and elements make use of a many-to-one relationship.

Summary (Dutch)

De factoren bepalend voor het succes van een project worden normaliter beschreven door de gouden driehoek van tijd, kosten en kwaliteit. Het nakomen van de bouwplanning is belangrijk, aangezien vertragingen het succes kunnen ondermijnen en resulteren in kostenoverschrijdingen en andere problemen. Opmerkelijk is dat negen van de tien grootschalige infrastructuurprojecten te maken heeft met aanzienlijke vertragingen en kostenoverschrijdingen. Oftewel, verandering is nodig.

Binnen de AEC-industrie (Architecture, Engineering en Construction) is het gebleken dat er moeilijk geleerd wordt van ervaringen en projecten. Als onderdeel van dit probleem kan een groot gedeelte van de bouwplanners geen gebruik maken van historische planningsdata. Als gevolg hiervan, is het gehele planningsproces voornamelijk gebaseerd op intuïtie en persoonlijke ervaringen, in plaats van op goed onderbouwde cijfers. Bovendien zijn vaak gebruikte methodes voor het bijhouden van de voortgang op de bouwplaats onnauwkeurig en tijdrovend doordat ze niet digitaal zijn en handmatige interactie vereisen. Ook is er een gebrek aan systematische feedback richting de planner, waardoor het continu verbeteren van de bouwplanningen wordt beperkt.

Het gebruik van BIM (Building Information Modelling) levert veel voordelen op, aangezien de gegevens hiervan de basis kunnen vormen voor informatie- en kennisverspreiding. Specifieke planning software tools kunnen 3D BIM-modellen gebruiken om zogenaamde 4D BIM-modellen te maken. Deze koppelen het BIM-model aan de bouwplanning. Bovendien wordt het gebruik van deze modellen in combinatie met BIM gebaseerde monitoringtechnieken momenteel meer geautomatiseerd. Op basis van de geïdentificeerde problemen en mogelijkheden heeft deze studie het doel BIM gebaseerde planningen in combinatie met methoden voor voortgangsmonitoring technieken te onderzoeken en om systematisch as-planned en as-built gegevens vast te leggen, op te slaan, te analyseren en te hergebruiken. Meer specifiek heeft deze studie het doel om de gegevens her te gebruiken om op deze manier taakduur te bepalen. Het doel van dit onderzoek is om een systeem te ontwikkelen dat betrouwbaar, nauwkeurig, toegankelijk en effectief bruikbaar is voor het bepalen van de taakduur voor toekomstige planningsprocessen. Dit heeft geresulteerd in de volgende onderzoeksvraag: “Hoe kan BIM gebaseerde as-planned en as-built planningsdata worden gebruikt om de taakduur voor toekomstige bouwprojecten in een pre-constructiefase te bepalen?”.

Een literatuurstudie is uitgevoerd voor de processen en state-of-the-art concepten gerelateerd aan BIM gebaseerde planningen, voortgangsmonitoring technieken en vertragingsanalyse, en het bepalen van de taakduur. Het fundamentele idee van BIM gebaseerd plannen is dat de elementen in het BIM-model de basis vormen voor het generen van de bouwplanning. De huidige techniek met betrekking tot het monitoren van de voortgang op de bouwplaats wordt als veelbelovend beschouwd. Echter presteert geen van de huidige technieken optimaal op basis van verschillende criteria zoals tijdsefficiëntie, nauwkeurigheid, automatiseringsniveau, voorbereiding en kosten van de techniek. Daarom is de keuze van de techniek afhankelijk van de situatie. Deze technieken kunnen ook gebruikt worden voor het vinden van afwijkingen in de planning, de duur en gerelateerde oorzaken.

De duur van een taak wordt beïnvloed door verschillende factoren die kunnen worden onderverdeeld in inputfactoren, factoren gerelateerd aan de interne omgeving en exogene factoren. Het tastbare karakter van elementen binnen een BIM-model maakt het gebruik van parametrische schattingen een geschikte methode. Deze methode is gebaseerd op een (statistische) relatie tussen variabelen. Aangezien de kennis en kwantificering van deze invloedsfactoren het mogelijk maakt de taakduur nauwkeuriger te bepalen, stelt deze studie voor om de as-planned en as-built planningen te verrijken met informatie over deze factoren. Bovendien is het voorstel om oorzaakinformatie toe te voegen aan taken die met een begin- of duurvariatie te maken hebben. Vervolgens kan deze informatie worden gebruikt om de oorzaken van een vertraging te achterhalen en risico's te identificeren.

Voor het ontwikkelen van het systeem is gebruik gemaakt van het V-model voor softwareontwikkeling dat bestaat uit een ontwikkelingsgedeelte en een testgedeelte. Eisen en specificaties zijn verzameld voor verschillende gebruiksscenario's (voor planners, projectmanagers en portfoliomanagers) om taakduur en eventueel ander gebruik van data te bepalen. Deze studie stelt het gebruik van een Plan, Do and Measure, Analyse and Improve, en (Re)use cyclus voor die continue verbetering mogelijk maakt. In de eerste fase wordt een op BIM gebaseerde as-planned planning gemaakt. In de tweede fase wordt de voortgang van de bouw bijgehouden. Dit resulteert in meerdere op BIM gebaseerde as-built planningen. Voor de opslag van de gegevens zijn SQL- en NoSQL-databases bestudeerd. SQL-databases zijn over het algemeen geschikt voor gestructureerde data, terwijl NoSQL voornamelijk wordt geprefereerd voor ongestructureerde data. Voor het 'proof of concept' is een relationeel databasemodel ontwikkeld in SQL. Een veel-op-een relatie tussen taken en BIM-elementen blijkt fundamenteel te zijn voor de mogelijkheid om gegevens over de duur van taken te kunnen hergebruiken, evenals het gebruiken van classificaties. BIM gebaseerde planningsdata die gegevens omtrent de BIM-elementen bevat, kan worden geëxporteerd als een Excel-bestand met behulp van Synchro 4D (BIM 4D-tool). Vervolgens kan een Python-script, dat in deze studie is ontwikkeld, worden gebruikt om de gegevens uit Excel te exporteren, transformeren en in te laden in de SQL-database.

De gegevens worden opgehaald met behulp van zoekopdrachten/query's nadat deze zijn opgeslagen. Over het algemeen kunnen gegevens in databases worden geanalyseerd met behulp van query's, dashboards en datamining technieken. In deze studie is de SQL-database verbonden met PowerBI, welke dient als een 'tijdelijke' gebruikersinterface. In PowerBI kunnen de gegevens op verschillende manieren worden doorzocht en gevisualiseerd. Vervolgens worden taakduurschattingen op drie manieren bepaald: 1) de duur kan direct worden bepaald met de zoekopdracht/query van een specifiek taakdeel voor een bepaald element, 2) een verhoudingsgetal kan worden afgeleid uit een bepaalde set van taken gerelateerd aan elementen, en 3) data-analyse kan worden uitgevoerd om een voorspellingsmodel af te leiden dat op basis van verschillende factoren een schatting kan maken over de taakduur.

Het systeem als onderdeel van de voorgestelde verbeteringscyclus is gevalideerd aan de hand van een fictieve casestudie. Deze casestudie geeft inzicht in de verschillende manieren waarop data kan worden opgevraagd, geanalyseerd en (her)gebruikt. Uit de casestudie blijkt dat het systeem relatief toegankelijk en bruikbaar is. Echter, zijn aparte gebruikersinterfaces nodig om het systeem volledig te benutten. De betrouwbaarheid en nauwkeurigheid van de schattingen van de taakduur kunnen niet voldoende worden gevalideerd omdat er geen gebruik is gemaakt van daadwerkelijke projectdata. In het bijzonder moet de bruikbaarheid van voorspellingsmodellen gerelateerd aan taken, elementen en invloedsfactoren verder worden onderzocht.

Over het algemeen is dit onderzoek erin geslaagd een systeem te ontwikkelen als een ‘proof of concept’ waarmee een productiviteitsdatabase kan worden opgebouwd. Deze database wordt gevoed door op BIM gebaseerde as-planned en as-built planningsdata om de duur van taken te bepalen. Deze studie raadt het gebruik van op BIM gebaseerde planningstechnieken, classificaties en standaarden aan, evenals het gebruik van uniforme definities. Ook wordt aangeraden gegevens te verzamelen over factoren welke invloed hebben op de taakduur tijdens de uitvoering van een project. Verder wordt aanbevolen om op BIM gebaseerde as-planned en as-built planningsdata op te slaan als onderdeel van een database waarbij taken en elementen gebruik maken van een veel-op-een relatie.

Abstract

The ability to exploit knowledge is essential for meeting business objectives, continuous improvement and avoiding the repetition of past mistakes. However, within the AEC (Architecture, Engineering and Construction) industry, the ability to learn from past experiences and projects is found difficult to be achieved. A significant number of construction planners are unable to use scheduling data of past projects. As a result, the overall scheduling process is primarily based on intuition and personal experiences, rather than well-founded figures. Closely intertwined is the current lack of systematic feedback towards the planner, which prevents the continuous improvement of construction scheduling. Building Information Modelling (BIM) can become the basis of information and knowledge distribution. When the elements of the BIM model are used as the foundation of creating a construction schedule, a so-called BIM-based schedule can be created.

This study focuses on the capture, storage, analysis and reuse of BIM-based as-planned and as-built scheduling data for the estimation of task duration. The objective of this research is to develop a system that is reliable, accurate, accessible and effectively usable for the estimation of task durations. To enable reliable and accurate estimations, this study proposes to enrich task data with information on factors that influence task duration. For this proof of concept, a relational database model is developed in SQL that allows storing BIM-based scheduling data. Subsequently, mainly key queries from different use case perspectives (planners, project managers and portfolio managers) are tested on a fictitious case project related to the pre-construction phase, construction phase and post-completion phase. As a result, the developed system enables to store and utilize BIM-based scheduling data to estimate task duration, although mostly at a fundamental database and process level without user interfaces.

Keywords: BIM 4D, BIM-based scheduling, Knowledge management, Information retrieval, predictive planning, delay analysis, databases, SQL

Table of contents

(page intentionally left blank)

List of abbreviations

ABS	Activity breakdown structure
AOA	Activity on arrow
AON	Activity on node
BIM 4D	Linking of the BIM model with schedule related information
BIM	Building Information Model
CBR	Case-based Reasoning
CP	Critical Path
CPM	Critical Path Method
EBS	Element breakdown structure
ERD	Entity Relationship Diagram
EVM	Earned Value Method
GUI	Graphical user interface
IDM	Information Delivery Manual
IFC	Industry Foundation Classes
KM	Knowledge Management
LOB	Line of Balance
LOD	Level of Detail
LODt	Level of Development
LPS	Last Planner System
NoSQL	Not only SQL
OBS	Object breakdown structure
ODBC	Open DataBase Connectivity
PCPSP	Resource-Constrained Project Scheduling
RDBMS	Relational database management system
SBS	System breakdown structure
SQL	Structured Query Language
UI	User interface
WBS	Work breakdown structure

List of figures

List of tables

(page intentionally left blank)

1. Introduction

The success of a project as part of the Architecture, Engineering and Construction (AEC) industry, as for any project, is measured against overall objectives. The factors that define the success of a project are commonly described by the golden triangle of time, cost, and quality. The management of a project compromises four functions: planning, organizing, control and leadership. All the major schools of thought view planning as the ‘core competence’ of construction management and one of the main factors of competitiveness within the construction industry (Baldwin & Bordoli, 2014; Jaśkowski & Sobotka, 2006). The accuracy of the project schedule is important since delays would lead to the incapability of achieving the schedule objectives and tend to result in cost overruns, client dissatisfaction, and other consequent problems (Hwang, Zhao, & Ng, 2013).

Berlin’s Brandenburg Airport and the California bullet train between San Diego and San Francisco are well-known examples of projects that faced severe schedule delays and cost overruns (Grushka-Cockayne, 2020). Another example is the Sea Lock IJmuiden project executed by BAM Infra (the graduation company) which is currently delayed with two years and is significantly over budget (Zoelen, 2019). According to Flyvbjerg et al. (2003), nine out of ten large-scale infrastructure projects face significant time delays and cost overruns. Moreover, an annual report by the EIB (Economisch Instituut voor de Bouw) indicated that between 2015-2019, infrastructure projects within the Netherlands were delayed between 15 to 20 percent (Groot, 2019).

A couple of decades ago it was found that humans tend to suffer from what is called the planning fallacy. Within its context, decisions are based on delusional optimism rather than on a rational weighting of gains, losses, and probabilities. Humans tend to overpromise and underdeliver by offering unrealistic forecasts of objectives and overlook the potential of mistakes and miscalculations (Lovallo & Kahneman, 2003; Grushka-Cockayne, 2020). No matter how detailed, the scenarios used in planning are generally found to be inadequate. According to Lovallo & Kahneman (2003), there is one simple reason: any complex project is subject to countless problems that are beyond the reach of the human imagination to be all foreseen from the start. As a result, the probability of things going wrong can be severely underestimated. These costs overrun or time extensions happen due to many reasons, such as design changes or errors, improper planning, economic conditions, resource availability and performance of project parties (Al Hammadi & Nawab, 2016).

A formal way to improve the reliability of forecasts would be the application of objective forecasting methods such as the use of analogous cases. Here, the outcome of a planned action is based on actual outcomes and experiences of similar past projects. However, within the AEC industry, the ability to learn from experiences and projects is found to be difficult to achieve (Tan, et al., 2010). This difficulty is closely related to the fragmented and predominantly project-based nature of the industry. As a result, knowledge cannot be effectively transferred leading to unnecessary reinventions, errors, and time waste (Wiewiora, Trigunarsyah, Murphy, & Liang, 2009). The importance of knowledge is emphasised by Tan et al. (2010) stating that “an organisation’s competitive advantage lies in the knowledge residing in the heads of its employees and the capability to harness the knowledge for meeting its business objectives, for continuous improvement and for avoiding the repetition of past mistakes". Therefore, the need for a systematic and organised attempt to store and use knowledge is recognized, also referred to as knowledge management (KM).

In the case of a planner, the possibility of reusing project scheduling knowledge and data is not self-evident either. It turns out that a significant number is unable to use scheduling data of past projects, as these records were either non-existent or inadequate (Winch & Kelsey, 2005). The limited and unsystematic transfer of knowledge within and between projects prevents adequate reuse and the opportunity of continuous improvement on these schedules. Not surprisingly, it was found that the overall scheduling process is primarily based on intuition and personal experiences, rather than well-founded figures (Büchmann-Slorup & Andersson, 2010).

In the AEC industry, valuable information is often lost because the information is still predominantly handed over in the form of drawings, as physical printed plots or in a limited digital format. The use of BIM (Building Information Models) allows information to be digitally stored, maintained and exchanged. As a result, the digital information can be reused, and laborious and error-prone work can potentially be avoided (Borrmann, König, Koch, & Beetz, 2018). BIM tools can utilize 3D models to create so-called 4D BIM models. 4D BIM involves linking 3D objects (elements) from a BIM model with project schedule information, including the project schedule and phasing, resources and quantities (Mubarak, 2015). The BIM 4D model allows for the virtual simulation of the construction. The fundamental idea of BIM-based scheduling is not to just link the BIM model and the schedule, but to use the BIM model as the basis for creating a schedule.

To control and manage construction projects, contractors typically require access to as-built schedule information. The timely and accurate on-site progress tracking can enhance the coordination and communication among participants (Tserng, Ho, & Jan, 2013). Also, it can bring awareness on specific issues that allow practitioners to take appropriate project control decisions on time (Yang & Teng, 2017). However, the current practice of progress tracking is often inconsistent, time-consuming, costly, and prone to errors (Son, Kim, & Cho, 2017). For instance, analysing the deviations between the as-built performance and construction plan (as-planned) primarily relies on the experience of on-site inspectors which makes it subjective and often error-prone (Yang & Teng, 2017). Some promising BIM-based progress monitoring techniques aim to facilitate the automated comparison between the actual and planned progress and indicate deviations. Studies show that use of data collection technologies such as radio-frequency identification (RFID) sensors, digital cameras, and laser scanners in combination with BIM can improve the efficiency of collecting as-built data (Son, Kim, & Cho, 2017). Subsequently, these techniques can be used to update the 4D BIM as-built model.

According to Deutsch (2015), learning to capture, analyse, and apply data is how many of us will take BIM to the next level. The shift to a more data-driven approach offers unprecedented opportunities for improvement; for instance, by answering a factual question, exploring relationships, discovering patterns, making a case for a decision, etc. Using BIM as a database yields an enormous number of benefits as this data can become the basis of information and knowledge distribution. As such, the data on experiences -and past projects- is allowed to be used as a searchable database (Deutsch, 2015). As part of construction scheduling, the use of historic planning data on projects' initial forecasted completion dates and the as-built information can be used to establish accuracy estimates. Subsequently, these accuracy estimates can be utilized as part of a data-driven approach to forecast, for instance, project completion dates and risks of delay (Grushka-Cockayne, 2020).

1. Problem definition

The importance of effective scheduling has become more evident in the last decades as the construction industry has become increasingly complex (Wang & Azar, 2019). The existence of a variety of problems related to the practices surrounding construction scheduling is recognized. A major problem is the delay of construction projects. Part of the delays are due to faults in the construction schedules. In this study, the problems related to construction scheduling are narrowed down to a selection of three problems: scheduling inefficiency, problems related to progress tracking, and errors and miscalculations. These individual problems are described below.

1.1. Scheduling inefficiency

While construction scheduling has come a long way, it remains a time-consuming and error-prone task when done manually. Most existing software products calculate the schedule using well-known methods, such as the critical path method. However, generally, they require time-consuming manual preparations, such as work breakdown structure (WBS) creation and determination of work-packages and their duration, sequences and estimation of required resources. Moreover, planners are often unable to reuse scheduling data and as a consequence, they have to spend time and resources to make schedules for similar and repetitive projects (Wang & Azar, 2019). As a result, many research efforts have been made to investigate how the process of schedule generation could be improved by automating activity generation, duration estimation and determining sequence logic (Kim et al., 2013).

1.2. Progress tracking

Conventional progress tracking methods depend on extensive manual interaction and rely upon pen and paper-based tracking methods, which is inaccurate, time-consuming, and labour-intensive even for small projects. This approach has been recognized as one of the major problems that cause project delays and cost overruns (Omar & Nehdi, 2016). Real-time progress tracking and monitoring of construction components is a vital part of project management. However, there is currently a lack of systematic evaluation and monitoring of construction projects (Winch & Kelsey, 2005). As a result, there is a scarcity of accurate, consistent and comprehensive data of executed past projects (Song & AbouRizk, 2008). Recently, BIM-based progress monitoring activities are becoming more automated and integrated.

1.3. Errors and miscalculations

The current scheduling process involves significant discrepancies between the overall and detailed levels of scheduling. The use of the Last Planner System during the construction phase generally allows for a well-stipulated approach on a detailed level involving all actors. However, the overall project planning and scheduling process are to a large extent characterized as being an individual endeavour, closely related to the knowledge, professional skills, role and identity of the actor who establishes the schedule (Büchmann-Slorup &

Andersson, 2010). Normally, a construction schedule for exactly one alternative is manually defined and other alternatives are often not investigated, justifications of decisions are not comprehensible and former experiences cannot be used by other planners (Tauscher et al., 2007). Hence, this process fails in the adopting of a more data-driven approach and therefore becomes an easy explanation for the planning fallacy. Furthermore, Winch & Kelsey (2005) found that a systematic review of project planning (and project review in general) was either rare or non-existent. Therefore, without systematic feedback, it is not clear how learning and continuous improvement of the planning process can take place. The three subproblems are visualized below.

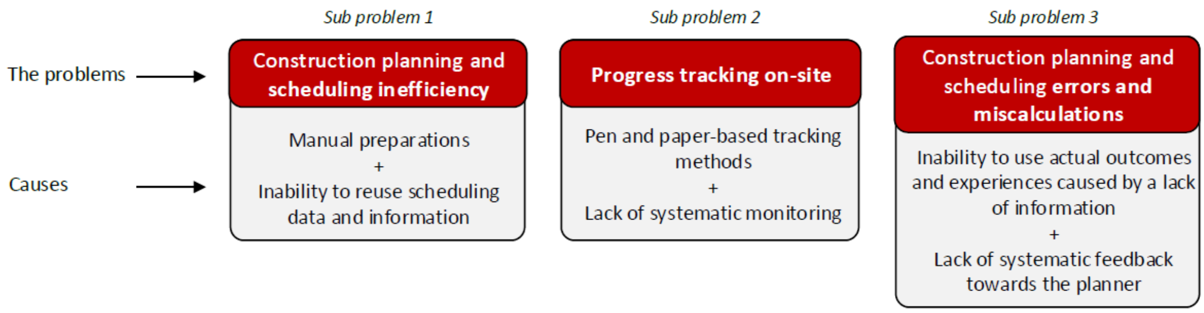


Figure 1: Visualization of the problems

2. Research objectives

Based on the identified problems, it is valuable to research the applicability of BIM-based scheduling in combination with BIM-based progress tracking methods to capture as-planned and as-built data in a systematic manner. Subsequently, it would be valuable to study potential ways to exploit this source of data such that it allows to be effectively reused. This BIM-based source of data could stimulate a more data-driven approach in future construction scheduling practices. For this study, the focus is directed to the utilization of BIM-based data for the primary purpose of estimating task durations for future projects. The overall objective and practical objective are:

Overall objective:
The development of a system that enables to retrieve reliable and accurate task duration scheduling data, as well as to make this data accessible and effectively useable in future planning processes. It hereby aims to optimize the process of construction scheduling.

Practical objective:
Creation of a system allowing a productivity database to be built, fed by BIM-based as-planned and as-built project scheduling data to estimate task durations.

As part of this objective, it is aimed to determine: 1) if historic BIM-based scheduling data can be used for the estimation of task durations for a construction schedule in the pre-construction phase, and 2) if the reliability and accuracy can be enhanced with the reuse of historical BIM-based as-planned and as-built scheduling data. The focus is on the aspect of time, the aspect of cost is not considered within this research.

Reaching the objective of BIM-based estimation of task durations could significantly reduce the time and effort needed to construct a part of the schedule. Furthermore, as an outcome of this study, the accuracy of the schedule could potentially be improved and the factors that cause delays identified.

3. Research questions

Based on the problem definition and research objectives, the following main research question is drafted:

How can BIM-based as-planned and as-built scheduling data be utilized to estimate task duration for future construction projects in a pre-construction phase?

In order to answer the main research question, several sub-research questions are formulated. These questions are divided over a theoretical part and a practical part.

Theory part

1. What does the process of BIM-based scheduling entail and what is the current state of the art?

- 1.1 Which planning and scheduling techniques, practices and methods exist?
- 1.2 How are the different fundamental aspects of a schedule determined (e.g. tasks, sequencing and duration)?
- 1.3 What is the state of the art regarding BIM-based construction scheduling?
- 1.4 What is the process of generating a BIM-based construction schedule?
- 1.5 How do the level of detail of a BIM model relate to the schedule?

2. What is the current state of practice regarding progress tracking and delay analysis?

- 2.1 What is the current state of practice regarding construction progress tracking?
- 2.2 What is the current state of practice regarding delay analysis?

3. How can the duration of a task be estimated?

- 3.1 Which methods can be used to estimate task duration?
- 3.2 What are the factors that influence task duration?
- 3.3 How can element related and unrelated task duration be estimated?
- 3.4 How can uncertainty and risks be coupled with the estimate of task duration?

Practical part

4. How can BIM-based as-planned and as-built scheduling data be captured and stored in a database?

- 4.1 What does the process look like for the capture and storage of BIM-based as-planned and as-built scheduling data?
- 4.2 How can a system be created to capture and store BIM-based as-planned and as-built scheduling data?

5. How can historical BIM-based as-planned and as-built scheduling data be analysed and (re)used for estimating task durations?

5.1 What does the process look like for analysing historical BIM-based as-planned and as-built scheduling data and its (re)use for estimating task durations?
5.2 How can a system be created to analyse historical BIM-based as-planned and as-built scheduling data and its (re)use for estimating task durations?

6. How well does this system perform in estimating task durations using BIM-based as-planned and as-built scheduling data?

4. Research design

As mentioned, this research will consist of two main parts, a theoretical part and a practical part. The theoretical part is related to sub-research questions 1,2 and 3 which are all related to the literature review. The literature review is mainly related to BIM-based scheduling, BIM-based progress tracking, delay analysis and the practice of task duration estimation. The practical part consists of three phases. The first phase is related to the methodology of this research and the proposed system. The second phase is related to the development of a database that aligns with the processes related to the capture, analysis and subsequent reuse of the data. The first two phases are aimed to answer research questions 4 and 5. In the last phase, a case study will be used to validate the system. Finally, conclusions will be drawn in order to answer the main research question and provide recommendations. The complete research design is shown in Figure 2.

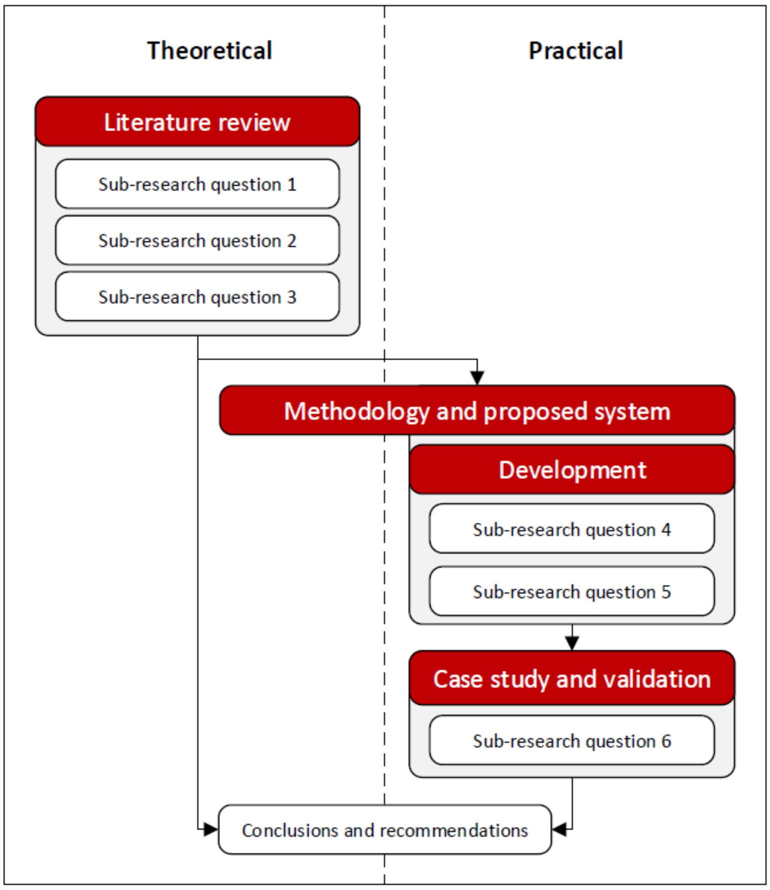


Figure 2: Research design

5. Research relevance

It is recognized that the construction industry is suffering from poor performance and the productivity (value-added per worker) has remained at a steady non-growing level for the last decade (Changali, Mohammad, & Nieuwland, 2015). A variety of problems exist related to the current practice of construction scheduling. As indicated, construction planners are faced with the inability to reuse former experiences and scheduling data, they need to conduct many manual preparations that are error-prone, and they cannot learn from mistakes due to the lack of systematic feedback. BIM as part of information management is crucial for improving effectiveness and efficiency in the AEC industry. Therefore, it is not a surprise that BIM is a graduation theme as part of the master CME (Construction Management and Engineering).

This research is beneficial from a scientific and practical perspective. It contributes to scientific knowledge in relation to construction management, automation in construction, BIM and BIM-based scheduling. More specifically, this research can provide a methodology for the utilization of historic BIM-based scheduling data for task duration estimations. In addition, it explores the usefulness of using as-built data to facilitate in continuous learning, and it can provide an incentive for the use of a more data-driven approach within construction scheduling.

This research is conducted as part of the business unit BAM Infra Multidisciplinary Contracts. This business unit is responsible for large-scale projects that usually cover a large period, have high costs, involve a wide range of risks and have multiple stakeholders. From a practical perspective, this study can improve the process of construction planning within BAM Infra and beyond. Furthermore, insights can be provided for the constructive role that reuse of data can play as part of construction management. Moreover, the relevance for BAM Infra is also shown by their ambition to fully embrace the digital ways of working and becoming the leader of digital construction.

6. Reading guide

This thesis is structured as follows. In the Chapter 2, the literature review is conducted for the theoretical part. In Chapter 3, the methodology and proposed system of this study are described. The practical part of the study is described in Chapter 4 and 5. As part of Chapter 4, different techniques for data storage, data analysis and subsequent reuse are investigated. Based on this chapter, the system is developed. In Chapter 5, the developed system is validated by means of a case study. Finally, in Chapter 6, the conclusions and recommendations of this research are presented.

(page intentionally left blank)

2. Literature review

The literature review is aimed to provide answers to sub-research questions one, two and three. The literature review provides a better understanding of the current problems as stated in Section 1.1 and explores potential solutions by studying state of the art techniques. Furthermore, it provides a foundation for the proposed system as described in Chapter 3. These individual research questions can be broadly divided into three main topics, namely: 1) BIM-based construction scheduling, 2) construction site progress tracking and delay analysis, and 3) task duration.

1. BIM-based scheduling

To get an understanding of BIM-based scheduling, the literature review starts with the current practices of construction planning and scheduling. Thereafter, an overview is presented of the state of the art regarding BIM-based construction scheduling. The last part hereby delves into the process of BIM-based scheduling. This part aims to answer sub-question one "What does the process of BIM-based scheduling entail and what is the current state of the art?".

1.1. Planning and scheduling techniques, practices and methods

The two terms, *planning* and *scheduling* are distinct but inseparable aspects. The process of planning primarily deals with selecting the appropriate policies and procedures to achieve the objectives of the project. "Scheduling converts the project action plans for scope, time cost and quality into an operating timetable." (Moylan, 2002). In short, planning refers to 'what' and 'how' whereas scheduling refers to 'when', graphically presented in Figure 3. Within the context of this research, the term planning is referred to as construction project planning. The Project Management Institute (PMI) defines a project as "a temporary endeavour undertaken to create a unique product, service, or result" (Project Management Institute, 2017, p. 4).

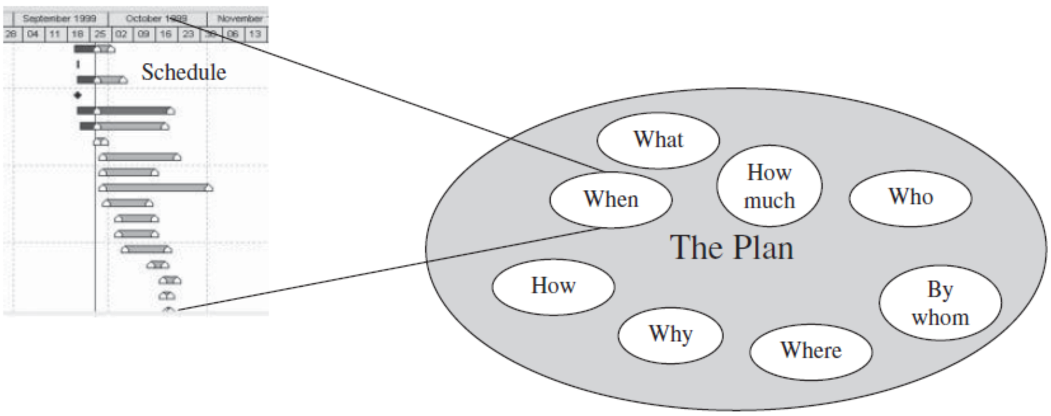


Figure 3: Concept of planning and scheduling (Mubarak, 2015, p. 5)

According to PMI, a Project Management Plan (PMP) is defined as "the document that describes how the project will be executed, monitored, and controlled." (Project Management Institute, 2017, p. 34). Traditionally the metrics of time, cost, scope and quality have been the most important factors in defining project success. The level of detail of the plan depends on several factors such as the purpose and timing of the plan and detail of available information.

Mubarak (2015) presented an overview of reasons for which project schedules are needed based on the perspectives of contractors and owners:

1. Calculate the project completion date;
2. Calculate the start or end of a specific activity;
3. Coordinate among trades and subcontractors, and expose and adjust conflicts;
4. Predict and calculate the cash flow;
5. Improve work efficiency;
6. Serve as an effective project control tool;
7. Evaluate the effect of changes;
8. Prove delay claims.

1.1.1. Bar Charts

Bar charts are the most familiar method for presenting a plan. A Gantt chart, which is a worldwide standard type of bar chart, lists the activities or work breakdown structure components to be performed on the vertical axis and time intervals on the horizontal axis. For each activity, time-scaled bar lines indicate the required time to complete the activity. Interdependencies between tasks can also be shown (Gantt, 1911). Many variations of bar charts are currently present, from minimalistic ones with only the start and end of each activity, to ones where resource and budget numbers are taken into account (Mubarak, 2015). Also, comparisons between the as-planned and as-built schedules are performed, an example is shown in Figure 4.

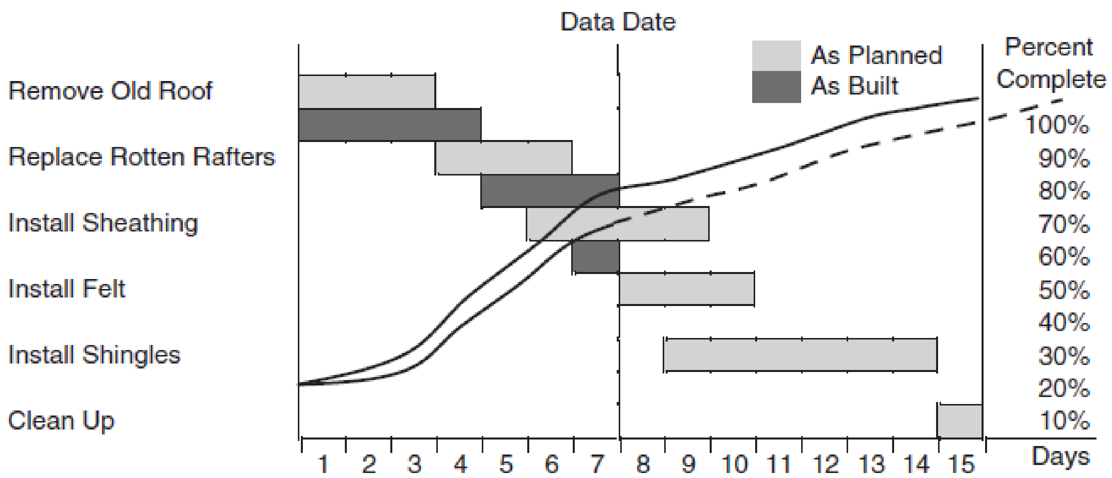


Figure 4: Bar chart example with a comparison between the percentage as-planned activities (thin line) and the as-built (actual) activities (thick line) (Mubarak, 2015, p. 19)

1.1.2. Network Diagrams

A network is a logical and chronological, graphic representation of the activities that enable the reader to understand the logic of the process and the alternative paths. There are two types: arrow networks and node networks. A flow diagram is a generic name for several methods, which includes the activity-on-arrow, precedence diagrams (an advanced form of node diagrams) and linked bar charts. Network analysis requires the construction of a diagram with tasks and the estimated duration of each task. Precedence diagrams became the choice for network scheduling (Baldwin & Bordoli, 2014; Mubarak, 2015).

Node networks, also called activity on node (AON) networks, use nodes and arrows to represent logical relationships (Figure 5). A purposely built-in waiting period between two activities is commonly referred to as a lag. An example of the representation of lag is shown between activity A and D in Figure 5. A lead refers to a negative time gap.

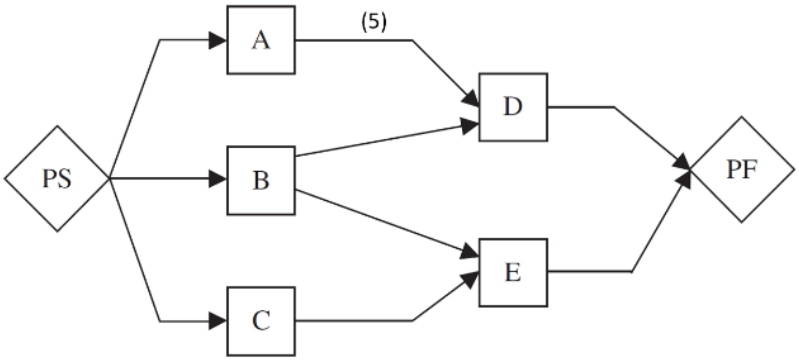


Figure 5: Example of an activity on node network (adapted from Mubarak, 2015, p. 33)

Precedence networks are node networks that allow for the use of four types of relationships: finish-to-start (FS), start-to-start (SS), finish-to-finish (FF), and start-to-finish (SF). The use of those relationships is useful when one deviates from the traditional FS relationship. Both SS and FF are common relationships used in scheduling. SS relationships allow a task to be started when another task starts, this is often in combination with a certain lag (e.g. laying pipes can only begin after a part of the excavation is executed). FF relationships relate to tasks that cannot finish until the other task is finished. Based on the float of tasks in a logical network between the possible starting and finishing times, the early start(ES), late start(LS), early finish (EF) and late finish(LF) can be calculated.

1.1.3. Linked bar charts

Linked bar charts try to combine bar charts with the logic of networks. An example of a linked bar chart is presented in Figure 6.

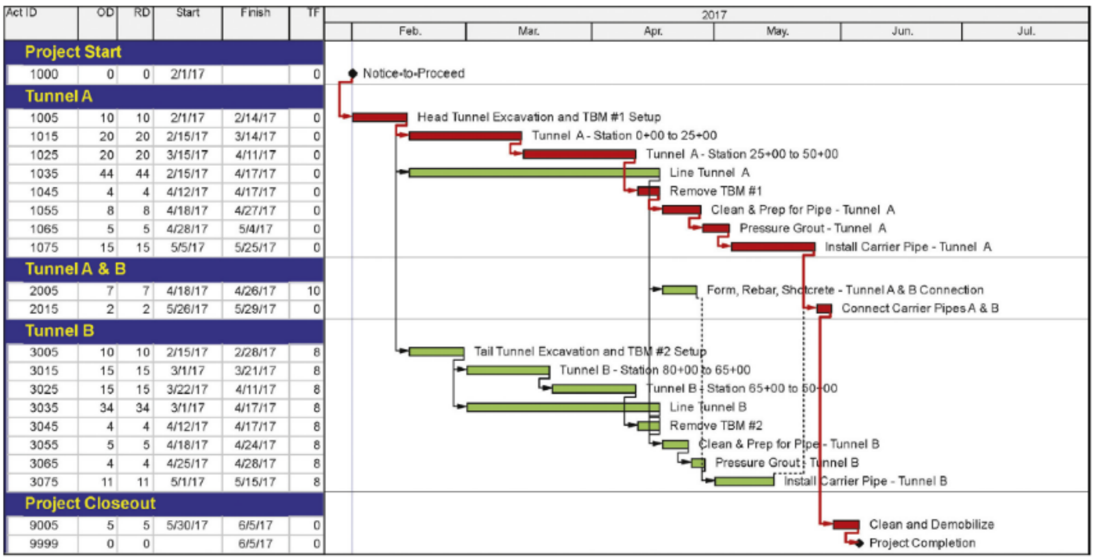


Figure 6: Example of a linked bar chart (red bars are critical tasks) (Nagata et al., 2018, p. 144)

1.1.4. The Critical Path Method (CPM)

Kelley & Walker (1961) developed the critical path method (CPM) as a solution that can appropriately allocate resources at a minimum total cost when the objective is to minimize the total duration. The critical path (CP) is the sequence of dependent activities that form the longest sequence in a project within the network. This is a deterministic model with the focus on the time-cost trade-off. The critical path method is commonly used in conjunction with the Program Evaluation and Review Technique (PERT), a technique initially developed for the US army (Hamilton, 1957). This technique makes use of a probabilistic model with the focus on time, used to manage uncertain activities of a project. It makes use of a duration frequency distribution for each activity.

To explain the concept of the CP more clearly, see Figure 7. Within the figure, the first two activities need to be completed before the last activity can be started, the first activity has some margin called float, whereas the second activity has no float and is therefore part of the critical path. A delay on the critical path has a direct effect on the start of the next activity and therefore delays the project. During a construction project, the critical path can change due to 1) activities being shortened or deleted, or 2) a non- or near-critical path becoming longer when tasks take up more time than originally planned.

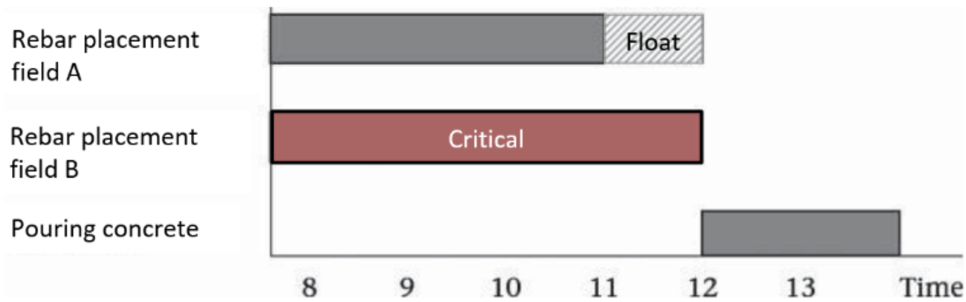


Figure 7: Example Critical Path (adapted from Mubarak, 2015, p. 46)

1.1.5. Other schedule methods and approaches

Furthermore, the commonly used bar charts and network diagrams, other methods exist. The Line of balance (LOB) method, is a graphical resource-driven technique used for scheduling repetitive construction tasks by determining the suitable crew size and number of crews to employ in each activity (Zhang & Zou, 2015).

Another approach, Critical Chain Project Management (CCPM), developed by Goldratt (1997) is frequently presented as a revolutionary new project management concept. It is a resource-based scheduling approach with the emphasis on resource constraints instead of tasks as in the CPM. It specifies the critical chain rather than the critical path. The chain is based on the critical path, but includes resource dependencies. Furthermore, it uses 50% probable activity times and buffers as a measurement tool to control the project schedule (Baldwin & Bordoli, 2014).

The Last Planner System (LPS) is currently a critical element of project production control in Lean construction and it is used to bridge the gap between long-term project planning and short-term execution planning. It includes a cooperative planning process as well as an analysis of incorrect planning (Heigermoser, De Soto, Abbott, & Chuad, 2019). The planning stages of the LPS system are presented in Figure 8.

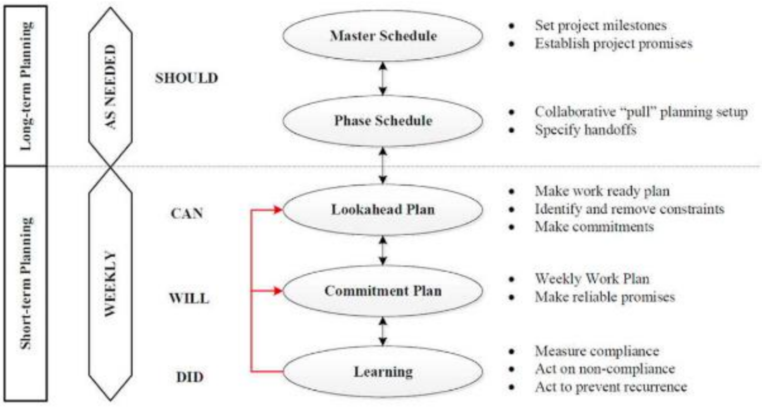


Figure 8: Planning stages of the LPS of Production Control (Heigermoser et al. 2019, p. 247)

The Earned value analysis (EVA) measures performance in terms of time, costs and scope. The technique may be used to assess progress on an individual activity, a group of activities, or the entire project and related costs. The most basic application is performed by quantifying the progress using planned value (PV) and earned value (EV).

1.2. The scheduling process

In this section, the process of construction scheduling is described. Figure 9 provides an overview that shows how the scheduling method, scheduling tool and project information interact to create a project schedule.

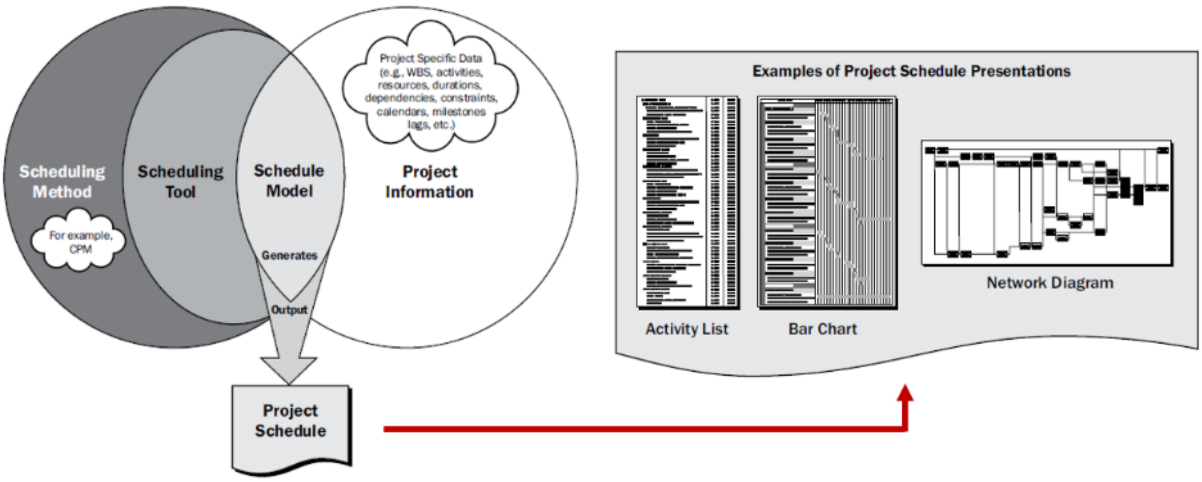


Figure 9: Scheduling overview (Project Management Institute, 2017, p. 176)

To describe the scheduling process, the most important steps to schedule a construction project concerning the execution phase are broken down. As part of the PMP, a schedule management plan needs to be conducted for which the following steps are required: 1) define activities, 2) sequence activities, 3) estimate activity resources, 4) estimate activity durations, 5) activity weightage distribution, 6) develop the schedule, and 7) progress curves development and update (PMI, 2017).

The CPM is predominantly used in construction in terms of contract claims related to delays that impact the project completion time (PMI, 2017, p. 52). The CPM method is used to describe the scheduling process. Table 1 outlines the steps required to develop a CPM schedule according to two sources.

Table 1: Steps required to schedule a project according to different sources (*= fundamental activities)

Step	**Develop CPM** (Mubarak, 2015)	**Develop CPM** (Oberlender, 2014)
1	Break down the project into work activities*	Develop a work breakdown structure (WBS) that identifies work items (activities)
2	Determine task durations*	Prepare a drawing (network diagram) that shows each activity in the order it must be performed to complete the project
3	Determine logical relationships*	Determine the time, cost, and resources required to complete each activity
4	Draw the logic network, and perform the CPM calculations*	Compute the schedule to determine start, finish, and float times
5	Review and analyse the schedule	Analyse costs and resources for the project
6	Implement the schedule	Communicate the results of the plan and schedule
7	Monitor and control the schedule	
8	Revise the database and record feedback	
9	Cost/resource allocation (or loading)	
10	Resource levelling	

The fundamental steps to creating a CPM schedule, according to Mubarak (2015) are described below.

1.2.1. Break down the project into work activities

A project, regardless of its size, should be subdivided into manageable tasks. There is not a correct or incorrect way of doing this and it purely depends on the planner at hand or the project team. There are two schools of thought regarding the breakdown of a project:

1. Restrict the number of tasks for the simplicity of the project: this results in tasks being "major" components (less fine-grained).
2. Break the project down into smaller activities: the approach allows for better control of the schedule during execution. Eventually, the work in progress will be recorded, monitored and reported upon this higher density level schedule (CIOB, 2010).

Subtasks or work packages should be codified to fit with the project cost coding system. This allows for relationships to be identified from an operational perspective down to an individual task level, and the linkages of cost allocation to each task (Lester, 2017). On another note, Mubarak (2015) argues that the perspective to break down a project alters between a cost estimator and a planner. Several factors influence how to break a project down, thus deciding which tasks are incorporated in the work breakdown structure (WBS). Ultimately to have a structure that is acceptable for project control (Lester, 2017). The factors influencing the way to break down a project are the nature of the work/homogeneity, location/floor/segment, size/duration, timing/chronology, responsibility/trade, phase, contractual restrictions, and level of confidence in the duration (Mubarak, 2015).

According to the Project Management Institute (2017) and the ISO 21511, the work breakdown structure (WBS) is "a hierarchical decomposition of the total scope of work to be carried out by the project team to accomplish the project objectives and create the required deliverables.” The NEN-ISO 21511 describes the WBS and other forms such as a product breakdown structure (PBS), organizational breakdown structure (OBS) and risk breakdown structure (RBS) (Koninklijk Nederlands Normalisatie-instituut, 2018). It is possible to integrate multiple work breakdown structures with different purposes. An example of a single WBS is presented in Figure 10.

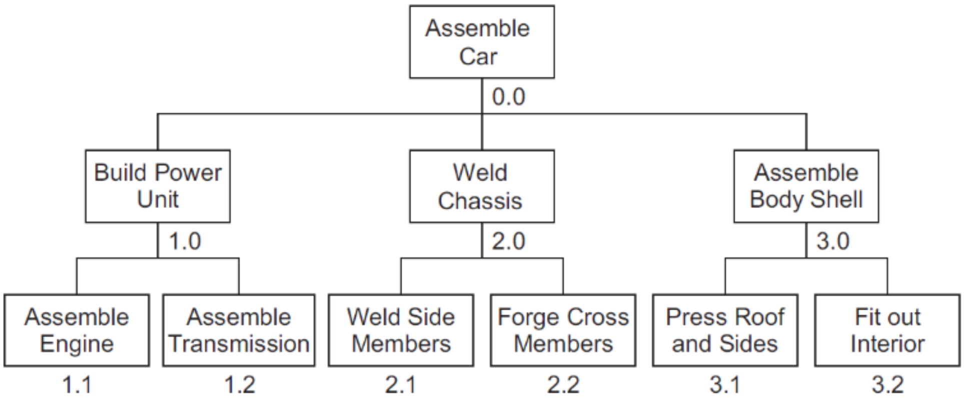


Figure 10: Example of a WBS (Lester, 2017, p. 55)

As part of the breakdown of tasks, Mubarak (2015) describes that every task should have a unique identity (ID), which is in many cases related to the company’s policy. It is recommended that a task title describe *what* (type of work) and *where* (location of the activity).

1.2.2. Determine task durations

There are different techniques for estimating task durations as well as they are depending on the situation at hand. As the estimation of task durations is a significant part of the thesis, it is part of separate sub-research question 3 in Section “2.3 Task duration”.

1.2.3. Determine logical relationships

The logical relationships between the different tasks, and thus the sequence of the tasks depend on the construction strategy and the technical feasibility (Mubarak, 2015; PMI, 2017). For the part of sequencing, the sequence of operations, construction methods, quantities of work and risk are considered (Baldwin & Bordoli, 2014). Two types of logic can be distinguished:

- Hard logic: the technological constraints and the physical relation between two tasks. The start of one task depends on the finish or start of another task (Kallantzis & Lambropoulos, 2004)
- Soft logic: in theory it is possible to execute the task at the same time, however, due to resource restriction (labour, equipment) this is not possible (Tamimi & Diekmann, 1988).

Furthermore, as part of the logic, the type of relationship between the task should be indicated such as start-to-start relationship.

1.2.4. Draw the logic network, and perform the CPM calculations

In this step, the logic of the network and the associated durations are used to perform the CPM calculations. This is generally is a straightforward process for which multiple scheduling programs can be used. The tasks, durations and logic (by indicating the successors and predecessors for every individual task) can be used to automatically perform the CPM calculations. As a result, the calculated finish date of the project, the critical path, and the available float for all noncritical tasks are obtained.

1.2.5. Steps 5 to 8

In the first step, the schedule can be reviewed and analysed to check for incorrect relationships. Part of step 6 and 7, the schedule is implemented for the execution of the project, as well as, monitored and controlled. In step 8, the past experiences related to time and cost are used to revise the database and record feedback. In addition, unusual events and adjustments that were made to the estimates should be recorded. For instance, the drop in productivity level because of extreme weather, equipment failure, etc.

1.2.6. Resource allocation

Step 9 and 10 may be implemented for a comprehensive scheduling approach. *Resource allocation* is the assignment of resources to a task, also called *resource loading*. In project management, all expenses of a construction project can be classified under three categories: labour, equipment, and materials. Labour can relate to ‘salaried staff’ or ‘hourly workers’. Equipment and materials can be classified into: ‘construction equipment and materials’ and ‘installed equipment and materials’. Construction equipment and materials are related to equipment that is needed for the construction process but not permanently installed, for example, bulldozers, cranes, power generators, and construction materials such as formwork and scaffolding. Installed equipment and materials relate to equipment and materials that are permanently installed for the project (Mubarak, 2015).

1.2.7. Resource levelling

As for the last step, resource levelling is executed to minimize the fluctuations in day-to-day resource use throughout the project. This is usually done by shifting noncritical tasks within their available float. For example, one would commonly aim for a similar labour force size every day, since varying a labour force every day is not practical and fluctuations in demand result in a part of the labour force not being utilized. Ultimately, the aim is to improve work efficiency and minimize the costs of the project.

1.3. BIM 4D

The emergence of building information modelling (BIM) has significantly improved building data modelling and information flow between stakeholders involved at all stages, resulting in an increase in efficiency by reducing the manual re-entering of information, which is the case for conventional paper-based workflow methods (Borrmann, König, Koch, & Beetz, 2018; Heigermoser, De Soto, Abbott, & Chuad, 2019). Important about BIM is the ability to represent physical and functional characteristics of building components (elements) that incorporate “knowledge” and “information”. Six important applications of BIM for a contractor are: 1) clash detection, 2) quantity take-off and cost estimation, 3) construction analysis and planning, 4) integration with cost and schedule control and other management functions, 5) offsite fabrication, and 6) verification, guidance, and tracking of construction activities (Eastman et al., 2011). BIM as a visualization, coordination, communication, and design tool utilizes 3D models to create 4D (including time), 5D (including cost), 6D (including sustainability) and 7D (including facility management applications) models. The ability to store information in the BIM model lays the foundation for using the model for any kind of analysis, such as a schedule planning analysis (Liu, Al-Hussein, & Lu, 2015; Borrmann et al., 2018).

1.3.1. Definition 4D BIM

Linking 3D objects from a BIM model with project schedule information, is called 4D BIM. The main idea of BIM-based scheduling is not to just link the BIM model and the schedule for a visualization of the building process, but to use it as a base for creating a schedule (Hartmann et al., 2012). 4D BIM incorporates a BIM model into a project timeline, including the project schedule and phasing, resources and quantities (Mubarak, 2015). 4D BIM provides the planner with the potential to represent the construction phase of the project visually, perform visual checks and detect clashes. It also allows to validate the construction sequence and the organization of site logistics. This results in a virtual simulation of the construction process from its beginning to its end. The individual 3D elements within the 4D simulation allow project stakeholders to view the what, where and how of the construction process from design, procurement and construction schedules (Butkovic et al., 2019). Those abilities of 4D BIM have the potential to minimize the probability of change orders, rework request, and time and cost overruns during a project (Eastman et al., 2011; Mubarak, 2015).

A variety of software packages exist for the process of 4D BIM. Butkovic et al. (2019) describe the functionalities of the different software packages, such as Navisworks, Visual Simulation, Synchro Pro, Powerproject BIM, Vico, iTWO 4.0 and usBIM gantt. Although the functionality of these tools varies, the ability to link 3D components with temporal data through either linking on an individual or group basis remains comparable. Bar charts are used for presenting the schedule. An example of a schedule in Synchro Pro is presented in Figure 11.

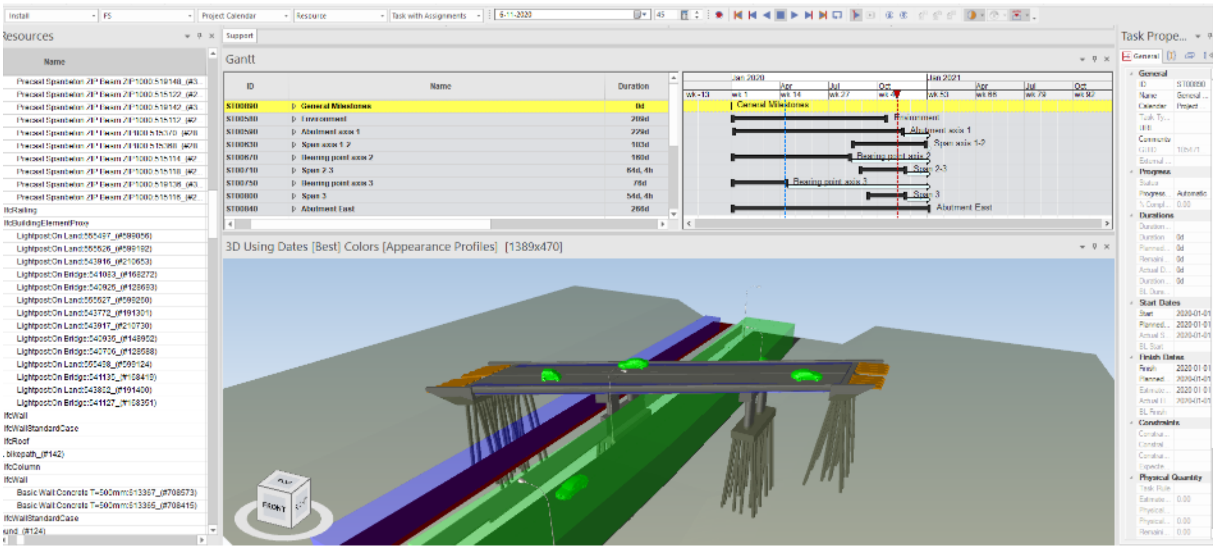


Figure 11: Example BIM 4D using Synchro Pro

The uses and benefits of 4D BIM mostly relate, but are not limited to 1) communicating the plan to the client, 2) as part of the project planning phase, 3) communication between contractors and subcontractors, 4) site logistics and space planning, 5) site safety briefings 6) compare schedules and track the construction progress, and others (Butkovic et al., 2019; Eastman et al., 2011).

1.3.2. Shortcomings of 4D BIM

Despite the reported advancements pertaining to BIM in previous studies, the extended use of BIM has not yet reached its full potential, especially not in relation to project planning and project monitoring. Currently, valuable information is lost because the information is still predominantly handed over in the form of drawings, either as physical printed plots on paper or in a digital but limited format (Liu, Al-Hussein, & Lu, 2015, Borrmann et al., 2018). Consequently, this is leading to the inability to (re)use the information. While BIM models have a huge potential, these models typically merely function as databases of 3D building components and provide only limited information on each component. Rich building information embedded in BIM models is not being fully utilized to facilitate the automatic generation of project schedules, entailing substantial manual work, especially in information exchanges between BIM modelling tools and scheduling tools (Liu et al., 2015).

Furthermore, in most current 4D applications the schedule is developed separately from the building model and the linking and sequencing of the construction processes are realized subsequently during a cumbersome manual procedure (Sigalov & König, 2017). This is also emphasized by Park & Cai (2015) describing that the challenge in current 4D modelling is the discrepancy between the element breakdown structure (EBS) of the BIM model and the work breakdown structure (WBS) of traditional construction schedules. Although generally undertaken manually, the functionalities of linking are evolving by the development of more automated algorithms that use attribute-based data attached to the 3D elements (Butkovic et al., 2019; Sigalov & König, 2017).

1.3.3. Research developments regarding BIM-based scheduling

Recently, there is an increasing number of studies on BIM-based scheduling i.e. BIM 4D. The studies mentioned hereafter are several endeavours that relate mostly to schedule automation in construction where BIM objects are linked with scheduling data to automatically generate construction schedules. De Vries & Harink (2007) presented an automated planning approach that uses an algorithm to determine the topology i.e. the construction order of the 3D CAD model. An external database is used for allocating equipment and labour, and formulas are used for the duration calculations. However, they found that the system lacked construction knowledge. Tauscher et al. (2007) present an approach to generate a construction schedule using IFC. Constraints from the IFC model and CBR (Case-based Reasoning) are used to specify tasks related to elements in the model. Hartmann et al. (2012) present an extended BIM-based scheduling concept, introducing abstract elements and the definition of states of elements. The processes, i.e. the states of elements, are stored in a CBR database. However, the study acknowledged a lack of quantitative information contained in the database. Kim et al. (2013) established a framework using ifcXML for automating the generation of construction schedules by using data (e.g. spatial, geometric, quantity, relationship and material layer set information) stored in BIM models. By using this data, tasks are created (activities associated with each element type have been predefined), sequencing rules applied, and activity production rates are used to compute task duration. Wang, Weng, Wang, & Chen (2014) propose a system that collects, stores, and transfers information among various software packages. It uses an interface system that uses the ability of BIM tools regarding quantity take-offs of materials to support site-level operations simulation, eventually leading to the generation of project schedules. It stores material quantity data within MS Access, runs a stroboscope simulation and uses probabilistic duration data. Liu et al. (2015) present a BIM-based integrated approach for automatic generation of detailed construction schedules under resource constraints, thereby making use of BIM product models with work package information, process simulations, and optimization algorithms. The optimization algorithm is employed to find a combination of the priority of work packages. Wang & Azar (2018) propose a system for BIM-based draft schedule generation in reinforced concrete-framed buildings, providing two schedules for each project, a sequential and an overlapped solution. Wang et al. (2019) present an integrated approach enabling data flow from the information model to the Resource-Constrained Project Scheduling Problem (RCPSP) model for construction scheduling. Also, a work package-based information model is proposed to capture all the required data of the RCPSP.

Furthermore, the study of Sigalov & König (2017) aims at the recognition of process patterns for BIM-based construction schedules. Identified process patterns can be subsequently generalized, supporting the design of process templates. Reusable process templates can significantly improve the efficiency of construction planning. Sigalov & König (2017) point out that automation is basically only possible on component level and therefore requires very detailed schedules.

1.4. Generating a BIM-based schedule

The steps for the generation of a BIM-based schedule are very much in alignment with the generation of a construction project schedule. However, the steps are different depending on the approach that is taken.

Below are the general steps for creating a BIM 4D model (Mubarak, 2015):

1. A BIM model should be created with intelligent objects and appropriate groupings of elements. Thus, the BIM model should be created based on hierarchical groupings. The selection three should be based on the work breakdown structure (WBS) of the BIM components and the project schedule.
2. A project schedule should be developed using separate scheduling software (e.g. MS-project, Primavera) or within the 4D modelling software (e.g. Synchro 4D, Navisworks).
3. Link the project schedule and the building elements by using the selection tree.
4. When linked, one can generate a 4D model showing the sequence of construction over time.

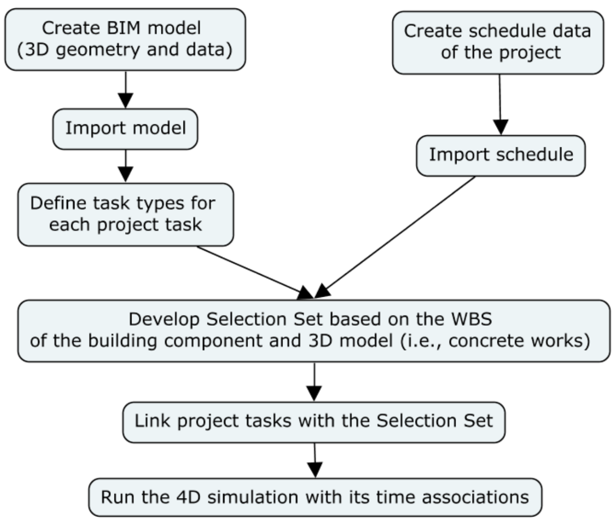


Figure 12: Example of possible workflow for creating a 4D model (adapted from Mubarak, 2015, p. 399)

The related workflow for creating a 4D model is shown in Figure 12. This workflow is based on the idea of developing a schedule separately from the BIM model and linking the two together afterwards. In other words, the schedule is not explicitly based on the BIM model. Oppositely, BIM-based scheduling or model-based scheduling, is the principle of generating tasks based on the 3D elements contained within the BIM model. Therefore, the schedule is directly based on the 3D elements of the BIM model and is thus altering the process of creating a 4D model. Note that the earlier indicated researches towards automated scheduling are schedules generated based on the elements within the BIM model.

1.4.1. The relation between tasks and elements

The relations between tasks and elements are fundamental for BIM-based scheduling. Tasks as part of a schedule or 4D model may represent the construction of one element or several elements; while, the construction of an element may require multiple tasks (Niknam & Karshenas, 2016). Combined, it will result in a many-to-many relationship. The two separate relationships are presented in Figure 13.

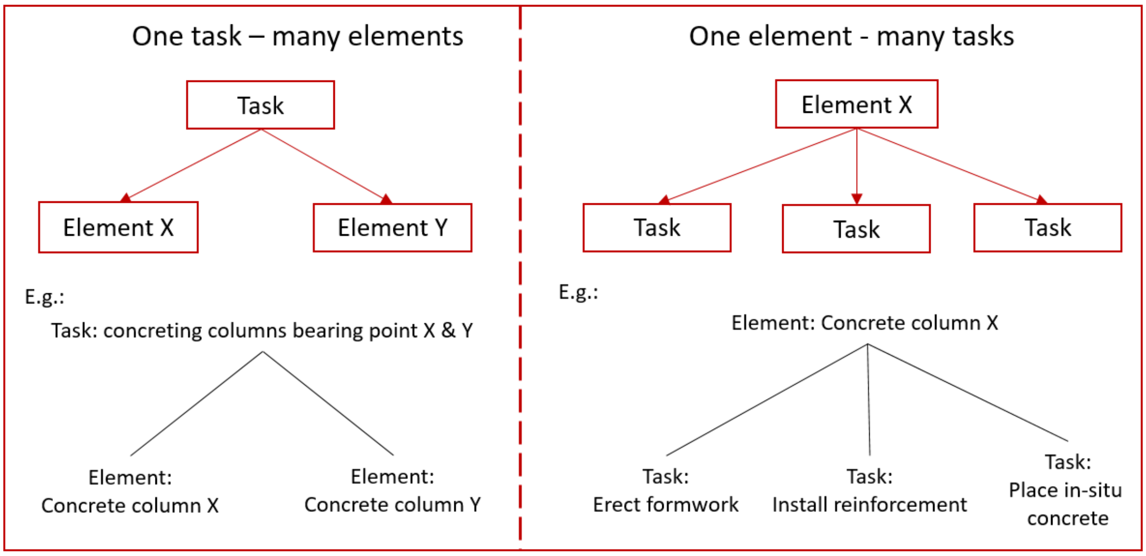


Figure 13: Example of the relationship between elements and tasks

The latter is the fundamental modelling approach where one or more processes (process patterns) or statuses can be linked to a single element (Hartmann et al., 2012; König et al., 2012). Sigalov & König (2017) used a similar approach where a construction task consists of an element, an activity and a resulting state. An example of a process pattern for a wall is shown in Figure 14. There may exist several process patterns for different elements. Alice, a construction planning tool that is able to generate different schedule options for a project, refers to these process patterns as recipes (Roedel, 2018; Heuton, 2020).

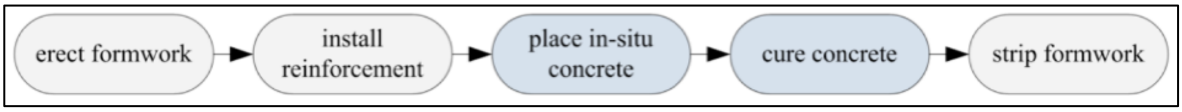


Figure 14: Example process pattern for the construction of a wall (König et al., 2012, p. 4)

1.5. The level of detail of the BIM model in relation to the schedule

1.5.1. BIM Maturity

BIM maturity stages provide a classification for the implementation of BIM. The BIM Maturity Model, developed by the UK BIM Task Group is widely used and defines four discrete levels of BIM (Figure 15). Level 0 describes a working practice based on 2D CAD and the exchange of paper-based drawings. Level 1 employs the concept of 3D, however, most of the design is still realized in 2D. Within level 2, BIM software products are used and various disciplines develop their own models. Coordination happens with periodic sessions and data exchange is realized based on files, however, both are managed within a Common Data Environment (CDE). The last level, level 3, is targeted on the concept of a fully integrated BIM, wherein ISO standards are employed for data exchange, digital models integrated throughout the entire lifecycle and cloud services are used to manage the project data (Borrmann et al., 2018; GCCG, 2011).

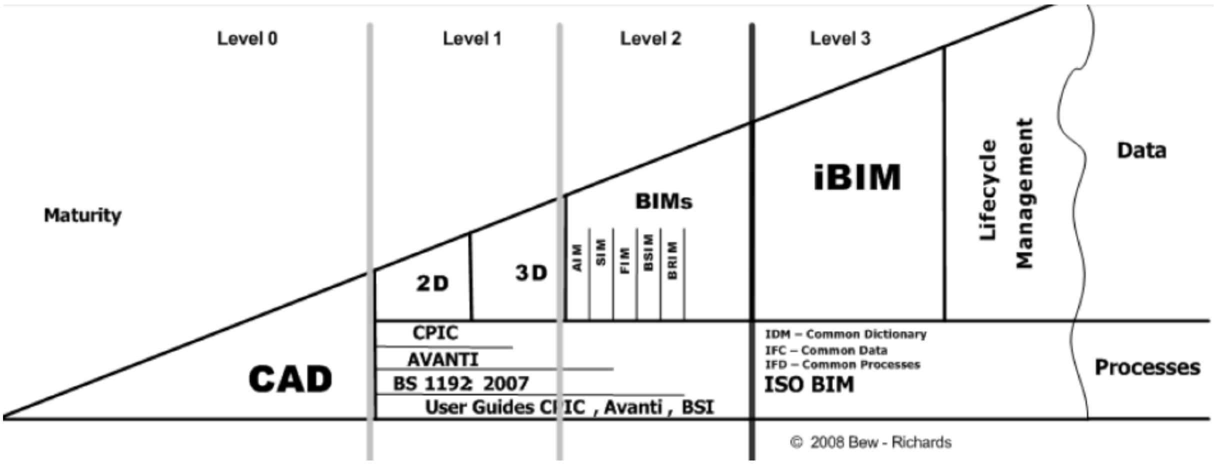


Figure 15: The BIM maturity of the UK BIM Task Group (adapted from GCCG, 2011, p. 16)

1.5.2. BIM Level of Development

Exchange information requirements (EIR) set out managerial, commercial and technical aspects for project information, including the information standard, production methods and procedures implemented by the delivery team (International Organization for Standardization , 2018). The Information Delivery Manual (IDM) can be used to capture and specify processes and information flow during the lifecycle of a facility.

As part of the exchange requirements, the concept of “Level of Development” (LOD) can be used to specify design and planning requirements. The assignment of a LOD to a BIM model or building component allows the recipient of the information to assess its reliability, representing the maturity and accuracy of the model. LOD definitions for infrastructure projects barely exist except for what is stipulated within the UK PAS 1192-2 (Tolmer, Castaing, Diab, & Morand, 2017). An example of different LODs according to the American Institute of Architects is presented in Figure 16.

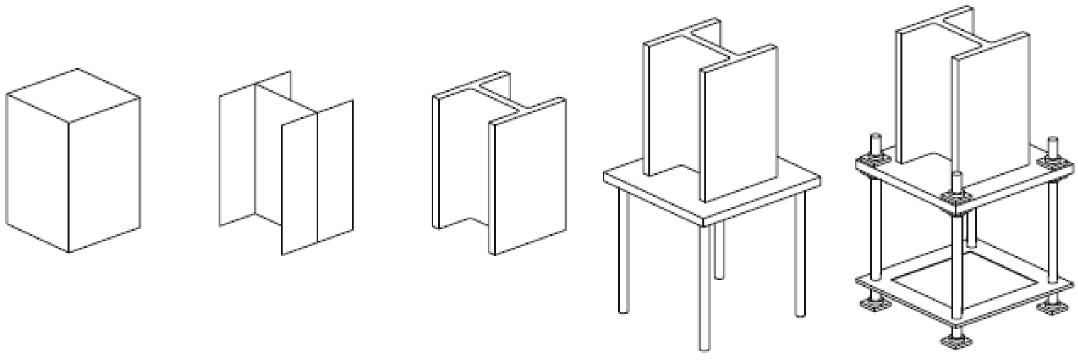


Figure 16: Example of different Levels of Development as defined by American Institute of Architects (Borrmann et al., 2018, p. 137)

1.5.3. LOD and BIM 4D

According to Aouad, Lee, Wu, and Onyenobi (2012) geometric composition of a 3D model needs to be created in such a way that it can support the 4D activities. Additionally, the 3D models must be organized into work components to match the level of detail in the schedule. Butkovic, Heesom, and Oloke (2019) present a framework for the development of dynamic 4D simulations incorporating discrete forms of LOD, able to manage both the graphical level of detail and the temporal level of information.

In relation to the stage of the project, the 4D-LODs can be different. The subdivision or aggregation of elements in relation to tasks provide the ability to better satisfy 4D LOD requirements (Guevremont & Hammad, 2018; Butkovic et al., 2019). This is also emphasized by Eastman, Teicholz, Sacks, and Liston (2011) since the way that architects or engineers organize a model is often insufficient for relating elements to tasks. The framework adapted from Butkovic & Heesom (2017) identifies the primary factors that impact the development of LOD_{4d} simulations (Figure 17).

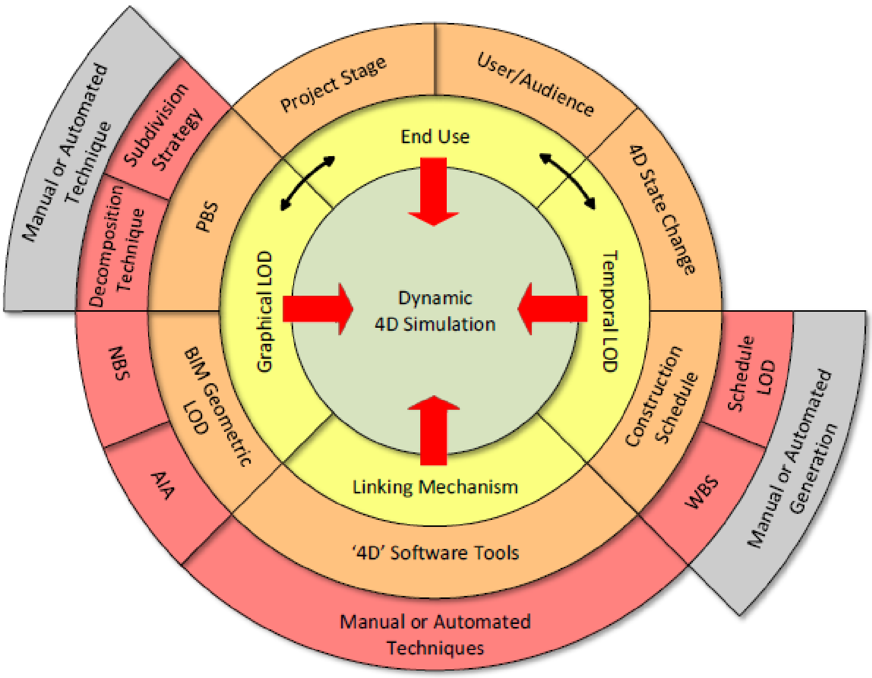


Figure 17: Conceptual Framework for dynamic 4D simulations (Butkovic et al., 2019, p. 265)

The framework shows that the *End Use* of the simulation will impact the LOD_{4d} in relation to the project stage and the use of the simulation. The *Temporal* LOD_{ti} comprises the temporal state change and the level of breakdown of the construction schedule. Future developments in the prospect of automated generation of schedules could impact the LOD_{ti} development. The *Graphical* LOD_g represent the amount of detail contained within the geometry of the BIM model in alignment with standards such as the AIA LOD as presented in Figure 16. Also, the geometric breakdown of the Product Breakdown Structure (PDS) relates to the strategy of the decomposition of the elements contained within the BIM model. The *Linking Mechanism* relates to the technology that provides the ability to link the graphical data and scheduling information (Butkovic et al., 2019). Butkovic et al. (2019) proposed another framework supporting the specification of the LOD of 4D simulations where the level of graphical LOD_g and LOD_{ti} are specified. The LOD_g makes part of the Geometric Level of Detail (LOD_{gem}) and Level of Detail of object Granularity (LOD_{gran}) describing how the object should be decomposed for the linking process. The framework provides the ability to specify LOD_{4D} for a specific purpose instead of a “one size fits all” approach many current 4D simulations use (Butkovic et al., 2019).

1.6. Conclusion

The first section of the literature review provides an answer to the first sub-research question regarding the process of BIM-based scheduling and the current state of the art. A variety of scheduling methods can be employed contributing to project success. Bar charts and network diagrams are the methods predominantly used. As part of 4D modelling, BIM-based scheduling hinges on generating tasks based on the 3D elements contained within the BIM model. Unfortunately, BIM models typically provide limited information on each element and usually require substantial manual work especially in the exchange of information between BIM modelling tools and scheduling tools. The use and definition of LOD_g and LOD_{ti} address the challenge of the discrepancy between the element breakdown structure (EBS) of the BIM model and the work breakdown structure (WBS) of traditional construction schedules. The state of the art regarding BIM-based scheduling is primarily directed at the extraction of data from the BIM model and linking it with other data or tools for the automated generation of construction schedules. In this case, the automation is basically only possible on element level and therefore requires very detailed schedules. Moreover, information on equipment and/or labour should be provided for resource-constrained scheduling to take place.

2. Progress tracking and delay analysis

In this section, a literature review is conducted regarding progress tracking and delay analysis. The literature review describes commonly used methods and techniques regarding progress tracking and delay analysis. Furthermore, state of the art concepts of progress tracking and delay analysis are described. As part of progress tracking, productivity measurement methods are described. This part aims to answer sub-question two i.e. “What is the current state of practice regarding progress tracking and delay analysis?”.

2.1. Progress tracking

Progress tracking is an important part of project control, for which schedules can be a useful tool. Once the project is started, it can, for example, deviate from the schedule or the scope can change. The concept of project control may cover different aspects (budget, schedule, quality, etc.). Within this thesis, the focus will be on schedule control (Mubarak, 2015).

The continuous process of project control has four basic functions:

1. Monitoring work progress;
2. Comparing it with the baseline schedule and budget (what both were supposed to be);
3. Finding any discrepancy, determining where they are and the extent of the discrepancy, and analysing them to discover the causes;
4. Taking corrective action to bring the project back on schedule and within budget.

It is always essential to know where you stand in comparison to what you planned (the baseline), which allows the project management team to analyse the deviations and come up with solutions to bring the project back on track. The baseline schedule is usually prepared before the start of the project by the contractor. Generally required as part of the contract is the submission of the as-planned schedule. For this thesis, it is assumed that the as-planned schedule is similar to the baseline schedule.

A project schedule that is properly and periodically updated once the project is started enables to track and measure a project’s progress. A schedule is usually updated according to some time interval, most common are weekly or biweekly updates. The information that is needed to update the schedule falls underneath two categories: past information and future information. Future changes in the schedule are of two types: *logic-driven changes* or *user’s changes*. Logic-driven changes are changes that occur as a result of changed dates or logic. User’s changes are the changes that are made by the planner not related to past events.

The as-built schedule reflects what actually happened, with real start and finish dates, disregarding any schedule logic, and may include actual resources. The as-built schedule is usually developed from project records and others. Actual durations create added value beyond the predicted durations. Actuals, if properly documented, can amongst others make clear why the original duration is wrongly estimated in the first place.

2.1.1. Progress monitoring techniques

The progress monitoring of construction projects is still predominantly conducted using traditional methods that rely on the manual collection of data performed through visual inspections (Navon, 2005). Thereafter, the collected data and status is for instance displayed on a paper-based large bar chart near the construction site (Figure 18).

Figure 18: Planning on the construction site (KSConsult, n.d.)

Lately, more and more advanced progress monitoring techniques are explored by researchers using 3D visualization and sensor-based automatic 3D modelling. BIM 4D enables the representation of the physical process and allows for the visual recognition of deviations when comparing actual versus the planned progress (e.g. Figure 19).

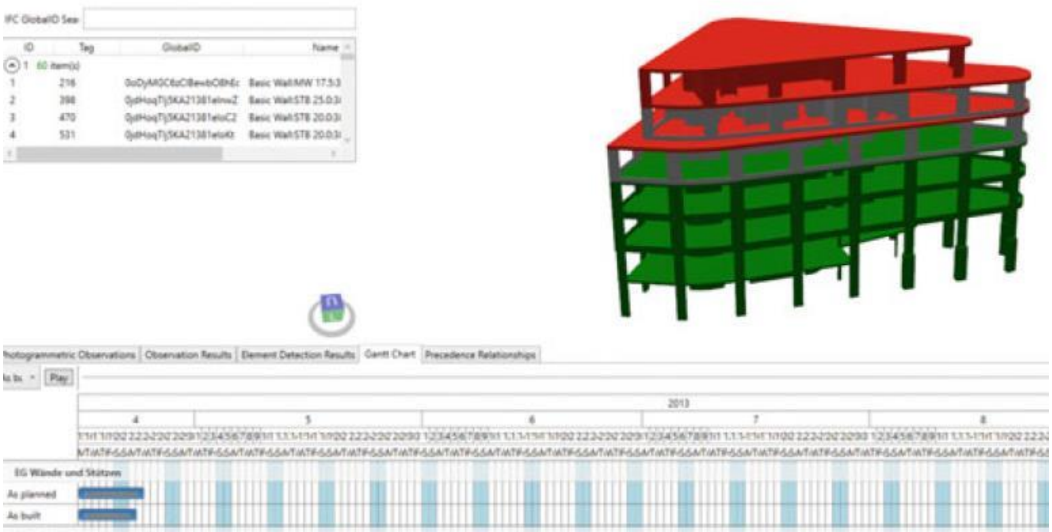


Figure 19: Example of as-planned vs. as-built comparison. Green elements are built ahead of schedule, red elements behind schedule (adapted from Borrmann et al., 2018, p. 474)

It is emphasized that, if tasks are measured differently, these tasks should not be combined. The process of BIM-based progress monitoring requires (automated/manual) updates of the status or percent complete of tasks related to the elements (foundation, walls, etc.) contained within the BIM model. As an example, Synchro 4D allows to connect the status (e.g. formwork placed, pored, formwork removed) to an element and use it to indicate percent complete of an element. Another way to represent element status is with: non-existent and completed. The modelling technique by Huhnt & Enge (2006) uses a predefined ordered set of status variables describing the construction process of an element (Figure 20).

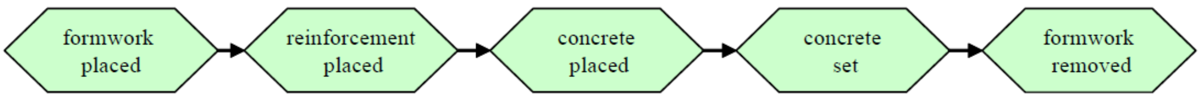


Figure 20: Example of status variable for a foundation (Huhnt & Enge, 2006, p. 551)

Herewith, state of the art techniques regarding progress monitoring are described. Son, Kim, & Cho, (2017) proposed a system that makes use of three-dimensional (3D) point-cloud data from construction sites and a four-dimensional (4D) model that includes an as-planned schedule to automatically update the schedule. Likewise, Turkan, Bosche, Haas, & Haas (2012) make use of laser scanning and describe a system that combines 3D object recognition technology with schedule information into a combined 4D object-oriented progress tracking system. Shahi, Cardona, Haas, & West (2012) developed an activity-based data fusion for automated progress tracking using Ultra Wide Band (UWB) positioning.

Kopsida, Brilakis, & Vela (2015) presented an overview of state of the art techniques in construction monitoring. The different techniques are assessed on certain aspects as shown in Figure 21.

	Mobile AR	Stationary AR	RFID	Laser Scanners	Vision Static Image	Vision Based Reconstruction	Ideal Case
Utility	multiple occasions	multiple occasions	multiple occasions	only spatial data	limited applications	only spatial data	multiple occasions
Time Efficiency	time spent on manual navigation of BIM	time spent on manual registration within registered view	instant information retrieval	time needed for scans	time spent on assigning manual information	time spent on the reconstruction	automatic information retrieval and assignment
Accuracy	AR registration, errors, subjective evaluation of progress	accurate AR registration, subjective evaluation of progress	subjective evaluation of progress	accurate	accurate for simple tasks	variable results, spatial data	accurate
Level of Automation	automated document management & data acquisition, no data analysis	automated document management & data acquisition, no data analysis	automated document management & data acquisition, no data analysis	automated document management & automated data analysis	partially automated data acquisition, automated data analysis	partially automated data acquisition, automated data analysis	fully automated
Required Preparation	minimal set-up required	set the equipment (<1h)	installation and maintenance of tags (>1h)	set the equipment (<1h)	minimal set-up required	minimal set-up required	minimum
Training Requirements	none	none	none	trained personnel for using the scanner	none	trained personnel for the reconstruction	none
Cost	consumer hardware	tracking cameras (£10000)	cost of installation & maintenance	laser scanner (£30000)	consumer hardware	consumer hardware	operates on commercial hardware
Mobility	handheld equipment	large equipment on tripod	handheld equipment	large & heavy equipment	handheld equipment	handheld equipment	handheld equipment

Figure 21: Performance of progress monitoring solutions (Kopsida, Brilakis, & Vela, 2015, p. 8)

The study concluded that none of the solutions achieves the performance of the ideal case and that the applicability of each technology depends on the situation at hand. Although the use of Augmented Reality (AR) is an efficient way to visualise the progress of a construction project, it is limited as it does not perform automated data analysis for progress tracking. Radio-frequency Identification (RFID) technologies allow one to automatically retrieve information by scanning an RFID tag using a smartphone or tablet. Additionally, it can create a link between the virtual model and the physical components in the construction process (Sørensen, Christiansson, & Svidt, 2010; Kopsida et al., 2015). Laser scanning and image processing techniques are deemed promising due to their accuracy; however, the costs are high and their applicability is inefficient for indoor environments. Moreover, the processing of defining objects with the use of algorithms as part of these scans is still under development (Kopsida et al., 2015).

2.1.2. Productivity measurement methods

Productivity measurement methods and/or techniques are of interest for both progress monitoring and task duration estimation. The collection of productivity data can be used for progress tracking and the identification of issues that result in a loss of productivity. Moreover, it can be used as a source to determine productivity rates and to estimate task duration. Moreover, the measurement could potentially improve productivity, see Figure 22 (Ghoddousi & Hosseini, 2012). Productivity is traditionally defined as the ratio of input/out. A more complete explanation can be found in Section "3.2.1 Productivity".

Figure 22: Productivity measurement leading to productivity improvement

There are many existing techniques for measuring productivity such as experience-based, mathematical, statistical models and computer-based applications. Unfortunately, there are also many uncertainties around the accuracy and applicability of these methods (Malisiovas, 2010). Productivity can be measured at different levels (industry, project, or activity/task) (Javed, Pan, Chen, & Zhan, 2018). Obviously, the initial focus in this thesis is on the task level. Different techniques for productivity measurement are described below.

Experience-based measurement

In the early stages of construction engineering, construction productivity was perceived and estimated based on the daily observations on-site. Still, many engineering companies use experience-based methods to measure productivity although they can be highly subjective and influenced by personal prejudice (Malisiovas, 2010). However, these methods can still contribute in ways the other methods cannot. For instance, by providing a deeper understanding of the productivity performance, related problems and opportunities to improve.

Milestones

Another method is measuring productivity using milestones, in which case the completion percentage of the milestones are discussed as part of weekly or biweekly meetings. Although the evaluation of the general work progress takes place, the method does not allow to identify root causes of productivity loss. Moreover, its usefulness is limited for defining on-site productivity as it is not allowing for numerical results (Malisiovas, 2010).

Input versus Output

Thomas, et al. (1990) differentiates between productivity measurement and work-study and distinguished three productivity measurement models: economic models, project-specific models and activity-oriented models. Equation 1 is an example of a project-specific measure.

$$Productivity = \frac{Output}{Labour + Equipment + Materials} \tag{1}$$

As part of activity-oriented models, labour productivity can be measured for instance with equations 2 and 3.

$$Labour\ productivity = \frac{labor\ costs\ or\ work - hours}{output} \tag{2}$$

or

$$Performance\ factor = \frac{estimated\ unit\ rate}{actual\ unit\ rate} \tag{3}$$

In clarification of Equation 2, labour productivity might be defined as the physical progress per person-hour e.g. person-hours per cubic meter of concrete poured (Dozzi & AbouRizk, 2011).

Work study

According to Barnes (1980), work-study is “the systematic study of work systems to find and standardize the least-cost method, determining standard times, and assisting in training in the preferred method. A work-study is sometimes called a time-and-motion study” (Barnes, 1980 as cited in Thomas, et al. 1990). Work-study is different from productivity measurement since productivity measurement generally does not involve characterizing the methods, skills, environment and other factors that influence productivity. Work-study models can be classified as delay, activity and task models.

Delay model

The delay model divides a workday into three parts and records the time spent in each mode (Figure 23). The productive time will be linearly related to the output if the productivity during the productive time stays constant. The output-productive time relationship is even for simple operations very complex. Therefore, the method is difficult to apply to most labour-intensive processes (Thomas et al., 1990).

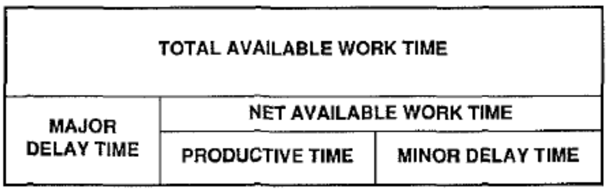


Figure 23: Work study - Delay Model (Bernard et al. 1973 as in Thomas et al., 1990, p. 712)

Activity model (work sampling)

The activity model is based on the work-measurement technique and measures the time engaged in various tasks (Thomas, et al., 1990). Work sampling is a statistical technique based on probability theory that requires a large number of observations. The sample determines the proportion of time in predefined categories, usually in three categories: direct work, support work and delays. Field rating is a similar method where the collected sample is categorized as either ‘working’ or ‘not working’. Figure 24 shows an example of an activity model.

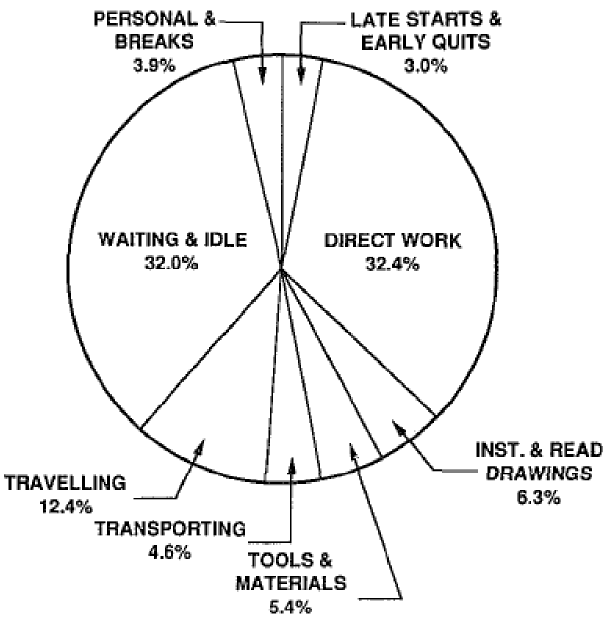


Figure 24: Example of an Activity Model (Thomas et al., 1990, p. 714)

Factor models

Thomas et al. (1990) found productivity measurement and work-study models are generally unsuitable productivity models since they do not model important external and management factors affecting productivity. They argue that the factor model might be the most suited tool to study the causes of productivity variations. If the factors can be quantified that disturb crew performance, one is left with an ideal productivity curve that can be used to forecast future performance (Thomas & Yiakoumis, 1987). Thomas & Yiakoumis (1987) achieved promising results by combining datasets of weather and learning curves to model productivity. Quantification of factors involves the statistical analysis of crew productivity and related factors with (Thomas, et al. 1990):

$$AUR_1 = IUR(q) + \sum_{i=1}^{m} A_i X_i + \sum_{j=1}^{n} F(y)_j \tag{4}$$

Where:

AUR	the actual (or predicted) crew productivity for time period *t;*
IUR	the ideal productivity for broad classifications of work performed under standard conditions. Where productivity improves because of repetition *IUR* is a function of the number of quantities installed *q.* The factors can be expressed as binary (zero-one), integer, or continuous variables.
M	Binary variables or factors
$A_i x_i$	*a_i = a constant representing the increase or decrease in productivity caused by factor i;* x_i = a zero-one variable denoting the presence of the factor
f(y)$_j$	y = factors in submodel *j as an integer or continuous variable to describe weather, crew size, absenteeism, etc.*

An example of the factor model is shown in Figure 25.

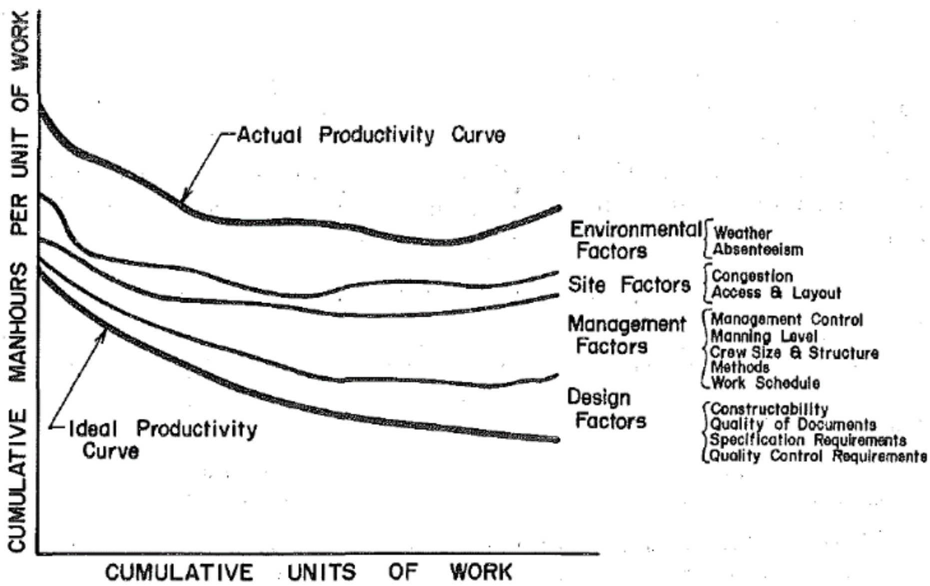


Figure 25: Factor model of Construction Productivity (Thomas & Yiakoumis, 1987, p. 627)

Alternatively, for the mathematical description of factors that influence productivity, fuzzy logic might be used. For instance, Malara, Plebankiewicz, & Juszczyk (2019) used fuzzy logic and proposed a formula to determine the productivity of construction workers. Moreover, Song & AbouRizk (2008) studied labour productivity using historical data and Artificial Neural Networks (ANN) and concluded that a productivity measurement method must be developed first since many contractors cannot take advantage of the productivity modelling approach due to the lack of accurate, consistent, and comprehensive data from past projects. Furthermore, the existence of numerous influencing factors can intensify the complexity of factor measurement, the interaction relationships between factors, and also between factors and resulting productivity (Heravi & Eslamdoost, 2015).

BIM-assisted productivity measurement

Lee, Park, Choi, & Han (2017) studied the use of a BIM-assisted labour productivity measurement method for structural formwork. The authors developed a method for field labour productivity (input versus output method) data acquisition by integrating a 3D model with associated information. The BIM-based productivity and progress information is collected daily into the related database. The process is shown in Figure 26.

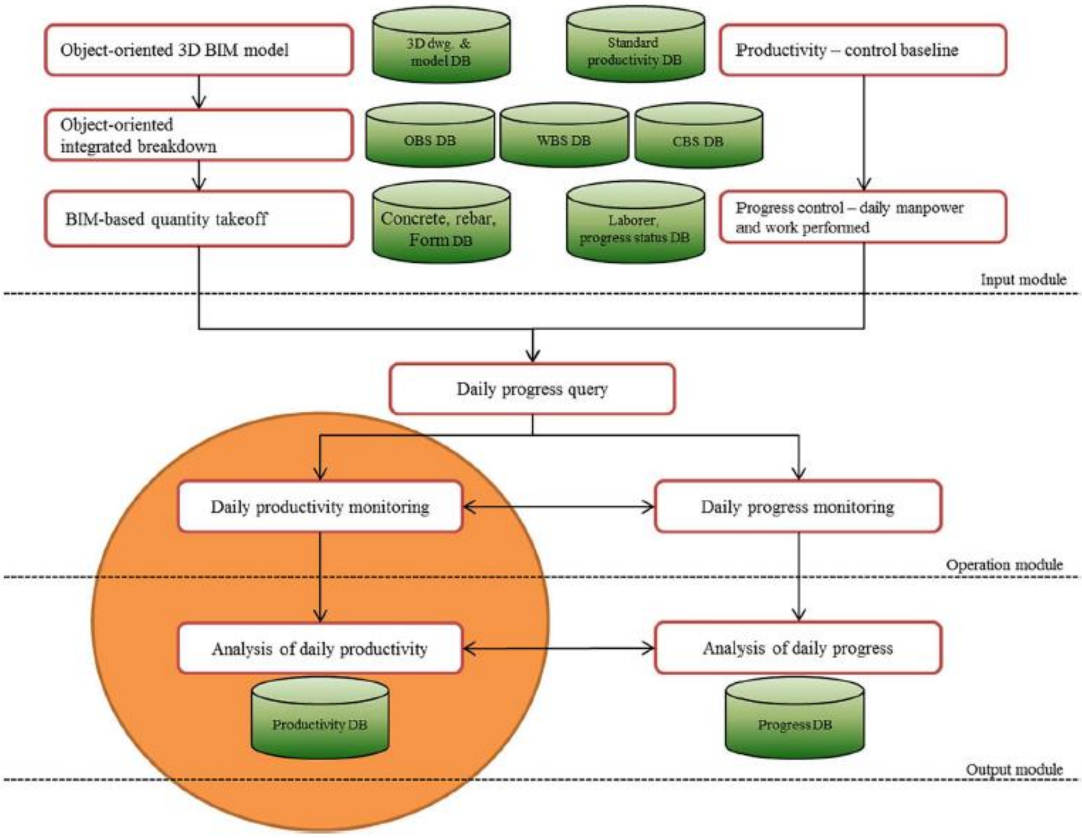


Figure 26: Productivity and progress monitoring process (Lee et al., 2017, p. 124)

The productivity and progress monitoring data are solely utilized for the purpose of labour productivity measurement for the project it is applied to. Moreover, causes analysis was manually performed when significant deviations from the assumed work volume were noticed.

2.2. Delay analysis

"Often, everything does not go according to plan". The failure to meet schedules can result in serious consequences with unprecedented cost implications. Unfortunately, even sophisticated practitioners can find it difficult to identify and quantify delays (Nagata et al., 2018). The components of a task are broken down according to Russel et al. (2014) showing the concept of a delay (Figure 27).

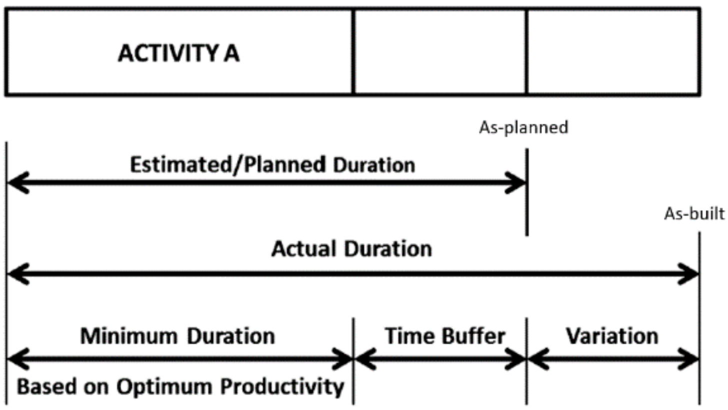


Figure 27: Task time breakdown (Russell et al. 2014, p. 2)

The planned task duration consists of a minimum duration with a time buffer. Adding a time buffer is used to compensate for uncertainty and to absorb the variations during execution. Variability can be defined as the standard deviation from an expected average, however, in this thesis, the definition of Russel et al. (2014) is adopted where task variation is defined as the difference in what was planned and what actually happened. This variation is therefore defined as a task delay.

Delay Analysis Techniques (DATs) give the means to study, analyse and quantify delay events and their impact on the project's duration (Hoda, Ossama, Khaled, & Rania, 2018). Important is that these techniques are directed at measuring the impact of individual delays upon the entire completion time of the project. However, primarily relevant for this study are the causes of individual task delays and their extent.

2.2.1. Delay claims

Project schedules being used as part of project control, are not only used to keep the project on track but also have a significant relation to claim management. A claim is "a request from one contract party to another party for additional compensation, an extension of time, or both". There is a variety of reasons for claims such as differing site conditions, design errors or omissions, changes in the owner's requirements, unusually adverse weather conditions or else. The importance of using CPM schedules and maintaining daily reports, journals and others is very much emphasized as its quality can result in winning or losing a delay claim (Nagata et al., 2018; Mubarak, 2015).

2.2.2. Types of delays

Important for the understanding of delay analysis, are the four basic delay categories:

1. **Critical or noncritical.** Whether a delay is critical or noncritical determines if the project is delayed or not. Delays that affect the project completion or a milestone date, are considered critical (derived from CPM). Delays that do not, are considered noncritical.

2. **Excusable or nonexcusable.** In general, an excusable delay is due to an unforeseeable event beyond the contractor's or the subcontractor's control. Examples of excusable delays (normally based on common general provisions) are:
    - Owner-directed changes
    - Errors and omissions in the plans and specifications
    - Differing site conditions or concealed conditions
    - Force majeure (a term usually used for events that are excusable such as earthquakes, hurricanes, labour strikes, or terrorism (Mubarak, 2015).

   Nonexcusable delays are delays that are within the control of the contractor. Examples are:
    - Late performance of subcontractors
    - Untimely performance by suppliers

3. **Compensable or noncompensable.** For a compensable delay, a contractor is entitled to a time extension or compensation, this is only the case for excusable delays. For a

noncompensable delay, the contractor is not entitled to a time extension or compensation.

4. **Concurrent or nonconcurrent.** A concurrent delay implies two or more separate delays to the 'critical path' occurring within the same time period.

An overview of how the different categories can be viewed is presented in Figure 28. Categorization of the delays is primarily important for delay claims; however, the categorization of individual delays can provide information for future estimates and possible risks that are involved.

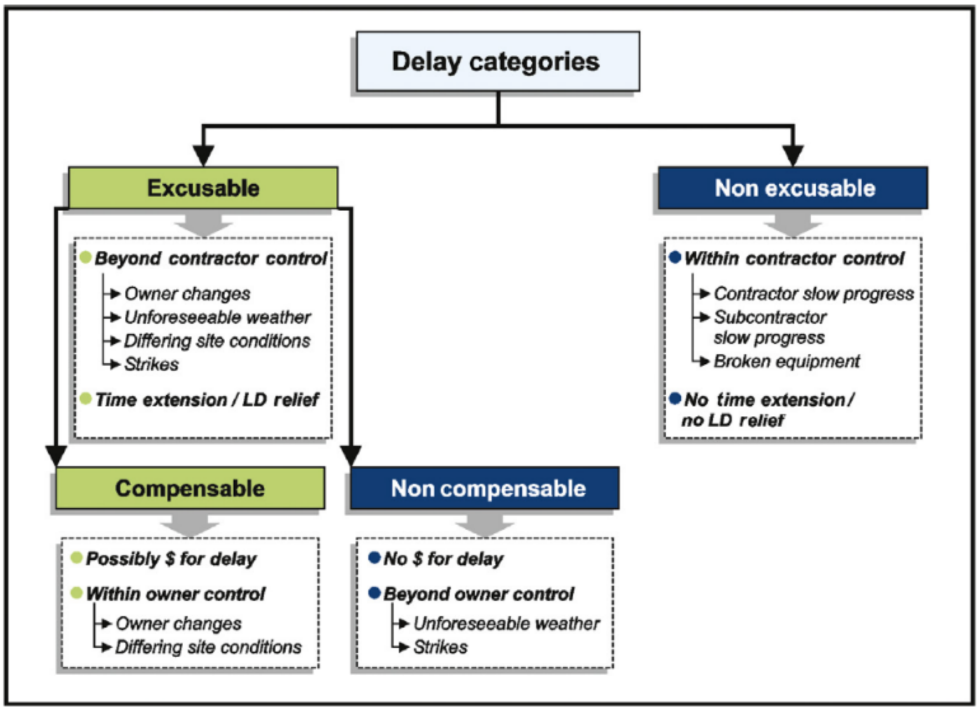


Figure 28: Overview of excusable, nonexcusable, compensable, and nonexcusable delays for a typical construction contract (Nagate et al., 2018, p. 74)

2.2.3. Delay measurement

Complimentary to the methods of progress monitoring and productivity measurement, are delay measurement models. Whereas methods such as work sampling and input/output techniques fail to pinpoint the cause of delays, delay measurement methods provide such data. There are a variety of DATs, the most common are: As-planned versus as-built, impacted as-planned, collapsed as-built, contemporaneous period analysis (window analysis) and Time Impact Analysis (TIA) (Hoda et al., 2018). The Impacted As-planned technique is preferred due to its lack of necessary updates and the TIA because it is widely accepted and provides accurate results (Hoda et al., 2018). Notably, both techniques do not require one to track all delays individually, rather they are used to show/analyse the effect of specific delays or scope changes on the project's completion time. It can be concluded that most DATs are sparsely focused on tracking all delays individually.

Alternatives to the DATs mentioned above, are techniques that collect information on the causes of delays as part of productivity measurement and/or productivity improvement. Dozzi & AbouRizk (2011) describe the use of field surveys and questionnaires to involve the foreman or craftsman in the evaluation and productivity improvement process. A foreman delay survey

(FDS) is a questionnaire filled out by the job foreman at the end of a working day to identify the number of hours lost due to delays. The results of the survey are converted from the person-hours lost into equivalent percentages per cause (Figure 29).

Typical survey form

Problem Causing Area	Person-Hours Lost		
	No. of Hours Lost	No. of Workers	Total Person- Hours
Redoing work (design error or change)			
Redoing work (prefabrication error)			
Redoing work (field error or damage)			
Waiting for materials (warehouse)			
Waiting for materials (vendor furnished)			
Waiting for tools			
Waiting for construction equipment			
Construction equipment breakdown			
Waiting for information			
Waiting for other crews			
Waiting for fellow crew members			
Unexplained or unnecessary move			
Other:			

Comments:

Example results

Problem-Causing Area	P-hs Lost	Percentage
Redoing work (design error or change)	122	2.3
Redoing work (prefabrication error)	24	0.5
Redoing work (field error or damage)	52	1.0
Waiting for materials (warehouse)	33	0.6
Waiting for materials (vendor furnished)	22	0.4
Waiting for tools	12	0.2
Waiting for construction equipment	56	1.1
Construction equipment breakdown	15	0.3
Waiting for information	12	0.2
Waiting for other crews	14	0.3
Waiting for fellow crew members	10	0.2
Unexplained or unnecessary move	20	0.4
Other	70	1.3
Total	462	8.9
Total work in person-hours	5 210	

Figure 29: Foreman Delay Survey (FDS) form and example results (Dozzi & AbouRizk, 2011, p. 8)

Another technique is the craftsman questionnaire (CQ) which is a simple questionnaire that relates to the productivity and motivation of a craftsman (Figure 30).

CQ (part)

Personal data	Check √ the appropriate box for YES or NO, or fill the box with the required information.	
Craft		
Location		
Type of work		
Other		
	YES	NO
Material		
Is material always available when you need it?		
How many hours do you estimate are lost per week due to material not being available?	______h	
Tools		
Are tools always available when needed?		
Are tools in acceptable shape?		
Are tools supplied always the right ones for the job?		
Are there any specific tools in short supply (please name)		
How many hours do you estimate are lost per week due to tools not being available or acceptable for the job?	______h	

Example results

Problem/Cause	P-hs lost per week	Percentage per week
Material not available or poorly located	5.2	13.0
Tools not available or suitable	3.2	8.0
Equipment not available or down for repair	2.0	5.0
Work redone	4.8	12.0
Management interference	2.1	5.3
Other	2.5	6.3
Total	19.8	49.5

Figure 30: Craftsman questionnaire (CQ) form and example results (Dozzi & AbouRizk, 2011, p. 9)

A third method, the Method Productivity Delay Model (MPDM), combines both time and productivity measurement. The method relies on an observer to collect data regarding the cycle time of a leading resource on the operation and to take notes of the nature of the delays. This results in a measure of productivity and identification of sources of delay and their relative contribution to the lack of productivity (Adrian and Boyer, 1976 as cited in Dozzi & AbouRizk, 2011). The MPDM cycle consists of:

1. Identification of the production unit, and the production cycle;
2. Identification of the leading resource;
3. Identification of the types of delay that can be encountered in the process;
4. Data collection;
5. Data processing, model analyse and recommendations.

An example of data that is collected and its results using the MPDM are shown in Figure 31.

MPDM data collection

Operation: Roof truss installation — Observer: SMA
Production unit: One truss — Unit of time: Second

Prod. Cycle	Cycle Time	Enviro. Delay	Equip. Delay	Labour Delay	Mat. Delay	Mngt. Delay	Processing column*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	354						12.83
2	465		x				98.17
3	343						23.83
4	445	x					78.17
5	504				x		137.17
6	470		x				103.17
7	395						28.17
8	345						21.83
9	360						6.83
10	400						33.17
11	460		x				93.17

Example results

Time Variance	Environment	Equipment	Labour	Material	Management
	(1)	(2)	(3)	(4)	(5)
No. of occurrences	2	5	1	1	3
Total added time	141.3	440.8	69.1	137.2	240.4
Probability of occurrence	0.09	0.22	0.04	0.04	0.13
Relative severity	0.17	0.21	0.17	0.33	0.19
Expected percentage of delay	0.01	0.05	0.01	0.01	0.03

Figure 31: MPDM data collection and example results (Dozzi & AbouRizk, 2011, p. 11)

The method can quantify the probability of occurrence, relative severity and the expected percentage of delay for the analysed operation according to the related delay categories (environment, equipment, labour etc.).

2.2.4. Delay causes

The variety of causes of delays have been studied extensively. Most studies rank the causes based on methods such as Relative Importance Index (RII), Frequency Index (FI), Severity Index (SI) and Importance Index (II) that rely on a pool of respondents to quantify the importance of the certain delays (Venkatesh & Venkatesan, 2017). Gündüz, Nielsen, & Özdemir (2013), Chan & Kumaraswamy (2002), Wambeke, Hsiang, & Liu (2011), Venkatesh & Venkatesan (2017), Ahmad, Ayoush, & Al-Alwan (2019) are all studies that conducted a detailed literature review regarding the different causes of delays. The magnitude of the different delay causes indicated is substantial, therefore an overview of the most important causes by the study mentioned above can be found under Appendix 1. The different delay causes provide information on the risks involved affecting task duration. Several authors categorized the different causes of delays (Table 2).

Table 2: Delay categories according to different sources

Gündüz, Nielsen, & Özdemir (2013) and Ahmad, Ayoush, & Al-Alwan (2019)	Wambeke, Hsiang, & Liu (2011)	Venkatesh & Venkatesan (2017)
Consultant-related factors	Prerequisite work	Owner/Consultant/Architect related
Contractor-related factors	Detailed design and work method	Contractor related
Design-related factors	Labour force	External
Equipment-related factors	Tools and Equipment	
Externality-related factors	Materials and components	
Labour-related factors	Work/job site	
Material-related factors	Management/supervision/information flow	
Owner-related factors	Weather/external conditions	
Project-related factors		

Wambeke, Hsiang, & Liu (2011) is by the knowledge of this author, yet the only study that differentiated between task-related variation causes for starting time variation and duration variation. The importance of this differentiation and level of analysis is also emphasized by Ballesteros-Pérez, et al. (2020) that conducted an empirical study and found that construction tasks do not end late on average, but instead, the large variability in task duration is the major factor causing significant project delays. Since the categorization by Wambeke, Hsiang, & Liu (2011) is aimed at task level, this categorization is concluded to be the most accurate categorization of delay causes for task duration analysis and estimation. However, the categorization fails to pinpoint the factor group (e.g. contractor, design-related, etc.).

2.3. Conclusion

The second section of the literature review provides an answer to the second sub-research question regarding the current state of practices of progress tracking and delay analysis. Alongside the monitoring of works, progress tracking can facilitate in finding discrepancies, their extent and discover causes. A variety of progress monitoring techniques can be employed for the creation of updated and as-built schedules. The current state of the art techniques regarding on-site progress tracking are deemed promising. However, it was found that based on different criteria such as time efficiency, accuracy, level of automation, preparation, costs of the technique, none of the current techniques performs ideally. Therefore, the choice of technique depends on the situation at hand.

Productivity measurement methods are employed to measure and to identify the main factors affecting productivity, as well as, to improve and potentially model productivity. The factor model is possibly the most suited tool to study the productivity variations. However, taking advantage of productivity measurement methods pivots on having accurate, consistent, and comprehensive data from past projects which most contractors do not have.

DATs give means to study, analyse and quantify delay events. Most DATs are directed at analysing the effect of delay events on the project’s completion time and providing information on delay claims. Other productivity measurement models can pinpoint the cause of delays and potentially allow to model their probability of occurrence and impact. The ability to model variability is important since it is found that tasks do not end late on average but the variability in task duration is a major factor for project delays.

3. Task duration

The last section of the literature review deals with the duration of tasks as part of a construction schedule. As part of task duration, different methods to estimate task duration, the factors that influence task duration, and the distinction between element- and non-element-related task duration are reviewed. Moreover, the association of risk and uncertainty with task duration is described. This part aims to answer sub-question three "How can the duration of a task be estimated?".

3.1. Methods

3.1.1. Duration estimation at the project level

The methods for estimating task duration can be conceived as a deeper level of estimation in comparison to the overarching estimation of project completion time in its entirety. As the total project completion time is derived from an agglomeration of tasks, some methods and/or factors mentioned below have the potential to be utilized for the estimation of task duration.

Many researchers aimed at predicting the project completing time using a variety of methods. In the early project stage, it is generally performed using a limited number of variables. In the 1960s, the Bromilow's Time-Cost (BTC) model was developed to predict construction project duration as a function of its construction costs. The method received criticism because of its sole reliance on one variable and the inability to use real project cost. Instead of the reliance on cost as an estimator for project duration, other methods are employed to study key factors impacting project completion time. Love, Tse, & Edwards (2005), Hoffman, Thal Jr., Webb, & Weir (2007) and Innocent et al. (2018) all used multiple regression analysis aimed at finding factors that influence or predict project duration. Alternatively, Attal (2010) studied the use of Neural Network Models for the prediction of Highway Construction Cost and Project Duration and found the model to generate accurate results.

In contrast to the above-described methods that generally use a small set of variables, several other methods and algorithms exist that relate to the area of construction schedule optimization (CSO). These methods aim to determine a feasible schedule of tasks to achieve certain predefined objectives, for example, the shortest project duration, lowest cost or highest profit (Zhou, Love, Wang, Teo, & Irani, 2013). Therefore, for CSO methods to be utilized, project information should be of a certain level e.g. tasks, durations, relations, costs and alternatives being specified. Zhou et al. (2013) classified the CSO problem into three methods: mathematical, heuristic and metaheuristic. Mathematical methods relate to the CPM and Integer/Linear programming, whereas heuristic methods relate to systems that rely on past experiences. Metaheuristics relate to methods used for solving combinatorial optimization problems. Most popular within construction is the use of genetic algorithm (GA), ant colony optimization (ACO) and particle swarm optimization (PSO) can be used.

3.1.2. Duration estimation at the task level

A single task part of a bar chart schedule from the case project is shown in Figure 32. As shown, the task consists of an ID, name, duration, start and finish date, and a connection to three 3D elements of the BIM model. Information regarding the methods, techniques and factors used and/or considered for the estimation of the task its duration are not present.

Figure 32: Representation of a single task of the BAM viaduct case project as part of a Bar chart

The estimation of task duration is not easy as pointed out by Winch & Kelsey (2005) referring to the inherent uncertainty of task durations and the associated opportunistic behaviour in estimating the true task durations. Subsequently, different tools and techniques can be used to estimate task durations. According to PMBOK (2013), the following techniques exist:

1. Expert judgment

For this method, the expertise from individuals or groups is used. Individuals' judgments are still heavily applied within the industry due to the uniqueness, complexity, and uncertainty involved in construction projects (Song & AbouRizk, 2008). Winch & Kelsey (2005) found that the development of a feel for task outputs and durations is the knowledge that allows planners to solve planning problems better. Also, within BAM Infra, task durations are usually derived from the knowledge that exists within the company, thus making use of expert judgment. Obviously, the accuracy and reliability of this approach are influenced by personal prejudice and can be highly subjective.

2. Analogous estimating

Analogous estimation relies upon the use of historical data from a similar task. The technique is considered a top-down estimation approach to yield rough estimates. Analogous estimation is primarily adopted when there is a limited amount of detailed information about the project. The method uses a combination of historical information and expert judgement.

3. Parametric estimating

This is an estimating technique for which an algorithm is used to calculate the duration based on historical data and project parameters (Project Management Institute, 2017). A regression model is the most common statistical model for productivity estimation when considering specific factors (Lee, 2017). However, parametric estimation as part of a productivity database is commonly applied by using a single parameter. The duration is then calculated by multiplying the quantity of work by the number of labour hours per unit of work. For instance, this can be done with BIM scheduling software where the expected duration of each task is computed from the quantities of work defined (units, surface areas, volumes, lengths) from objects according to the work type definition and dividing by the standard work rate of the task and crew size (Eastman et al., 2011). Thus, parametric estimation can only be applied when a formula, algorithm or statistical model can be devised. Tangible outputs are well-suited for parametric estimating; however, the method is less suitable for intangible outputs (e.g. judgmental tasks). Based on the literature reviewed as part of BIM-based scheduling, task duration is commonly calculated using a single-point estimate from some sort of productivity database. For instance, a database such as RSMeans data (North America) or GWWkosten (Netherlands) can be used. An example of RSMeans activity data is shown in Figure 33. For item 4700, a crew consisting of 1 carpenter has a daily output of 65,5 meters of Teak, nominal 25.4 x 25.4 in an 8 h working day or uses 0.037 labour-hours (8 h divided by 215 daily output in feet).

Table 2. Sample RS Means Activity Entry

Line number	06 22 13. 45 millwork-standard pattern wood trim-moldings, trim	Crew	Daily output	Labor-hours	Unit
4700	Teak, nominal 25.4 × 25.4 mm (1 × 1 in.)	1 Carpenter	65.5 (215)	0.037	LM (LF)
4800	Nominal 25.4 × 76.2 mm (1 × 3 in.)	1 Carpenter	61.0 (200)	0.040	LM (LF)
4900	Quarter round, stock pine, 6.4 × 6.4 mm (1/4 × 1/4 in.)	1 Carpenter	83.8 (275)	0.029	LM (LF)
4950	19.1 × 19.1 mm (3/4 × 3/4 in.)	1 Carpenter	77.7 (255)	0.031	LM (LF)
5600	Wainscot moldings, 26.8 × 14.3 mm (1-1/18 × 9/16 in.), 0.6 m (2 in.) high, minimum	1 Carpenter	7.1 (76)	0.105	SM (SF)
5700	Maximum	—	6.0 (65)	0.123	SM (SF)

Figure 33: Example of RSMeans activity data (Vereen, Rasdorf, & Hummer, 2016, p. 4)

The uniqueness affects the applicability of this method (using daily output), for instance, Song & AbouRizk (2008) describe: “even though productivity data in terms of workhours per ton are available from historical data, estimators seldom use these numbers to estimate new jobs due to the uniqueness of steel pieces”. In addition, the method has the disadvantage of being very simplistic and not able to take other factors into account (Malisiovas, 2010). Therefore, usually, when estimating task duration, standard work productivities are modified to account for the specific site and job characteristics, for example, done in the form: If a particular condition exists, *then* reduce productivity by *x%* (Hendrickson, Martinelli, & Rehak, 1987). An expert might look for patterns of conditions and immediately synthesize an appropriate productivity figure.

4. Three-point estimating

The single-point duration may be improved by considering uncertainty and risk. Three-point estimates define an approximate range of a task’s duration, by using the most likely(M), optimistic(O) and pessimistic(P) durations. The expected duration is commonly calculated using tE = (t0 + tM + tP) / 3 (triangular distribution) or using tE = (t0 + 4tM + tP) / 6 (beta distribution). Additionally, by using the standard deviation the level of confidence can be determined.

5. Bottom-up estimating

Bottom-up estimating is based on the aggregating of the estimates of the lower-level components of the WBS. When a task duration cannot be estimated, the task is decomposed into more detail. The opposite of the bottom-up estimating approach is the top-down estimating approach. Lester (2017, p. 56) describes this for the part of project cost, however, this method is also applicable to estimate durations as part of a schedule. Figure 34 shows a comparison between bottom-up and top-down estimation.

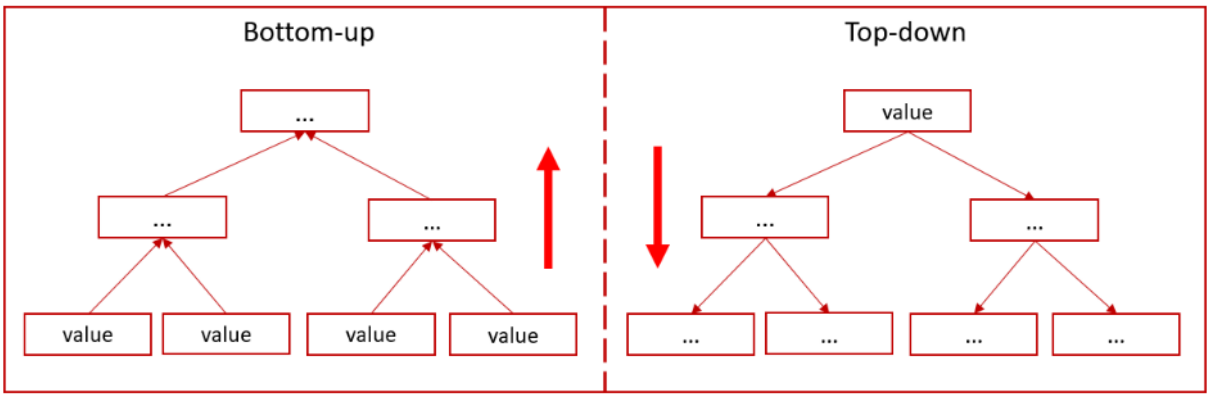


Figure 34: Bottom-up versus top-down estimation (adapted from Lester, 2017, p. 56)

6. Data analysis

Data analysis can be for example an alternatives analysis and reserve analysis. The latter is used to include contingency reserves to account for schedule uncertainty.

7. Group decision making

Estimating the duration using methods such as voting, which is often used in agile-based projects.

3.1.3. Comparison of estimation methods

Most commonly used techniques for time estimation are the deterministic (single-point) and the stochastic (three-points) estimating technique. Since the focus of the study is on the utilization of as-planned and as-built data, subjective methods such as expert judgment are not of interest. Bottom-up estimation can be perceived as a secondary estimation method for when lower estimates are known and are therefore not considered valuable for initial task duration estimates. Methods 2 to 6 rely on the use of historical data, directly or indirectly. Obviously, in this case, the collection of historical data is necessary as well as the definition of project parameters in the case of parametric estimating. Remark that these methods are not mutually exclusive and can often be used in combination with each other. Overall, the choice of method can be influenced by the available information, the stage of the project, etc. Since BIM-based scheduling allows for data extraction of tangible objects (elements of the BIM model with related parameters/attributes), parametric estimation is considered to be a suitable method at task level. Furthermore, this method aligns well in the prospect of automated scheduling for which data on element level is needed for its application.

3.2. Influencing factors

Estimating task durations uses information from the scope of work, required resource types or skill levels, estimated resource quantities, resource calendars and other factors (PMBOK, 2013). Even if the design is identical, there are differences such as the site, location, workforce and execution conditions that influence the project. In short, knowledge of factors that influence task duration and related productivity yield more accurate duration estimates. Moreover, these factors can contribute in the exploration of different schedule scenarios. Since a comprehensive overview of factors is presented that influence task duration, those factors are also applicable to the part of delay analysis. As part of this thesis, the influence of cost on task duration is not considered, however, the cost can certainly have a major impact on the viability of task durations.

3.2.1. Productivity

Factors that influence productivity, and especially factors causing the loss of productivity have been studied extensively. The definition of productivity is defined as the output divided by input. Although widely applied, there is no universally accepted productivity measurement standard for estimating purposes and a single industry measurement would be insufficient. The meaning of productivity for the same task may have a different meaning for different people (Song & AbouRizk, 2008; Thomas et al., 1990; Park, 2006). When multiple tasks are considered, the existence of numerous influencing factors can intensify the complexity of factors and the resulting productivity (Heravi & Eslamdoost, 2015). As pointed out earlier, individuals' judgments are still often used for estimation since the relationship between these factors and the resulting productivity is difficult to quantify. A simple representation of productivity measurement is presented in Figure 35.

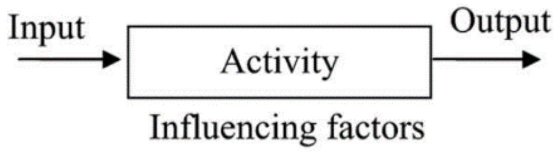


Figure 35: Productivity Measurement (adapted from Song & AbouRizk, 2008, p. 787)

The simplest model of a construction process is its representation as a closed conversion process, here all external factors are held constant and the input and output are known. Obviously, most construction processes are influenced by external factors and therefore subject to an open conversion process as shown in Figure 36 (Drewin as cited in Thomas et al., 1990).

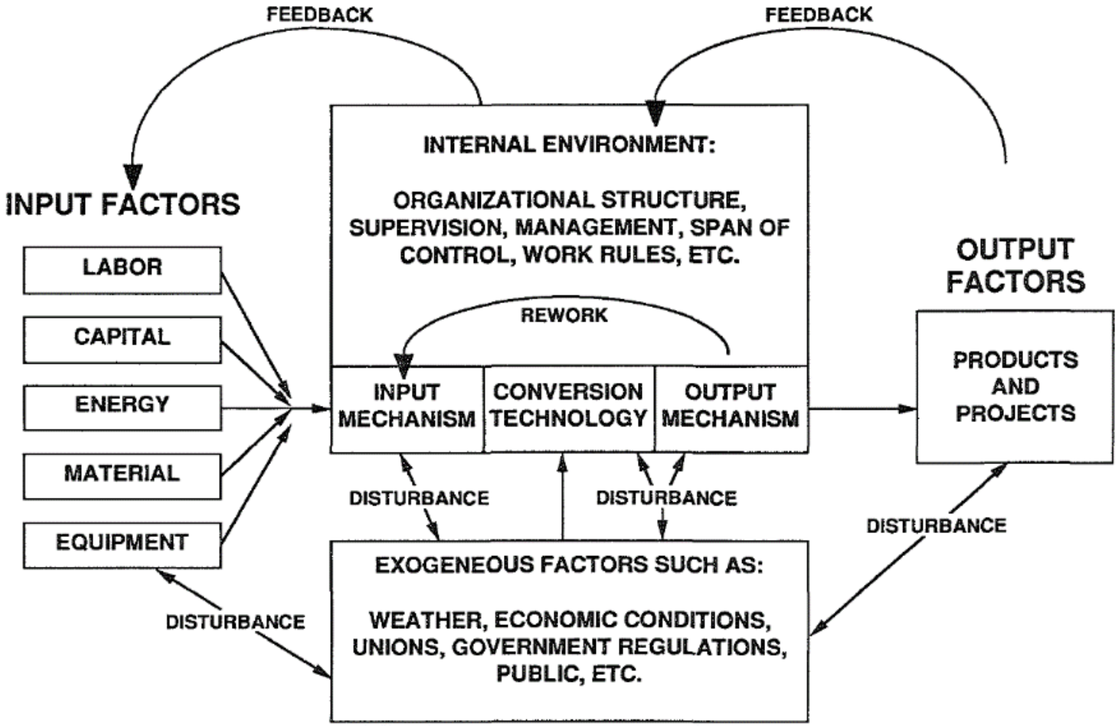


Figure 36: Construction as Open Conversion Process (Drewin, 1985 as cited in Thomas et al., 1990, 711)

Thomas & Zavrski (1999) defined baseline productivity as the optimum productivity unaffected by disruptions, thus being their best performance. However, Park (2006) proposed a model where baseline productivity is affected by both project environment factors and management factors, either positively or negatively. As shown in Figure 37, the baseline productivity can be affected positively (green arrow) or negatively (red arrow). The proposed factors by Park (2006) can be found in Appendix 2.

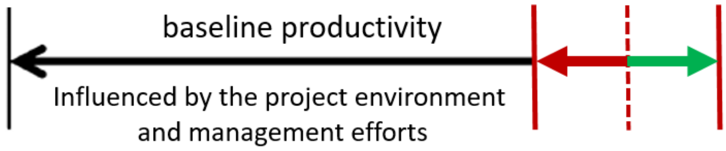


Figure 37: Expected productivity influenced by project environment and management efforts (adapted from Park, 2006)

A vast amount of studies tried to identify the most apparent factors affecting productivity or factors being able to improve productivity (Malisiovas, 2010; Hasan, Baroudi, Elmualim, & Rameezdeen, 2018; Arditi, 1985), as well as labour productivity (Heravi & Eslamdoost, 2015; Song & AbouRizk, 2008). The most important factors identified by those studies can be found in Appendix 2. Below a selection of often mentioned factors categorized according to the Open Conversion Process in Figure 36.

Input factors

Important input factors are labour, materials and equipment and tools. In terms of labour, factors such as labour skill, labour availability, crew size, absenteeism, long work hours and workers fatigue, motivation and communication affect productivity (Park, 2006; Hasan et al., 2018; Arditi, 1985, Heravi & Eslamdoost, 2015; Wambeke et al. 2011). Factors in terms of materials: material availability, prefabrication, standardization (Park, 2006; Hasan et al., 2018; Arditi, 1985, Heravi & Eslamdoost, 2015) but also the location and quality of the materials affect productivity (Wambeke et al. 2011). In terms of equipment: equipment availability, capacity, simplicity and utilization affect productivity (Hasan et al., 2018; Heravi & Eslamdoost, 2015; Wambeke et al. 2011).

Internal environment

As part of the internal environment, the following factors are found to be of importance. Project-related factors such as the site layout and conditions, level of project complexity, quality of drawings and specifications, frequency of change orders and contract type. Another factor is if the project allows for learning curves to take place. Generally, the expected productivity is represented as a constant parameter. However, due to learning curve effects one's productivity is likely to improve because of the familiarity with the process and more efficient application of tools and equipment (Thomas & Yiakoumis, 1987; Arditi, 1985).

Furthermore, factors in relation to organization and management, such as good supervision, coordination and communication, project team experience and turnover, proper pre-project planning, change order management and the use of technology can affect productivity (Park, 2006; Hasan et al., 2018; Arditi, 1985, Heravi & Eslamdoost, 2015).

Exogeneous factors

The weather is indicated in most studies to affect a project in one way or another (Park, 2006; Hasan et al., 2018; Arditi, 1985, Heravi & Eslamdoost, 2015; Thomas & Yiakoumis, 1987). Different components of the weather can influence the execution of tasks and one's productivity such as precipitation, wind, humidity, temperature (hot, cold, frost), and cloudiness. In addition, factors such as economic conditions, local codes, regulations and constraints can influence productivity.

3.3. Element related and unrelated task durations

The selection of a suitable estimation method whilst considering the influential factors allows one to estimate task duration. Since the fundamental principle of BIM-based scheduling relies on the link between tasks and elements contained within the BIM model, it is apparent to distinguish between tasks that are related to elements and tasks that are not related to elements.

3.3.1. Elements being modelled

The ability to link an element to a task allows visualizing the process in 4D, thus allowing for the virtual construction to be simulated (Figure 38).

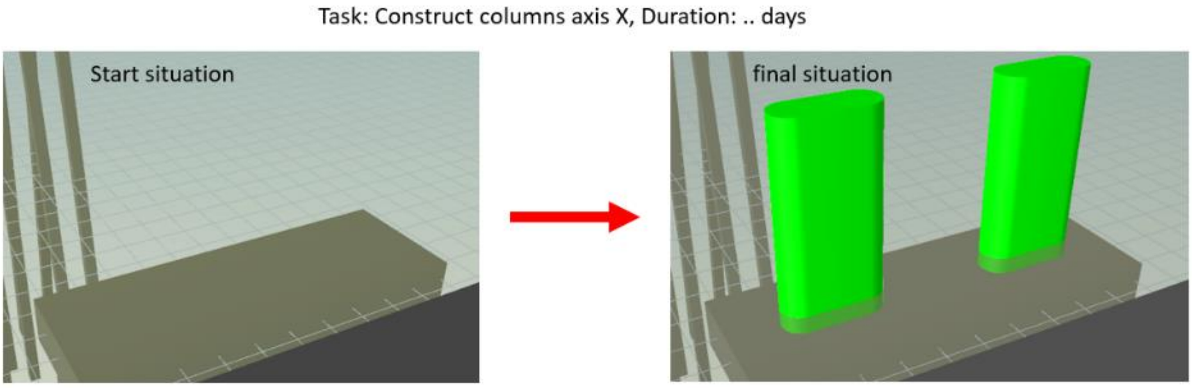


Figure 38: Elements being modelled - Construction of two columns (from BAM viaduct case)

Furthermore, when making use of parametric estimation, the duration of the task could be derived from the properties of the elements (e.g. size, weight, form, etc.).

3.3.2. Elements not being modelled

In contradiction, some tasks are not or cannot be represented by elements within the BIM model; therefore, they are also not allowing for the 4D visualization of the construction process to take place. An example is shown in Figure 39, where the excavation of the soil is necessary for the BAM viaduct case to be constructed. The task of excavation is not represented by elements part of the BIM model, and therefore is not visualized. However, Eastman et al. (2011) argue that ideally, a BIM model should provide information regarding temporary works (scaffolding, excavation), equipment and how the construction is phased (reach of the tower cranes, how the deck will be poured, etc.).

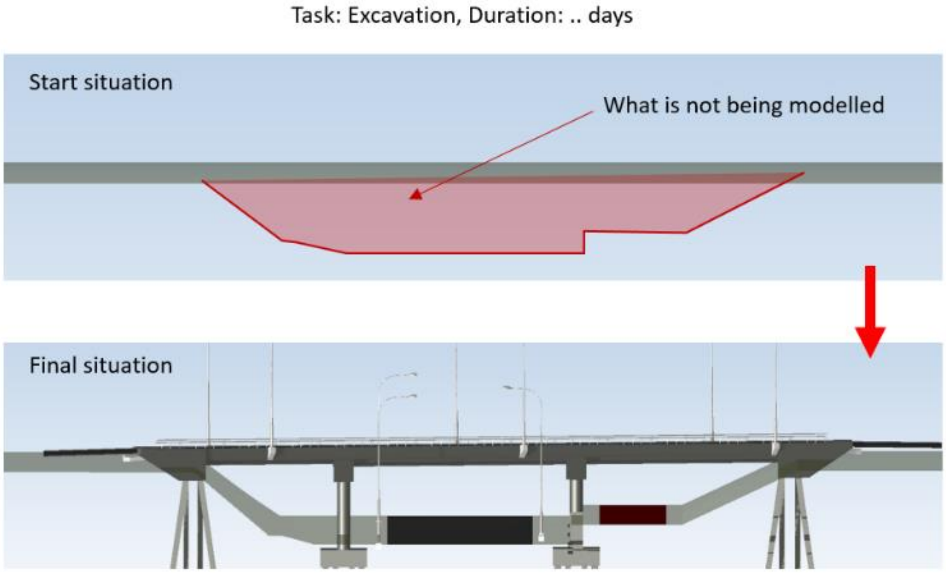


Figure 39: Elements not being modelled - Excavation (from BAM viaduct case)

When relying on parametric estimation the properties derived from the elements as part of the BIM model could be used; however, in this case, should be manually calculated.

3.3.3. Abstract elements

Abstract elements are elements that cannot be represented as elements within the BIM model. Since several planning processes cannot be defined by any kind of element, Hartmann et al. (2012) introduced the term of abstract elements. Within their research, they allow for the definition of abstract elements such as documents, materials, equipment, weather conditions, approvals or funding as part of the model. Hence, these abstract elements cannot be visualized and require another task estimation method than needed for tangible outputs.

3.4. Risk and uncertainty related to task duration

The project in its entirety, the project schedule and individual tasks can face some sort of risk and/or uncertainty. In terms of risk, “Risk management is the process of identifying, analysing, qualifying, and quantifying the risks, and developing a plan to deal with them.” (Mubarak, 2015).

3.4.1. Risks

Risks are often quantified by a combination of the impact and probability of occurrence. Generally, CPM schedules are not adjusted for risk since they are developed as one schedule and thus one alternative. Likewise, task durations are often calculated as a deterministic value based on productivity rates or an expert’s opinion, assuming those values being accurate. This assumption neglects duration variability and does most often not take into consideration factors such as crew composition, variable experience and working conditions (Mubarak, 2015; Ballesteros-Pérez, et al., 2020). A task may have a starting, duration, resource and production factor risk for example. Every potential delay cause or factor influencing the duration of a task as described in previous sections can thus form some sort of risk. Therefore, probabilistic values account for uncertainty, such as the continuous PERT beta distribution (Figure 40), or triangular method. The mean ($\bar{x}$) and variance (σ^2) of the activity durations are commonly calculated.

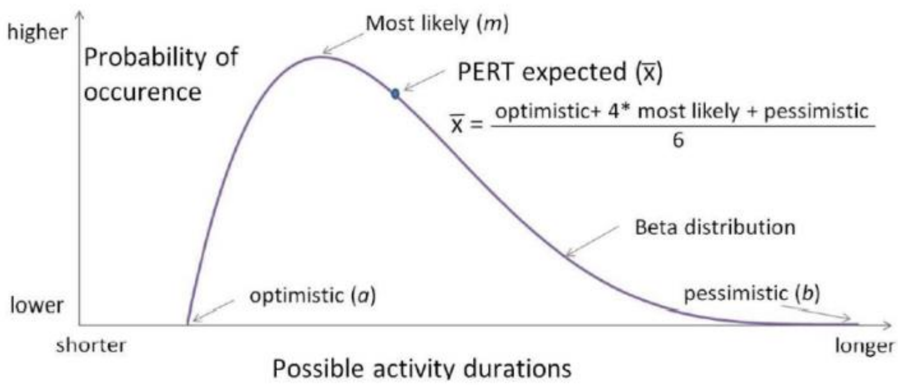


Figure 40: Typical function of PERT-beta distribution (Hajdu & Bokor, 2014, p. 767)

3.4.2. Uncertainty

In terms of uncertainty, the state of uncertainty can relate to the lack of knowledge or to inability to identify all possible events and their likelihood, while this can be done for risks (Toma, Chiriţă, & Şarpe, 2012). Uncertainty might be reduced, as an example: a planner makes a task duration estimate; unfortunately, during execution, the task took twice as long without disruptions nor negative risks influencing its duration. One might argue that the planner made the wrong estimation; on the contrary, one might say it was not wrongfully estimated since it represented the planner's knowledge at the time. The different states of knowledge and that proper knowledge would have reduced the uncertainty of the estimate is referred to as epistemic uncertainty. Aleatory uncertainty, however, relates to inherent randomness in the behaviour of the system being variable and irreducible (Helton, Johnson, Oberkampf, & Sallaberry, 2010).

3.4.3. Techniques

Many techniques have been proposed over the years to improve project and task duration estimates. Fuzzy logic (where linguistic variables are translated into mathematical measures), Monte Carlo simulations (simulations based on tasks having a statistical distribution), different variants of PERT (e.g. using different types of distribution), artificial intelligence (usually require training sets of similar construction projects) and extension of earned value management are all techniques that can be used (Ballesteros-Pérez, et al., 2020).

3.5. Conclusion

The third section of the literature review provides an answer to the third sub-research question regarding task duration estimation. It can be concluded that a variety of task duration estimation methods exist including subjective and data-driven methods. One method is not necessarily more sophisticated than another one and its choice depends on the situation at hand. For instance, the choice can be influenced by the available information, the stage of the project, etc.

When aimed at the estimation of task durations as part of BIM-based scheduling and the potential reuse of historical data, the method of parametric estimation arises as a favourable method. The tangible character of the elements in the BIM model, the link to the task and the ability of quantity take-off from the model form a source of data. The applicability of parametric estimation becomes more evident with the prospect of future developments regarding BIM-based automated scheduling. Furthermore, the knowledge and quantification of influential factors and the information related to risk and uncertainty allow task durations to be modelled more accurately.

3. Methodology and proposed system

In this chapter, the methodology of the research and the framework of the proposed system are described. The method and framework are used to reach the objective of this research as described in Section 1.2 i.e. "the development of a system that enables to extract reliable and accurate task duration scheduling data, as well as to make this data accessible and effectively useable in future planning processes."

1. Methodology: The V-model

The research design consists of a theoretical part and a practical part. The theory from the first part, related to research question one, two and three is used as input for the second part of this research, the practical part. The practical part is aimed to answer research questions four to six. To answer these questions, it is proposed to develop a system and test the system by means of a case study. The development of the system is planned to be executed according to the V-shaped software development life cycle model. This model consists of a development part and a testing part, also regarded as a verification and validation part (Singh, 2011), see Figure 41.

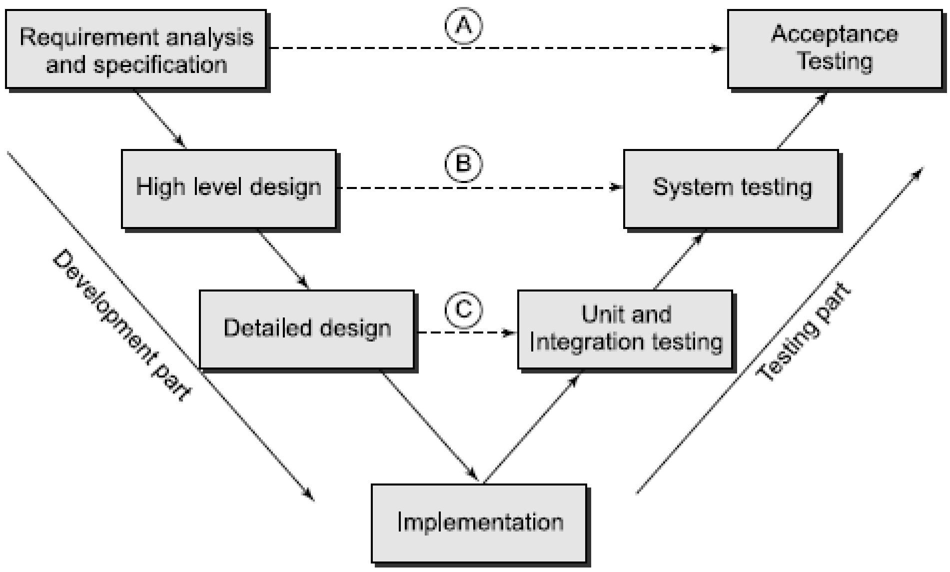


Figure 41: Software development and testing (Singh, 2011, p. 27)

These steps are used as guidance throughout the development of a system to extract reliable and accurate task duration scheduling data from. The data for this system is aimed to be gathered in the form of BIM-based scheduling data. The V-model outlines the processes and methodologies used in the different phases.

As described in Section "1.1 Problem definition', it is currently not possible or at least very difficult to reuse task duration data since no systematic access can be provided. Hence, the necessary information is mainly retrieved from other planners. In addition, potential information is often hidden in the form of separate planning files or documents that are stored over multiple locations. Therefore, the current process of generating a BIM-based schedule is investigated and whether that data can fulfil the requirements for it to be reused.

To make the data accessible structurally and centrally, it is studied how a database can facilitate in the storage of the BIM-based scheduling data and the subsequent reuse thereof. Hence, different types of databases and their applications are studied. In this case, the application of SQL (relational stores) or NoSQL (Not only SQL), see Section 4.1. Subsequently, the way these databases allow for the data to be analysed and reused is investigated. Executing specific searches, performing data analyses or data visualization are various options for achieving this. These possibilities are researched as part of Section 4.2.

The effective development of the system is focused on the database side and the way it allows for the data to be analysed and reused. An overview of a conceptual representation of the system to be developed and the associated research that needs to be conducted is shown in Figure 42. Data from historic BIM 4D models are aimed to be stored in a database. Subsequently, the process of data retrieval and analysis is studied to facilitate the reuse of task duration/productivity data.

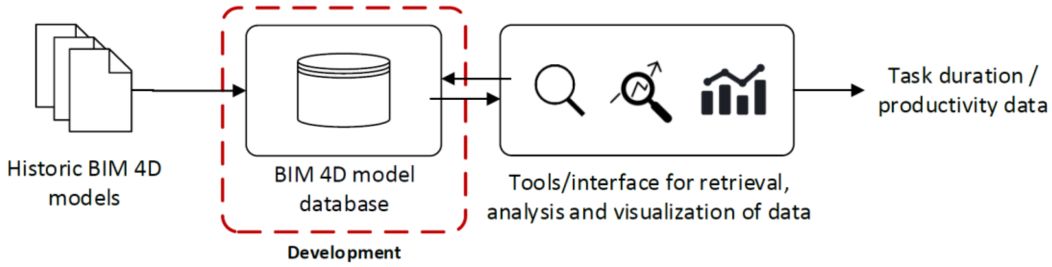


Figure 42: Concept development of the system

Returning to the V-model, first, different use cases are defined and applied for the analysis of the requirements and specifications of the system. Mapping these requirements and specifications are used to determine the functional and non-functional requirements for the output of the system. Based on these requirements, the development of the system and the required input can take place. Then, the high-level design focusses on system architecture and design. A conceptual framework for the proposed system is described as part of Section "3.3 The proposed system".

The detailed design focusses on the choice, implementation and the development of a database design as described above. The implementation of the system is followed by the system being tested through alpha testing. The system is tested within a controlled environment at the developer's site i.e. the author of this thesis. In addition, tests are conducted as part of the iterative development process. The alpha testing is conducted using the case study of a viaduct project. Thereafter, the results of the case are evaluated by a selection of experienced planners within BAM infra.

2. Development part

The main objective of the development part is the creation of a system that overcomes the problems as stated in the problem statement. This problem statement is translated into a vision where a system facilitates the use of historical BIM-based scheduling data to estimate task durations as part of a data-driven approach.

"The ability to manage the knowledge generated from the projects (including the capture of project knowledge and its subsequent transfer) not only can help to prevent the 'reinvention of the wheel' and the repetition of similar mistakes, but also serves as the basis for innovation and overall improvement." (Tan, et al., 2010, p. 2)

In support of the research aims, the development of the system is aimed to align and facilitate in the process in which the scheduling data can be captured, stored, analysed and reused. The main objective is transformed into a practical objective to concretise the development part. The practical objective:

Practical objective: Creation of a system allowing a task duration/productivity database to be built, fed by BIM-based as-planned and as-built project scheduling data to estimate task durations.

Before the development of a system can take place, it is important to understand how knowledge can be created and converted. In general, there are two types of knowledge: explicit and tacit knowledge (Collins, 2010). Explicit knowledge is formal systematic knowledge that can be codified in documents, reports, etc. Whereas, tacit knowledge relates to know-how from professional expertise, experience, etc. Tacit knowledge is the knowledge that is not explicated (Collins, 2010). In this research, the objective is to store embedded explicit knowledge from 4D BIM models and generate new explicit knowledge in the form of task duration estimations. Hence, separate explicit knowledge is converted to systematic explicit knowledge.

2.1. Requirement analysis and specifications

The ability of the system to function depends on the desired output to align with the information provided as input and the capability of the system to facilitate in that process. The requirement analysis and set specifications for the database to be designed are fed by findings as part of the literature study and open interviews that were held with planners and employees within BAM Infra. Moreover, the author spent a month in advance of the graduation project to get a better understanding of the current scheduling process and related difficulties that are faced within BAM Infra. The requirements of the system are primarily approached from the perspective of construction planners. Additionally, other users can benefit from the system such as project managers, portfolio managers, risk managers and foreman.

2.1.1. Use cases

How the system is intended the be used, is described within different use cases. For this study, four use cases are defined. Namely, the use of the system by the planner prior to a project, the planner at the construction site, the project manager and a portfolio manager. Different perspectives of how use cases interact with the system are shown in Figure 43. Within the figure, it is indicated with a symbol what their primary request of information will be (i.e. a detailed search/query, visualization of the data or data analysis/modelling).

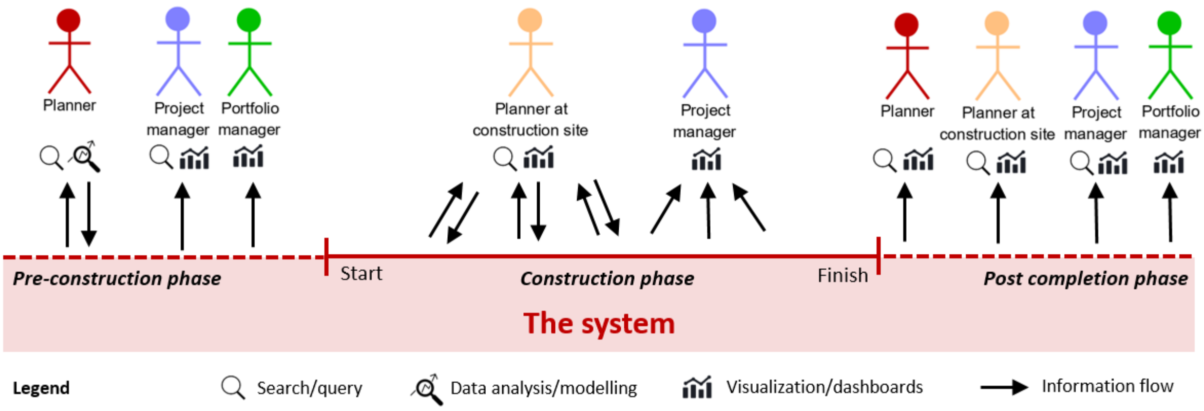


Figure 43: Possible use cases for the system

Pre-construction phase

Planner: the primary user of the system is initially interested in extracting accurate task duration information. When a planner receives a BIM model from the design team, a BIM 4D model in the pre-construction phase can be made. The planner can request historical task duration information from the system that relates to the elements within the BIM model (i.e. BIM-based task duration information). Hence, the planner wants to query data and possibly conduct data analysis to derive such figures. Finally, the planner can interact with the system by inserting the completed BIM-based as-planned schedule in the database.

Project manager: the project manager can request historical task duration information to gain insight into time-related risks related to the BIM-based as-planned schedule provided by the planner. As a result, a project manager might be able to take proactive actions to mitigate risks for the upcoming project. For instance, by investigating alternative options that involve lower risks.

Portfolio manager: the portfolio manager can use the historical task duration information to determine the ability to complete tasks on time. For instance, related to tasks that are associated with certain types of projects or certain phases within certain projects. The strengths and weaknesses related to these types of projects or phases within projects can complement the decision-making process for the definition of strategic objectives that can be pursued within the company.

Construction phase

Planner at the construction site: as BIM-based as-built schedules are consistently produced during the construction phase, the planner can use them to monitor to progress and identify task duration differences that have occurred. Subsequently, the planner can adjust the expected duration for similar tasks later in the project. In addition, the planner can submit alterations to the schedule to get the project back on track.

Project manager: during the construction phase, the project manager can use the system to monitor the progress of the project. In addition, the project manager can identify tasks that face a certain duration risk and communicate these risks within the team.

Post completion phase

Planner: based on the BIM-based as-built schedules, the planner who has drawn up the BIM-based as-planned schedule can use the system to obtain systematic feedback on the predicted task duration and gain insight into where deviations have arisen.

Planner at the construction site: the planner at the construction site can use the system to review the project and derive systematic feedback similar to the planner responsible for the BIM-based as-planned schedule.

Project manager: the project manager can use the BIM-based as-planned and as-built schedules as part of the pre-construction phase and construction phase for delay analysis and claim management.

Portfolio manager: the outcome of a project can be used by the portfolio manager to verify and update the strategical objectives within the company. Thereafter, the portfolio manager can use the system similar to its use in the pre-construction phase.

2.1.2. Desired output to estimate task durations

Hereby a concise description is given of what is desired from the perspective of a planner to be extracted from the system, being the extraction of BIM-based task durations.

The ability to extract:

- Task duration and related information e.g. equipment, labour, elements and element parameter/property data (e.g. length, width, weight and other variables)
- Task duration as part of a specific element, element category or class
- Task duration as part of a specific task, task category or class
- Task duration in the form of productivity data (e.g. units/h, m^2/person-hours)
- Task duration with specific conditions (e.g. location of an element or the equipment being used)
- Task duration as part of one or more projects or a certain type of project
- Deterministic and probabilistic duration of a task (min/most likely/max etc.)
- The risks and uncertainty related to the duration
- Influencing factors upon the duration of a task and its quantification

Note, the user requirements mentioned above are related to *what* and *who.* Functional requirements outline *what* the system should do i.e. the main ability to provide task duration data. Non-functional requirements relate to the system being accurate, reliable and easy to use. Furthermore, the effort of collecting, storing, retrieving data should be outweighed by the relative benefit of the information produced.

2.1.3. Requirements that enable to retrieve task and element information

The ability to retrieve task and element data from a database, depends on the quality and quantity of information, the level of detail, how the data is structured and if it is presented in a uniform way when it was entered into the database. For instance, the ability to retrieve multiple entries of tasks, labour, equipment and other data depend on the use of similar names and/or codes (classifications).

As described in Section 2.1.5.2, the Information Delivery Manual (IDM) standard can be used to determine which, when and to whom different kind of information has to be communicated within and between organizations. The standard organizes data exchange processes through a process map. Subsequently, exchange requirements (ER) define the information units required for each use-case-specific process for the exchange of information. Model View Definitions (MVD) are used for the technical implementation of the exchange requirements. To indicate why it is important to agree upon the processes and related exchange requirements, consider the following three examples:

1. A wall consisting of multiple material layers can be modelled as one single element. However, this limits the capability of attaching single tasks and duration data to all material layers. Hence, it will only be possible to attach a duration figure concerning the entire wall.

2. As described in Section 2.1.5.3, an important feature of BIM 4D modelling software is the subdivision or aggregation of elements related to tasks. However, once subdivisions or aggregation are made within the 4D modelling software, the element property data will probably not be updated (as this is the case as part of Synchro 4D). For instance, when a wall is split in two, the wall will still have the same dimensions listed and therefore result in inadequate data once loaded into the database. Hence, it might be necessary to subdivide or aggregate elements in the BIM modelling tool (e.g. Revit). Modelling guidelines can be set up in this regard.

3. Property Sets describe how sets of properties (usually defined by a name, value, unit triple) are associated with objects or object types (BuildingSmartAlliance, n.d.). Unfortunately, property sets have not yet been adequately organized to store information for a wide range of tools in a standard way and are currently left to users to set up (Eastman, Teicholz, Sacks, & Liston, 2011). Hence, the ability to retrieve reliable element data depends on including uniform properties and property names, having uniform dimension measurements etc.

These examples merely indicate why it is important to decide upon the processes and exchange requirements such as the level of information and level of detail in the model. Therefore, it is recommended to document these procedures and requirements, e.g. with the use of IDM. In addition, rather explicit modelling guidelines may be necessary to facilitate in the information exchange between BIM modelling tools (e.g. Revit), the BIM 4D modelling tools (e.g. Synchro 4D) and the database. As a result, the ability to retrieve task and element data from the database can be positively affected.

3. The proposed system

The proposed system described hereunder is part of the high-level design stage, here the *how* is defined on a conceptual level. The requirements and specifications of the previous phase serve as the basis for the development of the system. In addition, suitable methods are synthesized from the literature review that fits in the overall process of capturing, storing, analysing and reusing the data.

It is aimed to frame the proposed system into the perspective of the current way of construction planning and progress tracking within BAM Infra (Figure 44).

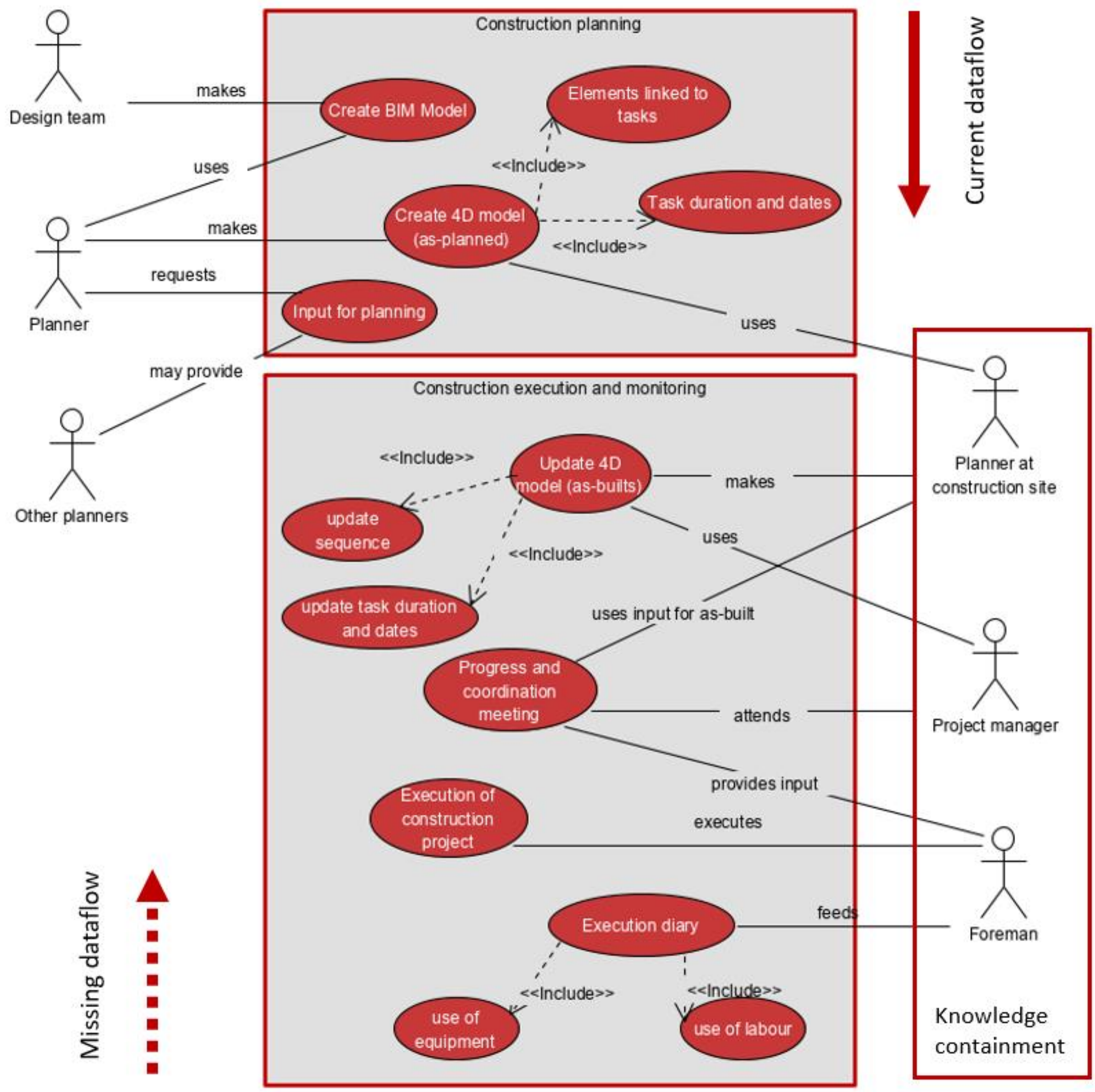


Figure 44: Simplified current use case of construction planning, execution and monitoring as part of BAM Infra

The diagram visualizes the interaction of the different actors for the creation of a construction planning and the actual execution and monitoring of the construction without a database to delve task durations from. The planner at the construction site and the planning stage is represented separately since a planner’s work is generally limited to the pre-tender, pre-construction stages or on-site (construction stage) but not simultaneously (Winch & Kelsey, 2005). Once the as-planned 4D model transfers to the execution stage, the planner at the construction site updates the 4D model accordingly. The use of equipment and labour is not linked to the updated 4D model as this is part of the execution diary which is fed by the foreman.

The knowledge regarding the differences between the as-planned situation and the as-built situation is reflected by a red bounding box representing the knowledge containment of the actors that are involved in the execution stage. The bounding box is further emphasized by the missing dataflow facing up towards the actors as part of the planning phase. When a planner estimates task duration for a new project it can request input from other planners. The knowledge of that person and the reliability of the information depends on the ability of the planner to recall the information. The methodology described below aims to systematically unlock the knowledge containment and enforce the missing dataflow.

Note, that the current way of planning follows the problems described as part of the problem statement related to the discrepancy between the overall schedule and detailed schedule and the inability to effectively reuse former experiences and well-founded figures. Storing BIM-based as-planned and as-built scheduling data is aimed to overcome this discrepancy. Continuous learning can take place since the expectations as formed in the as-planned situations are compared to the reality as part of the as-built situations. The information concerning the as-built situations is aimed to enhance the reliability and accuracy of future estimates. Hence, a system that stores BIM-based task duration data can be beneficial.

The method employed that suits the objective and processes of the system, are the steps mostly used in LEAN management and Six Sigma. The Plan-Do-Check-Act (PDCA) cycle, is a cycle for continuous improvement often used within LEAN management approaches. Six Sigma uses a set of proven quality principles and techniques for process improvement aiming to eliminate defects and reduce variation (Pyzdek & Keller, 2019). This data-driven technique employs a performance improvement model known as Define-Measure-Analyse-Improve-Control (DMAIC). As part of the process, critical factors on current defects are identified, measured and subsequently, the influential factors (causes) are analysed. The derived framework that is used for this thesis is shown in Figure 45.

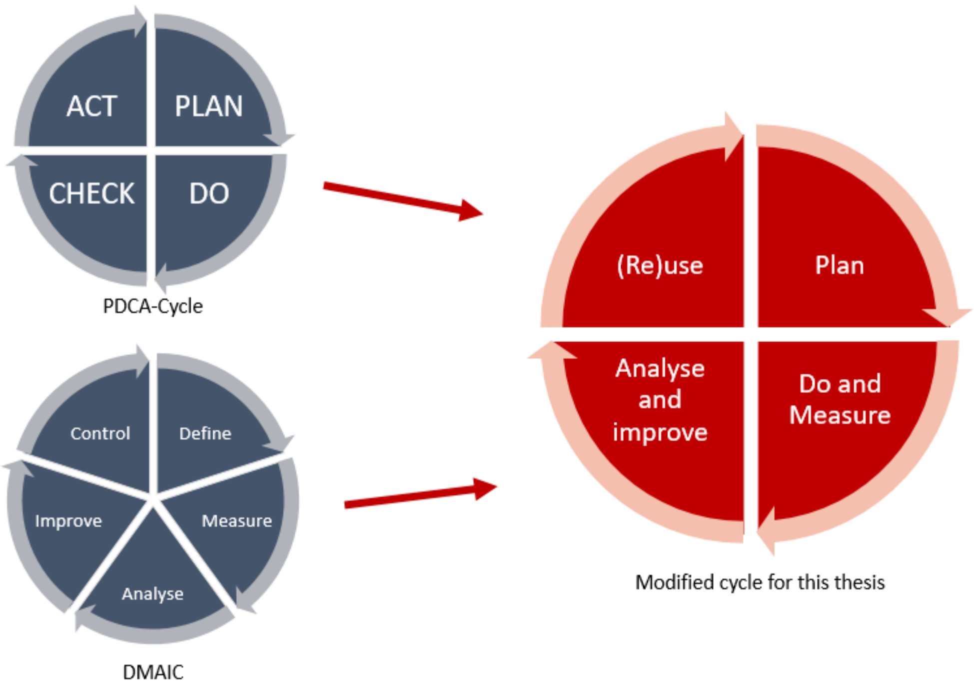


Figure 45: Methodology that is translated into the different stages incorporated in the proposed system (derived from PDCA-cycle and DMAIC)

The focus is on the ability to (re)use the data that is collected and analysed in the adjoining phases. The consecutive paragraphs describe the cycle per phase. The implications of the literature in Chapter 2 on the methodology are also described.

3.1. Plan

The Plan phase consists of the generation of a BIM-based schedule as part of the pre-construction phase, resulting in a 4D BIM-based as-planned construction schedule. Since the concept of BIM-based scheduling evolves around elements being complemented with task information, the link in between is fundamental. The construction schedule is made by a planner that receives a BIM model modelled by a design team. The process of construction scheduling according to the CPM (as described in Section 2.1.2) is chosen to be applied within this phase as this is also an often-used method within BAM infra.

It can be argued that, when limited information is provided on a construction task and its related duration, its usability and reusability will be limited. A construction task where the input factors and related conditions and context are limited is represented as 'simple task representation' in Figure 46. In this case, the schedule is not linked to the BIM model. Since the duration of a task is most often a composite of multiple influencing factors, the information used to make a task duration estimate is aimed to be explicitly embedded to a certain extent in the schedule, i.e. information of input factors, internal environment and exogenous factors represented as 'enhanced task representation' in Figure 46. Obviously, this involves the link between the elements within the BIM model.

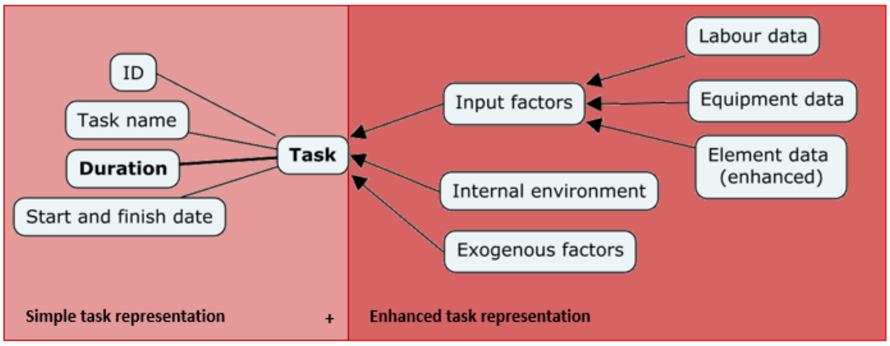


Figure 46: Simple and enhanced task representation

Concluded from the literature and in the prospect of the 4D modelling, the method of parametric estimation is primarily applied. Therefore, the data of past projects should allow being reused in a manner that aligns with parametric estimation. Hence, the data as presented in Figure 46 should be linked and stored.

Moreover, besides deterministic duration as part of a CPM schedule, stochastic task duration information can identify risks and uncertainty that comes with a specific task duration. It is argued that the ability to extract data of the factors and related risks and uncertainty for the duration of a task and its quantification allows for more accurate and reliable task duration estimates. The context given to the estimation of a task can be used in the consecutive phases. The desired in- and output of the *plan* phase is shown below.

Input:
- BIM model
- BIM-based schedule
- Task duration/productivity data

Output:
- As-planned schedule

3.2. Do and Measure

As part of the do and measure phase, the actual construction is executed (do) according to the as-planned model. Simultaneously, the related progress is being tracked (measured). Important for this phase is the measurement and collection of data that are relevant for productivity measurement and delay analysis. As part of Section 2.2.1, progress tracking, productivity measurement methods and delay analysis techniques are reviewed. Although a variety of methods exists, it is sought to employ a productivity measurement and a delay analysis technique that allows for use and reuse of the gathered data. Since many factors influence the duration of a task, it should be sought to measure/note these factors during construction. Hence, the factor model for productivity measurement is considered to be a suitable method since it aims to analyse and quantify the factors that influence productivity. Furthermore, it aligns with the enhanced task representation in the Plan phase. Subsequently, the parametric estimation technique can be employed for task duration estimation. Complementary, as part of delay analysis, it is sought to record delay/changes and causes on task level. The understanding of the discrepancy between condition changes and delays, and related information at task level are vital for the measurement within this phase. Consider the example in Figure 47:

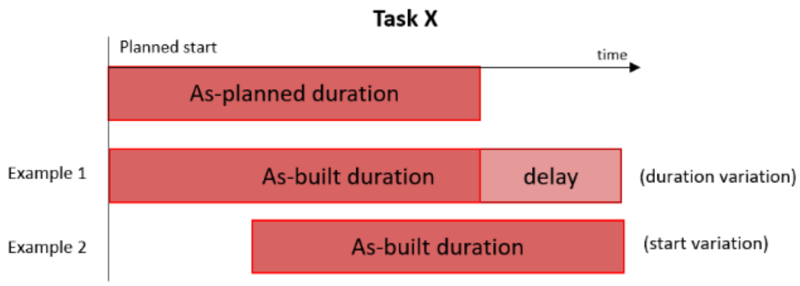


Figure 47: Start duration and duration variation

Task X as part of example 1 faced a certain delay leading to the extension of the total task duration time i.e. a duration variation. Example 2 shows that the starting time is delayed; however, the duration itself did not change. Thus, the task did not have a duration variation but only a starting time variation. Assume that task X represents the construction of a certain column and faced a delay as in example 1. The question is, should the planner schedule more time for the construction of that column next time? The question cannot be answered since the information does not satisfy the purpose of future reuse. The representation lacks the context of the duration estimate and related delay cause(s). Subsequently, consider the following example where the representation of the task is enhanced (Table 3).

Table 3: Enhanced representation of as-planned and as-built task duration (factors and data are for indicative purpose only)

Ex.	Duration	Input factors			Internal environment	Exogenous factors
	Task X : Construct Column	*Element data:*	*Labour data:*	*Equipment data:*	*Site layout:*	*Weather:*
		Concrete column	Labour workers	Equipment type	Distance to materials	Temperature
	As-planned duration	1x (Size X)	3	Type A	100 m	Comfortable
1	As-built duration / delay	1x (Size X)	3	Type A	100 m	Comfortable
2	As-built duration	1x (Size X)	2	Type B	20 m	Comfortable
3	As-built duration	1x (Size X)	6	Type B	100 m	Cold

The as-planned and as-built task duration are enriched with information about input factors, internal environment and exogenous factors representing the conditions on which the duration is based. The adjusted factors in the as-built situation reveal the renewed conditions for the duration of the task. Example 1 represents a situation wherein the conditions remained equal; however, the task still faced a delay. Complementary information is required to pinpoint the cause of the delay. Example 2 and 3 respectively represent an equal and accelerated (duration variation) duration compared to the as-planned duration. Even though example 2 and 3 did not face a delay, the conditions changed where the duration of the task was based upon. The information can be used to feed a database whereby the duration of the task is complemented with the related conditions. As part of the as-built duration, it is sought to measure the conditions for every task unrelated to the presence of a delay. For example, the conditions of example 3 likely contain valuable data for the task being accelerated.

Section 2.2.1.2 describes delay measurement techniques i.e. the Foreman Delay Survey, the Craftsman Questionnaire and the Method Productivity Delay Model (MPDM). The latter is considered most suitable since it enables to identify the sources, the probability of occurrence and severity of delays. The philosophy of the MPDM can be used to enrich task information with the causes of delays. It is therefore proposed to enrich a negative duration variation (a delay) or a negative start variation (started later than planned) with cause information (Figure 48). Also, it is proposed to enrich tasks that started earlier than planned and had a shorter duration. Cause information for start variations relates to direct variations only.

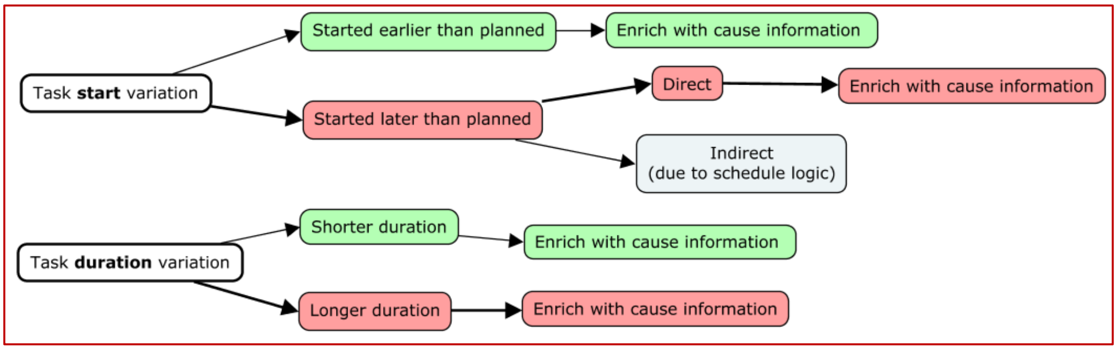


Figure 48: Cause information on start and duration variations

Although the altered conditions may provide a direction for the cause of a variation, actual cause information allows pinpointing the cause of the delay by which the usability and reliability of the as-built duration can be enhanced for future projects. In clarification, if the delay as part of example 1 was due to a client putting the task on hold, one might not schedule more time in the future. On the contrary, if the delay was due to the as-planned duration not being realistic, one might schedule more time. Furthermore, it allows for continuous improvement since delays might be better anticipated upon in future projects, thus enabling risk mitigation, as well as, providing insight into the accuracy of the task duration and related risk. Obviously, the individual causes of the delays can indicate the causes of a project being delayed in total as well as be used for delay claims. Additionally, the information on the input factors can facilitate in the process of resource-based scheduling.

The as-planned schedule is updated regularly according to the requirements that are set as part of the project and the progress tracking method that is employed. Assuming this is a weekly or biweekly update, an updated version of the as-planned schedule will be generated accordingly. Finally, when the construction is complete, the last as-built schedule will be made. Note, how this data is captured is of minor importance for this study. The desired in- and output of the do and measure phase is shown below.

Input:
- As-planned schedule

Output:
- As-built schedules (as-built in progress schedules and final as-built schedule)

3.3. Analyse and improve

In the analysis and improve phase, the as-planned and as-built schedules as part of the first two phases are used as input. First, these schedules need to be stored inside a database comprising as-planned and as-built schedules (described in Chapter 4). In Chapter 4, the applications of databases are studied and how they intersect with its subsequent abilities to analyse the data.

Once stored, the database can be analysed in various ways. The primary focus of this study on the use of queries (searches) to retrieve data from the database. The database is usually connected to a user interface (front-end) that allows to search the database and gain insight into the data. Furthermore, the data in a database can be visualized and analysed, for instance, with the use of dashboards. And lastly, as a subset in the process of Knowledge Discovery in Databases (KDD), data mining techniques can be employed that cover a variety of descriptive and predictive analysis techniques (described in Chapter 4).

At the same time, it is aimed that this data supplements the database allowing to delve from a larger dataset. Hence, it improves its estimation capabilities for future projects. The main ability of the database is to allow one to search for task duration data. Added data can provide new or adjust known insights of influencing factors on task durations and related variations. Furthermore, although not the focus of this thesis, the data stored in the database might provide information that is found to be useful for project managers and others. This can be, for instance, the visualization of data as metrics on a dashboard for the part of project control.

Input:
- As-planned schedules
- As-built schedules (as-built in progress schedules and final as-built schedule)

Output:
- Database inserts
- Search task data
- Analyse data
- Visualize data

3.4. (Re)use

In the last phase, the (re)use phase, the output that is generated in the previous step can be (re)used. The BIM-based scheduling database consisting of past projects allows being reused to estimate task durations, while simultaneously accounting for factors influencing the durations, causes of variations and risks. In short, the data provided as output in the previous phase can be synthesized by a planner to estimate task durations. Alternatively, the data can be (re)used to suit other use cases as defined in Section 2.1.1, such as providing systematic feedback towards the planner, providing progress information to the project managers for project control and risk events that can be used by a portfolio manager to determine strategic objectives.

In the case of estimating task duration, parametric data can be extracted as part of the properties of the elements contained within the BIM model. For instance, element-related data such as surface (m^2) of a concrete slab that is linked to the duration (hours) for the task of erecting formwork could be stored and provide a productivity figure (e.g. hours/m^2) to be reused (see example 1 in Figure 49). However, this productivity figure does not take specific influencing factors into account.

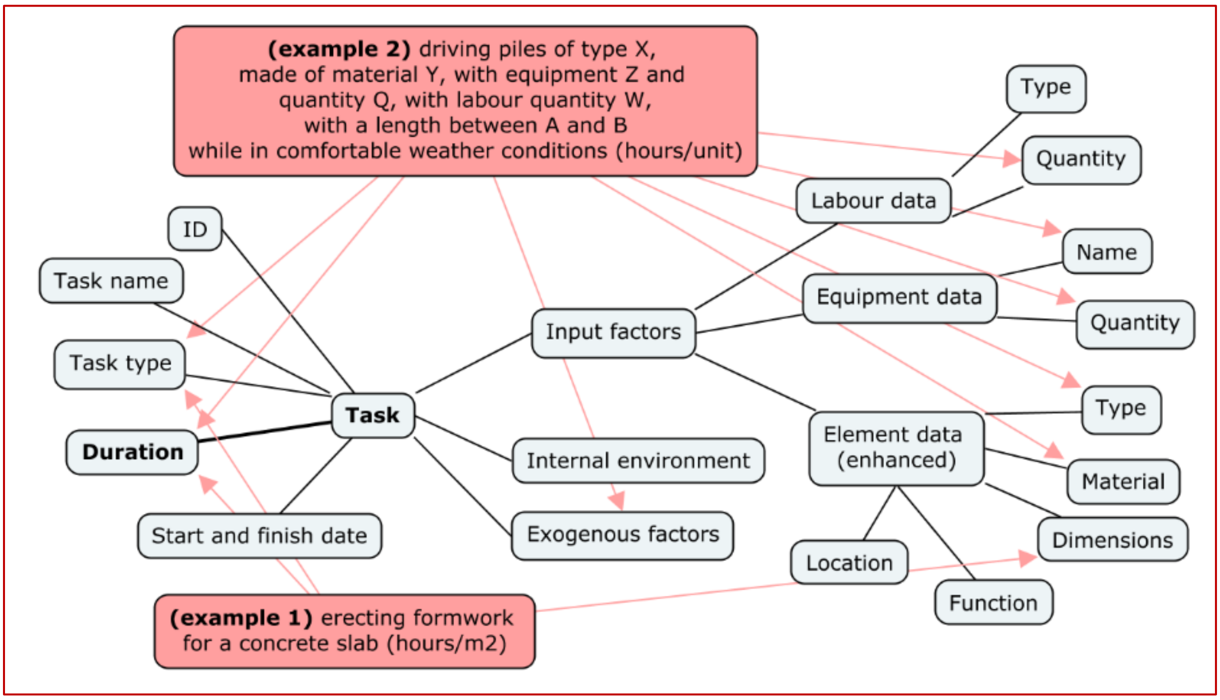


Figure 49: Examples for the reuse and extraction of task duration/productivity data

Alternatively, example 2 in Figure 49 shows the extraction of task duration data of driving piles where multiple conditions and thus influencing factors are considered. Accurate quantity take-off from the BIM model, linking the data and systematically storing the input factors, factors related to the internal environment and exogenous factors all determine the ability to extract such figures, and use it as part of the parametric estimation process. The enhancement of element data with properties about the type, function, location, dimensions, connections etc. allows to further distinguish and analyse the task duration with the properties of the elements.

Input:
- Query data
- Analyse data
- Visualize data

Output:
- Task duration data/productivity data
- Other data usage (project control and else)

3.5. The framework of the proposed system

The complete methodological framework is presented in Figure 50. The framework shows how the different phases are linked to each other based on the in- and outputs of every phase. The framework is employed every time a new plan for a construction project is made and the construction project is being executed.

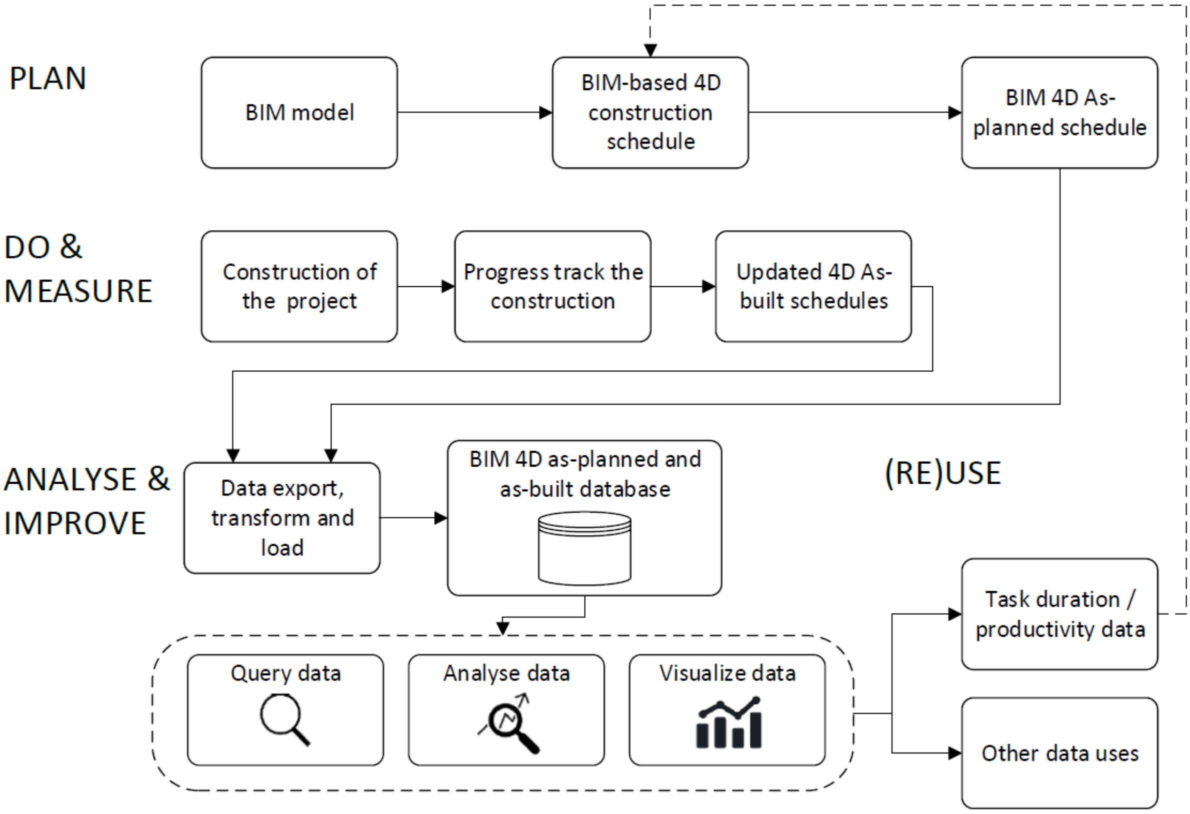


Figure 50: The framework for the proposed system

The implications of the methodology are again placed into the perspective of the use case of construction planning, execution and monitoring within BAM Infra (Figure 51). The execution diary provides information about the use of equipment and labour for the input factors. For both the as-planned and as-built 4D models, information on the input factors, internal environment and exogenous factors are provided. The implemented framework facilitates in an upstream dataflow through the use of a database.

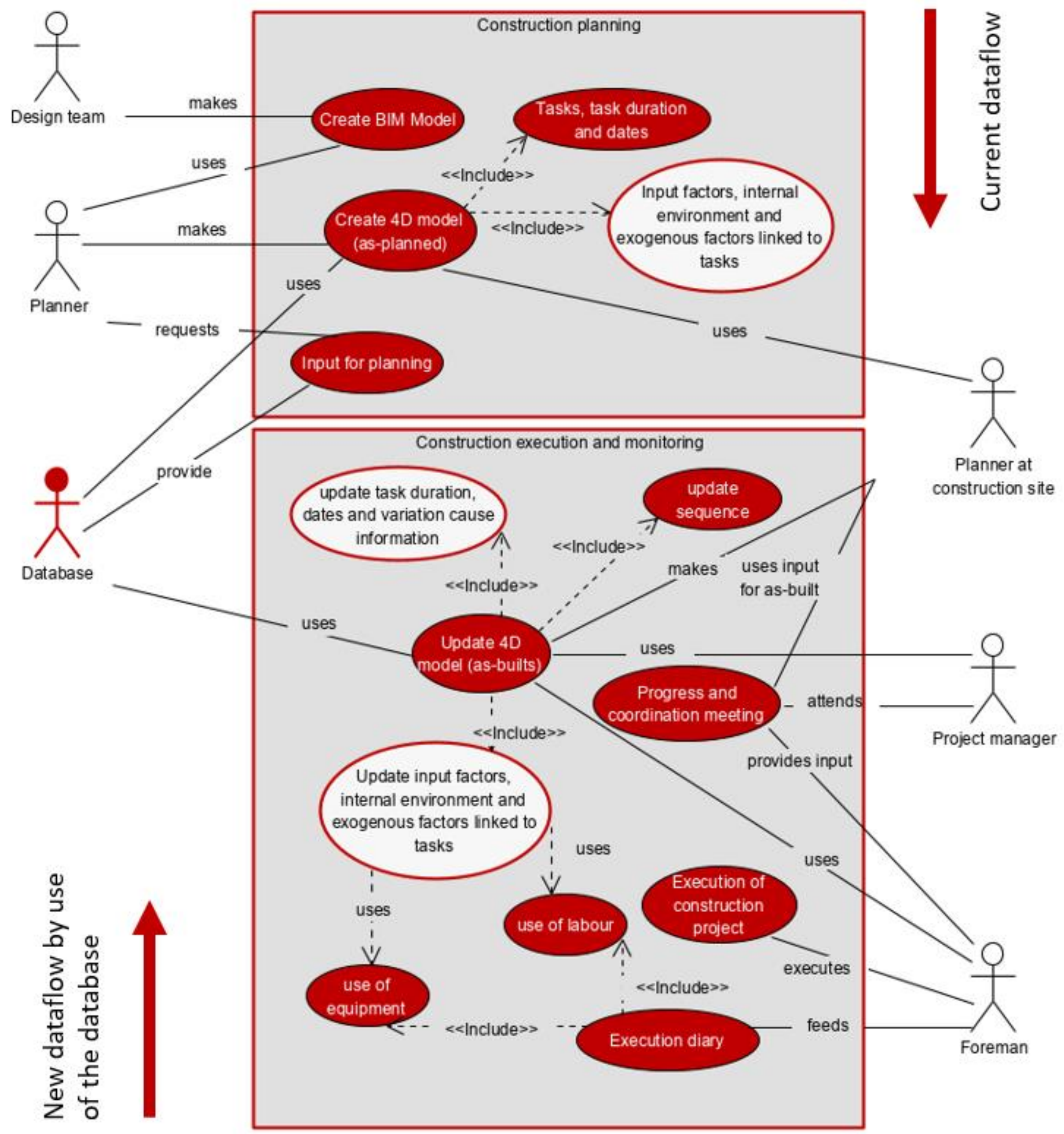


Figure 51: Simplified updated use case of construction planning, execution and monitoring as part of BAM Infra (the white circles identify the changes compared to the original use case)

4. Conclusion

In this chapter, the methodology of this research and the framework of the proposed system are outlined to reach the objective of this research i.e. "the development of a system to extract reliable and accurate task duration scheduling data from, as well as to make this data accessible and effectively useable in future planning processes." The V-shaped software development life cycle model which consists of a development part and testing part is adopted to reach this objective. The system will be tested using a case study subsequently to the development of the system as described in Chapter 4. A framework is derived from the PDCA cycle method and DMAIC performance measurement method consisting of four phases that fit in the overall process of the proposed system. This framework is aimed to enrich the BIM-based as-planned and as-built scheduling data generated as part of the *plan* phase, and *measure and do* phase with information regarding the input factors, internal environment and exogenous factors that drive and influence the duration of a task. These factors can give parametric meaning to the duration. Additionally, knowledge of these factors can potentially be utilized in future developments regarding automated scheduling. In the *do* phase, it is aimed to include the causes of task start and task duration variations. In doing so, the reusability of the data can be improved by identifying risks and possibly mitigate those risks within future practices. In the analyse and improve phase, and the re(use) phase, the BIM-based scheduling data is stored in the database and queried, analysed and visualized for it to be (re)used.

4. System development

The chapter of system development relates to the detailed design level of the V-model and consists of two parts: 1) capture and data storage, and 2) analysis and (re)use. Both parts affect the in- and output requirements for one another. This chapter describes the development of a system as a proof of concept to store BIM-based as-planned and as-built scheduling data for the main purpose of estimating task durations. This chapter answers sub-research question 4 “How can BIM-based as-planned and as-built scheduling data be captured and stored in a database? and sub-research question 5 “How can historical BIM-based as-planned and as-built scheduling data be analysed and (re)used for estimating task durations?”

1. Capture and data storage

In this section, techniques for data storage are reviewed. Furthermore, the development of an actual database is described that allows to store BIM-based as-planned and as-built scheduling data. The way data is captured and stored depends on the way it is aimed to be reused.

1.1. Data modelling

The data within the domain of AEC is represented using so-called data modelling concepts. First, a part of reality is mapped into a data model. Second, actual data is stored in a file or database (Figure 52). Semantics describe the meaning of data or information. For instance, within construction the sole reliance upon geometric data is not sufficient; semantic data such as construction methods, materials and functions need to be considered in addition (Borrmann et al., 2018).

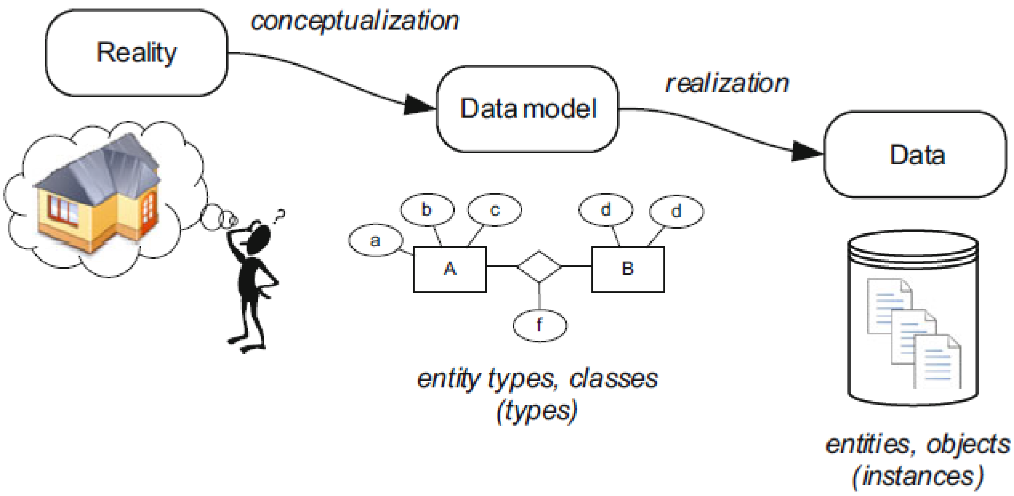


Figure 52: Procedure of data modelling (Borrmann et al., 2018, p. 45)

A data model comprises entities, attributes and relationships. An entity can be an item in the real world, such as a wall or a column; alternatively, it can also be non-physical such as a load or a task. An entity type (or class) groups entities that have similar structures and characteristics. An example is a certain concrete column (entity) being part of columns (entity type). Attributes represent the properties of an entity and its information (Borrmann et al., 2018).

1.2. SQL and NoSQL

To develop a database, knowledge of databases is required. A storage model is a model that captures the physical aspects and features for data storage, such as file-based storage on a Windows pc. A data model captures the logical representation and structures for data processing and management, such as the databases described in this section. A data model also represents the relations among different data elements. A data model can be categorized in SQL (Structured Query Language) (relational) stores and NoSQL (Not only SQL) stores (Zomaya & Sakr, 2017). The past decades, relational database management systems (RDBMS) have been the dominant solution for most applications. In addition, they have a well-established community and integrate well with other programs. SQL is the language used for storing, manipulating and retrieving data in relational databases. A query is a request for data from a database table or a combination of database tables. SQL allows complex queries to be executed. Relational databases organize data into one or more predefined tables. Each table has columns (attributes) and rows (tuples) whereby each row is identified with a unique key (the primary key). An example of two tables is shown in Figure 53.

attribute

Task table

*Task_ID	Taskname	Duration days	Startdate	Finishdate	Egp_ID
T1	Task X	1	1-1-2020	1-1-2020	E1
T2	Task Y	2	2-1-2020	3-1-2020	E2
T3	Task Z	3	4-1-2020	6-1-2020	E1

tuple

Equipment table

*Eqp_ID	Equipment name
E1	Equipment X
E2	Equipment Y

Figure 53: Example of tables as part of a RDBMS

The task table has Task_ID as the primary key (PK), the equipment table has Eqp_ID as the primary key. Primary keys are used to establish relationships from one table to another, as such Eqp_ID from the equipment table is inserted as a so-called foreign key (FK) into the task table (Figure 53).

Since a RDBMS uses predefined schemas, the related capability of dealing with large scale and semi-/unstructured data is limited. Therefore, SQL databases typically work well with structured data. Data integrity is an important aspect of SQL databases. Data integrity refers to the accuracy and consistency of data over its lifecycle. Within SQL, data integrity is ensured by with entity, referential, domain and user-defined integrity. For example, as part of entity integrity, every tuple should be unique which is guaranteed with the PK.

NoSQL data storage systems have emerged, since they provide higher scalability, more flexibility and do not require predefined schemas. NoSQL aims to provide an answer to the increasing growth in data size and data variety (Zomaya & Sakr, 2017). NoSQL systems are less robust and trade reliability and consistency for high availability and scalability. NoSQL systems are typically well-suited for semi-/unstructured data, such as big data. Key-values, documents, extensible-records and graphs are common NoSQL data model types. Relational models typically lack native support for particular data structures such as graphs and trees. The use of graphs is appropriate if data is well presented as a graph and one wants to obtain information on how the topology of a network evolves or how entities relate to each other (Sakr &

Pardede, 2012). The Resource Description Framework (RDF) is a widely adopted data model used for manipulating graph-like data designed for representing data in the Web. This mechanism is a major element of the Semantic Web that aims to make data on the web machine-readable. An example of an ontology as part of a scheduling knowledge base developed in RDF/OWL format is shown in Figure 54.

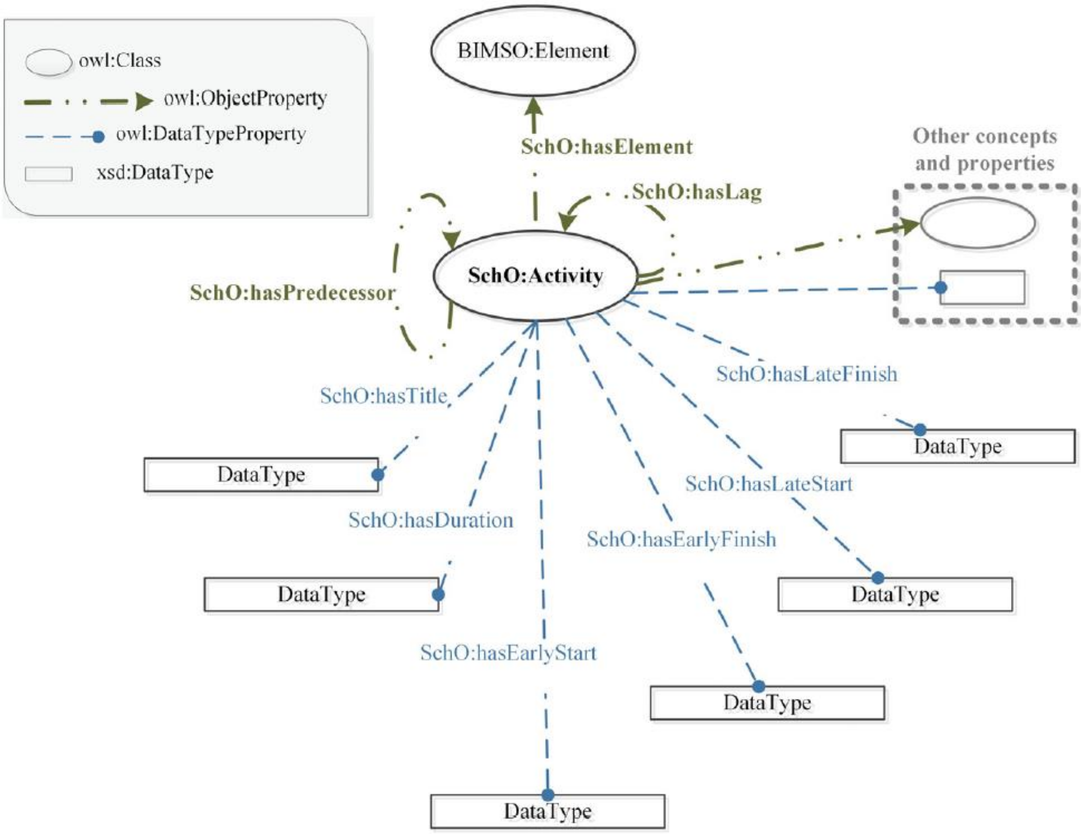


Figure 54: Schedule ontology for the concept of an activity/task (Niknam & Karshenas, 2016, p. 693)

1.3. Current data representation and export capabilities of used tools

The current use of BIM model design software and BIM 4D scheduling software within BAM Infra are used as the foundation for the development of the system. Therefore, Autodesk Revit modelling software and Synchro 4D software are used. Factors such as structured or unstructured data, flexibility, variety and size of the data influence the choice of a database system, as well as the options to access and export data as part of the software tools used to make a BIM model and BIM 4D model. Therefore, the data export capabilities of Revit and Synchro 4D are assessed. Revit can export data as gbXML, IFC, ifcXML and to an ODBC (Open DataBase Connectivity) database. The ODBC database exports model types and model instances according to categories within Revit to a SQL server. IFC data can also be stored in a SQL environment; however, the development of the SQLite version is still in an experimental stage (Building Smart International, n.d.). Synchro 4D allows exporting data as XML (P6, MS Project and Asta PowerProject), SDEF, IFC and MS Excel. An extended package allows for data to be retrieved as relational data. The data exported as MS Excel contains IDs that enables to make the relationship between, for example, tasks and elements. The ability of both software tools to represent the data as part of relational stores (SQL) makes the use of SQL more apparent. Moreover, SQL being a popular standard for storing, manipulating and querying data makes it suitable to be used as part of this thesis.

1.4. Detailed design

The requirements and specifications set as part of the previous phase are used in the design phase. The design at a higher level corresponds with the methodology described in Chapter 3. Within the detailed design, the parts of the workflow are defined more explicitly.

The BIM model and the BIM 4D model data can be exported and stored separately (e.g. direct export using the ODBC in Revit) and from the 4D model (e.g. export using XML to MS Project in Synchro 4D). However, Synchro 4D can export the property data of the BIM model similar to what can be retrieved when exported from Revit. The export function to Excel in Synchro 4D allows to export properties as well. Hence, relevant data can be stored in a single Excel file. An Excel file (Synchro export) is generated when a schedule is made (as-planned), when it is being updated as part of progress tracking (as-built-progress) and when the construction is finished (as-built). Thus, every single project has multiple Excel files (see Figure 55 and Figure 56). The complete process of which steps should be taken to generate different schedules within Synchro 4D and insert them into the database can be found in Appendix 6.

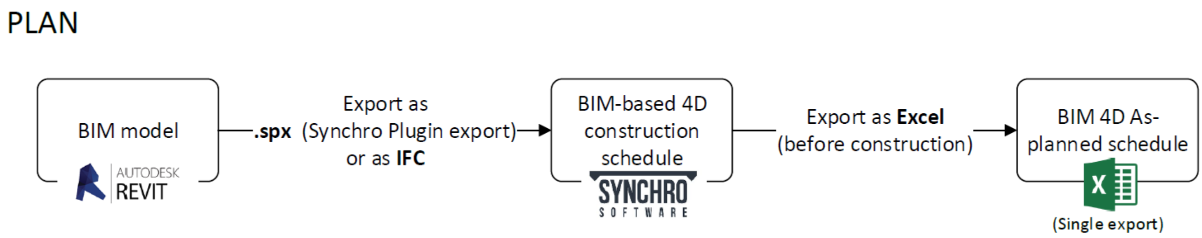


Figure 55: Creation of BIM 4D as-planned schedule (part of the Plan phase)

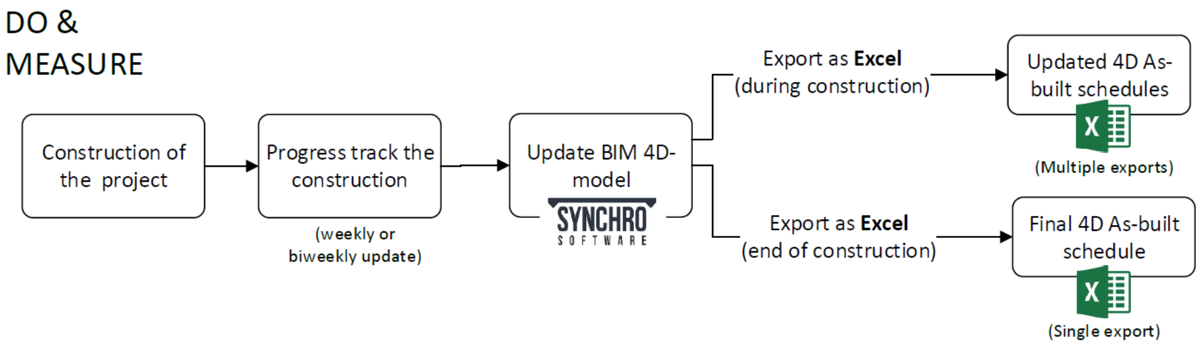


Figure 56: Creation of Updated BIM 4D as-built and final BIM 4D as-built (part of the Do & Measure phase)

The generated multiple Excel exports are extracted, transformed and loaded (ETL) to the relational database using a Python Script. Python is a programming language that is amongst many other things able to transform and load Excel data to MySQL. The Python script for the ETL procedure can be found in Appendix 5. MySQL is a popular open-source relational database. The data from the Excel file moves to the MySQL database according to the predefined table structure. The process from Excel file to data in MySQL is shown in Figure 57.

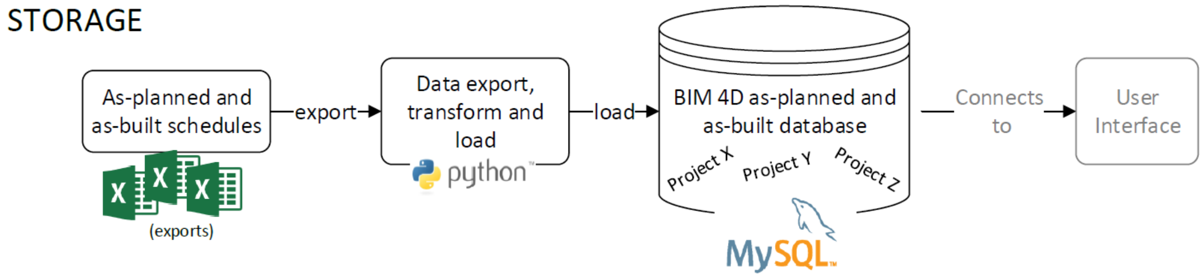


Figure 57: Workflow of storing data in MySQL (part of the Analysis and Improve phase)

Note that the choice of data export, transformation and data storage is merely an option chosen within this study. For example, the SQL export option as part of Synchro 4D could make the data transformation redundant. It is therefore recommended to explore the possibilities of storing the data within one's organization and in relation to the tools being used. Also, alternative to the SQL database used here, other types of databases (incl. NoSQL databases) could be used as well.

1.5. Classifications, taxonomies and ontologies

"Dictionaries, classifications and ontologies are an important facilitator for unambiguous definitions and data exchange. Textual descriptions of products written for humans are usually only available in semi-formal and natural language formats. These textual descriptions of products are severely limited in terms of machine readability which results in limited search and query capabilities" (Borrmann et al., 2018, p. 157). Well-known classification systems are NL-SfB, OmniClass and UniClass. A classification hierarchy of specialization (classes and subclasses) is represented by a type-of relation. Whereas, a composition hierarchy is represented by a part-of relationship (Figure 58). For instance, the part-of relationship can be the system breakdown structure of a construction project/object. Whereas the type-of relationship can be used to classify the execution of works related to the work breakdown structure for example. Ontologies can be used to represent several relationship types and aspects, whereas classification tables usually have a single relationship type (commonly the type-of relationship). Recently, the NTA (Nederlandse technishe afspraak) 8035 was published that aims for clarity in the application of semantic W3C languages for the exchange or sharing, and integration of data in the built environment (NEN, 2020).

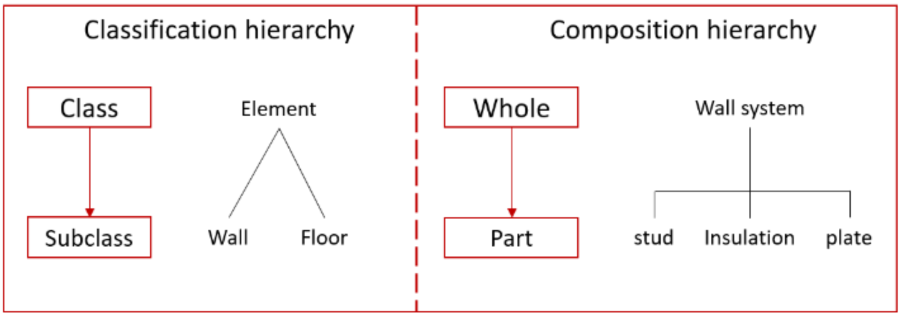


Figure 58: Classification hierarchy and composition hierarchy (adapted from Koninklijk Nederlands Normalisatie-instituut, 2015, p. 8)

To illustrate the example of classification for BIM element data, consider the element structure generated by IFC when a Revit to IFC export is made. Here, the family name, type and element tag are incorporated in the complete element name, allowing to make a distinction on different levels (Figure 59).

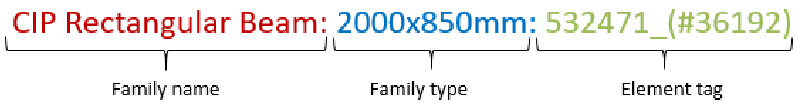


Figure 59: Example of element name when imported in Synchro part of an IFC model

When within Revit a Revit file is exported using the Synchro plugin, the classification of the element data will be incorporated as defined in Revit. Hence, the type of export determines the classification of the elements. Within Synchro 4D, Revit and IFC files can be imported. Although Synchro 4D can export IFC files, in this study only Excel export will be used.

1.6. ERD

Entity Relationship Diagrams, also known as ERD, graphically describe Entity Relationship Models (ERM) for the use in database design. An ERD is mostly developed for designing relational databases in terms of concept visualization and physical database design. Therefore, in order to develop the database in MySQL, an ERD is drawn up. As part of the ERD, fundamental decisions are taken of how the data is represented and how it relates to each other. Figure 60 shows the ERD of the tables in Figure 53, showing the entities (Task table and Equipment table), entity attributes (Task_ID, Taskname etc.), primary keys (Task_ID and Eqp_ID), foreign key (Equipment table_Eqp_ID) and relationship. Furthermore, the datatypes are indicated such as INT (integer) and VARCHAR(6) (Variable Character Field).

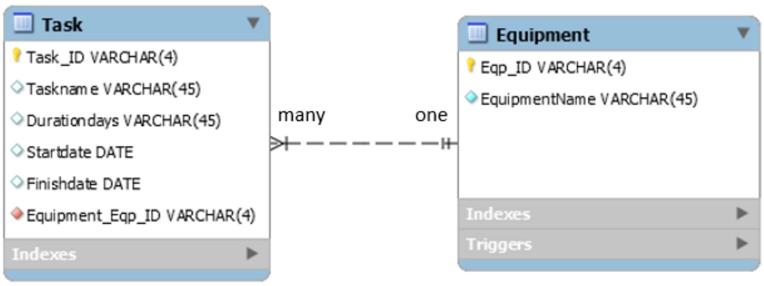


Figure 60: Example of ERD of the tables part of Figure 53

Alongside that the relationship shows that two entities are related (e.g. tasks and equipment), its cardinality is also defined. Cardinality defines the possible number of occurrences between rows in one table associated with the number of occurrences of rows in another table. Three common cardinalities are one-to-one, one-to-many, and many-to-many. An example of one-to-many is shown in Figure 60, where one piece of equipment can relate to many tasks; whereas, one task can only relate to one piece of equipment. The ERD developed for the storage of BIM-based as-planned and as-built data is shown in Figure 61, a readable version can be found under Appendix 3. It is chosen to develop an own ERD since it provides more flexibility to be tailored for the proof of concept of this study.

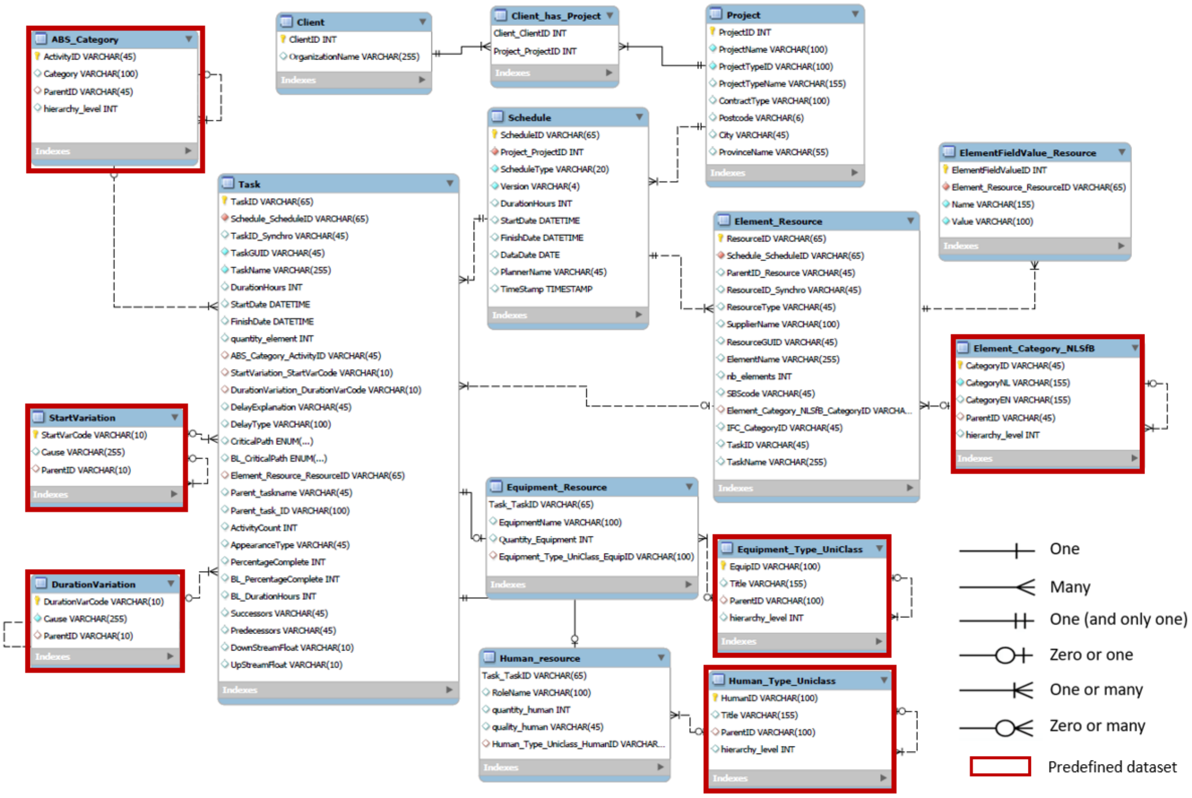


Figure 61: Proposed ERD for BIM-based as-planned and as-built database design (image is for indicative purposes, full ERD can be found in Appendix 3, the red squares indicate the predefined tables)

The ERD is constructed in MySQL Workbench 8.0 CE, one can forward engineer the ERD to SQL code for the creation of these tables, associated relations and properties that come with it into a RDBMS. The SQL code can be found in Appendix 4.

Since classifications enhance search and query capabilities, predefined classification tables are defined as part of the ERD. The ERD contains classification tables that relate to the elements in the BIM model, tasks, start and duration variation causes, roles (labour) and equipment.

It is chosen to classify the elements according to the NL-SfB in Synchro 4D with a classification code. Depending on the BIM model that is imported into Synchro 4D, the elements will be automatically classified with the classifications of the particular modelling tool that is used. As a result, the data can be retrieved and analysed according to those categories. Task names, on the other hand, are not automatically classified. Hence, it is chosen to classify the task according to a standard Activity (task) Breakdown Structure (ABS), in this case, defined by BAM Infra. Classification tables from the UniClass are applied to the roles and equipment assigned to tasks. For the start and duration variation causes, tables are predefined according to Table 12 in Appendix 1. The classification of all aspects related to tasks, thus the categorization of the activities (task names), roles, equipment and variation causes are applied at the deepest level of the schedule where one element relates to one task. Ultimately, task information can be queried according to its associated classifications. Within this process, it is necessary to manually assign classification codes within Synchro 4D.

1.7. ERD tables

As described earlier, every table should have a primary key that uniquely defines every row. A single key such as the TaskID, TaskGUID, ResourceID or ResourceGUID (that is generated within Synchro 4D) is by itself not a suitable primary key, since this key is not changing when a schedule is adjusted and exported again for the same project. Therefore, a unique schedule identifier (ID) is created every time a schedule is loaded to MySQL using the Python Script. The primary keys of the Task, Element_Resource, Equipment_Resource and Human_Resource tables consist of a single primary key that is a combination of the schedule ID and the original ID created in Synchro 4D. For instance, the Schedule ID equals 1234 and the original Task ID in Synchro 4D equals 0123, then the primary key in MySQL (i.e. the TaskID) equals 12340123.

The ERD is made up from many different tables, hereunder they are ordered to some extent and an explanation is given. Several tables have been created for predefined classification datasets. These classifications only need to be inserted once into the MySQL database. Furthermore, it is explained how the data is imported in MySQL using the Python script.

First, the path is defined for the Excel file that needs to be inserted into MySQL. Also, a connection is established with MySQL using the ‘mysql.connector’. A snippet of the Python script is shown below.

Python

```python
import openpyxl as xl
import mysql.connector
import re
from datetime import datetime
import uuid

#get and insert path of workbook
filename = ('C:\\Users\\path\\BAM-viaduct_As-Planned.xlsx') #<-- Excel workbook

wb = xl.load_workbook(filename, read_only=True, data_only=True)

# Establish a MySQL connection
#db = mysql.connector.connect(user='scott', password='password', host='127.0.0.1',
database='schedules')
db = mysql.connector.connect(user = "root",
                                db = 'bim_schedules') #<-- name of database


# Get the cursor, which is used to traverse the database, line by line
cursor = db.cursor()

# Creates Unique QUID for every new inserted schedule
ScheduleID = str(uuid.uuid4())
```

Four basic functions that can be performed on a relational database are the Create, Read, Update and Delete (CRUD) functions. To insert data into MySQL, the Create (insert) function is used. A simple example of an insert statement is shown below as part of MySQL:

MySQL

```sql
INSERT INTO table_name (column1, column2, ...)
VALUES (value1, value2, ...);
```

General information on the project and the schedule are imported into four SQL tables (Table 4). An example of data that is defined within Synchro 4D, then exported to an Excel file and inserted into MySQL as part of the Client table, is described below.

Table 4: General tables part of the ERD

Tables	Explanation
Client	The Client table provides data on the client of a project.
Client_has_Project	This table maps the Clients and Projects with a many-to-many relationship (a project can have more than one client and vice versa)
Project	Provides data on the project with attributes as Project ID, Project Name, Project Type and Postcode. The project type ID is similar to the code ID used in the NL-SfB classification table.
Schedule	Provides data on the schedules that are part of the database. With attributes as Projects ID, Schedule type (As-planned/As-built-Progress/As-built), version, duration of the entire schedule in hours, start date and finish date.

Project information can be listed as part of the activity codes tab in Synchro 4D as shown in Figure 62.

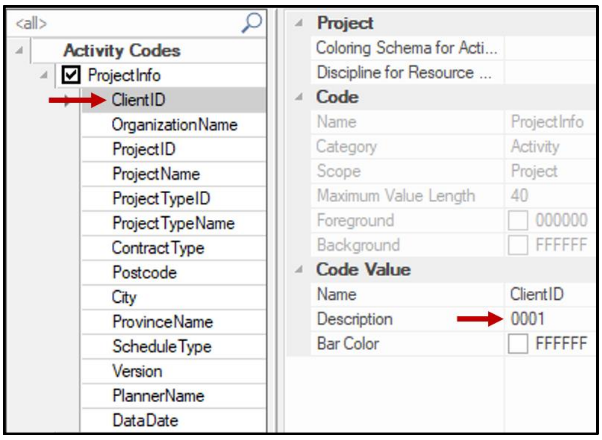


Figure 62: Example of Project data defined in Synchro 4D

As part of the Excel export, this specific data is retrieved from the 'codes' sheet in Excel. Subsequently, the data for the Client table is inserted into MySQL using the Python script as shown below.

Python

```python
ws = wb ['Codes']

# SQL Insert statement
add_client = ("INSERT IGNORE INTO Client "
               "(ClientID, OrganizationName) "
               "VALUES (%s, %s)")

# loops through the rows in Excel and loads the data into MySQL
for row in ws.iter_rows(min_row=2):
    for cell in row:
        if cell.value == "ClientID":
            ClientID = ws.cell(row=cell.row, column=6).value
        if cell.value == "OrganizationName":
            OrganizationName = ws.cell(row=cell.row, column=6).value

            data_client = (ClientID, OrganizationName)
            # load data for every row into MySQL by replacing %s with values
            cursor.execute(add_client, data_client)
```

Result of the table in MySQL:

ClientID	OrganizationName
1	Eindhoven University of Technology
NULL	NULL

A comparable approach, like the one described above, is used to insert the data from Excel into MySQL for the other tables except for the predefined classifications tables (which are only inserted once).

Table 5: Task-related tables part of the ERD (predefined tables indicated in grey)

Tables	Explanation
Task	The task table contains the tasks created as part of the Synchro 4D schedule. The task table has attributes such as TaskID, TaskGUID, TaskName, Duration in hours, start and finish data, critical path (yes/no), percentage complete and appearance type. Foreign keys within the table relate to the schedule table, Element_resource table, ABS category table, start and duration variation tables.
ABS category	ABS (Activity Breakdown Structure) category is a predefined table that contains data of standard Activity codes (task codes) on various levels that can be linked to a task (e.g. AC001: "Pour concrete").
StartVariation & DurationVariation	The start and duration variation tables are predefined tables with causes for task starting & duration variations according to Table 12 in Appendix 1.

To provide a better understanding of the actual data as part of a table, a section of data belonging to the Task table and how it refers to the ABS category and DurationVariation table is shown in Figure 63.

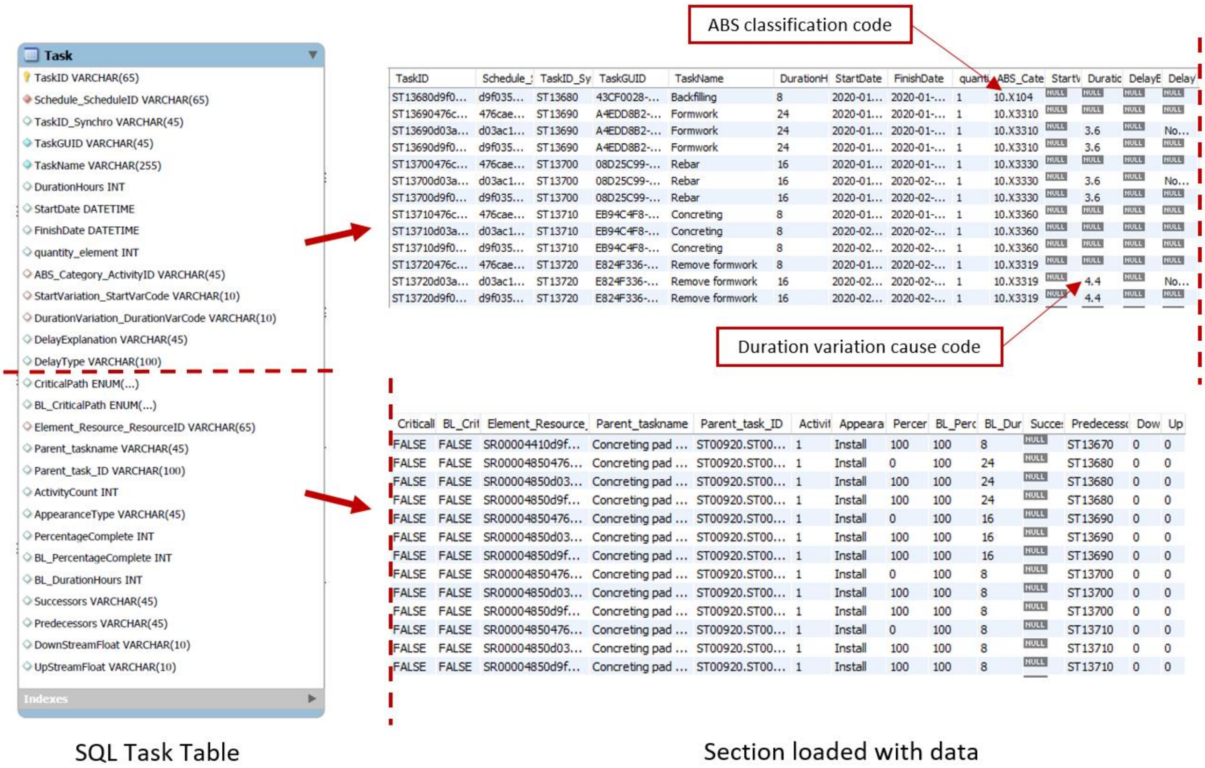


Figure 63: Section of data in the Task table as part of MySQL (blue line indicates MySQL table)

Table 6: Human-related tables part of the ERD (predefined table indicated in grey)

Tables	Explanation
Human_Resource	Provides data on the human/labour role name, quantity and quality that is employed for a specific task. Human resources have a zero-to-one or one-to-one relationship with the task table.
Human_Type_UniClass	A predefined table of Roles from UniClass which is converted to suit the structure of ID, name and Parent ID.

Table 7: Equipment-related tables part of the ERD (predefined table indicated in grey)

Tables	Explanation
Equipment_Resource	Provides data on the equipment name and quantity that is employed for a specific task. Equipment resources have a zero-to-one or one-to-one relationship with the task table.
Equipment_Type_ UniClass	A predefined table of Equipment from UniClass which is converted to suit the structure of ID, name and Parent ID.

Table 8: Element-related tables part of the ERD (predefined tables indicated in grey)

Tables	Explanation
Element_Resource	The Element_Resource table contains the elements (e.g. walls, columns etc.) as part of the BIM model. The table has attributes such as ResourceID, Element name, the number of elements, supplier name and a system breakdown structure (SBS) code, also referred to as the Object Breakdown Structure. A single SBS code (e.g. 1.1.1.1) can relate to one or more elements. At a higher level (e.g. 1.) many elements relate to one code. Every element can have assigned one SBS code. This code differs for every project since every project is broken down differently. Foreign keys within the table relate to the NL-SfB element classification table. The NL-SfB classification code is embedded as part of the properties per element in the BIM model.
ElementFieldValue_ Resource	Provides data on every element contained in the BIM model in an entity-attribute-value structure (e.g. entity,*"length","1000"*).
Element_Category_ NL-SfB	Provides a predefined dataset of the NL-SfB categories with their ID, name and hierarchy level.

To insert the data into the ElementFieldValue_Resource table, a slightly different approach is taken. In general, the data inserted as part of a Synchro 4D Excel export into MySQL can be well represented using individual attributes (i.e. single columns) that are predefined in MySQL and only apply to one specific variable. However, the property data (such as length, width etc.) as part of the elements in the BIM model are different for every single model. Therefore, it is difficult to predefine every attribute as part of an SQL table. As an example, when a Revit export file is exported to MySQL using the ODBC, the file automatically creates 243 tables and related attribute columns to the element family type and name.

The element property data can be exported as part of the Synchro 4D Excel export 'resource' sheet. Every resource ID is related to an element in the BIM model. To insert the data into MySQL, the data is transposed and loaded into an entity-attribute-value (EAV) model. The transformation is indicated in Figure 64.

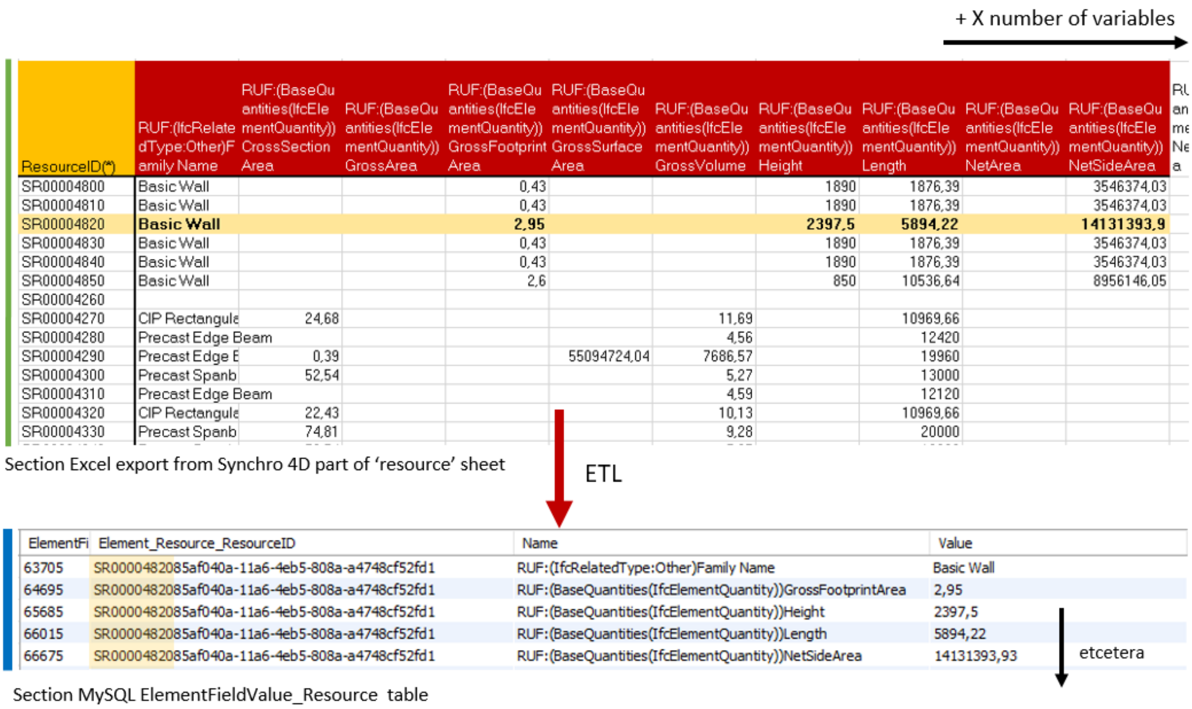


Figure 64: Transpose transformation from Excel to MySQL (simplified)

The following Python script is made to load the element-related properties and other created user fields into MySQL:

Python

```python
add_ElementFieldValue_Resource = ("INSERT IGNORE INTO
                                  ElementFieldValue_Resource "
                                 "(Element_Resource_ResourceID, Name, Value) "
                                 "VALUES (%s, %s, %s)")

AmountColumn = ws.max_column - 12
#sets amount of columns equal to number of variables

d = {}
for i in range(AmountColumn):
    d["variable_name" + str(i)] = ws.cell(row=1, column=13+i).value
#retrieves 'Name' and stores them in a dictionary

for i in range(AmountColumn):
    for row in ws.iter_rows(min_row=2):
        ResourceID_Synchro = row[1].value
        variable_data = row[12+i].value
        VariableNameInput = 'variable_name' + str(i)

        data_ElementFieldValue_Resource= (ResourceID_Synchro + ScheduleID,
       d.get(VariableNameInput), variable_data)
        cursor.execute(add_ElementFieldValue_Resource,
       data_ElementFieldValue_Resource)
```

It would be beneficial to also insert the unit (e.g. m^2 or kg) and type in the ElementFieldValue_Resource table. However, this data is not exported as part of the Excel export. Although the EAV model is widely applied, it should be recognized that they do not allow for referential integrity and require to be stored as a VARCHAR. Since the property names (e.g. RUF:ifcGUID as in Figure 64) are not predefined as attributes of a particular column, it is important for the property names to be uniform. The uniformity in the name allows to group data of a similar property more easily. Again, the use of IDM and modelling guidelines can help in this regard. Within this process, the required attributes and related attributes types can be defined.

1.8. Fundamental relationship within the ERD

Getting into more depth on the ERD provides a better understanding of how the database has been set up. Two important tables of the ERD are the Task table and the Element_Resource table, the Task table has a many-to-one relationship with the Element_Resource table. The existence and type of relationship are both fundamental for the extraction of task data. Without a relationship between tasks and elements, it is not possible to query meaningful task data. For instance, the task name of example 1 in Figure 66 is “construct slab & columns” with a duration of 10 days. If this data is stored without a connection to the three elements and one queries this task from the database, one will be able to query the task name and duration without any knowledge of the properties of the elements the task relates to. As a result, the data will be unusable. When the connection is in place, the type of relationship (cardinality) is furthermore of interest. To illustrate this, consider the following three examples in Figure 65 and Figure 66.

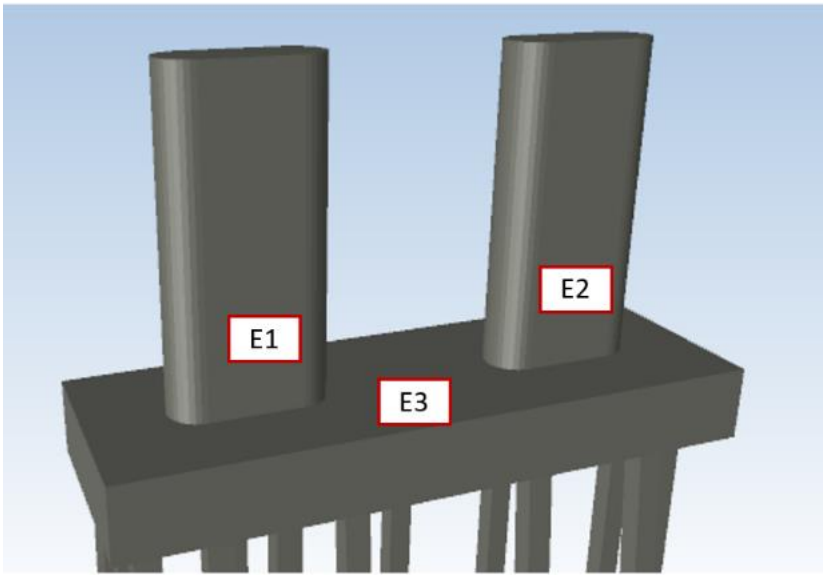


Figure 65: Three elements that are part of a 4D schedule

1 Task <-> many Element_Resource

Example 1

*Task ID	Taskname	Duration days	3D Resource
T1	Construct slab & columns	10d	E1, E2, E3

Many Tasks <-> 1 Element_Resource

Example 2

*Task ID	Taskname	Duration days	3D Resource
T1	Prepare construction slab	1d	E3
T2	Construct slab	1d	E3

1 Task <-> 1 Element_Resource

Example 3

*Task ID	Taskname	Duration days	3D Resource
T1	Construct columns	8d	
T2	Construct column 1	4d	E1
T3	Construct column 2	4d	E2
T4	Construct slab	2d	E3

Figure 66: Task vs Element_resource relationship (one-to-many, many-to-one and one-to-one)

For example, one wants to estimate the duration to construct a slab based on the entry of example 1 into the database. When trying to query the data, one will only be able to query the duration that relates to three elements (entities). Therefore, it is not possible to attach a specific duration to a specific element. Dividing the duration by three results in a value that is inaccurate since the elements are significantly different and may have other factors driving the duration of the task. This furthermore creates a problem for the comparability with other elements of the same element type (entity type). While, when multiple elements of the same element type are stored in the database (e.g. columns are stored separately), it becomes possible to query an average duration based on the different durations stored in the database per unit, per square meter, per meter, etc.

Furthermore, if the primary key is representing one element in the Element_resource table and it is used as a foreign key in the task table, then it can only contain one foreign key for each row. In example 1, a single row contains three foreign keys (E1, E2 and E3), hence this is not possible. In conclusion, storing the data with a one-to-many relationship between tasks and elements is not suitable.

Example 2 and 3 represent a many-to-one and one-to-one relationship that are both suitable possibilities for storing the data. For instance, if the data of example 2 is stored in the database and queried, one will find the duration of an element belonging to a specific task. When multiple entries of the same element type are made in the database with similar tasks, one can query the average duration. Example 3 also contains a summary task for more than one element. However, underneath the summary task, the elements are ultimately represented with a many-to-one relationship. The proposed fundamental relationships are in line with the concept of BIM-based scheduling where the elements form the basis to generate a construction schedule.

1.9. Limitations of the ERD design

The ERD design has implications on the way exports are generated from Synchro 4D. To insert the data into the MySQL database correctly, the attributes in Synchro 4D should match the attributes in the ERD design. If one wants to add an attribute (e.g. a cost attribute in the Task table), it would be necessary to alter the table and update the Python script. Hence, in this case, it is important to well-define the required attributes as part of the MySQL tables beforehand.

The ERD design itself also has some limitations. The design allows to only assign a single equipment name and type, a single role name and type and a single start and duration variation cause to a single task. While in practice more than one type of equipment or role can be required for a task to be executed. As a solution, one-to-many and many-to-many relationships can be formed, however, the current structure of the Excel export made this solution difficult to realise. Furthermore, the current design did not include or predefine attributes of factors related to the internal environment and exogenous factors. Regarding the latter, a system could be built on top of the database that allows to consume an external data source that includes weather data. Alternatively, a linked data approach could be tried.

1.10. Conclusion

The first section of this chapter provides an answer to the fourth sub-research question regarding the way BIM-based as-planned and as-built scheduling data can be captured and stored in a database. For the representation of data in the AEC, so-called data modelling concepts are used. Generally, a data model can be categorized as relational stores (SQL) or NoSQL stores. The last decades, relational stores have been the dominant solution for most applications. SQL databases make use of predefined schemas and are particularly suited for structured data. In addition, they deal well with data integrity and have the ability to perform complex queries. Highly scalable, more flexible and less robust NoSQL stores emerged as a response to the growth in data size and data variety while compromising on reliability and consistency. NoSQL stores are typically suited for semi-/unstructured data. Graph databases are part of NoSQL stores and designed to store linked data. The same principle is used for the storage of data part of the semantic web.

The choice of database, therefore, depends on how the data is currently represented and the consequent analysis i.e. the initial objective of storing the data. Considering the current export capabilities of the tools used for the proof of concept, the use of a relational store (MySQL) has been chosen. As part of Synchro 4D, Excel exports of the BIM-based schedule are made for the situations: as-planned, as-built in progress and the final as-built. Subsequently, a Python script is developed to extract, transform and load the data from Excel into MySQL. The cardinality as part of the ERD between the task table and BIM model element table is found to be fundamental to extract meaningful task duration data. Moreover, since classifications ensure unambiguous definitions and enhance search and query capabilities, several predefined classification tables are included as part of the ERD.

The ERD that is developed for the proof of concept has some limitations. The model incorporates a one-to-one relationship between the task and labour table and between tasks and equipment table. To assign multiple types of labour and equipment to a single task, the ERD should be adjusted with a many-to-one or many-to-many relationship. Furthermore, it is recognized that the process of exporting an Excel and using a Python script to load the data into MySQL can be cumbersome if adjustments need to be made upon the tables, attributes or relations within MySQL. For instance, when one wants to add a certain attribute, the script should be adjusted accordingly. However, the process of using a Python script can become redundant if a more sophisticated system is developed.

2. Analysis and (re)use

This section deals with the second part of the system development, the analysis and (re)use of the stored BIM-based as-planned and as-built scheduling data. As part of this chapter, the process and requirements for the analysis and (re)use of the data are outlined. This section provides the answer to sub-research question 5 “How can historical BIM-based as-planned and as-built scheduling data be analysed and (re)used for estimating task durations?”. Once the data is stored inside the database, the data can be analysed and (re)used in a variety of manners. The proposed system in Chapter 3 outlines that the data should allow being queried, analysed and visualized (Figure 67). Hereafter, these three items are discussed.

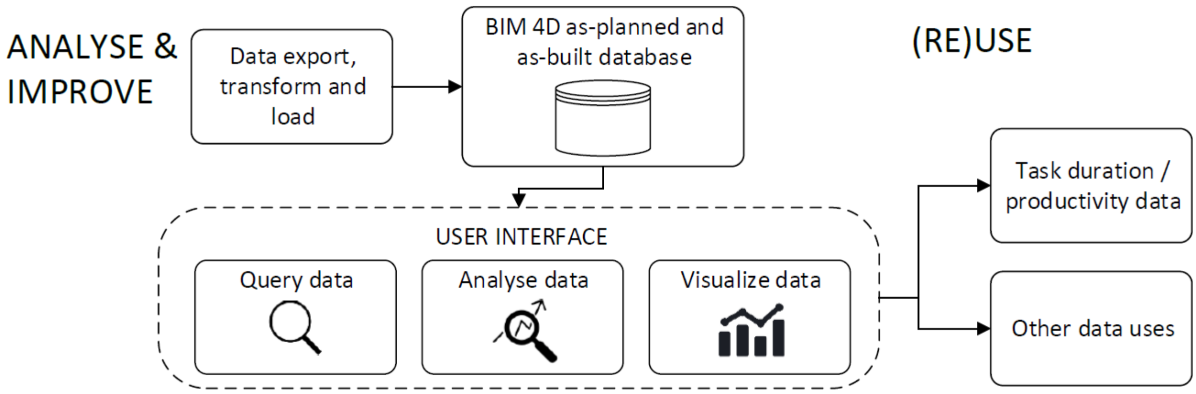


Figure 67: Analysis and improve phase

2.1. Query

The data query language (DQL) is used to query (select) data from a SQL relational database i.e. the MySQL database in this case. Remark that this is the Read (C**R**UD) function as part of the four basic functions. Specific data can be retrieved from one or more tables using an SQL-statement. A basic SQL-statement is:

```sql
SELECT column1, column2, …
FROM table_name;
```

The query returns the data for the columns specified as part of a specified table. There are many SQL keywords similar to *SELECT* and *FROM* that can be used to perform operations in the database such as *WHERE, ORDER BY, GROUP BY, JOIN, LIKE, COUNT, MIN, MAX, SUM, AVG* etc. As an example, a query is performed on the data of the tables in Figure 53. The goal of the query is to retrieve the average task duration in days when equipment X is being used. The query:

```sql
SELECT AVG(Durationdays)
FROM task
INNER JOIN equipment ON task.Equipment_Eqp_ID = equipment.Eqp_ID
WHERE equipmentname = 'Equipment X'
```

The result of the query:

	AVG(Durationdays)
▶	2

The average function (AVG) is applied to the column *durationdays*, the *INNER JOIN* allows to combine rows from two or more tables and the *WHERE* clause is used to only extract records that fulfil a specified condition, in this case, a certain equipment name. Thus, queries can be written to suit the retrieval of specific data (customized information retrieval) i.e. what is required by a planner for the estimation of task durations. For inexperienced users, it is hard to pose SQL queries, as they are required to be proficient in SQL syntax and have a thorough understanding of the underlying schema (Fan, Li, & Zhou, 2011; Kimball & Ross, 2013). Therefore, although outside the scope of this study, a user interface (UI) needs to be built on top of the database. The use of the interface allows to retrieve the data in a user-friendly manner. Queries are fundamental for retrieval and subsequent other uses. Hence, when one wants to visualize or analyse the data, the data needs to be queried in advance. In this study, the distinction is made between queries, analysis and visualization. However, queries can also be considered as a form of analysis. Performing queries on the data is the primary objective of this study.

2.2. Analysis and visualization

Besides using queries for information retrieval, other analysis methods can be applied. In general, the data can be visualized (and analysed) e.g. with the use of dashboards or the data can be analysed using data analysis techniques.

2.2.1. Data mining

Data mining is a subset within the process of Knowledge Discovery in Databases (KDD). As part of this process, the aim is to utilize tools that assist humans in extracting previously unknown and potentially useful information (knowledge) from large volumes of data (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). The process is shown in Figure 68.

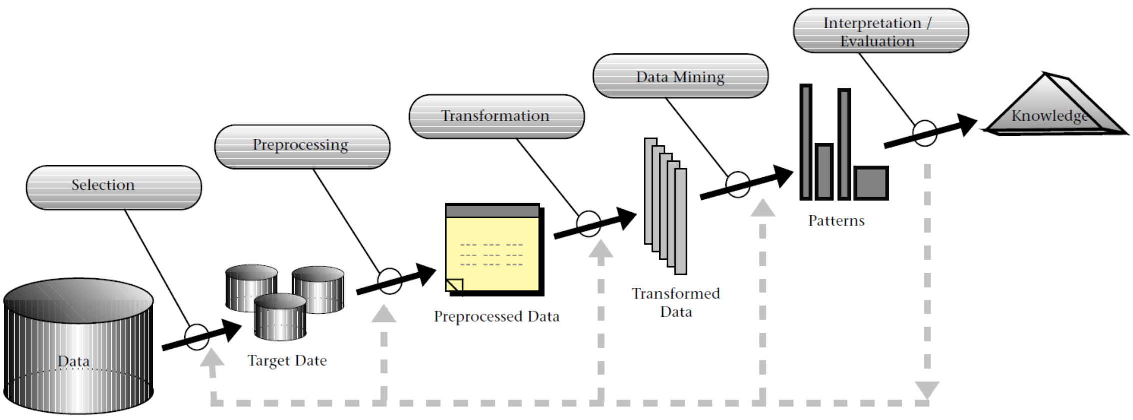


Figure 68: Overview of the steps within the KDD process (Fayyad et al., 1996, p. 41)

Depending on the type of database e.g. SQL or a specific type of NoSQL database, different tools can be used. There are several data mining functionalities. The interdisciplinary nature of data mining led to the adoption of many techniques such as machine learning, statistics, data warehouses, visualization etc. These data mining tasks can be classified as descriptive (unsupervised) or predictive (supervised). Descriptive techniques characterize the dataset, whereas predictive techniques aim to make predictions based on the dataset. These techniques are described below.

Characterization and discrimination

Data can be associated with classes or concepts. For instance, data can be characterized by summarizing all task durations related to wall elements (element type). Data discrimination is aimed at the comparison of comparative classes. An effective method to summarize and characterize data is with the use of OLAP (Online analytical processing) cubes (Han, Kamber, & Pei, 2012). These OLAP cubes are usually constructed as part of a data warehouse. OLAP systems allow users to interactively query and automatically aggregate the data (Kimball & Ross, 2013). The use of OLAP cubes allows one to slice and dice the data to discover meaningful patterns. A schematic representation of an OLAP cube is shown in Figure 69. The output of the OLAP operations can be presented in bar charts, pie charts, graphs etc.

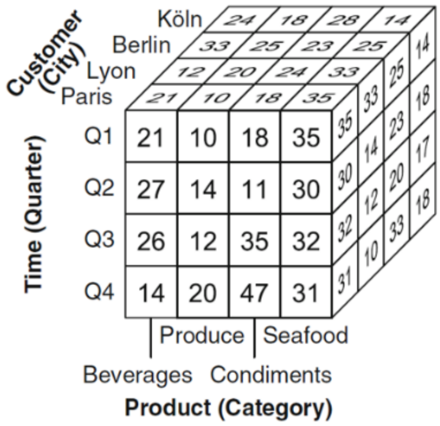


Figure 69: OLAP cube (Vaisman & Zimányi, 2014, p. 60)

Other techniques

Exploratory data analysis is mostly used to get insight into a dataset using visual and interactive tools, for example using a scatterplot. The aim of *frequent pattern mining* is aimed at the discovery of interesting associations and *correlations* within data. For instance, frequent occurring sequences of tasks can be mined. *Classification* is the process of finding a model or function that describes and distinguishes data classes. This process predicts the value of a variable based on other variables. Examples are the use of a decision tree, neural networks, support vector machines (Han et al., 2012). *Regression***,** on the other hand, models continuous-valued functions. *Clustering* analyses data without taking their labels into account. Often, this is the case for big data where these labels are simply not available. Clustering is aimed at finding high similarity between one cluster and high similarity between another cluster. *Outlier analysis* is conducted to find data that does not find the general behaviour of the data (Han et al., 2012).

2.2.2. Dashboards and KPIs

Dashboards are popular visualization tools in business intelligence. A dashboard usually consists of multiple visuals such as KPI's, charts and are mostly used to effectively measure, monitor and manage business performance (Vaisman & Zimányi, 2014). KPI's are quantifiable measurements used to estimate the effectiveness and performance of the processes and business strategies within an organization. Usually, KPI's are included in dashboards and typically have a target value, a threshold and minimum value.

2.3. Data repository: the data warehouse

Generally, databases are designed to record data whereas data warehouses are designed to analyse data. A data warehouse is a common repository of data that is collected from different sources and reduced through extraction, transformation, integration, and cleansing processes to a form that can be used for certain reporting and analysing purposes (Vaisman & Zimányi, 2014; Maheshwari, 2015). Data warehouses are fundamental for business intelligence (BI). "Data analytics is the process of exploiting the contents of a data warehouse in order to provide essential information to the decision-making process" (Vaisman & Zimányi, 2014, p. 7). The architecture of a typical data warehouse is presented in Figure 70.

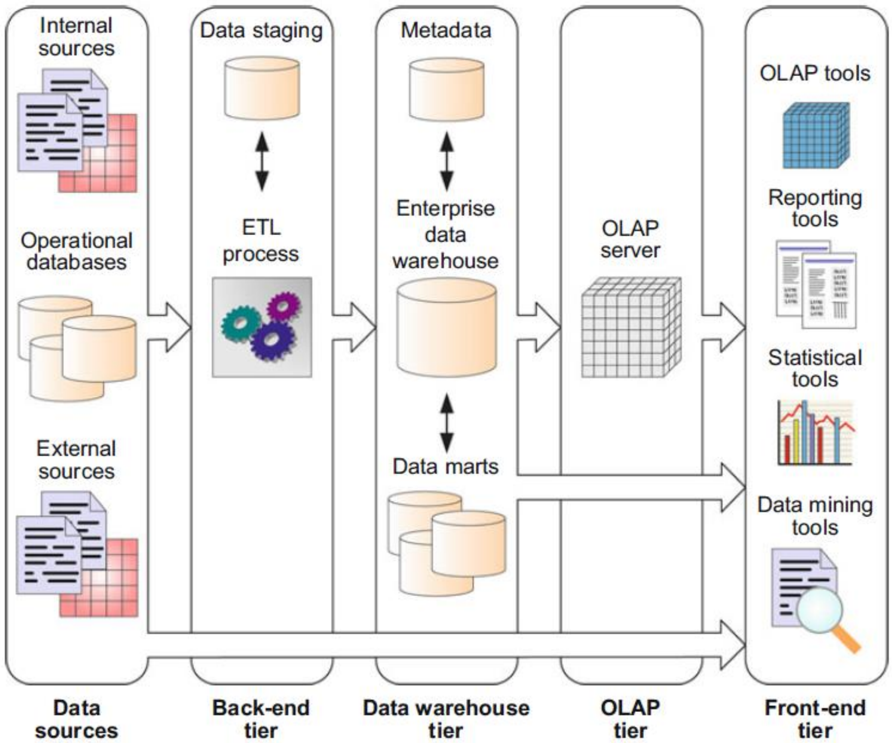


Figure 70: Typical data warehouse architecture (Vaisman & Zimányi, 2014, p. 77)

OLAP operations mostly rely on a multidimensional database. Dimensional modelling is a widely accepted technique for presenting analytic data that is understandable to a business user and delivers fast query performance (Kimball & Ross, 2013). These models often only include specific data from the original data source or databases that are needed for the analysis to take place. Furthermore, these models do not need to deal with the insert, update or delete operations that the operational databases need to deal with. Hence, they are only required to read the data.

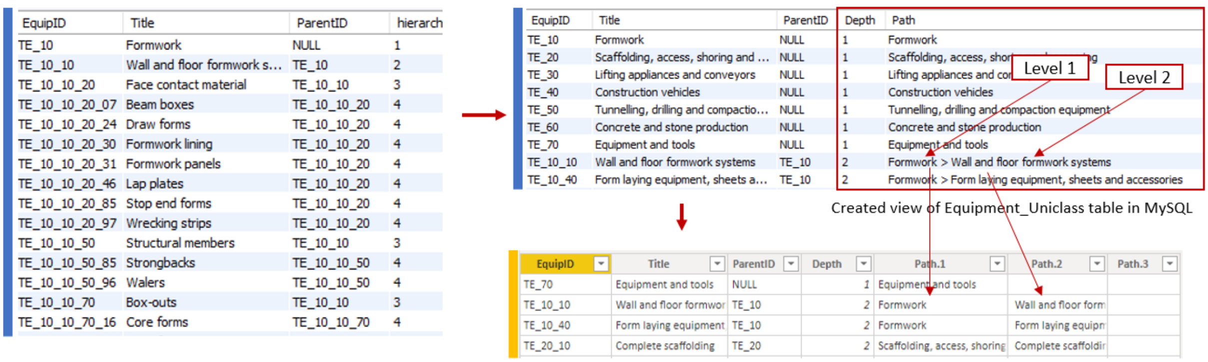
2.4. Application of queries, analysis and visualizations

The data query language (DQL) can be used to query data from the database. To query the data in a user-friendly manner, a user interface (UI) is built on top of the database. Within this study, it is out of scope to develop such a user interface.

Since data warehouses typically provide the means for data analysis, it would be of interest to perform ETL processes on the designed database as part of Section 4.1 and construct a data warehouse that is optimized for data analytics. However, for the proof of concept of this study, it is out of scope to construct a separate data warehouse. Moreover, the developed ERD of the database already has pseudo dimensions included in the form of the predefined classification tables. Hence, these dimensions can be used to roll-up and drill-down the data somewhat like the use of an OLAP cube. Furthermore, the developed ERD is directly designed with the purpose to derive task/productivity data from. Nevertheless, it would be of interest to study how such a database could be implemented into an organizational-wide data warehouse system.

In order to derive task/productivity data, insight into the BIM-based scheduling database should be provided. At this stage, the aim is on the data as it exists in the database. SQL queries and dimensional slicing, filtering, and pivoting using OLAP can be used for the retrieval of this data. It is chosen to use PowerBI as a medium (user interface) to retrieve the data from the database. However, to fully utilize the data it is recommended to develop a custom user interface. PowerBI can be utilized as an interactive tool that allows for the data to be retrieved (queried) and visualized. The software is capable of coping with data from different data sources and relies mostly on structured data. Moreover, it is largely capable of handling relationships between tables similar to the relationships that are established as part of SQL. A direct connection can be made between MySQL and PowerBI. However, some data needs to be transformed to make it usable in PowerBI. As part of SQL, so-called 'views' or 'stored procedures' can be used for repetitive actions in the database. Multiple views are created to make the data useable in PowerBI. Two types of views are created:

1. The predefined classification tables are originally stored within an Adjacency List Model (ID (child), Name, Parent) relationship. PowerBI does not allow for roll-up and drill-down operations using this list. Therefore, a view in MySQL is created that allows PowerBI to perform these operations. To do so, a so-called flattened table is made. Thus, the structure of the data is adjusted for it to be used in PowerBI. This process is shown in Figure 71.



EquipID	Title	ParentID	hierarch
TE_10	Formwork	NULL	1
TE_10_10	Wall and floor formwork s...	TE_10	2
TE_10_10_20	Face contact material	TE_10_10	3
TE_10_10_20_07	Beam boxes	TE_10_10_20	4
TE_10_10_20_24	Draw forms	TE_10_10_20	4
TE_10_10_20_30	Formwork lining	TE_10_10_20	4
TE_10_10_20_31	Formwork panels	TE_10_10_20	4
TE_10_10_20_46	Lap plates	TE_10_10_20	4
TE_10_10_20_85	Stop end forms	TE_10_10_20	4
TE_10_10_20_97	Wrecking strips	TE_10_10_20	4
TE_10_10_50	Structural members	TE_10_10	3
TE_10_10_50_85	Strongbacks	TE_10_10_50	4
TE_10_10_50_96	Walers	TE_10_10_50	4
TE_10_10_70	Box-outs	TE_10_10	3
TE_10_10_70_16	Core forms	TE_10_10_70	4

EquipID	Title	ParentID	Depth	Path
TE_10	Formwork	NULL	1	Formwork
TE_20	Scaffolding, access, shoring and ...	NULL	1	Scaffolding, access, shor...
TE_30	Lifting appliances and conveyors	NULL	1	Lifting appliances and co...
TE_40	Construction vehicles	NULL	1	Construction vehicles
TE_50	Tunnelling, drilling and compactio...	NULL	1	Tunnelling, drilling and compaction equipment
TE_60	Concrete and stone production	NULL	1	Concrete and stone production
TE_70	Equipment and tools	NULL	1	Equipment and tools
TE_10_10	Wall and floor formwork systems	TE_10	2	Formwork > Wall and floor formwork systems
TE_10_40	Form laying equipment, sheets a...	TE_10	2	Formwork > Form laying equipment, sheets and accessories

EquipID	Title	ParentID	Depth	Path.1	Path.2	Path.3
TE_70	Equipment and tools	NULL	1	Equipment and tools		
TE_10_10	Wall and floor formwor	TE_10	2	Formwork	Wall and floor form	
TE_10_40	Form laying equipment	TE_10	2	Formwork	Form laying equipn	
TE_20_10	Complete scaffolding	TE_20	2	Scaffolding, access, shoring	Complete scaffoldir	

Original Equipment_Uniclass table in MySQL

Table in PowerBI with individual columns for each level

Figure 71: Process steps to allow for drill-down and roll-up operations in PowerBI (blue lines indicate the MySQL tables and the yellow line indicates the PowerBI table)

Within PowerBI, a matrix table can be used to drill-down different levels. An example of the Equipment_Uniclass table is shown in Figure 72.

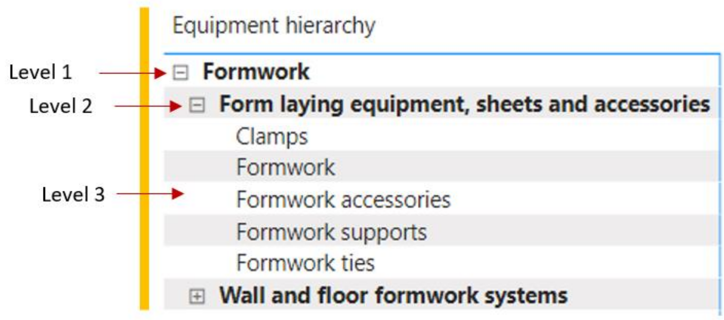


Figure 72: Drill-down of equipment hierarchy within PowerBI

2. A view is created for the ElementFieldValue_Resource table that transposes the entity-attribute-value structure into a table where every attribute has its own column. It should be noted that as the volume of the database generally increases, this operation will result in countless columns.

The SQL code can be found in Appendix 4. The other steps that need to be taken to import the data in PowerBI are described in Appendix 7.

Within PowerBI, several slicing and dicing operations can be executed; for instance, sorting, grouping and filtering the data. As part of filtering, specific attribute data can be searched for. Thus, task duration data can be filtered based on the input factors, internal environment and exogenous factors when contained in the database. Related to exogenous factors, it would be of interest to utilize weather data from an external source. A system that would be built on top of the database could allow to also consume this external source. Alternatively, a linked data approach could be tried. Moreover, after slicing and dicing the data, descriptive statistics (aggregate functions) such as sum, minimum, maximum, average (arithmetic mean), standard deviation, median and count can be derived. It is also possible to perform exploratory analysis on the data to find patterns and relationships, for instance by using box plots or histograms. Additionally, user-defined calculations can be performed in PowerBI using the measure (formula) function. In doing so, productivity measures can be derived from the data, such as unit/person-hour, m^2/person-hour etc.

In later stages, other data analysis methods can be exploited. Thomas & Yiakoumis (1987) used regression analysis to derive the factor model from the data (described in Section 2.2.1.2). Generally, reliable statistical analyses require a large dataset. The influence of the independent variables (input factors, internal environment and exogenous factors) on the dependent variable (i.e. productivity/duration) can be determined from the model. Subsequently, the model can be used to predict the productivity/task duration. A more preliminary version of the factor model would be to derive a productivity figure based on the statistical relationship between two variables, x (independent) and y (dependent). For instance, between person-hours and diameter of a concrete column for the erection of formwork. Moreover, statistical associations between multiple variables can be indicated with a correlation matrix. Besides regression analysis, many more data analysis methods could be applied. Data analyses could be leveraged on multiple scales for the quantification of these influential factors associated with different types of elements and related tasks.

In future developments, those factors could potentially be connected to the BIM-based scheduling software, how this can be done is outside the scope of this study. When the use of the proposed system is deemed valuable, Synchro 4D developers could be consulted for possibilities to integrate the system. Since the proof of concept is relying upon a fictitious case study, the validation of the factor model is outside the scope of this study. An overview of the tools applied, and their main functionalities are presented in Figure 73.

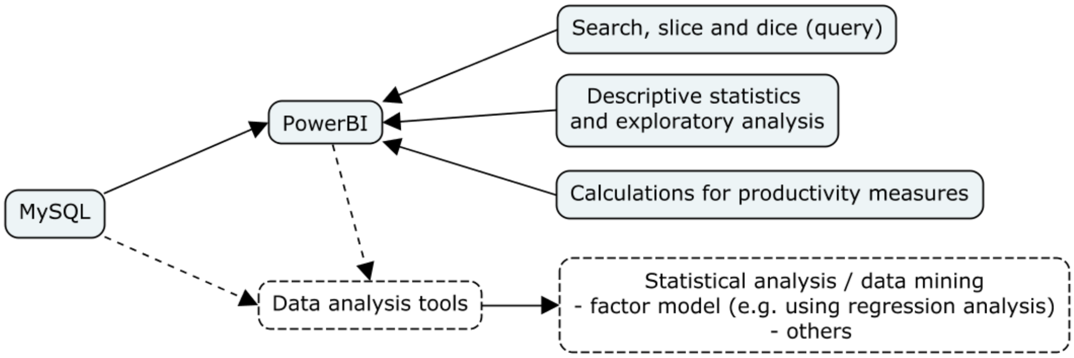


Figure 73: Tools applied for querying, visualizing and analysing the data

2.5. (Re)use

The data can suit several (re)use purposes. When creating a new BIM-based schedule, it is assumed that a BIM model is already present. The family and type name from the elements in this new BIM model can be used to retrieve similar historical element-related task information. Alternatively, the different element classifications (i.e. NL-SfB, IFC and Revit) and task classification can be used to find similar task information when elements of the new BIM model are not present in the database. The filter options as part of PowerBI allow filtering based on multiple attribute values. Moreover, the reliability of the estimates can be determined based on the discrepancies between the as-planned and as-built data, and associated causes for those variations.

A planner that made the as-planned schedule can use the database as the medium to receive feedback on its estimations. Once the as-built data is entered into the database, the planner can review the discrepancies between the actual durations and estimated durations, and the causes of those variations. Before and during the construction of a new project, a construction manager or foreman can use the database to identify tasks that have a high probability to be delayed, its impact and potential causes. According to this information, it might be possible to take mitigating actions.

2.6. Input requirements

Based on the previous section, the general input requirements for the system to function are listed below:

- Semantically rich 4D BIM-based construction schedules should be created.
- The philosophy of BIM-based scheduling should be adopted, where the elements from the BIM model form the basis of the 4D schedule.
- Exchange requirements and modelling guidelines are predefined and adopted. For instance, IDM can be used and set up within the company. These requirements can relate to: the subdivision and aggregation level of the elements in the BIM model relate

to the required level of detail in the schedule, uniform classifications and related codes are applied, variables are uniformly defined, the LOD of the BIM model allows for quantity take-off, etc.)
- Tasks should comprise information that gives meaning to the estimate of the duration. Thus, most likely include information on the input factors such as labour and equipment, internal environment and exogenous factors.
- Elements, tasks, labour, equipment, and start and duration variations are classified with a classification code according to the chosen predefined classifications.
- A many-to-one relationship is adopted between elements and tasks at the deepest level of the schedule.

2.7. Conclusion

The second section of this chapter provides an answer to the fifth sub-research question regarding the way BIM-based as-planned and as-built scheduling data can be analysed and (re)used as part of the database.

The data query language (DQL) is used to query data from the database. A user interface (UI) is built on top of the database that allows the database to be queried in a user-friendly manner. By doing this, one can retrieve the information and gain insight into the data. Within this study, it is out of scope to develop such a user interface. Yet, in this study, the primary goal is to retrieve meaningful data using queries.

Besides making use of a UI to analyse the data, the data can be analysed using the Knowledge Discovery in Databases (KDD) process. As described earlier, a subset of this process is data mining. Data mining comprises a large set of techniques to perform either descriptive or predictive tasks on the data in the database. As part of the descriptive tasks, the data can, for instance, be characterized by the use of OLAP cubes or with frequent pattern analysis. In this study, the predefined classifications tables allow roll-up and drill-down operations similar to OLAP cubes to take place. As part of the predictive tasks, techniques such as classification and regression can be used.

To visually characterize and explore the data, dashboards can be used. Dashboards and KPIs are often used as part of business intelligence (BI). Datawarehouse can be constructed that perform ETL processes on operational databases or other data source and combine the data to perform data analyses as described above. For this study, it was not found necessary to migrate the data in the database towards a data warehouse. To retrieve and reuse the data from MySQL, a direct connection between MySQL and PowerBI is made. PowerBI allows to search, slice and dice the data (by performing indirect queries), derive descriptive statistics, perform exploratory analysis and calculate productivity measures.

When sufficient data is collected, data mining tools can be exploited. For instance, factor models can be derived that measure and model the influence of input factors, factors related to the internal environment and exogenous factors upon productivity. In future developments, those factors could potentially be used to predict task duration as part of BIM-based scheduling software.

5. Case study

In this chapter, the developed system is tested by means of a case study as a proof of concept. The case study shows the different steps that are taken according to the workflow as described in the previous two chapters. It is aimed to validate the capability of extracting task durations or productivity data from the case study for future reuse. In addition, the application of the data for other use cases is validated. Thus, this chapter provides an answer to sub-research question 6 “How well does this system perform in estimating task durations using BIM-based as-planned and as-built scheduling data?”.

1. The case and Plan phase

The case study is used to validate the primary objective of estimating task durations using the BIM-based as-planned and as-built scheduling database, as well as, to validate other use cases of the system. However, the database is currently empty and will only be populated with data from this case. Hence, the database is not used to estimate task duration for this case in advance. Therefore, it is first shown how the data as part of the Plan phase, and the Do and Measure phase is generated within Synchro 4D and subsequently stored in the MySQL database. Then, as part of the Analyse and Improve, and the (Re)use phase it is described how the data could have been used according to the different use cases as defined in Section 3.2.1.1 (see Figure 74).

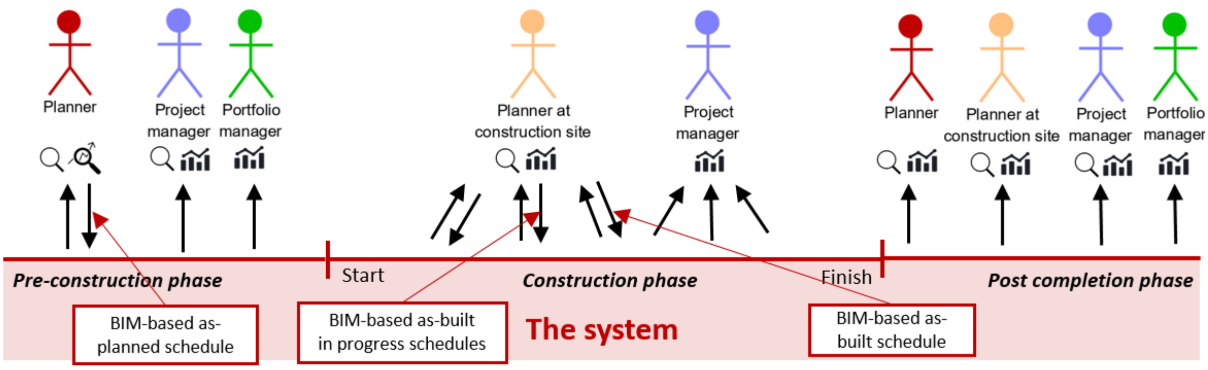


Figure 74: Use cases for the system

The use cases relate to three phases, namely the pre-construction, the construction and the post-completion phase. Three types of BIM-based schedules are generated for the case study that represent the input by the planner in the pre-construction phase and the input by the planner at the construction site in the construction phase. Hence, a BIM-based as-planned schedule, a BIM-based as-built in progress schedule and a final BIM-based as-built schedule is made.

BAM Infra made a fictitious case and Synchro 4D schedule for the construction of a new viaduct that will also be used in this study. The BIM model was made in Revit and exported with the Synchro plugin. Subsequently, it was opened into Synchro 4D where the BIM-based as-planned schedule was constructed. The situation before and after construction is shown in Figure 75. Originally the model contained 397 elements and 96 tasks.

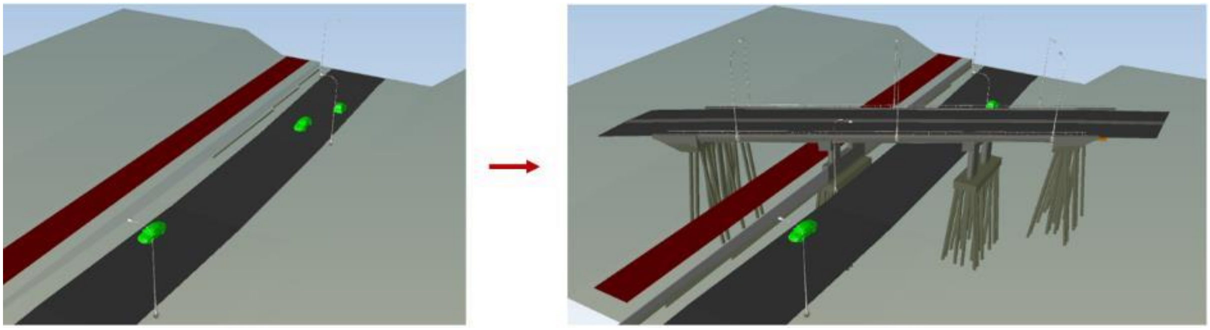
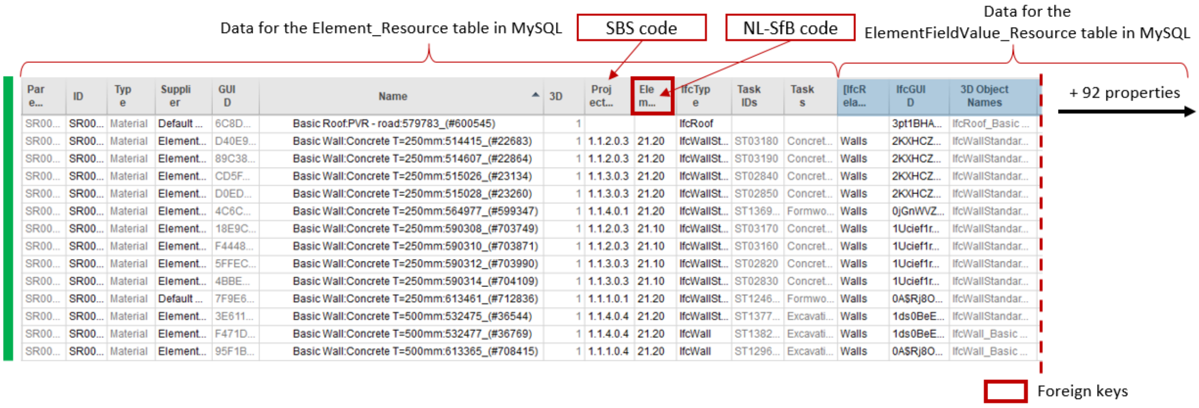


Figure 75: Situation before and after construction

The original BIM-based as-planned schedule does not fulfil the requirements that are set according to the proposed workflow and subsequently developed system. Hence, the model is adjusted. The steps that are taken are described below. A detailed process of creating Synchro 4D as-planned and as-built schedules, exporting them as Excel files and inserting them into the MySQL database using the Python script can be found in Appendix 6.

1.1. Elements in the BIM model

The 'Resource' tab within Synchro 4D contains the data to populate the Element_Resource table and ElementFieldValue_Resource table in MySQL. After studying the different element properties (called user fields in Synchro 4D), it became clear that the naming of those properties fails to uniformly represent the data that is attached to specific elements. For instance, the user field names contain a mix of the Dutch and English language. Also, the Synchro plugin in Revit does not allow to export base quantities. Since the uniform extraction is fundamental for the reuse of the data, an IFC export within Revit is made instead of using the Synchro plugin. Thereafter, the IFC is imported into Synchro 4D to replace the original import of the Revit file that used the Synchro plugin. The IFC is able to represent base quantities uniformly. Furthermore, for every BIM element, an NL-SfB classification code and a system breakdown structure (SBS) code is added. Since the NL-SfB code is a foreign key, the name of the class can be retrieved using SQL query once the data is inserted into the database. The same applies to the other foreign keys that relate to predefined classifications tables. In total, 95 element properties variables are included as part of the 'Resource' tab in Synchro 4D. A section of the BIM-based as-planned data as part of the 'Resource' tab is shown in Figure 76.



Par e...	ID	Typ e	Suppli er	GUI D	Name	3D	Proj ect...	Ele m...	IfcTyp e	Task IDs	Task s	[IfcR ela...	IfcGUI D	3D Object Names
SR00...	SR00...	Material	Default ...	6C8D...	Basic Roof:PVR - road:579783_(#600545)	1			IfcRoof				3pt1BHA...	IfcRoof_Basic ...
SR00...	SR00...	Material	Element...	D40E9...	Basic Wall:Concrete T=250mm:514415_(#22683)	1	1.1.2.0.3	21.20	IfcWallSt...	ST03180	Concret...	Walls	2KXHCZ...	IfcWallStandar...
SR00...	SR00...	Material	Element...	89C38...	Basic Wall:Concrete T=250mm:514607_(#22864)	1	1.1.2.0.3	21.20	IfcWallSt...	ST03190	Concret...	Walls	2KXHCZ...	IfcWallStandar...
SR00...	SR00...	Material	Element...	CD5F...	Basic Wall:Concrete T=250mm:515026_(#23134)	1	1.1.3.0.3	21.20	IfcWallSt...	ST02840	Concret...	Walls	2KXHCZ...	IfcWallStandar...
SR00...	SR00...	Material	Element...	D0ED...	Basic Wall:Concrete T=250mm:515028_(#23260)	1	1.1.3.0.3	21.20	IfcWallSt...	ST02850	Concret...	Walls	2KXHCZ...	IfcWallStandar...
SR00...	SR00...	Material	Element...	4C6C...	Basic Wall:Concrete T=250mm:564977_(#599347)	1	1.1.4.0.1	21.20	IfcWallSt...	ST1369...	Formwo...	Walls	0jGnWVZ...	IfcWallStandar...
SR00...	SR00...	Material	Element...	18E9C...	Basic Wall:Concrete T=250mm:590308_(#703749)	1	1.1.2.0.3	21.10	IfcWallSt...	ST03170	Concret...	Walls	1Ucief1r...	IfcWallStandar...
SR00...	SR00...	Material	Element...	F4448...	Basic Wall:Concrete T=250mm:590310_(#703871)	1	1.1.2.0.3	21.10	IfcWallSt...	ST03160	Concret...	Walls	1Ucief1r...	IfcWallStandar...
SR00...	SR00...	Material	Element...	5FFEC...	Basic Wall:Concrete T=250mm:590312_(#703990)	1	1.1.3.0.3	21.10	IfcWallSt...	ST02820	Concret...	Walls	1Ucief1r...	IfcWallStandar...
SR00...	SR00...	Material	Element...	4BBE...	Basic Wall:Concrete T=250mm:590314_(#704109)	1	1.1.3.0.3	21.10	IfcWallSt...	ST02830	Concret...	Walls	1Ucief1r...	IfcWallStandar...
SR00...	SR00...	Material	Default ...	7F9E6...	Basic Wall:Concrete T=250mm:613461_(#712836)	1	1.1.1.0.1	21.20	IfcWallSt...	ST1246...	Formwo...	Walls	0A$Rj8O...	IfcWallStandar...
SR00...	SR00...	Material	Element...	3E611...	Basic Wall:Concrete T=500mm:532475_(#36544)	1	1.1.4.0.4	21.20	IfcWallSt...	ST1377...	Excavati...	Walls	1ds0BeE...	IfcWallStandar...
SR00...	SR00...	Material	Element...	F471D...	Basic Wall:Concrete T=500mm:532477_(#36769)	1	1.1.4.0.4	21.20	IfcWall	ST1382...	Excavati...	Walls	1ds0BeE...	IfcWall_Basic ...
SR00...	SR00...	Material	Element...	95F1B...	Basic Wall:Concrete T=500mm:613365_(#708415)	1	1.1.1.0.4	21.20	IfcWall	ST1296...	Excavati...	Walls	0A$Rj8O...	IfcWall_Basic ...

Figure 76: Section of BIM model element data as part of the 'Resource' tab in Synchro 4D (green line indicates the use of Synchro 4D)

1.2. Tasks

The 'Task/Gantt' tab within Synchro 4D contains the data to populate the Task, Equipment_Resource and Human_Resource tables in MySQL. At the deepest level of the original schedule, the tasks did relate to more than one element. Since a many-to-one relationship between tasks and elements is required (as described in Section 4.1.8), all tasks are converted as shown in Figure 77.

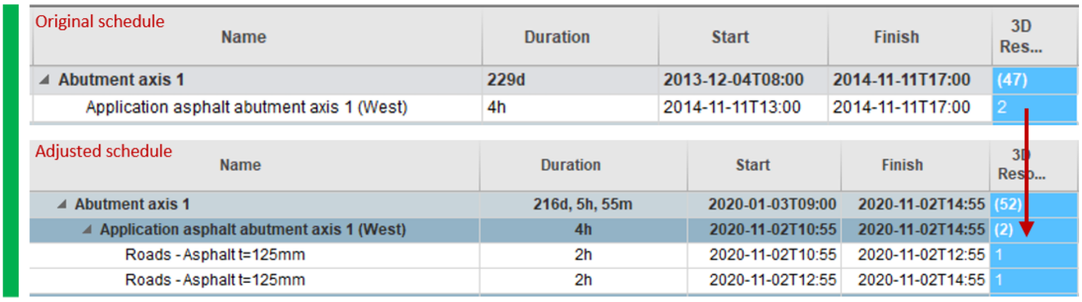


Figure 77: Creating a many-to-one relationship with summary tasks in Synchro 4D

The original schedule did not have input factors regarding labour and equipment assigned to the tasks. Therefore, these input factors are defined with the help of a planner within BAM Infra and added to the schedule. Role names, the quantity, quality and ID according to the Uniclass are added related to the use of labour. Similarly, the equipment name, the quantity of equipment and the ID according to the Uniclass are added. Furthermore, a uniform task code is added for every task according to the standard ABS (Activity Breakdown Structure). An example of a section of this data for the concreting of a pad foundation is shown in Figure 78. The attributes used as foreign keys are indicated within the red squares. The columns in the figure align with the attributes as part of the MySQL tables.

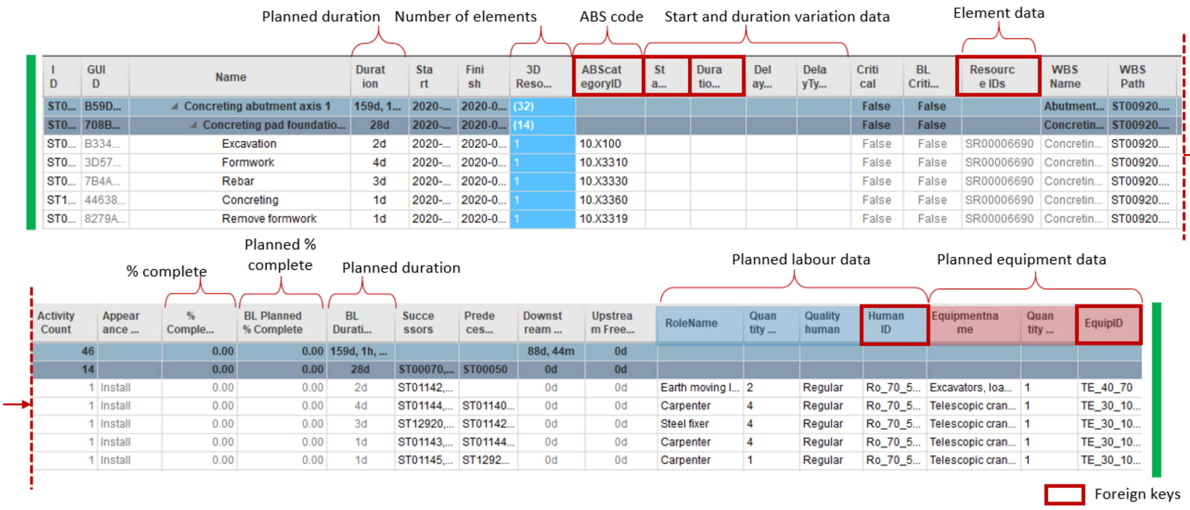


Figure 78: Example of section BIM-based as-planned schedule and input factors in Synchro 4D

The adjusted schedule contains 278 elements, 547 tasks excluding summary tasks, 8 equipment types, 9 labour types. The project is estimated to take 254 working days (2030 hours) to complete. The adjusted schedule is used as the as-planned schedule. Factors concerning the internal environment such as site layout and conditions, level of project complexity, quality of drawings and specifications, frequency of change orders are not considered on task level as part of this case study. The influence of exogenous factors is defined as part of a duration variation code as part of the as-built schedules. To summarize, the input and output of this phase are shown below.

Input:

- BIM Model constructed in Revit, exported as IFC, and subsequently imported into Synchro 4D

Output:

- BIM-based as-planned schedule exported as an Excel file from Synchro 4D

2. Do and Measure phase

In the do and measure phase, the on-site progress of the project is tracked. As described earlier, two additional fictitious schedules are created to validate the use cases as part of the construction phase and post-completion phase. Therefore, an as-built-progress schedule is made after a couple of months from the start of construction. At that point, 21% of the project is complete, see Figure 79. Furthermore, a final as-built schedule is made that represents the point where the project is completed. These two schedules reflect the input of the planner working at the construction site. During an actual project, it is recommended to continuously generate as-built-progress schedules; however, this is out of scope as part of this case study.

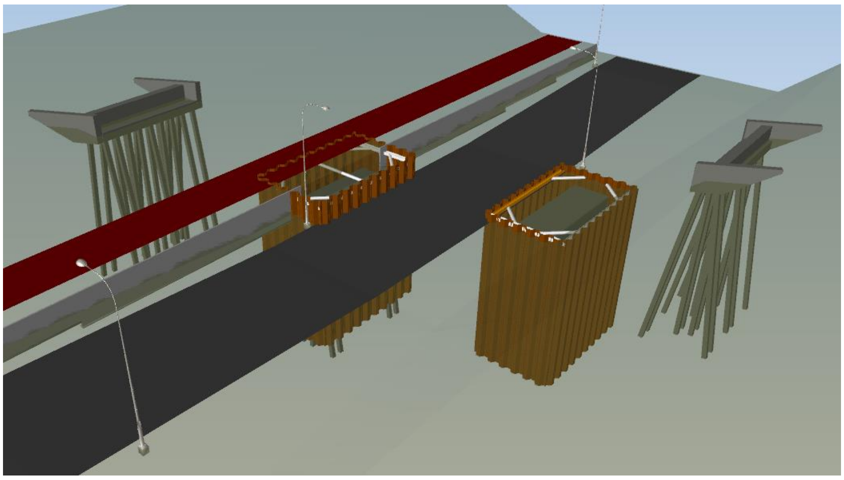


Figure 79: Construction status after three months of construction (21% complete)

For both schedules, the actual task durations are reflected in the schedules and complemented with duration variation codes indicating the cause of the duration difference (as described in Section 3.3.2). Similarly, this is done for direct starting variations. Additionally, the delay types are classified as excusable and compensable, excusable and non-compensable or non-excusable. Also, the task input factors of equipment and labour resources are adjusted according to their actual utilization. Figure 80 displays the start and duration variation data, the actual labour and actual equipment data attached to four tasks. The two tasks of executing formwork and rebar faced a duration delay and thus have a duration variation code and delay type attached to it. For instance, the task of executing formwork was delayed with one day. The delay is associated with duration variation code 3.6. This code is used as a foreign key from the DurationVariation table in MySQL and represents the cause of the crew size being inadequate. Hence, the labour quantity is adjusted. In this case, two people instead of four people completed the task, which was the planned quantity as part of the as-planned schedule (see Figure 78).

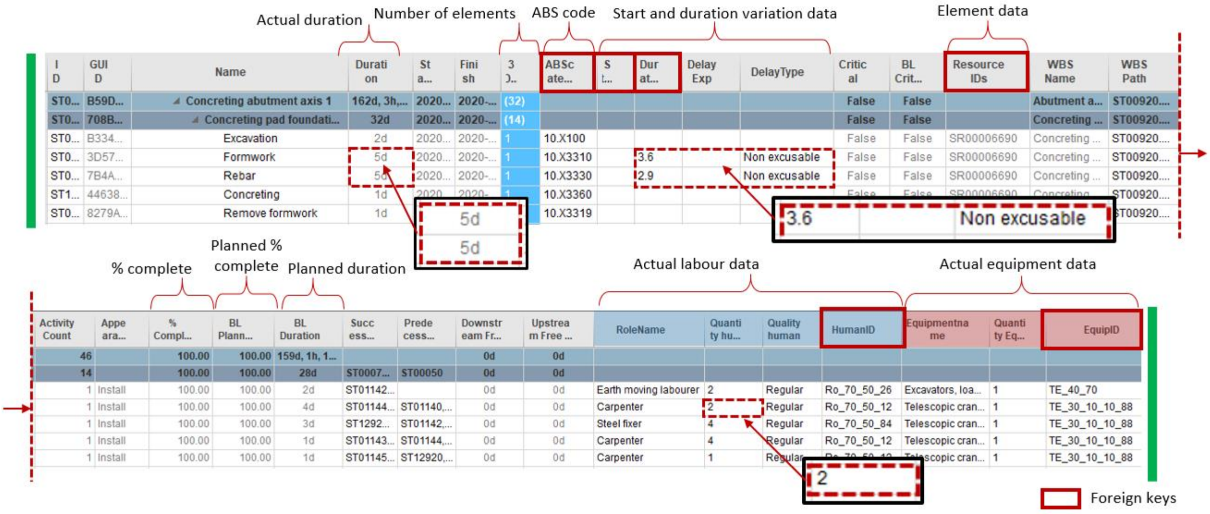


Figure 80: Example of task, element, labour, equipment and delay data related to tasks in Synchro 4D

In total, 37 tasks faced a duration variation (delay) and 1 task faced a start variation as part of the as-built-progress schedule. As a result, the project is delayed with 32 hours (= 4 days). As part of the as-built schedule (including the variations of the as-built-progress schedule), 48 tasks faced a duration variation and 4 tasks a start variation. The actual construction duration increased towards 2190 hours (≈274 days) with a total delay of 20 days. To summarize, the input and output of this phase are shown below.

Input:

- BIM-based as-planned schedule as Synchro 4D file

Output:

- BIM-based as-built-progress schedule (21% completed) made in Synchro 4D and exported as an Excel file
- BIM-based as-built schedule (100% completed) made in Synchro 4D and exported as an Excel file

3. Analyse and Improve phase

In the previous two phases, it is shown how the BIM-based as-planned, as-built in progress and as-built schedules are generated within Synchro 4D. Now, as part of the analyse and improve, and the subsequent (re)use phase, it is described how the data could have been used according to the different use cases and related phases of the project. Within Synchro 4D, the as-planned, as-built-progress and as-built schedules are exported to single Excel files and imported into the MySQL database using the Python script (the detailed process can be found in Appendix 6). Sections of data within different tables in MySQL can be found in Appendix 8. An overview of all three schedules as part of the Schedule table in MySQL is shown in Figure 81.

ScheduleID	Project_ProjectID	ScheduleType	Version	DurationHours	StartDate	FinishDate	DataDate	PlannerName	TimeStamp
1b195be8...	1	As-planned	1	2030	2020-01-01 09:00:00	2020-12-21 14:55:00	2020-01-01	Max Wessel	2020-07-22 23:44:26
4cded4bc...	1	As-built-progress	1	2062	2020-01-01 09:00:00	2020-12-25 14:55:00	2020-04-14	Max Wessel	2020-07-22 23:12:46
ab821430...	1	As-built	1	2190	2020-01-01 09:00:00	2021-01-18 14:55:00	2021-01-19	Max Wessel	2020-07-22 23:44:01

Figure 81: Overview of the schedules as part of the MySQL Schedule table

Note that every time a schedule in Synchro 4D is exported as an Excel file during the project, it can be inserted into the MySQL database using the Python script. Hence, the data can directly be analysed and (re)used. With every insert, the newly stored data enriches the existing data in the database and improves its estimation capabilities.

3.1. Introduction of analysing the MySQL database in PowerBI

Now the data is stored inside the database, the data can be queried, analysed and visualized. First, the process is explained of how this can be done in PowerBI. Then, two queries are shown to get an understanding of how the system functions. Thereafter, multiple queries are executed that relate to the different use cases.

As described earlier, a direct connection is made between MySQL and PowerBI. Once new data is inserted into the database, the data connection between MySQL and PowerBI can be refreshed. Subsequently, indirect queries can be performed within PowerBI on the MySQL database. A possible layout of PowerBI is presented in Figure 82. Remark that in a more ideal scenario, a custom user interface will be made. In PowerBI layout, specific searches can be entered on the left side and the corresponding results are displayed on the right side as tables or other visualizations. An example of using the search filter for the attribute 'elementname' as part of the Element_resource table is shown in Figure 83.

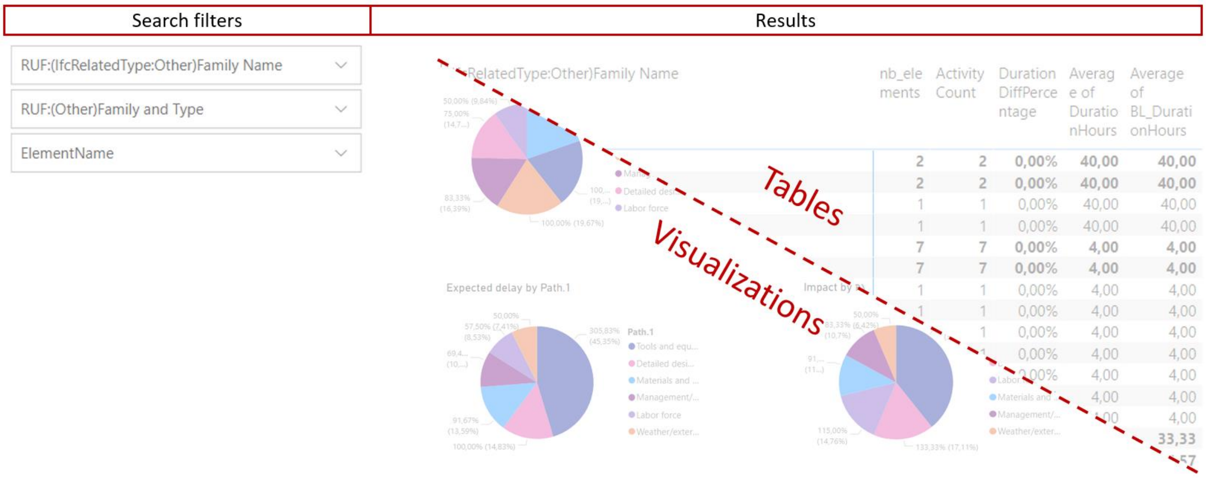


Figure 82: Example of possible layout in PowerBI

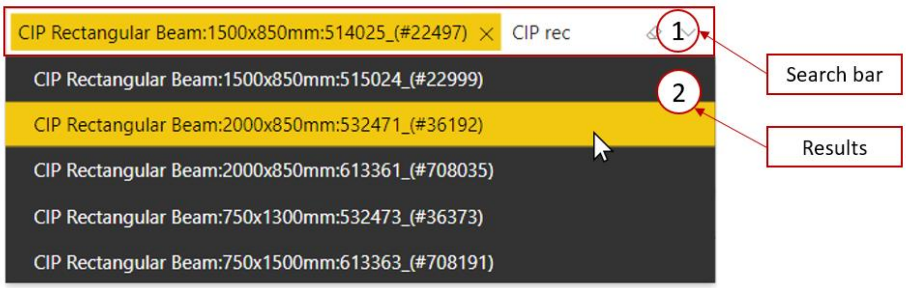


Figure 83: Example search filter Elementname [element_resource] part of PowerBI

For the understanding of the capabilities of the system, the presented inputs and outputs are gradually advanced. The attributes from the different MySQL tables can be simply dragged and dropped to create tables or other visualizations in PowerBI. The used input attributes and functions (such as sum, min, max) applied upon them are described as:

Attribute (aggregate function) [Table] *(e.g. Elementname (sum) [Element_resource])*

All used attributes, values, general filters, search filters and measures are presented as part of the input. As a first example, the element duration without idle time for the selected element is presented in Figure 83.

Query 1 - Single element duration

The fundamental premise is based on using the as-built times to better predict the task duration for future projects. Therefore, a general filter is applied for the schedule type being the 'As-built' schedule, thus presenting the as-built duration. Furthermore, the number of elements and the underlying task count is shown in the table. To put the input as part of PowerBI into perspective, the same query is listed as part of MySQL.

Query in MySQL:

```
SELECT E.Elementname, E.nb_elements, SUM(T.ActivityCount), SUM(T.DurationHours)
FROM element_resource AS E
    INNER JOIN task AS T ON T.element_resource_resourceID = E.ResourceID
    INNER JOIN schedule AS S ON S.ScheduleID = E.Schedule_ScheduleID
WHERE S.scheduletype = 'As-built'
    AND E.ElementName = 'CIP Rectangular Beam:2000x850mm:532471_(#36192)';
```

Output in MySQL Workbench:

Elementname	nb_elements	SUM(T.ActivityCount)	SUM(T.DurationHours)
CIP Rectangular Beam:2000x850mm:532471_(#36192)	1	6	104

Query in PowerBI:

Attributes	*Elementname, nb_elements [element_resource], ActivityCount (sum), DurationHours (sum) [Task]*
General filters	*ScheduleType [schedule] = "As-built"*
Search filter	*Elementname [element_resource] = "CIP Rectangular Beam: 2000x850mm:532471_(#36192)"*

Output in PowerBI:

ElementName	nb_elements	Task Count	Duration Hours
CIP Rectangular Beam:2000x850mm:532471_(#36192)	1	6	104
Total	**1**	**6**	**104**

SUM of underlying task durations without idle time (waiting time)

Query 2 - Element and related tasks

Here, the individual tasks and Activity (task) Breakdown Structure (ABS) category as part of the element in the previous query are retrieved as well; thus, showing the individual durations for each task. Note that the task sequence cannot be retrieved from this query. To obtain task sequence as well, start and finish dates need to be included in the query.

Input:

Attributes	*Elementname [element_resource], Taskname [task], Category [abs_category], nb_elements [element_resource], ActivityCount (sum), DurationHours (sum) [Task]*
General filters	*ScheduleType [schedule] = "As-built"*
Search filter	*Elementname [element_resource] = "CIP Rectangular Beam: 2000x850mm:532471_(#36192)"*

Output:

ElementName	TaskName	ABS Category	nb_elements	Task Count	Duration Hours
CIP Rectangular Beam:2000x850mm:532471_(#36192)	Backfilling	Grond leveren en aanvullen	1	1	8
CIP Rectangular Beam:2000x850mm:532471_(#36192)	Concreting	Verwerken beton - algemeen	1	1	8
CIP Rectangular Beam:2000x850mm:532471_(#36192)	Excavation	Grondwerk - algemeen	1	1	24
CIP Rectangular Beam:2000x850mm:532471_(#36192)	Formwork	Bekisting - algemeen	1	1	32
CIP Rectangular Beam:2000x850mm:532471_(#36192)	Rebar	Wapening - algemeen	1	1	24
CIP Rectangular Beam:2000x850mm:532471_(#36192)	Remove formwork	Ontkisten - algemeen	1	1	8
Total			**1**	**6**	**104**

3.2. Analysis for the (re)use in the pre-construction phase

In this section, it is described how the data within the database could have been analysed and (re)used if it was filled with data, for the use cases within the pre-construction phase. Thus, related to the planner, project manager and the portfolio manager.

First, the query capabilities are accessed according to the requirements that are set as described in Section 3.2.1 from the perspective of a planner. In this phase, the planner receives a BIM model from the design team and uses the BIM model to develop a BIM-based as-planned schedule. The planner can use the element names and/or element types from the BIM model (e.g. Revit or IFC names) to retrieve historic task information from the database. For example, when the BIM model is made within Revit, the BIM element family name e.g. 'CIP Rectangular Beam' can be used to retrieve the family types and related elements in the database (see Figure 84). In the figure, each family type has two elements in the database.

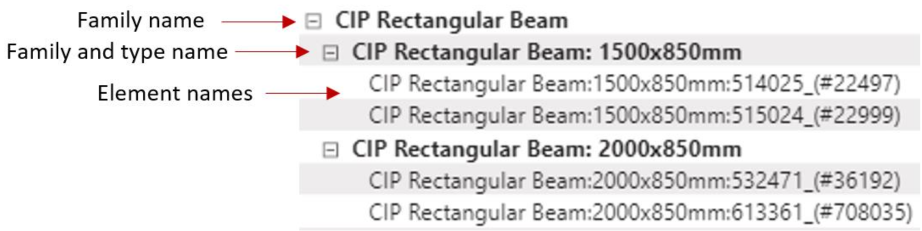


Figure 84: Example retrieval of element family, types and names from the database in PowerBI

Furthermore, the NL-SfB classification can be used to query and group BIM model elements and related task information (see Figure 85). Note that the use of element family names is limited to models that are created in Revit. Hence, the NL-SfB classifications are fundamental when grouping and retrieving data as part of different modelling software.

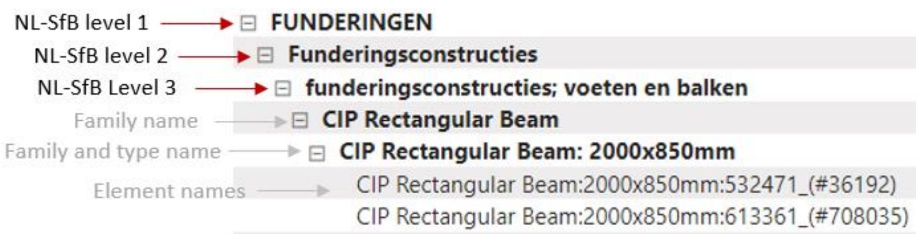


Figure 85: Example retrieval of elements with the NL-SfB hierarchy from the database in PowerBI

Query 3 – Task duration based on the ABS classification

For this example, the planner requests task information as part of the planning process for the element in Figure 86. The planner uses the NL-SfB classification hierarchy and element family name and related type as part of a matrix table to query the information. Furthermore, the planner uses the assigned ABS classification to compare the retrieved elements that are part of the same task category. For instance, allowing to query multiple elements for the task of erecting formwork. The query below shows the associated durations for each task category.

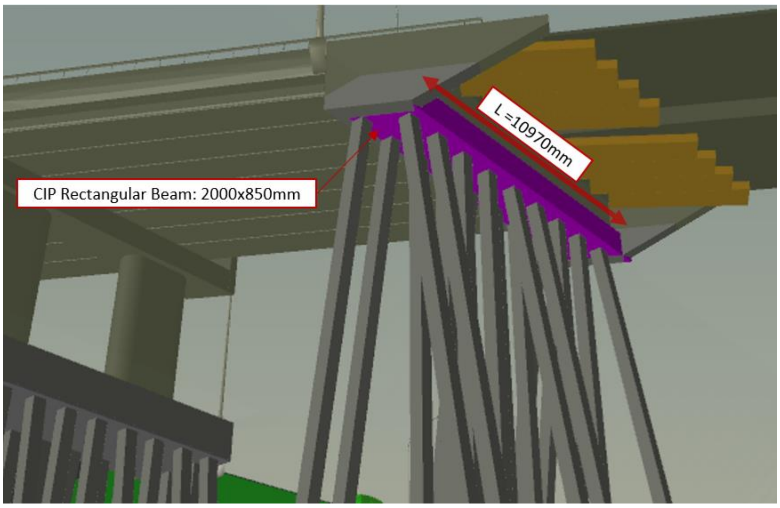


Figure 86: CIP Rectangular Beam: 2000x850mm

Input:

Attributes (rows)	*Path.1, Path.2, Path.3 [Elementnlsfbpath], RUF:(IfcRelatedType:Other)Family Name, RUF:(Other)Family and Type [elementfieldvalue_resource_col], Category[abs_category], Elementname [Element_resource]*
Values	*DurationHours (sum, min, avg, max, st.def) [Task]*
General filters	*ScheduleType [schedule] = "As-built"*
Search filter	*Path.1 = "Funderingen", Path.2 = "Funderingsconstructies", Path.3 = "funderingsconstructies; voeten en balken", RUF:(IfcRelatedType:Other)Family Name = "CIP Rectangular Beam", RUF:(Other)Family and Type = "CIP Rectangular Beam: 2000x850mm"*

Output:

Element_SfB_Levels	Duration Hours	Min of Duration Hours	Average of Duration Hours	Max of DuratioNHours	Standard deviation of Duration Hours
FUNDERINGEN	**368**	**8**	[illegible]	[illegible]	[illegible]
Funderingsconstructies	**368**	**8**	[illegible]	[illegible]	[illegible]
funderingsconstructies; voeten en balken	**368**	**8**	[illegible]	[illegible]	[illegible]
CIP Rectangular Beam	**368**	**8**	**18,40**	**40**	**11,62**
CIP Rectangular Beam: 2000x850mm	**[illegible]24**	**8**	**18,67**	**40**	**12,36**
Bekisting - algemeen	**72**	**32**	**36,00**	**40**	**4,00**
CIP Rectangular Beam:2000x850mm:532471_(#36192)	32	32	32,00	32	0,00
CIP Rectangular Beam:2000x850mm:613361_(#708035)	40	40	40,00	40	0,00
Grond leveren en aanvullen	**16**	**8**	**8,00**	**8**	**0,00**
CIP Rectangular Beam:2000x850mm:532471_(#36192)	8	8	8,00	8	0,00
CIP Rectangular Beam:2000x850mm:613361_(#708035)	8	8	8,00	8	0,00
Grondwerk - algemeen	**40**	**16**	**20,00**	**24**	**4,00**
CIP Rectangular Beam:2000x850mm:532471_(#36192)	24	24	24,00	24	0,00
CIP Rectangular Beam:20[illegible]	16	[illegible]	[illegible]	[illegible]	[illegible]
Ontkisten - algemeen	**16**	[illegible]	[illegible]	[illegible]	[illegible]
CIP Rectangular Beam:2000x850mm:532471_(#36192)	8	[illegible]	[illegible]	[illegible]	[illegible]
CIP Rectangular Beam:2000x850mm:613361_(#708035)	8	8	8,00	8	0,00
Verwerken beton - algemeen	**16**	**8**	**8,00**	**8**	**0,00**
CIP Rectangular Beam:2000x850mm:532471_(#36192)	8	8	8,00	8	0,00
CIP Rectangular Beam:2000x850mm:613361_(#708035)	8	8	8,00	8	0,00
Wapening - algemeen	**64**	**24**	**32,00**	**40**	**8,00**
CIP Rectangular Beam:2000x850mm:532471_(#36192)	24	24	24,00	24	0,00
CIP Rectangular Beam:2000x850mm:613361_(#708035)	40	40	40,00	40	0,00

Element (SfB) Categories
Element family
Element family and type
Task Category (ABS)
Element names
Average duration for erecting formwork (bekisting – algemeen)
Single task durations associated with its category (32+8+24+8+8+24 = 104)

For the understanding of the output, it is shown that the sum of the single task durations per task category is equal to the total element duration as shown in query 1. The minimum, average, maximum and standard deviation for the duration of tasks are calculated using formulas/measures related to the drill-down level.

In this case, two elements within the database can be used when the planner wants to estimate the duration related to the task category 'erecting formwork (bekisting – algemeen)'. The average duration of the task equals 36 hours (see output). However, this duration cannot be reused directly since several factors on which the duration is based are still missing.

While the width and height of the element are indicated with the type name (i.e. 2000x850mm), the length of the elements may vary and are not considered yet. Logically, the length of the elements needs to be accounted for to derive an accurate duration estimate. In addition, the number of labourers, the equipment that is used and other factors are not shown. In other words, the complete context of the input factors, internal environment and exogenous factors that influence the duration of the task is not fully present.

Query 4 – Element dimension variables

This query is shown to indicate the possibility to extract dimensions from the database. Multiple variables can represent the same dimension which complicates the extraction process. Moreover, within the case data, it was found that the value for a single dimension may be displayed as part of two different dimension variables (e.g. length and height).

Input:

Attributes (rows)	*Path.1, Path.2, Path.3 [Elementnlsfbpath], RUF:(IfcRelatedType:Other)Family Name, RUF:(Other)Family and Type [elementfieldvalue_resource_col], Elementname, [element_resource]*
Values	*nb_elements [element_resource], length, height, width, thickness attributes (sum) [elementfieldvalue_resource_col]*
General filters	*ScheduleType [schedule] = "As-built"*
Search filter	*Path.1 = "Funderingen", Path.2 = "Funderingsconstructies", Path.3 = "funderingsconstructies; voeten en balken"*

Output:

		Length			Height				Width			Thickness		
Element_SfB_Levels	nb_elements	RUF:(BaseQuantities(IfcElementQuantity))Length	RUF:(Dimensions)Length	RUF:(IfcRelatedType:Dimensions)Length	RUF:(BaseQuantities(IfcElementQuantity))Height	RUF:(Dimensions)Height	RUF:(IfcRelatedType:Dimensions)Height	RUF:(IfcRelatedType:Dimensions)Hoogte	RUF:(BaseQuantities(IfcElementQuantity))Width	RUF:(IfcRelatedType:Dimensions)Width	RUF:(Dimensions)Width	RUF:(Dimensions)Thickness	RUF:(IfcRelatedType:Dimensions)Thickness	RUF:(IfcRelatedType:Dimensions)Breedte
Basic Wall: Foundation Wall	**1**	**10392**	**10392**	**0**	**2750**	**0**	**0**	**0**	**1000**	**0**	**0**	**0**	**0**	**0**
CIP Rectangular Beam:	**2**	**21940**	**21940**	**0**	**0**	**0**	**0**	**1700**	**0**	**0**	**0**	**0**	**0**	**4000**
CIP Rectangular Beam:2000x850mm:532471_(#36192)	1	10970	10970	0	0	0	0	850	0	0	0	0	0	2000
CIP Rectangular Beam:2000x850mm:613361_(#708035)	1	10970	10970	0	0	0	0	850	0	0	0	0	0	2000
CIP Rectangular Beam:	**1**	**10970**	**10970**	**0**	**0**	**0**	**0**	**1300**	**0**	**0**	**0**	**0**	**0**	**750**
CIP Rectangular Beam:750x1300mm:532473_(#36373)	1	10970	10970	0	0	0	0	1300	0	0	0	0	0	750
CIP Rectangular Beam:	**1**	**10970**	**10970**	**0**	**0**	**0**	**0**	**1500**	**0**	**0**	**0**	**0**	**0**	**750**
CIP Rectangular Beam:750x1500mm:613363_(#708191)	1	10970	10970	0	0	0	0	1500	0	0	0	0	0	750
Foundation Slab: Concrete t=1200mm	**2**	**0**	**20784**	**0**	**0**	**0**	**0**	**0**	**7000**	**0**	**7000**	**2400**	**0**	**0**
Foundation Slab:Concrete t=1200mm:531445_(#35887)	1	0	10392	0	0	0	0	0	3500	0	3500	1200	0	0
Foundation Slab:Concrete t=1200mm:616800_(#713103)	1	0	10392	0	0	0	0	0	3500	0	3500	1200	0	0

Query 5 – Task duration and related input factors

Here, the relevant input factors are queried that provide the planner with the information on which the duration of the task is based. Thus, the use of labour and equipment are queried, as well as the dimensions of the elements. This information allows the planner to synthesize a figure that can be used to estimate task duration. This is shown in the next query.

Input:

Attributes (rows)	*RUF:(IfcRelatedType:Other)Family Name, RUF:(Other)Family and Type [elementfieldvalue_resource_col], category [abs_category], Elementname, [element_resource]*
Values	*DurationHours (avg) [task], RoleName, Quantity_human (avg) [human_resource], Equipmentname, Quantity_equipment (avg) [equipment_resource], length, width, height, volume attributes (avg) [elementfieldvalue_resource_col]*
General filters	*ScheduleType [schedule] = "As-built"*
Search filter	*RUF:(IfcRelatedType:Other)Family Name = "CIP Rectangular Beam"*

Output:

		Labour data		Equipment data		Dimensions			
RUF:(IfcRelatedType:Other)Family Name	Average of Duration Hours	First RoleName	Average of quantity_human	First EquipmentName	Average of Quantity_Equipment	Average of RUF:(Dimensions)Length	Average of RUF:(IfcRelatedType:Dimensions)Breedte	Average of RUF:(IfcRelatedType:Dimensions) Hoogte	Average of RUF:(Dimensions)Volume
⊟ CIP Rectangular Beam	**18,67**	**Carpenter**	**2,67**	**Excavators, loaders and dredgers**	**1,00**	**10.969,66**	**2.000,00**	**850,00**	**15,46**
⊟ CIP Rectangular Beam: 2000x850mm	**18,67**	**Carpenter**	**2,67**	**Excavators, loaders and dredgers**	**1,00**	**10.969,66**	**2.000,00**	**850,00**	**15,46**
⊟ Bekisting - algemeen	**36,00**	**Carpenter**	**3,00**	**Telescopic cranes**	**1,00**	**10.969,66**	**2.000,00**	**850,00**	**15,46**
CIP Rectangular Beam:2000x850mm:532471_(#36192)	32,00	Carpenter	4,00	Telescopic cranes	1,00	10.969,66	2.000,00	850,00	15,46
CIP Rectangular Beam:2000x850mm:613361_(#708035)	40,00	Carpenter	2,00	Telescopic cranes	1,00	10.969,66	2.000,00	850,00	15,46

The output indicates that for one element four persons were used, while for the other element two people were used. All the other factors are equal. Hence, the quantity of labour is taken into account to derive a task duration estimate.

An alternative approach to query relevant information is the use of certain conditions. Conditions such as equal to, higher than, or lower than can be applied. In that regard, the task duration can be queried based on conditions such as the number of people working on the task, the type and quantity of equipment, etc. Furthermore, these conditions can be used to exclude certain delays from a dataset; for instance, delays that are caused by a change in the scope of work.

Query 6 – Productivity measure

In this query, a productivity measure is derived for the CIP Rectangular Beam: 2000x850mm such that the planner can estimate the task duration for executing formwork for the element in Figure 86. As the length may vary related to the element family and type, it is necessary to perform a productivity calculation that can be used to calculate the duration of a task for an element that has other dimensional properties.

For this query, calculations have been made to determine the total number of person-hours used for the task, as well as to determine the productivity per meter of the element per person-hours. The equations are listed below.

$$Total\ person\ hours = task\ duration * crew\ size\ (quantity\ of\ labour) \quad (5)$$

$$Productivity\ per\ meter\ per\ person\ hour = \frac{Quantity\ (length)}{Total\ person\ hours} \qquad (6)$$

Input:

Attributes (rows)	*RUF:(IfcRelatedType:Other)Family Name, RUF:(Other)Family and Type [elementfieldvalue_resource_col], Elementname, [element_resource], category [abs_category]*
Values	*m1 (length) per hour according to input factors (measure), Meter (length) per person-hour (measure), Total_Manhours (measure), DurationHours (sum, avg) [task], Quantity_Equipment (avg) [equipment_resource], Quantity_human [human_resource], length, breedte, hoogte attributes [elementfieldvalue_resource_col]*
General filters	*ScheduleType [schedule] = "As-built", AppearanceType [task] = "Install"*
Search filter	*Category [abs_category] = "Bekisting – algemeen", RUF:(Other)Family and Type [elementfieldvalue_resource_col] = "CIP Rectangular Beam: 2000x850m"*
Measures	*- m1 (length) per hour according to input factors = divide(SUM('bim_schedules elementfieldvalue_resource_col'[RUF:(BaseQuantities(IfcElementQuantity))Length]),1000)/SUM('bim_schedules task'[DurationHours])* *- Meter (length) per person-hour (calculation) = divide(SUM('bim_schedules elementfieldvalue_resource_col'[RUF:(BaseQuantities(IfcElementQuantity))Length]),1000)/[Total_Manhours]* *- Meter (length) per person-hour = AVERAGEX(KEEPFILTERS(VALUES('bim_schedules element_resource'[ElementName])), CALCULATE([Meter (length) per person-hour]))* *-Total_Manhours = SUMX('bim_schedules human_resource','bim_schedules human_resource'[quantity_human]*[SUM_Duration])*

Output:

RUF:(IfcRelatedType:Other)Family Name	Meter (length) per hour according to input factors	Meter (length) per person-hour	Total_Manhours	DurationHours	Average of DurationHours	Average of Quantity_Equipment	Average of quantity_human	RUF:(BaseQuantities(IfcElementQuantity))Length	RUF:(IfcRelatedType:Dimensions)Breedte	RUF:(IfcRelatedType:Dimensions)Hoogte
CIP Rectangular Beam	0,305	0,111	208	72	36,00	1,00	3,00	21940	4000	1700
CIP Rectangular Beam: 2000x850mm	0,305	0,111	208	72	36,00	1,00	3,00	21940	4000	1700
Bekisting - algemeen	0,305	0,111	208	72	36,00	1,00	3,00	21940	4000	1700
CIP Rectangular Beam:2000x850mm:532471_(#36192)	0,343	0,086	128	32	32,00	1,00	4,00	10970	2000	850
CIP Rectangular Beam:2000x850mm:613361_(#708035)	0,274	0,137	80	40	40,00	1,00	2,00	10970	2000	850

Element family name; Element family and type; Task category (formwork); Average productivity per meter per person-hour

Equation 6 can be used to determine the necessary person-hours to complete the task. For instance, the average productivity per meter per person-hour of the two elements equals 0,111. The length of the CIP Rectangular Beam: 2000x850mm in Figure 86 is 10,970 meters, which result in: 10,970 / 0,111 (rate) ≈ 99 person-hours. If a crew of three people is used, then the duration of the task can be estimated with: 99 (person-hours) / 3 (crew size) = 33 hours.

As part of this system, the planner must determine the unit of measure for all productivity rate calculations. In other words, it must be determined per element or element category and/or task category if, for instance, the number of units, length, area or volume of the element is used.

Query 7 – Productivity measure unconnected to a specific element type

Previous queries all rely on one single element type to query task information. Yet, when the planner wants to derive a duration estimate for a certain element type that is not part of the database, the information should be queried differently. For example, when one wants to query information on a CIP Rectangular Beam: 1000x500mm instead of the CIP Rectangular Beam: 2000x850mm.

The elements from different types can be compared to derive an estimate. However, in this case, the length of the elements will no longer be a valid unit of measurement since the width and height of the elements are different as well. Therefore, it is necessary to use another unit of measurement, such as the volume of the elements. The difference between the estimation of a single element type or the estimation based upon multiple element types is graphically presented in Figure 87.

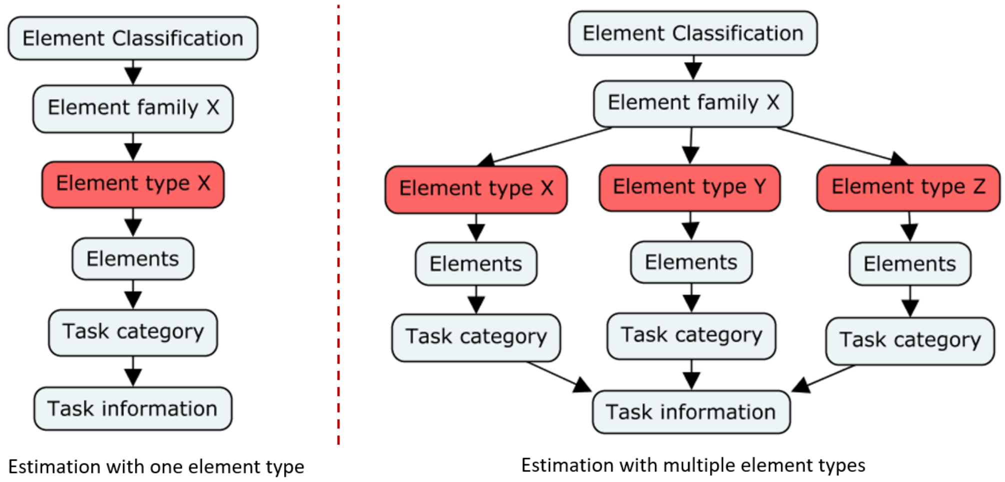


Figure 87: Difference between the estimation of single and multiple elements from the same type

For this query, all elements related to the element family of CIP Rectangular Beam part of the task category 'erecting formwork (bekisting -algemeen)' are queried.

Attributes (rows)	*RUF:(IfcRelatedType:Other)Family Name, RUF:(Other)Family and Type [elementfieldvalue_resource_col], Elementname, [element_resource], category [abs_category]*
Values	*Volume (m3) per person-hours (measure), Total_Manhours (measure), DurationHours (sum, avg) [task], Quantity_Equipment (avg) [equipment_resource], Quantity_human [human_resource], volume, length, breedte, hoogte attributes [elementfieldvalue_resource_col]*
General filters	*ScheduleType [schedule] = "As-built", AppearanceType [task] = "Install"*
Search filter	*Category [abs_category] = "Bekisting – algemeen", RUF:(Other)Family Name = "CIP Rectangular Beam"*
Measures	*- Volume (m3) per person-hour (calculation) = SUM('bim_schedules elementfieldvalue_resource_col'[RUF:(Dimensions)Volume])/[Total_Manhours]* *- Volume (m3) per person-hour = AVERAGEX(KEEPFILTERS(VALUES('bim_schedules element_resource'[ElementName])), CALCULATE([Volume (m3) per person-hour]))* *- same as previous*

Output:

Average productivity per volume (m^3) per person-hour

RUF:(IfcRelatedType:Other)Family Name	Volume (m3) per person-hour	Total_Manhours	DurationHours	Average of DurationHours	Average of Quantity_Equipment	Average of quantity_human	Average of RUF:(Dimensions)Volume	Average of RUF:(BaseQuantities(IfcElementQuantity))Length	Average of RUF:(IfcRelatedType:Dimensions)Breedte	Average of RUF:(IfcRelatedType:Dimensions)Hoogte
CIP Rectangular Beam	**0,1039**	**704**	**216**	**36,00**	**1,00**	**3,33**	**11,57**	**10681**	**1417**	**1033**
CIP Rectangular Beam: 1500x850mm	**0,0695**	**240**	**80**	**40,00**	**1,00**	**3,00**	**8,34**	**10104**	**1500**	**850**
Bekisting - algemeen	**0,0695**	**240**	**80**	**40,00**	**1,00**	**3,00**	**8,34**	**10104**	**1500**	**850**
CIP Rectangular Beam:1500x850mm:514025_(#22497)	0,0695	120	40	40,00	1,00	3,00	8,34	10104	1500	850
CIP Rectangular Beam:1500x850mm:515024_(#22999)	0,0695	120	40	40,00	1,00	3,00	8,34	10104	1500	850
CIP Rectangular Beam: 2000x850mm	**0,1570**	**208**	**72**	**36,00**	**1,00**	**3,00**	**15,46**	**10970**	**2000**	**850**
Bekisting - algemeen	**0,1570**	**208**	**72**	**36,00**	**1,00**	**3,00**	**15,46**	**10970**	**2000**	**850**
CIP Rectangular Beam:2000x850mm:532471_(#36192)	0,1208	128	32	32,00	1,00	4,00	15,46	10970	2000	850
CIP Rectangular Beam:2000x850mm:613361_(#708035)	0,1933	80	40	40,00	1,00	2,00	15,46	10970	2000	850
CIP Rectangular Beam: 750x1300mm	**0,0791**	**128**	**32**	**32,00**	**1,00**	**4,00**	**10,13**	**10970**	**750**	**1300**
Bekisting - algemeen	**0,0791**	**128**	**32**	**32,00**	**1,00**	**4,00**	**10,13**	**10970**	**750**	**1300**
CIP Rectangular Beam:750x1300mm:532473_(#36373)	0,0791	128	32	32,00	1,00	4,00	10,13	10970	750	1300
CIP Rectangular Beam: 750x1500mm	**0,0913**	**128**	**32**	**32,00**	**1,00**	**4,00**	**11,69**	**10970**	**750**	**1500**
Bekisting - algemeen	**0,0913**	**128**	**32**	**32,00**	**1,00**	**4,00**	**11,69**	**10970**	**750**	**1500**
CIP Rectangular Beam:750x1500mm:613363_(#708191)	0,0913	128	32	32,00	1,00	4,00	11,69	10970	750	1500

The output shows that the average volume (m^3) per person-hour of all elements is equal to 0,1039. Say, the CIP Rectangular Beam: 1000x500mm has a length of 5000mm, then its volume would be equal to 2,5 m^3. 2,5 (volume) / 0,1039 (rate) ≈ 24 person-hours. If a crew of two people is used, then the duration of the task can be estimated with: 24 (person-hours) / 2 (crew size) = 12 hours. The accuracy of such a measure can be questioned since not all factors are accounted for, e.g. the shape of the element.

Query 8 – Distribution of productivity data using the Box-and-Whisker plot

As part of query 3, the minimum, average, maximum and standard deviation have been expressed, thus providing information on the distribution (and accuracy) of the duration data related to different elements and / or task categories. In the category of descriptive statistics, Box-and-Whisker plots can be used for the visual representation for the degree of dispersion (spread) and skewness in the data.

The productivity measures (volume (m^3) per person-hour) from the previous query are visualized as part of a Box-and-Whisker plot.

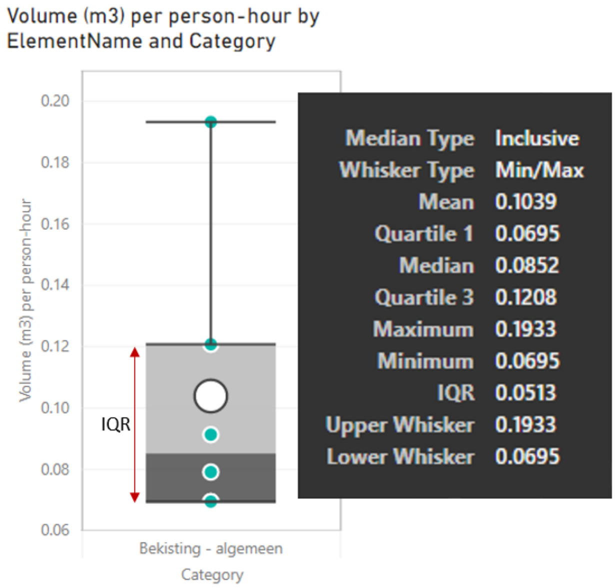


Figure 88: Box-and-whisker plot of volume m^3 per person-hour

The plot shows that the data is positively skewed as the mean is closer to the bottom of the box. It also shows that the lower whisker is relatively close to the median, whereas the upper whisker is much further away. Therefore, most productivity values are close to the lower whisker, thus have a low productivity and need more time to complete the task. In comparison, the productivity will be higher for the higher whisker, however, its occurrence is less likely. Based upon this plot, it can be argued that instead of using the mean as a productivity measure (as this is done in query 7), it would be better to use median value as the occurrence of this value would be more likely. Furthermore, it can be stated that the chance of severe delays is minimal since this would mean that lower whisker would be much further away from the interquartile range (IQR).

Query 9 – Prediction with the use of linear regression

Prediction models can be used to determine the influence of an independent variable on a dependent variable. At this moment, the database is only populated with fictitious data of a single project. Therefore, it is not possible to validate the usability of a prediction model. Purely illustrative, it is shown how a regression model can be used for the estimation of task durations.

For this example, a scatter plot is presented showing the relationship between the volumes of all the elements in the database related to the durations of the task category of erecting formwork (bekisting – algemeen). Regression analysis tries to fit a linear function that best fits the data. The linear regression function is listed below.

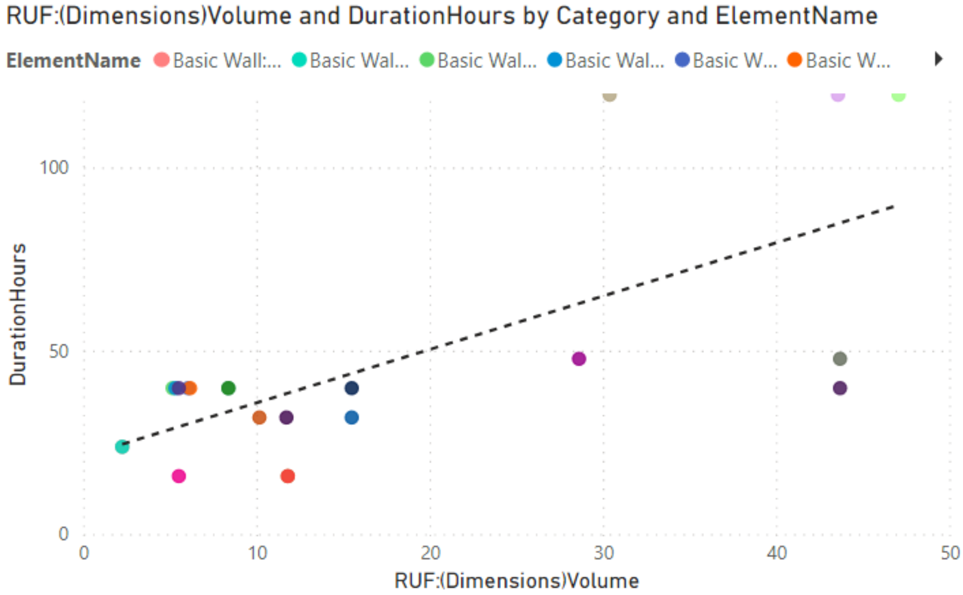
$$\hat{Y}_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1 \quad (7)$$

Where β_0 represents the intercept (i.e. a constant value), β_1 the coefficient (i.e. the slope of the line) and ε_1 the error term.

Input:

Details	*Category [abs_category]*
Legend	*Elementname [element_resource]*
X Axis	*RUF:(Dimensions)Volume [elementfieldvalue_resource_col]*
Y Axis	*DurationHours [task]*
General filters	*ScheduleType [schedule] = "As-built"*
Search filter	*Category [abs_category] = "Bekisting – algemeen"*

Output:



A line can to a certain extent be fitted within the scatter plot. The line shows that when the volume of the elements increases, the task duration will increase as well. The data from this graph has been exported to perform regression analysis. The regression function that approximates the line in the graph is shown below.

$$Estimated\ Duration\ hours = 21{,}50 + 1{,}455 * X_1(Volume) \quad (8)$$

If this regression function was used to estimate the task duration as in query 7, the result would be: 21,50 + 1,455*2,5 (volume) ≈ 25 hours (without considering the quantity of labour).

As described in Section 4.2.2, many more prediction models can be used. For instance, multiple linear regression can be applied for a regression function with multiple independent variables. Performing data analysis can potentially quantify the influence of the various factors upon the duration of a task and subsequently be used to predict task duration. By achieving this, manually determining the unit of measurement (as described in query 6) may become redundant. To illustrate this, multiple regression is conducted for the same dataset that is used for simple linear regression. However, instead of using the volume of the elements, the length, width and height are used as the independent variables. The multiple regression function is shown below.

$$Estimated\ Duration\ Hours = 11{,}79 + 0{,}00141 * X_1(Length) + 0{,}00525 * X_2(Width) + 0{,}00059 * X_3(Height) \quad (9)$$

The estimated duration of a CIP Rectangular Beam: 1000x500mm with a length of 5000mm has the following result: 11,79 + 0,00141*5000mm + 0,00525*1000mm + 0,00059*500mm ≈ 24 hours to erect the formwork.

Query 10 – Expected delay with as-planned and as-built durations

At this point, the use case from the perspective of the project manager is considered. The project manager is interested in historical task duration information to gain insight into time-related risks related to the BIM-based as-planned schedule provided by the planner. As a result, the project manager might be able to take mitigation measures to potentially prevent their reoccurrence for the current project.

The difference between the historical as-planned and as-built task durations per group of elements, per element and related tasks or task categories can be used to calculate measures derived from the Method Productivity Delay Model (MPDM). As such, the *probability of occurrence* of a delay and the *impact* that the potential delay has on the as-planned duration can be determined. Then, the *expected delay* can be calculated related to the as-planned durations. This calculation can provide insight into potential risks. It should be recognized that historical data cannot link a complete risk profile to certain elements and related tasks since the data only includes the actual risks that have arisen in past projects. Thus, risks that have never occurred but may have a major impact, are not included. The equations are listed below.

$$Probability\ of\ occurence = \frac{count\ of\ variations\ (delays)}{total\ task\ count} \tag{10}$$

$$Impact = \frac{as\ built\ duration - as\ planned\ duration}{as\ planned\ duration} \tag{11}$$

$$Expected\ delay = Probabilty\ of\ occurence * Impact \tag{12}$$

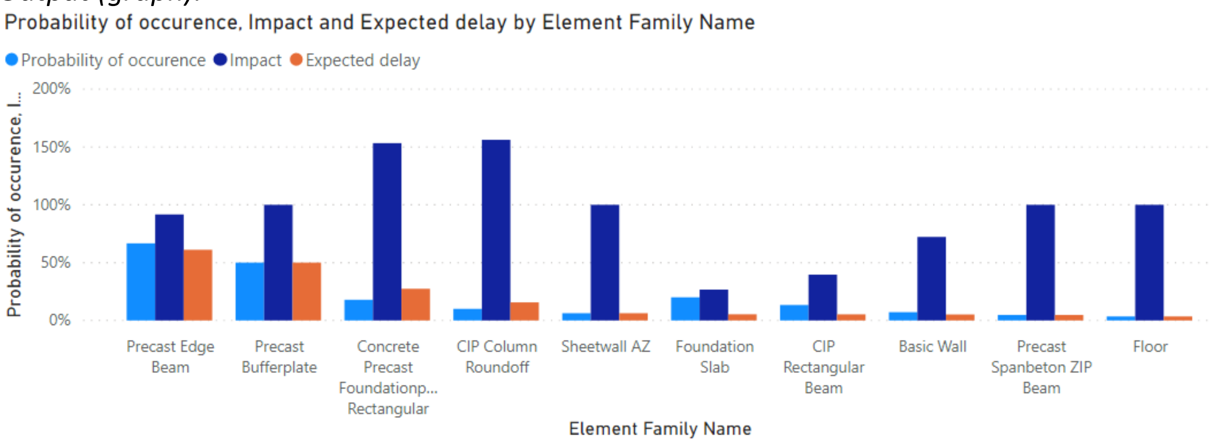
Input:

Attributes (rows)	*RUF:(IfcRelatedType:Other)Family Name, RUF:(IfcRelatedType:Other)Family and Type [elementfieldvalue_resource_col], category (abs_category), Elementname [element_resource]*
Values	*Nb_elements [element_resource], ActivityCount [task], DurationVarCount (measure), Probability of occurrence (measure), Impact (measure), Expected Delay (measure)*
General filters	*ScheduleType [schedule] = "As-built"*
Measures	*- DurationVarCount = countx(filter('bim_schedules task',[DurationDiffPercentage]>0),[DurationDiffPercentage])* *- Probability of occurence = [DurationVarCount]/SUM('bim_schedules task'[ActivityCount])* *- Impact = AVERAGEX(filter('bim_schedules task',[DurationDiffPercentage] >0),[DurationDiffPercentage])* *- Expected delay = [Probability of occurence]*[Impact]*

Output (table):

RUF:(IfcRelatedType:Other)Family Name	nb_elements	Task Count	DurationVar Count	Probability of occurence	Impact	Expected delay
Precast Edge Beam	6	6	4	66,67%	91,67%	61,11%
Precast Bufferplate	16	16	8	50,00%	100,00%	50,00%
Concrete Precast Foundationpile Rectangular	84	84	15	17,86%	153,33%	27,38%
CIP Column Roundoff	4	20	2	10,00%	156,25%	15,63%
Sheetwall AZ	64	64	4	6,25%	100,00%	6,25%
Foundation Slab	2	10	2	20,00%	26,67%	5,33%
CIP Rectangular Beam	6	30	4	13,33%	39,58%	5,28%
Basic Wall	17	42	3	7,14%	72,22%	5,16%
Precast Spanbeton ZIP Beam	21	21	1	4,76%	100,00%	4,76%
Floor	20	29	1	3,45%	100,00%	3,45%
Arcelor Mittal Corner C14	7	7				
Circular Hollow Sections	10	10				
H-Wide Flange Beams	8	8				
Lightpost	11	11				
Total	**276**	**358**	**44**	**12,29%**	**109,26%**	**13,43%**

Output (graph):



The output shows the probability of occurrence, impact and expected delay grouped by the BIM model element family names that exist in the BIM-based as-planned schedule provided by the planner. Alternatively, the data could be queried based on the NL-SfB categories, ABS categories, specific element types etc. The 'expected delay' column is colour coded based on its percentage: green < 5%, yellow between 5% and 25% and red > 25%. The scale is arbitrarily chosen.

The output indicates that based on historical task duration data, tasks related to the element families of 'Precast Edge Beam', 'Precast bufferplate' and 'Concrete Precast Foundationpile Rectangular' have the highest expected delay. At this point, the project manager can drill-down the data to retrieve the causes of the delays. Thereafter, it might be possible to take mitigating measures or explore alternative options to construct certain elements. In addition, a critical path filter can be applied that relates to the BIM-based as-planned schedule that allows the project manager to focus on delays that could impact the overall project duration the most. In general, this data can also be useful for a planner to get a better insight into the reliability of the duration estimates.

Query 11 – The probability of occurrence, impact and expected delay by cause category

Based on the previous query, the causes for the variation (delay) as part of the variation codes can be retrieved. In this way, it can be made clear which specific cause or cause category is responsible for the probability of the occurrence, impact and expected delay. Hence, the drill-down capacity allows to do a root cause analysis on the duration variations. First, for this example, a drill-down is performed on the element family and type 'Concrete Precast Foundationpile Rectangular: 400x400mm - Vertical' that is part of the top 3 with the highest expected delay from the previous query.

Output:

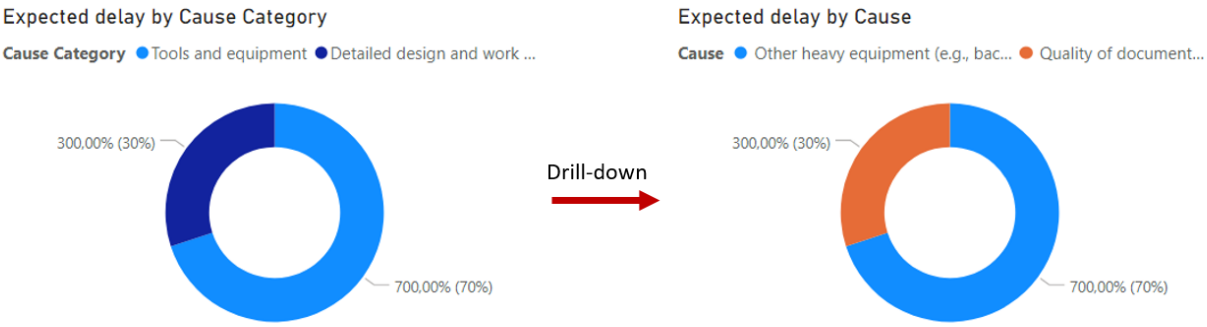
RUF:(IfcRelatedType:Other)Family Name	nb_elements	Task Count	Duration as-planned	Duration as-built	Probability of occurence	Impact	Expected delay
⊟ Concrete Precast Foundationpile Rectangular	**44**	**44**	**88**	**122**	**6,82%**	**566,67%**	**38,64%**
⊟ Concrete Precast Foundationpile Rectangular: 400x400mm - Vertical	**44**	**44**	**88**	**122**	**6,82%**	**566,67%**	**38,64%**
Concrete Precast Foundationpile Rectangular:400x400mm - Vertical:532545_(#39383)	1	1	2	16	100,00%	700,00%	700,00%
Concrete Precast Foundationpile Rectangular:400x400mm - Vertical:613413_(#710416)	1	1	2	16	100,00%	700,00%	700,00%
Concrete Precast Foundationpile Rectangular:400x400mm - Vertical:499854_(#8564)	1	1	2	8	100,00%	300,00%	300,00%
Concrete Precast Foundationpile Rectangular:400x400mm - Vertical:499110_(#1498)	1	1	2	2			
Concrete Precast Foundationpile Rectangular:400x400mm - Vertical:499114_(#1687)	1	1	2	2			
Concrete Precast Foundationpile Rectangular:400x400mm - Vertical:499118_(#1840)	1	1	2	2			
Concrete Precast Foundationpile Rectangular:400x400mm - Vertical:499122_(#1993)	1	1	2	2			
Concrete Precast Foundationpile Rectangular:400x400mm - Vertical:499126_(#2146)	1	1	2	2			

The output shows that three delays occurred of the 44 tasks that were performed. Based on this dataset, the project manager can retrieve the causes and related categories responsible for the delays.

Input:

Legend	*Path.1, Path.2 [durvarpath]*
Values	*Expected delay*
General filters	*ScheduleType [schedule] = "As-built"*
Search filter	*RUF:(IfcRelatedType:Other)Family and Type = "Concrete Precast Foundationpile Rectangular: 400x400mm – Vertical"*
Measures	*Same as previous*

Output:



From the output, it can be noted which cause and the related category was responsible for the delays part of the tasks related to the 'Concrete Precast Foundationpile Rectangular: 400x400mm – Vertical'. The cause category 'detailed design and work method' and related cause 'other heavy equipment (e.g., backhoe, loader, dump truck) not available' accounted for the most severe delays of 700 percent. For instance, the project manager can use this information to anticipate on the delay by making sure the required equipment is ordered well in advance for the upcoming project.

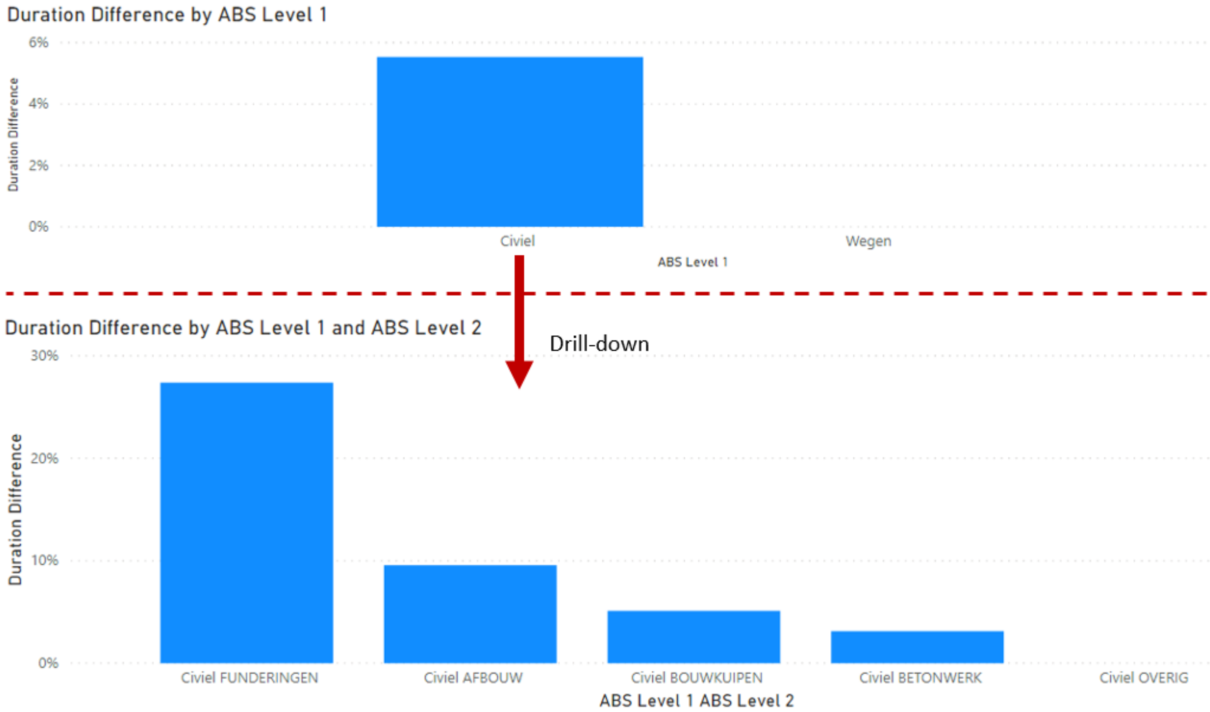
Query 12 – Performing tasks on time

Here, the use case from the perspective of the portfolio manager is considered. The portfolio manager can use the historical task duration information to determine the ability to complete tasks on time. For instance, related to tasks that are associated with certain types of projects or certain phases within certain projects. In this case, the difference in percentages between the as-planned and as-built duration are retrieved as part of the main ABS categories.

Input:

Axis	*Path.1, Path.2 [abstypepath], RUF:(Other)Family and Type [elementfieldvalue_resource_col], elementname [element_resource]*
Values	*Duration difference*
General filters	*ScheduleType [schedule] = "As-built"*
Measures	*Duration difference = DIVIDE(SUM('bim_schedules task'[DurationHours])-SUM('bim_schedules task'[BL_DurationHours]);SUM('bim_schedules task'[BL_DurationHours]))*

Output:



The output shows the historic duration delays as part of the civil constructions (civiel) category and none as part roads (wegen) category related to the first level of the ABS. The drill-down of the civil category allows retrieving the duration difference related to the categories of the second level of the ABS. The result shows that the foundation (funderingen) category historically differed the most from the as-planned durations. Thus, based on this information, the portfolio manager can say something about the ability to complete tasks on time, related to the different categories. The portfolio manager then might use this information to pursue projects that have a lower historical duration difference, such as road construction projects.

3.3. Analysis for the (re)use in the construction phase

In this section, it is described how the data within the database could have been analysed and (re)used related to the use cases within the construction phase. Thus, related to the planner at the construction site and the portfolio manager.

In this phase, the planner at the construction site uses the BIM-based as-planned schedule from the pre-construction phase to track the progress of the construction and generate the corresponding BIM-based as-built schedules. When the construction is finished, the final as-built schedule is generated. As described earlier, for the case study a BIM-based as-built in progress schedule is made after three months of construction.

Query 13 and 14 – Feedback on the estimated durations

From the perspective of the planner, the difference of task duration between the as-planned and as-built situation can be retrieved and possibly used to adjust the expected duration for similar tasks later in the project. Hence, the duration difference is retrieved for tasks related to the elements part of the BIM-based as-built in progress schedule. For this example, the element family names are used as the basis. However, task categories can be used as well.

Input:

Rows	*RUF:(Other)Family Name, RUF:(Other)Family and Type [elementfieldvalue_resource_col], Category [abs_category], elementname [element_resource]*
Values	*Duration As-planned Complete (measure), Duration As-built Complete (measure), Duration difference (measure), Duration, BL_DurationHours (sum), Durationhours (sum), PercentageComplete (avg) [task]*
General filters	*ScheduleType [schedule] = "As-built-progress"*
Measures	*- Duration As-planned Complete = CALCULATE(SUM('bim_schedules task'[BL_DurationHours]); 'bim_schedules task'[PercentageComplete] IN { 100 })* *- Duration As-built Complete = CALCULATE(SUM('bim_schedules task'[DurationHours]); 'bim_schedules task'[PercentageComplete] IN { 100 })* *- Duration Difference Completed = CALCULATE([Duration difference]; 'bim_schedules task'[PercentageComplete] IN { 100 })*

Output:

RUF:(IfcRelatedType:Other)Family Name	Duration As-planned Completed	Duration As-built Completed	Duration Difference Completed %	Total Duration As-planned	Total Duration As-built	Total Tasks Complete %
Floor	16	24	50,00%	1076	1084	6,90
Concrete Precast Foundationpile Rectangular	168	214	27,38%	168	214	100,00
CIP Rectangular Beam	336	368	9,52%	688	720	66,67
Sheetwall AZ	256	272	6,25%	448	464	50,00
Foundation Slab	304	320	5,26%	304	320	100,00
Basic Wall	1040	1056	1,54%	1400	1416	71,43
Arcelor Mittal Corner C14	28	28	0,00%	49	49	50,00
Circular Hollow Sections	43	43	0,00%	70	70	55,00
H-Wide Flange Beams	32	32	0,00%	56	56	55,19
Lightpost	40	40	0,00%	88	88	9,09
CIP Column Roundoff				384	384	0,00
Precast Spanbeton ZIP Beam				240	240	0,00

The output shows the element families within the BIM 4D model of which the underlying tasks are completed and the related duration difference. In addition, the total as-built and as-planned duration and the total percentage complete are shown. The elements related to the underlying tasks that faced a duration delay are visualized as part of the BIM 4D model in Figure 89.

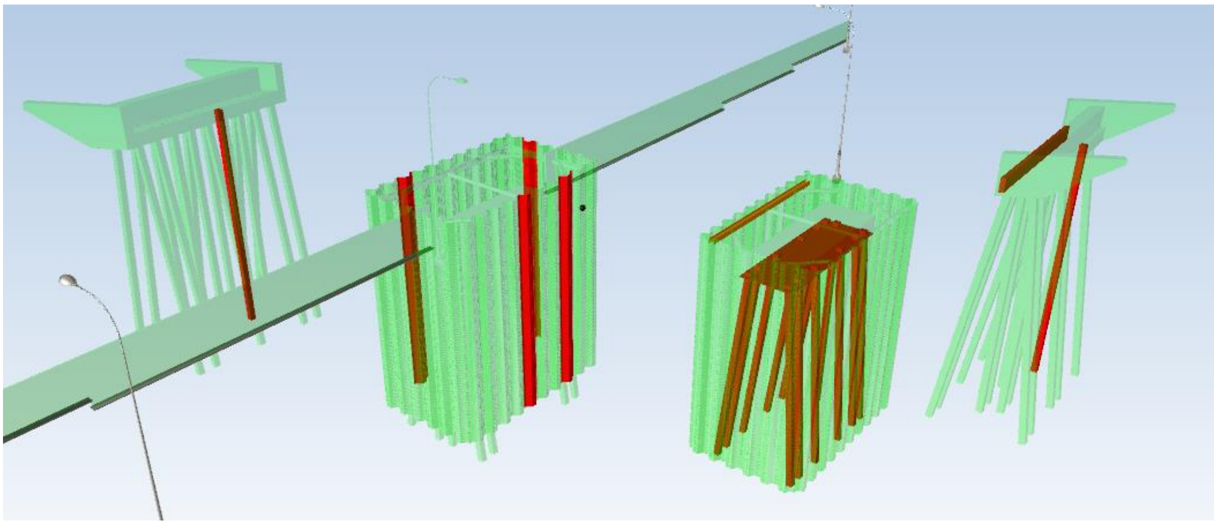


Figure 89: Elements of which its tasks faced a duration delay (red: task with a duration delay, green: tasks completed within the as-planned duration)

Tasks related to the element families of 'Floor', 'Concrete Precast Foundationpile Rectangular' and 'CIP Rectangular Beam' faced the largest duration delays up until this moment. Information related to the 'Concrete Precast Foundationpile Rectangular' cannot be used for later in the project since all tasks related to this element family are all completed.

To indicate how the information could be used, tasks related to the element family of 'CIP Rectangular Beam' are used as an example. The tasks indicate a duration delay for which 66,67% of the tasks are already complete. Therefore, it is of interest to analyse the related causes and possibly use the information to take mitigating measures and/or adjust the upcoming task durations. The elements that still need to be constructed as part of 'CIP Rectangular Beam' element family are shown in Figure 90.

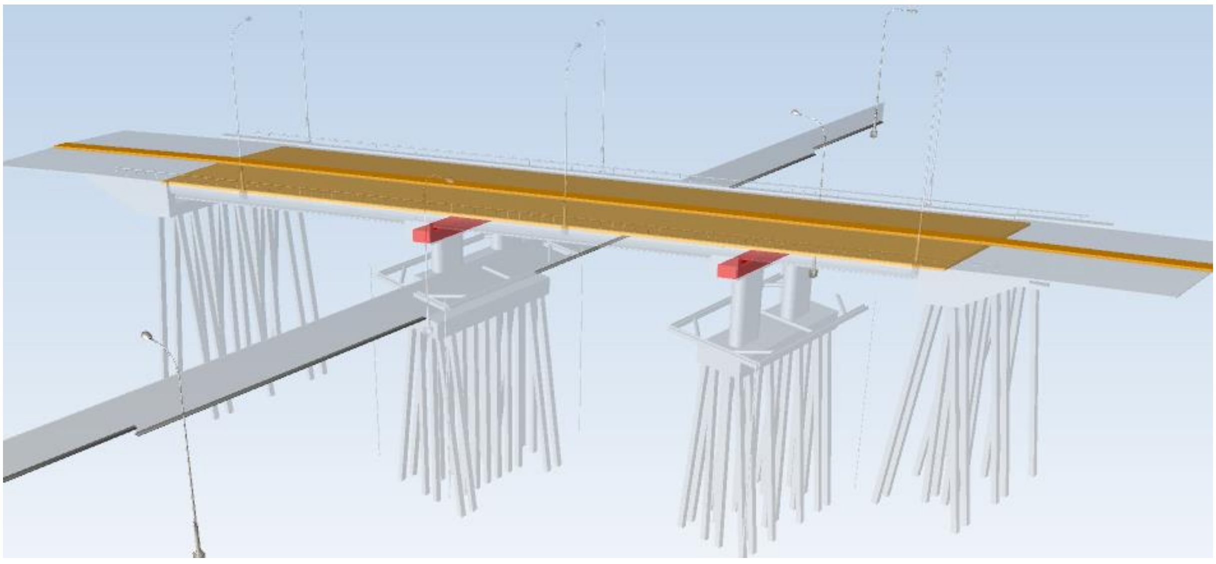


Figure 90: Elements that still need to be constructed for which similar elements and tasks already faced a duration delay (CIP Rectangular Beams indicated in red)

The planner can drill-down the data to retrieve the task categories similarly as this is done in query 3. Also, the planner can retrieve the causes responsible for the delays. The output is visualized below.

Input (bar chart):

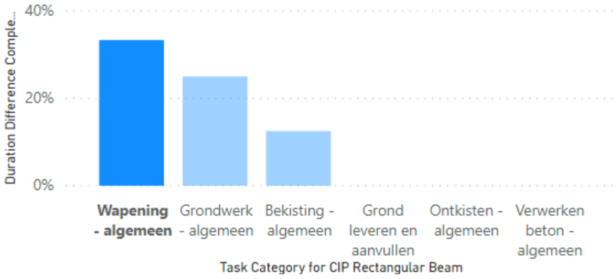
Axis	*Category [abs_category]*
Values	*Duration Difference Complete % (measure)*
General filters	*ScheduleType [schedule] = "As-built-progress"*
Search filter	*RUF:(IfcRelatedType:Other)Family Name = "CIP Rectangular Beam"*
Measures	*Same as previous*

Input (donut chart):

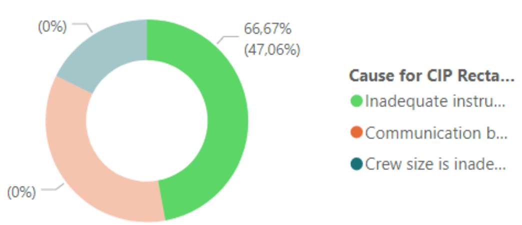
Legend	*Cause [durationvariation]*
Values	*Duration Difference Complete % (measure)*
General filters	*ScheduleType [schedule] = "As-built-progress"*
Search filter	*RUF:(IfcRelatedType:Other)Family Name = "CIP Rectangular Beam"*
Measures	*Same as previous*

Output:

Duration Difference Complete % by Task Category for CIP Rectangular Beam



Duration Difference Complete % by Cause for CIP Rectangular Beam



The output shows that the task category of rebar (wapening – algemeen) part of the 'CIP Rectangular Beam' element family faced the most significant duration difference. Related to this category, one cause (i.e. inadequate instruction on detailed working method) was responsible for the delays as is shown in the donut chart. As a result, the planner can suggest doing certain checks on the working methods related to tasks that still need to be performed. In addition, the planner could extend the duration of the upcoming tasks relative to the duration difference of tasks that already have been completed or by calculating the expected delay similar to query 10.

Query 15 – Progress track

During the construction phase, the project manager can utilize the database to monitor the progress based on the BIM-based as-built schedules inserted by the planner. In addition, the project manager can identify task duration risks similar to query 10 and 13. Since the data is stored in a central database, it becomes possible to retrieve the status of multiple ongoing projects at one place.

Every time as-built schedules are produced, the related logic in the schedule and the subsequent calculation of the critical path will produce a new completion date for the project. An overview of the construction progress in this stage is presented below.

Input:

Values	*ProjectName [project], Scheduletype, StartDate, FinishDate, DataDate [schedule], As-planned duration days, As-built duration days, Day delayed, Duration Difference (percentage delayed) (measures), BL_percentageComplete, PercentageComplete [task], Difference percentage complete (measure)*
General filters	*ScheduleType [schedule] = "As-built-progress"*
Search filter	*Taskname [task] = "BAM Viaduct"*
Measures	*- As-built duration days = DIVIDE(sum('bim_schedules task'[DurationHours]);8)* *- As-planned duration days = DIVIDE(sum('bim_schedules task'[BL_DurationHours]);8)* *- Days delayed = [As-built duration days] - [As-planned duration days]* *- Difference percentage complete = [As-built Percentage Complete] - [As-planned Percentage Complete]*

Output:

ProjectName	ScheduleType	StartDate	FinishDate	DataDate	As-planned duration days	As-built duration days	Days delayed	Percentage delayed	As-planned Percentage Complete	As-built Percentage Complete	Difference percentage complete
BAM Viaduct	As-planned	1-1-2020 9:00:00	21-12-2020 14:55:00	woensdag 1 januari 2020	253,75	253,75	0	0,00%	0,00%	0,00%	0,00%
BAM Viaduct	As-built-progress	1-1-2020 9:00:00	25-12-2020 14:55:00	dinsdag 14 april 2020	253,75	257,75	4	1,58%	37,00%	29,00%	-8,00%

The output shows that due to the current delays, the project estimated completion date is four days later compared to the as-planned completion date. Furthermore, it is indicated that the actual percentage of completed tasks is 29% while in the as-planned situation the tasks should already have been completed for 37%. The fact that this percentage is not relative to the total delay of the project (i.e. four days) can be explained by some delays that were not part of the critical path. Therefore, those delays, while within its float, will not extend the overall completion date. At the data date, the tasks of the following elements are behind schedule (Figure 91). Again, the drill-down possibilities allow retrieving information on the specific tasks that are behind of schedule.

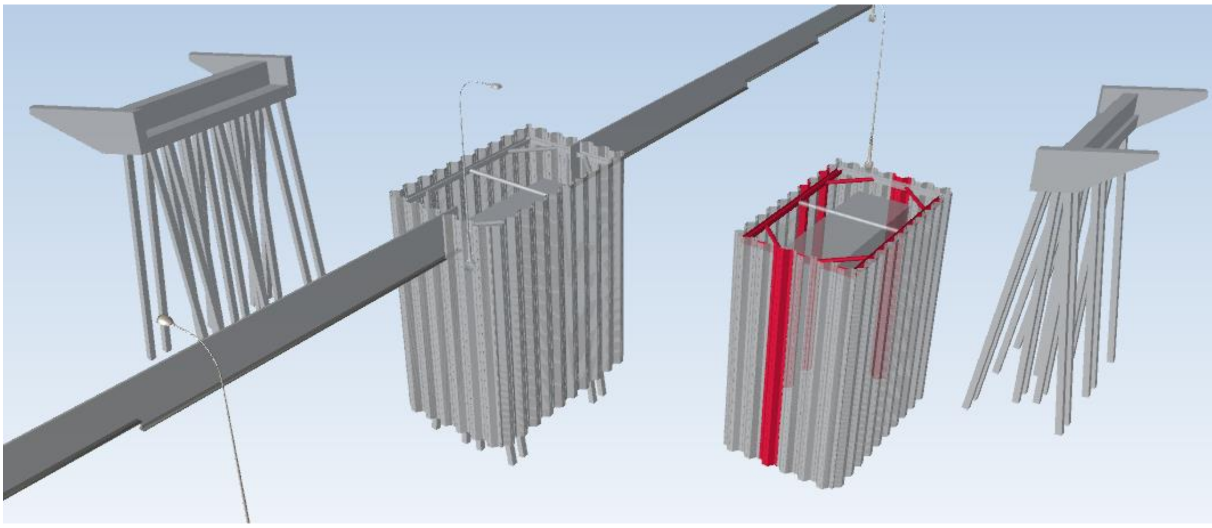


Figure 91: Elements and related tasks that are behind schedule indicated in red

Foremost, the project manager can use this data to access if the project is in control and where the project is behind of schedule. In addition, the project manager can propose or investigate possibilities to accelerate the project in a later stage to get the project back on track. For instance, by investigating the effect of a larger crew size, the use of other construction methods or by altering the sequence of certain tasks that are part of the critical path.

3.4. Analysis for the (re)use in the post-completion phase

Related to the post-completion phase, it is described how the data within the database could have been analysed and (re)used for the related use cases. Thus, related to the planner, planner at the construction site, the project manager and the portfolio manager. The portfolio manager will use the system similar to its use in the pre-construction phase.

Query 16 – Systematic feedback on the estimated durations

The planner that made the as-planned schedule, as well as, the planner that made the as-built schedules at the construction site, can use the database to retrieve systematic feedback on the accuracy of the estimated task durations. The planner can find where the duration has been exceeded and the reasons for it. Thus, it provides an opportunity for systematic learning. This can be done similar to query 14. The graph below shows the difference between the as-built and as-planned duration in percentages per task category and related causes. The drill-down functionality allows retrieving the specific tasks and related elements.

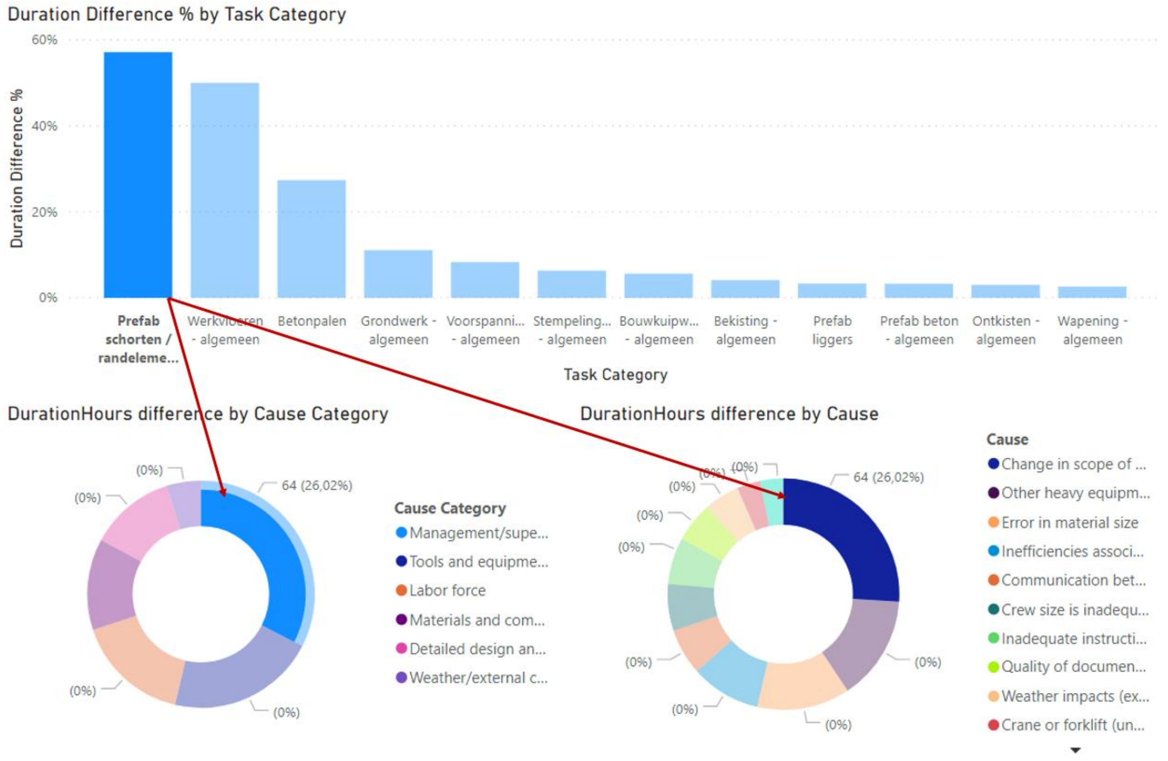
Input (bar chart):

Axis	*Category [abs_category]*
Values	*Duration Difference (measure)*
General filters	*ScheduleType [schedule] = "As-built"*
Measures	*Same as query 14*

Input (donut charts):

Legend	*Cause Category (path.1) [durationvarpath]*
Values	*Duration Difference (measure)*
General filters	*ScheduleType [schedule] = "As-built"*
Measures	*Same as query 14*

Output:



The output shows that the most significant duration delays are part of the task category prefabricated edge elements (prefab shorten / randelementen). The elements of which the related tasks faced a duration delay are visualized in Figure 92. The delays related to the task category prefabricated edge elements are caused by a 'change in scope of work' part of the cause category 'management/supervision/information flow' as can be seen in the figure above. Every other task category and related causes can be analysed in a similar fashion. The cause of a 'change in scope of work' might not be particularly useful from a learning perspective of a planner. However, many other causes might such as 'work sequence or method is not well planned', 'crew size is inadequate', etc.

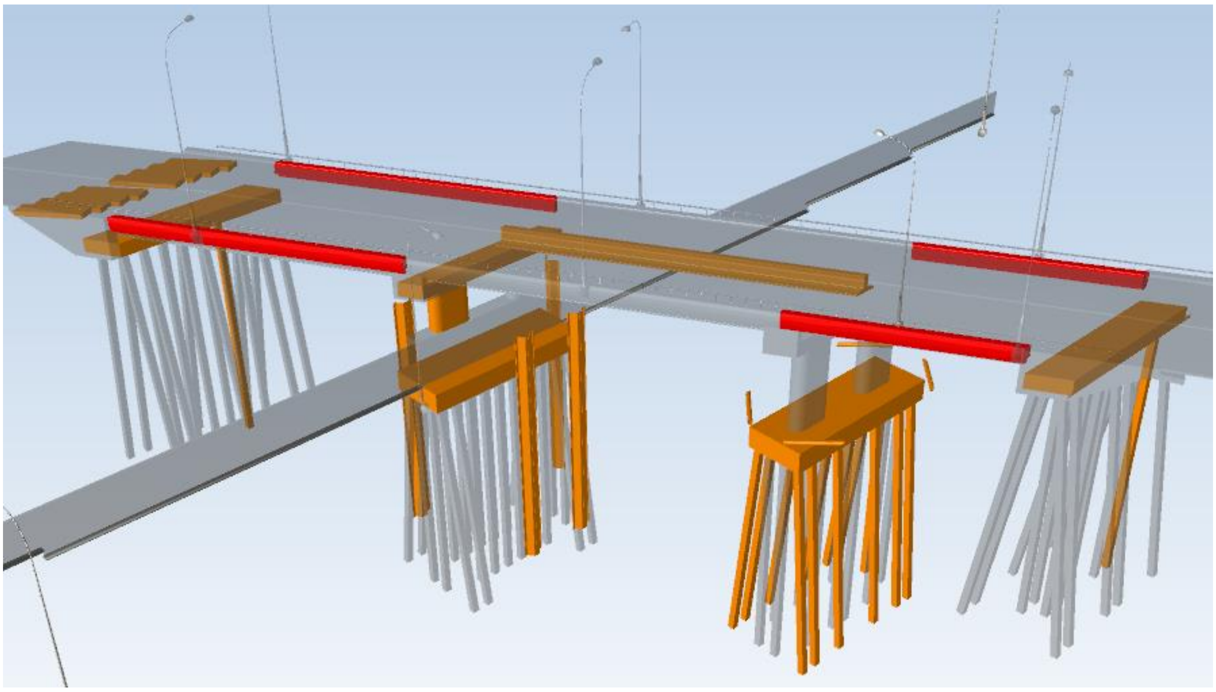


Figure 92: Elements of which the related tasks faced a duration delay (most significant duration differences % are indicated in red)

Query 17 – Project delay and claim management

The project manager can also use the information from the previous query to analyse the extend, impact and causes of certain task duration delays upon the total completion time of the project. This can be done by applying a critical path filter. In addition, the information can play a role in the process of claim management. Filters can be applied on the type of delay and the name of suppliers or subcontractors can be retrieved for this process. An example is shown below. The output shows the duration difference (delay) between the as-planned and as-built duration for the task categories and related elements part of the critical path.

Input:

Attributes (rows)	*RUF:(Other)Family and Type [elementfieldvalue_resource_col], Category [abs_category], Elementname [element_resource]*
Values	*Nb_elements (sum) [element_resource], ActivityCount (sum) [task], DurationHours difference (measure), SupplierName [element_resource], DelayType [task], BL_CriticalPath [task]*
General filters	*ScheduleType [schedule] = "As-built", Path.1 [Elementnlsfbpath] = is not blank, BL_CriticalPath [task] = True*
Measures	*Same as query 16*

Output:

Element family and type — Task category — Element name

RUF:(Other)Family and Type	nb_elements	Activity Count	Duration Hours difference	SupplierName	DelayType	CriticalPath
⊟ **CIP Column Roundoff: Concrete 2500x1000mm_R=500mm**						
⊟ **Bekisting - algemeen**						
CIP Column Roundoff:Concrete 2500x1000mm_R=500mm:488121_(#477)	1	2	24	ElementSupplier1	Non excusable	TRUE
⊟ **Wapening - algemeen**						
CIP Column Roundoff:Concrete 2500x1000mm_R=500mm:488121_(#477)	1	1	8	ElementSupplier1	Non excusable	TRUE
⊟ **Basic Wall: Concrete T=250mm**						
⊟ **Prefab beton - algemeen**						
Basic Wall:Concrete T=250mm:590310_(#703871)	1	1	16	ElementSupplier2		TRUE
⊟ **Precast Edge Beam: Edge Beam**						
⊟ **Prefab schorten / randelementen**						
Precast Edge Beam:Edge Beam:539453_(#75728)	1	1	8	ElementSupplier3	Excusable.Non compensable	TRUE
Precast Edge Beam:Edge Beam:539553_(#109220)	1	1	8	ElementSupplier3	Excusable.Non compensable	TRUE
⊟ **Sheetwall AZ: AZ18 - temporary**						
⊟ **Bouwkuipwanden - algemeen**						
Sheetwall AZ:AZ18 - temporary:586752_(#634467)	1	1	4	ElementSupplier2	Non excusable	TRUE
Sheetwall AZ:AZ18 - temporary:586758_(#637527)	1	1	4	ElementSupplier2	Non excusable	TRUE
Sheetwall AZ:AZ18 - temporary:586772_(#644017)	1	1	4	ElementSupplier2	Non excusable	TRUE
Sheetwall AZ:AZ18 - temporary:586784_(#649487)	1	1	4	ElementSupplier2	Non excusable	TRUE
⊟ **CIP Rectangular Beam: 1500x850mm**						
⊟ **Voorspanning - algemeen**						
CIP Rectangular Beam:1500x850mm:514025_(#22497)	1	1	8	ElementSupplier3	Non excusable	TRUE

4. (Re)use phase

The previous section described the various options for retrieving BIM-based scheduling data from the MySQL database using PowerBI as part of the case study. Roughly, the data can be used in three ways, namely before a new project (pre-construction phase), during the construction phase and when a project is completed.

4.1. Pre-construction phase

Focusing on the perspective of the planner, the data can be reused to determine the task duration related to elements from historical BIM 4D models in the database. As described in Chapter 3, it is assumed that a BIM model is available before a schedule is created. Thus, the names of the elements (family and type name) and associated properties from the BIM model can be used to search the database. Furthermore, it is shown that the elements and task data can be queried based on the applied classifications and the related hierarchical structure. In the current situation, task information can be retrieved related to disparate elements of the BIM model one at the time.

There are roughly three possibilities for reusing the data to estimate task durations as became clear from this case study. The first focuses on directly reusing the task duration as it is associated with a task and element in the database. In other words, if the input factors (element, the quantity of labour, equipment etc.), the internal environment and exogenous factors are the same as the task and related element in the new situation, then the task duration can be reused directly. When there are several comparable elements in the database with the same input factors, an average as-built task duration can be determined to increase the reliability of the duration. This first possibility can be considered as the retrieval of cases, where each task and associated element are represented as a single case. Subsequently, based on a new situation, the corresponding case can be looked up. This approach is somewhat similar to the CBR system proposed by Tauscher et al. (2007) that makes use of feature logic.

The second possibility is aimed at performing a productivity rate calculation. It is shown that productivity rates can be determined based on a specific element family and type or based on multiple types related to a single element family. The unit of measurement that is used for a productivity calculation, can be determined based on the comparability of certain characteristics of the elements (dimensions, etc.), the task category and other input factors (e.g. quantity of labour). As a result, a productivity value such as m^2/hour and m^3/hour can be determined relative to the number of person-hours. Alternatively, a productivity measure can be calculated without considering the number of person-hours. Both forms are considered as parametric estimation.

The third possibility is aimed at performing data analysis per task category and element family and type or as part of an umbrella category to determine the influence of the different factors upon task duration. In this way, it is possible to determine which factors are or are not important for the duration of a task. Different data analysis techniques could be used to predict task duration based on the given input. Linear regression and multiple linear regression are performed as illustrative examples to indicate this principle. The significance and weights that are assigned to the different units of measurement (length, width, height, weight, etc.) based on the analysis can potentially make the manual determination for the unit of measurement redundant. To achieve this, it is necessary to perform multiple data analyses. The currently proposed system could potentially be extended such that a system can automatically run a data analysis over certain elements and/or element groups and associated tasks.

The project manager can use the database to retrieve risks related to the duration of tasks. The probability of occurrence, impact and expected delay can be determined based on the historical difference between the as-planned and as-built task durations. The information can be retrieved according to the task classifications, element classifications, element family and types, and causes depending on how the output is structured. The project manager might be able to take mitigating measures to potentially prevent those risks from reoccurring within the current project.

The portfolio manager can utilize the roll-up and drill-down functionalities associated with the applied classifications to determine the ability to complete tasks on time, related to certain types of projects, certain main task categories or element categories.

4.2. During the construction phase

During the execution of a project, the planner at the construction site can retrieve feedback on the estimated durations for the tasks that have been completed. Subsequently, the affected tasks can be compared with similar task categories or elements that still need to be constructed. For the respective tasks, the duration might be adjusted by calculating the expected delay. In addition, the planner can suggest taking certain actions based upon the duration delay causes to prevent their reoccurrence.

The project manager can monitor the progress of the construction project based on the as-built schedules inserted into the database. Figures related to the new completion date and the percentage of tasks that are complete can be compared to the as-planned completion date and percentage of completed tasks. In this way, the status can be determined, and consequent actions can be taken.

4.3. Post completion phase

After completing the project, the data can be used to systematically provide feedback towards the planner on the task duration estimates made as part of the as-planned schedule. Similarly, it can provide feedback towards the planner at the construction site as part of the as-built schedules. As described earlier, the planner can find where task durations have been exceeded and its related causes. Therefore, it provides an opportunity for systematic learning. The project manager can also utilize this data for the part of delay claims, whereby it can be determined which tasks have been delayed for what reason and if these tasks were on the critical path. Hence, it can be demonstrated which task was responsible for a delay in the overall project. The division into the delay categories helps to determine whether a delay is compensable or not. The portfolio manager can use the as-built schedules to update the strategical objectives similar to its use in the pre-construction phase.

5. Validation

As part of the method in Chapter 3, it has been described that the validation of the system takes place as an iterative process throughout the development of the system and through alpha testing which is a type of acceptance testing. The alpha test aims to assess the functioning and usability of the system related to the requirements and specifications that are set as part of Chapter 3. For this, the results of the case study and the related process are used. Furthermore, some planners within BAM Infra have examined the input and output related to the case study and provided feedback. The test consists of a black-box test related to the input and output of the system, and a white-box test related to the internal operations of the system.

5.1. The input and internal process

Since data classifications enable the effective retrieval of task duration data, the planner must consistently apply classification codes for elements, tasks, labour, equipment and causes for variations. The same applies for the use of uniform property sets related to BIM model elements. Furthermore, the many-to-one relationship between elements and tasks should be applied. Hence, all the requirements needed to store and effectively (re)use the task duration data should be documented within some sort of information delivery manual. Complying to the requirements of the system can provide guidance for the structures that need to be maintained within the planning process.

Also, variables related to the internal environment and exogenous factor should be defined and included. However, these factors are not included in the current model. Therefore, the quantification of these factors and their implication needs further investigation. The whole process of defining all these factors is considered to be more labour-intensive compared to only describing the task name and linking the name to an element. Hence, the feasibility of this manual and detailed element tagging procedure should be further investigated.

Currently, the associated classification codes must be entered manually. For the export to MySQL to run smoothly, it is necessary to set the structure of the columns in Synchro 4D according to the defined order. Accidentally deviating from this order leads to an incorrect reading of the data in MySQL. When a planner wants to include new variables for instance related to tasks, equipment and labour, it is necessary to adjust the structure of the tables in

MySQL and the related Python script. Hence, in that respect, the system is not very flexible. Therefore, it is important to define the required variables correctly in advance.

Currently, graphical user interfaces (GUIs) are both missing for the ETL procedure executed with the Python script, as well as, an interactive user interface that connects to the MySQL database. In that regard, responsive web pages can be built that connect the user at the front-end (using a web page including buttons etc.) with underlying queries and scripts to the back-end (i.e. the database in MySQL or Python). Furthermore, it would be interesting to extend the tools to a more advanced integrated system.

The value of the system is also determined by the consistency of tracking the actual task durations and related factors during the execution of the project. This requires consistent monitoring of construction activities. This part has been out of scope for this project, and thus needs further investigation of course.

5.2. The output

The results of the case study showed that the information in the database can be utilized related to different use cases within the pre-construction, construction and post-completion phase of a project. The desired retrieval of information as specified in Section 3.2.1.2 can be met with the system, except for the limitations described below.

It is shown that historic BIM-based as-planned and as-built schedules can indeed be utilized by the planner to estimate task duration related to BIM model elements of a new project within the pre-construction phase. Hence, the system is limited to the application of tangible task duration data. The incorporation of the context on which the durations of tasks are based (i.e. the input factors, internal environment and exogenous factors) allows retrieving meaningful and useable data. This is accomplished with a database design that allows inserting data of BIM model elements and related properties, tasks and related inputs factors (i.e. labour and equipment), multiple classifications and information of delays causes and associated relations. However, factors concerning the internal environment and exogenous factors are currently not taken into account as part of the database design and this case study. Applying classifications codes in the BIM 4D model makes it possible to roll-up/summarize, group and filter the data in PowerBI. This allows to effectively search and calculate task duration. Furthermore, filters can be applied upon the attributes of the database that allow retrieving data under specific conditions.

Currently, for every disparate element in the BIM model, task duration should be queried individually. Task duration data can be transformed into a productivity measure based on another variable. As part of this process, the unit of measurement should be manually determined within PowerBI. This can be overcome by predefining the unit of measurement for all element types and related task categories. It may be that the necessary data is spread over several property variables which complicate the process. It should be noted that as the shape of an element becomes complex, the retrieval of data related to the dimensions of an element and the associated ability to calculate productivity data is hindered. Furthermore, it was found that in some cases it is difficult to grasp the meaning of certain values concerning the dimensions of the elements. As a result, it can become difficult to select the correct unit of measurement for the productivity calculation. Therefore, the ability to visualize the elements as part of the database would be beneficiary.

Data analysis should be conducted to determine and model the influence of multiple factors upon the duration of a task. It is shown that these factors can potentially be quantified using data mining tools, such as multiple regression analysis. Successfully modelling the influence of multiple factors could potentially yield more accurate task duration estimates and make the manual selection for the unit of measurement redundant. However, the results of the regression analysis were purely indicative as the case study relied upon a fictitious project. To validate such a model, it would require more detailed tests with actual data.

In the pre-construction phase, the project manager can identify risks by utilizing the database by retrieving information on the historical differences between the as-built and as-planned duration. It is shown, that based on the historical information, the probability of occurrence, impact and expect delays can be determined which can be used to take mitigation measures to potentially prevent their reoccurrence or alternatives can be investigated. A portfolio manager can use the roll-up functionalities to retrieve the duration differences between the as-built and as-planned duration.

During the construction phase and post-completion phase, it is shown that the planner at the construction site and the planner that generated the as-planned schedule can retrieve systematic feedback from the database upon the difference between the as-built and as-planned task durations and their related causes. During the construction phase, the planner can use this information to possibly adjust the duration of upcoming similar tasks that already faced a duration delay during the project. The project manager can use the system in a similar way. In addition, the project manager can track the progress of the project based on the percentage of tasks that are complete and the overall project completion date as part of the as-built schedules compared to the as-planned schedule.

Overall, more in-depth tests are required that use (fictitious) data that resemble an actual case. Thereupon, the added value of the system can be further validated in relation to different use cases scenarios. Once this proof is provided, a more sophisticated system can be developed.

5.3. The objectives

It is aimed to relate the current value of the system to the general and practical objective of this study. Thus, to evaluate the ability to extract reliable and accurate task duration data, as well as, the extent of the accessibility and effective usability of the system (i.e. the overall objective in this study). Similarly, the practical objective of this study is evaluated i.e. "creation of a system allowing a task duration/productivity database to be built, fed by BIM-based as-planned and as-built project scheduling data to estimate task durations".

It can be considered that the practical objective of this research has been reached, albeit mostly at a fundamental database and process level (without user interfaces). Furthermore, the database design can facilitate in the extraction of task duration data from different use case perspectives.

Regarding the overall objective, the database design allows to store information on the input factors that influence task duration; however, factors related to the internal environment and exogenous factors are not yet considered. Therefore, the current system cannot yet fully facilitate in storing all the data that can be utilized for the estimation of reliable and accurate task duration data. Moreover, the reliability and accuracy of the output depend on the input. For example, the output is negatively impacted when the progress of tasks and related influencing factors are not consistently tracked during the execution of a project. Also, the reliability and accuracy of the system to estimate task duration should be further investigated since the validation of the system relied upon a fictitious case study. The usability of the system is influenced by multiple factors such as the accessibility, performance and the ease of use. Currently, PowerBI is used as a medium to retrieve the contents of the database in an accessible manner. However, the effective usability and accessibility mainly dependent on the front-end (i.e. the GUIs), which this system currently lacks. Hence, separate GUIs are required to be built on top of the database and the ETL procedure. Moreover, integrating the system with other software tools will enhance usability. The level of effort required for collecting, storing and retrieving data also affect the usability. However, this is not validated within this study and needs to be further investigated.

6. Conclusion

The case study shows how 4D BIM-based as-planned and as-built models are created and how the related input factors are processed. The case study provides examples of how the data that is stored as part of the Plan phase, and the Do and Measure phase can be analysed and subsequently (re)used for different use cases related to the pre-construction phase, construction phase and post-completion phase. This chapter provides an answer to the sixth sub-research question indicating how well the system performs to estimate task durations using historical BIM-based as-planned and as-built scheduling data. The case study was used to perform an alpha test on the system to validate its functioning, which in this case mostly consists of (1) testing key queries and (2) a qualitative overall evaluation.

The system is found to fulfil the practical objective of this study being the creation of a system that allows task duration/productivity database to be built, fed by BIM-based as-planned and as-built project scheduling data to estimate task durations. Although, being mostly at a database and process level. The overall objective of the development of a system that allows extracting reliable and accurate task duration data, as well as, being accessible and effective usable, is partially reached. The ability to provide reliable and accurate task duration estimates depend on taking the influencing factors into account. Within the current database design, the input factors are taken into account; however, factors related to the internal environment and exogenous factors are not yet considered within this design. The effective usability and accessibility are mainly dependent on the front-end (user interfaces) of the system. Currently, PowerBI is used as a medium to retrieve the contents. However, to fully utilize the system, separate GUIs should be built on top of the database and related ETL procedures. Furthermore, it is recognized that applicability of the system depends on the compliance with the necessary input and the ability to consistently collect the required task progress data during the execution of the project. Further improvements need to be made in this regard when aiming at final and fully functional tools, which requires more elaborate programming and user interfaces.

To estimate task durations, it is found that the data can be reused in roughly three ways: 1) the task duration can be retrieved as it is stored and linked to an element and associated input factors, 2) in the form of parametric estimation, where a productivity calculation can be made based on the relationship between the task duration and the selected unit of measurement, and 3) with the use of data analysis, such as regression analysis, to determine the influence of the various factors on the duration and subsequently calculate the duration of a task. The last option could not be validated properly as the case study relied upon a fictitious project (and related factors). To further validate this possibility, a case must resemble actual project data.

Furthermore, it is found that the system can facilitate in providing systematic feedback towards the planner during the project and after completing the project. The system can also provide information on the status of a project and risks that impact the duration of tasks from the perspective of the project manager and portfolio manager.

(page intentionally left blank)

6. Conclusions and recommendations

In this chapter, conclusions for the study will be drawn by answering the sub-research questions that were formulated at the beginning of this research. Together, these conclusions provide an answer to the overall research question. After answering the research questions, recommendations for practical implementation and future research opportunities will be discussed.

1. Conclusion

This paragraph will answer the sub-research questions and conclude the main research question that was formulated based upon the problem statement, the overall objective and practical objective of this study. How well the research achieved the objectives will also be discussed.

1.1. Conclusion to sub-research questions

The first three sub-research questions relate to the theoretical part of this study. The last three sub-research questions relate to the practical part.

1. What does the process of BIM-based scheduling entail and what is the current state of the art?

BIM-based scheduling hinges on the definition of tasks based on the 3D elements contained within the BIM model. This process differs from a process for which a separate planning is made and thereafter an attempt is made to link the planning to the BIM model. The latter approach is mostly used in current applications; however, this process involves a cumbersome manual procedure of linking the BIM model and construction schedule, as well as, the sequencing of the construction processes. The related discrepancy between the element breakdown structure (EBS) of the BIM model and the work breakdown structure (WBS) of traditional construction schedules is considered as one of the major challenges. On the contrary, the current state of the art regarding BIM-based scheduling has shown that data can be extracted from the BIM model and linked with other data or tools for the automated generation of construction schedules. BIM-based scheduling is considered to be fundamental in the pursuit to automate the scheduling process, which is basically only possible on an element level, where element data is complemented with task data.

2. What is the current state of practices regarding progress tracking and delay analysis?

BIM-based state of the art progress tracking methods can facilitate in the creation of BIM-based as-built schedules. These techniques include Augmented Reality (AR), radio-frequency identification (RFID) chips, laser scanners and digital cameras. Although these techniques are deemed promising, it is found that based on different criteria such as time efficiency, accuracy, level of automation, preparation, costs of the technique, none of the current techniques performs ideally. Therefore, the choice of technique depends on the situation at hand. Alongside the monitoring of works, progress tracking can facilitate in finding discrepancies between the as-planned and as-built situation, their extent and discovery of the related causes.

Most delay analysis techniques (DATs) are solely aimed at analysing the effect of certain delays upon the total duration of a project. Therefore, DATs are closely related to claim management. Alternatively, productivity measurement methods can identify the main factors affecting productivity, as well as, be used to improve and potentially model productivity. Some methods, such as the Method Productivity Delay Model (MPDM), can pinpoint the cause of delays and subsequently allow to model their probability of occurrence and impact on productivity. The factor model is considered to be a suitable data-driven productivity measurement method since it aims to analyse and quantify the factors that influence productivity.

3. How can the duration of a task be estimated?
Deterministic (single-point) and the probabilistic (three-points) techniques are mostly used to estimate task duration. A variety of task duration estimation methods exist including subjective and data-driven methods. The choice of technique depends on the situation at hand. For example, the choice can be influenced by the available information and the current stage of the project. In the perspective of BIM-based scheduling and the potential reuse of historical BIM-based scheduling data, the method of parametric estimation arises as a favourable method. The tangible character of elements, the ability of quantity take-off and the linkages between tasks and elements in the BIM 4D model combined, form a potentially rich source of data. The factors that influence task duration can be roughly categorized by 1) input factors, 2) factors related to the internal environment, and 3) exogenous factors. The knowledge and quantification of these influential factors can potentially allow task duration to be modelled more accurately. The importance that these factors should be quantified becomes clearer in the perspective of the future developments surrounding BIM-based automated scheduling. For example, these factors can contribute to the exploration of different schedule scenarios.

4. How can BIM-based as-planned and as-built scheduling data be captured and stored in a database?
The capture of data in the AEC is done using so-called data modelling concepts. Data models can be categorized as relational stores (SQL) or NoSQL stores. SQL stores are mostly suited for structured data and rely upon predefined schemas. NoSQL stores provide more flexibility and scalability for the growth in data size and variety. However, NoSQL stores compromise on reliability, consistency and the ability to perform complex queries. Graph databases are part of NoSQL stores and designed to store linked data (the same principle applies for the storage of data in the semantic web).

The choice of a database depends on how the data is currently represented and the consequent analysis i.e. the initial objective of storing the data. For the proof of concept, it was chosen to develop a database schema in MySQL to reuse the data for task duration estimation. The cardinality between the task table and element table (i.e. many-to-one relationship) in the database as part of the schedule was found to be fundamental to extract meaningful task duration data. Furthermore, it is recognized that modelling guidelines and exchange requirements should be set up beforehand to ensure the (re)usability of the data. In addition, classifications should be applied within the BIM-based schedule to ensure unambiguous definitions and enhance search and query capabilities as part of the reuse phase. Within Synchro 4D, the BIM-based as-planned and as-built schedules can be exported as an Excel file that includes element property data. A Python script is developed that can be used to load the data into the MySQL database.

For the capture and storage of BIM-based scheduling data, this thesis proposes to enrich the task data with information regarding the input factors, internal environment and exogenous factors. During execution, these factors should be adjusted accordingly, as well as, be enriched with cause information for variations in the start and/or durations of tasks.

5. How can historical BIM-based as-planned and as-built scheduling data be analysed and (re)used for estimating task durations?
SQL syntax can be used to perform queries on the database. This allows for retrieving specific data that is requested. In general, the data can be analysed using reporting tools such as dashboards, statistical tools and data mining tools. Within this study, it was outside the scope to develop a separate user interface (UI) on top of the database that allows querying the database in a user-friendly manner. As a temporary solution in this study, the MySQL database is directly connected to PowerBI and used for the analysis and reuse of the data. In PowerBI one can search, slice and dice the data, perform exploratory analysis and calculate productivity rates. The predefined classifications tables allow for roll-up and drill-down operations to take place. Once sufficient BIM-based scheduling data is collected, data mining tools such as regression analysis can potentially be exploited to estimate task duration.

Since the elements in a BIM model form the basis of developing a new BIM 4D model, one can query task duration data based on the elements in the BIM model. The results of the case study showed that task duration can be estimated in roughly three ways: 1) the task duration can be retrieved as it is stored and linked to an element and associated input factors, 2) in the form of parametric estimation, where a productivity calculation can be made based on the relationship between task duration and the selected unit of measurement, and 3) with the use of data analysis, such as regression analysis, to derive a model/function that takes the input factors, internal environment and exogenous factors into account. Subsequently, the model can be used to calculate task duration.

6. How well does this system perform in estimating task durations using BIM-based as-planned and as-built scheduling data?
The case study showed how data that is stored as part of the Plan phase, and Do and Measure phase can be analysed and subsequently be (re)used. Several use cases were considered from the perspectives of a planner, a planner at the construction site, a project manager and a portfolio manager related to the pre-construction phase, construction phase and post-completion phase. An alpha test is used to validate the system that consists mostly of testing key queries and a qualitative overall evaluation. The current database design partially features for the storages of task duration data as the input factors are taken into account; however, factors related to the internal environment and exogenous factors are not yet considered. Overall, the system is found to enable the estimation of task durations using BIM-based as-planned and as-built scheduling data. Although, being mostly at a database and process level. The applicability of the system depends on the compliance with the necessary input and the ability to consistently collect the required task progress data during the execution of the project. The effective usability and accessibility mainly dependent on the front-end (user interfaces) which the current system currently lacks. Thus, separate GUIs should be built on

top of the database and related ETL procedures to utilize the system. The accuracy and reliability of task duration estimates mostly rely upon taking influencing factors into account. Within the case study, the option of using data analysis to estimate task duration could not be validated properly as the case study relied upon a fictitious project (and related factors). To further validate this possibility, it is necessary for a case to resemble actual project data.

Furthermore, it was found that the system can facilitate in providing systematic feedback towards the planners on the estimated durations during the construction of the project and in the post-completion phase. Project managers and portfolio managers can mainly utilize the system for the identification of risks related to task duration and to monitor the progress of the construction.

1.2. Conclusion to the main research question

In this paragraph, the conclusion will be drawn to answer the main research question of this study. The main research question is formulated as follows:

"How can BIM-based as-planned and as-built scheduling data be utilized to estimate task duration for future construction projects in a pre-construction phase?"

It was found that multiple methods can be used to determine task duration. However, parametric estimation is considered to be a well-suited method because of the tangible nature of elements in a BIM model and associated tasks. The duration of a task is influenced by input factors, the internal environment and exogenous factors. The documentation of these factors as part of the BIM-based as-planned and as-built schedules is considered to increase the (re)usability since it provides context on which the duration of the task is based and subsequently allows to estimate/model task duration. Moreover, ascribing cause information to task deviations in the construction phase is considered to increase the reusability and enables systematic learning.

How BIM-based scheduling data is created and stored is fundamental for its subsequent effective utilization. The classification of tasks, elements and other factors are important to ensure unambiguous definitions and enhance search and query capabilities. Also, exchange requirements and modelling guidelines are required. In that regard, it is found that a many-to-one relationship between tasks and elements is required. The separate explicit knowledge that is embedded in BIM-based as-planned and as-built scheduling data can be converted to systematic explicit knowledge when the data is stored in a single source i.e. a database. Hence, it can facilitate in systematically unlocking the knowledge that was previously divided over multiple individuals, documents and/or files.

In this study, the data of the BIM model elements and related properties, tasks and associated influencing factors are stored in a MySQL database. Subsequently, the data can be queried, visualized and analysed with the use of varies tools. In this study, PowerBI was used for the reuse of the BIM-based scheduling data. It is shown that one can query and derive a productivity measure to estimate task duration based on element names, element classifications and related task classifications. Furthermore, the different factors can potentially be utilized to derive prediction models with the use of data mining tools, such as the factor model. Subsequently, these models can be used to estimate task duration.

1.3. Discussion on the research objective

The objective of this study was based on the identified causes and problems that relate to inefficiency, and errors and miscalculations as part of construction scheduling, as well as, problems related to tracking the progress on-site. The overall objective of this research was to develop a system that enables to retrieve reliable and accurate task duration scheduling data, as well as to make the data accessible and effectively useable in future planning processes. It is concluded that the developed system can partially facilitate in the extraction of reliable and accurate task duration data since the input factors are taken into account; however, factors related to the internal environment and exogenous factors are not yet considered within this design. Moreover, the applicability of estimation models such as regression functions could not be validated as the case study relied upon a fictitious project (and related factors). The system is found to be useable from different use case perspectives related to the pre-construction, construction and post-completion phase. The effective usability and accessibility mainly dependent on the front-end (user interfaces) which the current system currently lacks. Hence, the current usability and accessibility from the perspective of an end-user are limited.

2. Practical recommendations

The findings of this study can be potentially useful for construction companies when they want to reuse BIM-based scheduling data. Therefore, practical recommendations are given for the implementation of a system that allows for BIM-based scheduling to be reused. Since this study is conducted in collaboration with BAM Infra, short-term and long-term (after a year) recommendations are provided based on their current situation. Fulfilling the short-term recommendations is a precondition for realizing the long-term recommendations.

2.1. Short-term

1. Adopt the BIM-based scheduling philosophy: Adopting the philosophy where the elements of the BIM model form the basis of the construction schedule is considered fundamental for the potential reuse of the data. To do so, it is key to enforce the required cardinality between BIM elements and tasks.

2. Make use of classifications, exchange requirements and modelling guidelines: The use of standards/classifications ensures unambiguous definitions and enhances search and query capabilities. It is recommended to use classifications (NL-SfB, IFC, …) for information it can be applied to, such as elements, tasks, labour, equipment and cause variation information. Primarily, the classification of tasks is important. Furthermore, it is recommended to establish modelling guidelines and define exchange requirements (e.g. with the use of IDM) to ensure the (re)usability of the data. Hereby, the uniform definition and documentation of element variables as part of the BIM model should be ensured.

3. Define and collect influential factors upon task duration in a standardized manner: For the ability to (re)use data, it is recommended to standardize the way in which influential factors (i.e. input factors, factors related to the internal environment and exogenous factors) upon task duration are defined and documented as part of the BIM-based as-planned schedule. Likewise, it is recommended to standardize how these factors and causes of duration variations are monitored and documented during the execution process as part of the as-built schedules.

4. Store the BIM-based as-planned and as-built scheduling data in a database: Systematically storing the BIM-based data in a database such as SQL allows for fact-based (re)use of the data from a single source. It is recommended to well-define the structure of the database in a way that allows being reused on a detailed level by planners, as well as, to be (re)used by project managers and others in a more overarching manner.

2.2. Long-term

1. Conduct more in-depth case studies for different use case scenarios to validate the added value of the system: Conducting more in-depth tests that use (fictitious) data that resemble actual cases can provide a more thorough validation. Which can be used to prove the added value of the system in relation to different use cases scenarios. Once it is found that the system brings the required value, a more sophisticated system can be developed.

2. Explore the application of data mining tools and determine their usability: The second recommendation is to explore the application of different data mining tools or machine learning algorithms. Subsequently, the usability and accuracy of these tools should be determined for task duration estimation or other purposes.

3. Develop a more sophisticated system: The last recommendation is to develop a more sophisticated system (including separate user interfaces) that enhances the usability. It is recommended to explore the possibilities to integrate the system with the used planning tools, (management) reporting systems and other administrative systems (e.g. a foreman diary system).

3. Research recommendations

In this paragraph, recommendations for future research opportunities are discussed.

The proposed system in this study is based on specific BIM modelling and 4D modelling software and its related capabilities of representing the accompanying data. However, to avoid being limited to certain software applications and improve the sharing of information, it would be of interest to study the application of openBIM standards, such as IFC. Therefore, this study proposes to study how IFC data files can become the fundamental source of information within this process, as well as, how the related influencing factors can be linked or stored with the use of this non-proprietary format.

The validation of the system has been carried out by means of a fictitious case study. Since no actual data was used, it was not possible to validate all aspects of the system properly. Especially, this research could be extended to validate the applicability of conditional task duration estimates that can currently be derived from the database, as well as, the use of prediction models to estimate task duration. In both cases, it would be of interest to estimate task duration based upon BIM-based as-planned and as-built scheduling data. Subsequently, a comparative study can be executed that compares those estimates with task duration estimates that are based upon expert’s judgement or parametric estimation using static ratios (e.g. with the use of the RSMeans database).

This study has shown that task duration for each task related to an element can be queried individually based on task classification categories and related elements. Further research can explore how the elements in the BIM model can be (semi-)automatically linked to the associated tasks and durations within a BIM 4D tool based on the information of a BIM-based as-planned and as-built database. For estimation of task duration, it would be of interest to study how for instance certain rules, formulas or other functions can be automatically ascribed to certain tasks. And subsequently, how task duration can be automatically calculated based upon defined input factors.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Ai Communications: The European Journal on Artificial Intelligence, 7(1)*, 39-59.

Aggarwal, C. C., & Wang, H. (2010). *Managing and Mining Graph Data.* New York: Springer.

Ahmad, H., Ayoush, M., & Al-Alwan, M. (2019). Causes of delay to public infrastructure projects according to engineers representing different contract parties. *Built Environment Project and Asset Management, 10*(1), 153-179. doi:10.1108/BEPAM-03-2019-0026

Akhavian, R., & Behzadan, A. H. (2016). Smartphone-based construction workers' activity recognition and classification. *Automation in Construction, 71*, 198-209. doi:https://doi.org/10.1016/j.autcon.2016.08.015

Al Hammadi, S., & Nawab, M. (2016). Study of Delay Factors in Construction Projects. *International Advanced Research Journal in Science*, Vol. 3, Issue 4, 87-93. doi:10.17148/IARJSET.2016.3420

Aouad, G., Wu, S., Lee, A., & Onyenobi, T. (2012). *Computer Aided Design Guide for Architecture, Engineering and Construction.* London: Routlegde. doi:10.4324/9780203878750

Arditi, D. (1985). Construction Productivity Improvement. *Journal of Construction Engineering and Management, 111*(1), 1-14.

Attal, A. (2010). *Development of Neural Network Models for Prediction of Highway Construction Cost and Project Duration (Master's thesis).* Civil Engineering, Russ College of Engineering and Technology of Ohio University.

Azaaj, T. (2018). *Improving the accuracy of project duration with artificial intelligence .*

Azeem S.A., A., Hosny, E. H., & Ibrahim, A. H. (2014). Forecasting project schedule performance using probabilistic and deterministic models. *HBRC Journal*, 10, 35–42. doi:10.1016/j.hbrcj.2013.09.002

Baldwin, A., & Bordoli, D. (2014). *Handbook for Construction Planning and Scheduling.* John WIley & Sons, Ltd.

Ballesteros-Pérez, P., Sanz-Ablanedo, E., Soetanto, R., González-Cruz, M., Larsen, G., & Cerezo-Narváez, A. (2020). Duration and Cost Variability of Construction Activities: An Empirical Study. *Journal of Construction Engineering and Management, 146*(1), 04019093. doi:https://doi.org/10.1061/(ASCE)CO.1943-7862.0001739

Borrmann, A., König, M., Koch, C., & Beetz, J. (2018). *Building information modeling : Technology foundations and industry practice.* Cham, Switzerland: Springer. doi:10.1007/978-3-319-92862-3

Briggs, D. (2017). Comparison between Deterministic and Stochastic Time Estimating Techniques. *International Journal of Advanced Research in Science, Engineering and Technology, 4*(7), 4302-4311.

Büchmann-Slorup, R., & Andersson, N. (2010). BIM-based scheduling of Construction - A comparative analysis of prevailing and BIM-based scheduling processes. *CIB W78 2010: 27th International Conference*, (pp. 1-10). Cairo, Egypt.

Building Smart International. (n.d.). *IFC Formats*. Retrieved from buildingsmart: https://technical.buildingsmart.org/standards/ifc/ifc-formats/

BuildingSmartAlliance. (n.d.). *4.4.2 Property Sets*. Retrieved from buildingsmartalliance: http://docs.buildingsmartalliance.org/MVD_WSIE/schema/templates/property-sets.htm

Butkovic, B., & Heesom, D. (2017). Towards A Framework For Multi-LOD 4D BIM Simulations. *International Research Conference 2017*, (pp. 578-586).

Butkovic, B., Heesom, D., & Oloke, D. (2019). The Need For Multi-LOD 4D Simulations In Construction. *Journal of Information Technology in Construction (ITcon), 24*, 256-272. Retrieved from http://www.itcon.org/2019/14

Chan, D., & Kumaraswamy, M. (2002). Compressing construction durations: lessons learned from Hong Kong building projects. *International Journal of Project Management, 20*, 23-35.

Changali, S., Mohammad, A., & Nieuwland, M. (2015). *The construction productivity imperative*. Retrieved from Mckinsey: https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/the-construction-productivity-imperative

Chartered Institute of Building. (2010). *Guide to Good Practice in the Management of Time in Complex Projects.* Chichester: John Wiley & Sons.

Collins, H. M. (2010). *Tacit and explicit knowledge.* University of Chicago Press.

Czarnigowska, A., & Sobotka, A. (2013). Time-Cost Relationship for Predicting Construction Duration. *Archives of Civil and Mechanical Engineering, 13(4)*, 518-526.

De Vries, B., & Harink, J. (2007). Generation of a construction planning from a 3D CAD model. *Automation in Construction, 16(1)*, 13-18. doi:10.1016/j.autcon.2005.10.010

Deutsch, R. (2015). *Data-Driven Design and Construction: 25 Strategies for Capturing, Analyzing, and Applying Building Data.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Dozzi, S., & AbouRizk, S. (2011). *Productivity in Construction.*

Drucker, P. (1998). *The coming of the new organization.* Harvard Business Review on Knowledge Management, pp. 1–19. Harvard Business School Press, Boston.

Dülger, S. (2020). *Using point cloud to automatically update the BIM 4D-Model (master's thesis).* Retrieved from https://www.ofcoursecme.nl/?mdocs-file=5356

Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors.* John Wiley & Sons.

Faghihi, V. N., Reinschmidt, K., & Kang, J. (2015). Automation in construction scheduling: a review of the literature. *The International Journal of Advanced Manufacturing Technology, 81(9-12)*, 1845-1856. doi:10.1007/s00170-015-7339-0

Fan, J., Li, G., & Zhou, L. (2011). Interactive SQL Query Suggestion: Making Databases User-Friendly. *27th International Conference on Data Engineering, ICDE 2011*, (pp. 1-13). Hannover.

Farghal, S., & Everett, J. (1997). Learning curves: accuracy in predicting future performance. *Journal of Construction Engineering and Management, 123*(1), 41-45.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine, 17*(3), 37-51.

Flyvbjerg, B., Bruzelius, N., & Rothengatter, W. (2003). *Megaprojects and risk : An anatomy of ambition.* United Kingdom: Cambridge University Press.

GCCG. (2011). *A Report for the Government Construction Client Group – BIM Working Party Strategy Paper.*

Ghoddousi, P., & Hosseini, M. (2012). A survey of the factors affecting the productivity of construction projects in Iran. *Technological and Economic Development of Economy, 18*(1), 99-106.

Goldratt, E. (1997). *Critical chain.* Aldershot: Gower.

Gopinath, S., Varadharajan, R., & Thavva, R. (2018). Activity Duration Forecast Model for Construction Projects in Accordance with Standard Tools and Techniques. *Journal of Advanced Research in Dynamical and Control Systems, 10, 08*, 1071-1077.

Groot, P. (2019). *Infrastructuurmonitor - Voortgang projecten Infrastructuurfonds en Deltafonds 2018-2019.* EIB.

Grushka-Cockayne, Y. (2020). *Use Data to Revolutionize Project Planning*. Retrieved from Harvard Business Review: https://hbr-org.cdn.ampproject.org/c/s/hbr.org/amp/2020/02/use-data-to-revolutionize-project-planning

Guevremont, M., & Hammad, A. (2018). Multi-LOD 4D Simulation in Phased Rehabilitation. *Proceedings of the 17th International Conference on Computing in Civil and Building Engineering (ICCCBE)*, (pp. 724-731). Tampere.

Gündüz, M., Nielsen, Y., & Özdemir, M. (2013). Quantification of Delay Factors Using the Relative Importance Index Method for Construction Projects in Turkey. *Journal of Management in Engineering, 29*(2), 133-139. doi:10.1061/(ASCE)ME.1943-5479.0000129.

Hajdu, M., & Bokor, O. (2014). The Effects of Different Activity Distributions on Project Duration in PERT. *Procedia - Social and Behavioral Sciences 119*, pp. 766-775. doi:10.1016/j.sbspro.2014.03.086

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3rd ed.). Waltham, US: Morgan Kaufmann.

Hartmann, V., Beucke, K. E., Shapir, K., & König, M. (2012). Model-based Scheduling for Construction Planning. *14th International Conference on Computing in Civil and Building Engineering*, (pp. 1-8). Moscow, Russia.

Hasan, A., Baroudi, B., Elmualim, A., & Rameezdeen, R. (2018). Factors affecting construction productivity: a 30 year systematic review. *Engineering, Construction and Architectural Management, 25*(7), 916-937.

Heigermoser, D., De Soto, B. G., Abbott, E. L., & Chuad, D. K. (2019). BIM-based Last Planner System tool for improving construction project. *Automation in Construction 104*, 246–254. doi:10.1016/j.autcon.2019.03.019

Helton, J., Johnson, J., Oberkampf, W., & Sallaberry, C. (2010). Representation of analysis results involving aleatory and epistemic uncertainty. *International Journal of General Systems, 39*(6), 605-646.

Hendrickson, C. T., & Au, T. (1989). *Project management for construction: Fundamental concepts for owners, engineers, architects, and builders.* Englewood Cliffs, N.J: Prentice Hall.

Hendrickson, C., Martinelli, D., & Rehak, D. (1987). Hierarchical rule-based activity duration estimation. *Journal of Construction Engineering and Management, 113*(2), 288-301.

Heravi, G., & Eslamdoost, E. (2015). Applying Artificial Neural Networks for Measuring and Predicting Construction-Labor Productivity. *Journal of Construction Engineering and Management, 141(10)*, 04015032. doi:10.1061/(ASCE)CO.1943-7862.0001006

Heuton, J. (2020, June). *Recipe Fundamentals*. Retrieved from Alice Technologies: https://support.alicetechnologies.com/hc/en-us/articles/360001262594-Recipe-fundamentals

Heylighen, A., & Cavallin, H. (2005). Knowledge exchange in the wild. Building stories' attempt to unlock the knowledge capital of architectural practice., (p. 11).

Hoda, A., Ossama, H., Khaled, N., & Rania, E. (2018). Delay Analysis Techniques in Construction Projects. *Canadian Society of Civil Engineers*, (pp. 1-10).

Hoffman, F. O., & Hammonds, J. (1994). Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability. *Risk Analysis, 14*(5), 707-712.

Hoffman, G., Thal Jr., A., Webb, T., & Weir, J. (2007). Estimating Performance Time for Construction Projects. *Journal of Management in Engineering*, 193-199.

Hoogeveen, N. (2015). *Building Knowledge Modelling - From BIM data to Artificial Intelligent Knowledge Systems (Master's thesis).*

Hu, X. (2015). *BIM-Based Data Mining Approach to Estimating Job Man-Hour Requirements in Structural Steel Fabrication.*

Huhnt, W., & Enge, F. (2006). Can algorithms support the specification of construction schedules? *Electron J. Inf. Technol. Constr. 11*, (pp. 547-564).

Hwang, B.-G., Zhao, X., & Ng, S. Y. (2013). Identifying the critical factors affecting schedule performance of public housing projects. *Habitat International 38*. doi:http://dx.doi.org/10.1016/j.habitatint.2012.06.008

Innocent, M., Wasek, J. S., & Franz, A. (2018). Predicting Military Construction Project Time Outcomes Using Data Analytics. *Engineering Management Journal*, 30:4, 232-246. doi:10.1080/10429247.2018.1490995

International Organization for Standardization . (2018). *ISO 19650-1: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) - Information management using building information modelling - Part 1: Concepts and principles.*

Jaśkowski, P., & Sobotka, A. (2006). Scheduling Construction Projects Using Evolutionary Algorithm. *Journal of Construction Engineering and Management*, 861-870.

Javed, A., Pan, W., Chen, L., & Zhan, W. (2018). A systemic exploration of drivers for and constraints on construction productivit enhancement. *Built Environment Project and Asset Management, 8*(3), 239-252. doi:10.1108/BEPAM-10-2017-0099

Jongeling, R., Kim, J., Fischer, M., Mourgues, C., & Olofsson, T. (2008). Quantitative analysis of workflow, temporary structure usage, and productivity using 4D models. *Automation in Construction 17*, 780-791. doi:10.1016/j.autcon.2008.02.006

Jung, Y., & Kang, S. (2007). Knowledge-Based Standard Progress Measurement for Integrated Cost and Schedule Performance Control. *Journal of construction engineering and management*, 10-21.

Kallantzis, A., & Lambropoulos, S. (2004). Critical path determination by incorporation dminimum and maximum time and distance constraints into linear scheduling. *Engineering, Construction and Architectural Management, 11(3)*, 211-222.

Kelley, J. (1961). Critical-path planning and scheduling: Mathematical basis. *Operations Research, 9(3)*, 296-320.

Kim, H., Anderson, K., Lee, S., & Hildreth, J. (2013). Generating construction schedules through automatic data extraction using open BIM (building information modeling) technology. *Automation in Construction*, 1-11. doi:http://dx.doi.org/10.1016/j.autcon.2013.05.020

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit - The definitive guide to dimensional modeling* (3rd ed.). Wiley & Sons.

König, M., Koch, C., Habenicht, I., & Spieckermann, S. (2012). Intelligent BIM-based construction scheduling using discrete event simulation. *Proceedings - Winter Simulation Conference*, (pp. 662-673). WSC 2012 Berlin, DEU. doi:10.1109/WSC.2012.6465232

Koninklijk Nederlands Normalisatie-instituut. (2015). *Building construction - Organization of information about construction works - Part 2: Framework for classification (ISO 12006-2:2015.,IDT).*

Koninklijk Nederlands Normalisatie-instituut. (2018). *Work breakdown structures for project and programma management (ISO 21511:2018,IDT).* https://www.iso.org/standard/69702.html.

Koninklijk Nederlands Normalisatie-instituut. (2020). *NTA 8035: Semantic modeling of information in the built environment.*

Kopsida, M., Brilakis, I., & Vela, P. (2015). A Review of Automated Construction Progress Monitoring and Inspection Methods.

KPMG. (2000). *Knowledge Management Research Report 2000.* KPMG Management Consulting.

Lee, J., Park, Y.-J., Choi, C.-H., & Han, C.-H. (2017). BIM-assisted labor productivity measurement method for structural formwork. *Automation in Construction, 84*, 121-132. doi:https://doi.org/10.1016/j.autcon.2017.08.009

Lester, A. (2017). *Project management, planning and control: managing engineering, construction, and manufacturing projects to pmi, apm, and bsi standards* (Seventh ed.). Oxford: Butterworth-Heinemann.

Lin, J., Hu, Z., & Zhang, J. (2013). BIM oriented intelligent data mining and representation. *30th CIB W78 International Conference*, (pp. 280-289).

Liu, H., Al-Hussein, M., & Lu, M. (2015). BIM-based integrated approach for detailed construction under resource constraints. *Automation in Construction 53*, 29-43. doi:http://dx.doi.org/10.1016/j.autcon.2015.03.008

Lovallo, D., & Kahneman, D. (2003). *Delusions of Success: How Optimism Undermines Executives' Decisions*. Retrieved from Harvard Business Review: https://hbr.org/2003/07/delusions-of-success-how-optimism-undermines-executives-decisions

Love, P., Tse, R., & Edwards, D. (2005). Time–Cost Relationships in Australian Building Construction Projects. *Journal of Construction Engineering and Management, 131(2)*, 187-194. doi:10.1061/(ASCE)0733-9364(2005)131:2(187)

Maheshwari, A. (2015). *Data Analytics Made Accesible.*

Malara, J., Plebankiewicz, E., & Juszczyk, M. (2019). Formula for Determining the Construction Workers Productivity Including Environmental Factors. *Buildings, 9*(12), 1-14. doi:https://doi.org/10.3390/buildings912024

Malisiovas, A. (2010). Construction Productivity: From measurement to improvement. *PM-05 - Advancing Project Management for the 21st Century*, (p. 8). Heraklion, Crete.

Moylan, W. A. (2002). Planning and scheduling: the yin and yang of managing a project. *Paper presented at Project Management Institute Annual Seminars & Symposium.* San Antonio, TX. Newtown Square: Project Management Institute.

Mubarak, S. (2015). *Construction project scheduling and control* (3 ed.). Upper Saddle River, N.J: Wiley.

Nagata, M., Manginelli, W., Lowe, S., & Trauner, T. (2018). *Construction delays* (Third ed.). Oxford, United Kingdom: Butterworth-Heinemann is an imprint of Elsevier.

Navon, R. (2005). Automated project performance control of construction projects. *Automation in Construction, 14*(4), 467-476.

Niknam, M., & Karshenas, S. (2016). Integrating BIM and Project Schedule Information Using Semantic Web. *Construction Research Congress*, (pp. 689-697). doi:10.1061/9780784479827.070

Oberlender, G. (2014). *Project management for engineering and construction* (3rd ed.). New York: McGraw-Hill Education LLC.

Olawale, Y. A., & Sun, M. (2010). Costs and time control of construction projects: inhibiting factors and mitigating measures in practice. *Construction Management and Economics, 28(5)*, 509-526. doi:10.1080/01446191003674519

Omar, T., & Nehdi, M. L. (2016). Data acquisition technologies for construction progress tracking. *Automation in construction, 70*, 143-155. doi:10.1016/j.autcon.2016.06.016

Park, H.-S. (2006). Conceptual framework of construction productivity estimation. *KSCE Journal of Civil Engineering, 10*, 311-317.

Park, J. &. (2017). Wbs-based dynamic multi-dimensional bim database for total construction as-built documentation. *Automation in Construction, 77*, 15-23. doi:https://doi.org/10.1016/j.autcon.2017.01.021

Park, J., & Cai, H. (2015). Automatic Construction Schedule Generation Method through BIM Model Creation. *Computing in Civil Engineering*, (pp. 620-627). doi:https://doi.org/10.1061/9780784479247.077

Project Management Institute. (2016). *Construction extension to the PMBOK guide.* Newtown Square, Pennsylvania: Project Management Institute.

Project Management Institute. (2017). *A guide to the project management body of knowledge (PMBOK Guide)* (7 ed.). Newtown Square, Pa: Project Management Institute.

Pyzdek, T., & Keller, P. A. (2019). *The six sigma handbook* (5 ed.). McGraw-Hill Education.

RazaviAlavi, S., & Abourizk, S. (2017). Site Layout and Construction Plan Optimization Using an Integrated Genetic Algorithm Simulation Framework. *Journal Comput. Civ. Eng. 31(4)*, 1-10. doi:10.1061/(ASCE)CP.1943-5487.0000653.

Rekveld, Y. (2017). *BIM based cost estimation knowledge management - The application of a Knowledge Management System by combining BIM models and the data mining process to estimate the costs of residential buildings (Master's thesis).*

Roedel, H. (2018). *Why is scheduling in construction broken?* Retrieved from Medium: https://medium.com/alice-technologies/why-is-scheduling-in-construction-broken-9839ff8ad084

Russell, M. M., Hsiang, S., Liu, M., & Wambeke, B. (n.d.). Causes of Time Buffer and Duration Variation in Construction Project Tasks: Comparison of Perception to Reality. *Journal of Construction Engineering and Management, 140(6)*, 04014016. doi:10.1061/(ASCE)CO.1943-7862.0000819

Sakr, S., & Pardede, E. (2012). *Graph data management : techniques and applications.* Information Science Reference.

Shahi, A., Cardona, J., Haas, C., & West, J. C. (2012). Activity-based data fusion for automated progress tracking of construction projects. *Construction Research Congress*, (pp. 838-847). doi:http://dx.doi.org/10.1061/9780784412329.085

Sigalov, K., & König, M. (2017). Recognition of process patterns for BIM-based construction schedules. *Advanced Engineering Informatics 33*, 456-472. doi:10.1016/j.aei.2016.12.003

Singh, Y. (2011). *Software testing.* Cambridge University Press.

Son, H., Kim, C., & Cho, Y. (2017). Automated Schedule Update using AS-Built Data and 4D Building Information Model. *ASCE Journal of Management in Engineering, Vol. 33, Issue 4*. doi:dx.doi.org/10.1061/(ASCE)ME.1943-5479.0000528

Song, L., & AbouRizk, S. (2008). Measuring and Modeling Labor Productivity Using Historical Data. *Journal of Construction Engineering and Management, 134(10)*, 786-794. doi:10.1061/(ASCE)0733-9364(2008)134:10(786)

Sørensen, K. B., Christiansson, P., & Svidt, K. (2010). Ontologies to Support RFID-Based Link between Virtual Models and Construction Components. *Computer-Aided Civil and Infrastructure Engineering, 25*, 285-302. doi:10.1111/j.1467-8667.2009.00638.x

Storm, K. (2019). *Industrial process plant construction estimating and man-hour analysis.* Gulf Professional Publishing.

Tamimi, S., & Diekmann, J. (1988). Soft Logic in Network Analysis. *Journal of Computing in Civil Engineering, 2(3)*, 289–300. doi:10.1061/(ASCE)0887-3801(1988)2:3(289)

Tamimi, S., & Diemann, J. (1988). Soft Logic in Network Analysis. *Journal of Computing in Civil Engineering, 2*(3), 289-300.

Tan, H. C., Anumba, C., Carrilo, P., Bouchlaghem, D., Kamara, J., & Udeaja, C. (2010). *Capture and Reuse of Project Knowledge in Construction.* Wiley-Blackwell. doi:10.1002/9781444315448

Tauscher, E., Mikulakova, E., König, M., & Beucke, K. (2007). Generating Construction Schedules with Case-Based Reasoning Support. *Computing in Civil Engineering*, 119-126. doi:10.1061/40937(261)15

Thomas, H. R., & Yiakoumis, I. (1987). Factor Model of Construction Productivity. *Journal of Construction Engineering and Management, 113*(4), 623-639.

Thomas, H. R., & Zavrski. (1999). Construction Baseline Productivity: Theory and Practice. *Journal of Construction Engineering and Management, 125*(5), 295-303.

Thomas, H., Maloney, W., Horner, M., Smith, G., Handa, V., & Sanders, S. (1990). Modeling Construction Labor Productivity. *Journal of Construction Engineering and Management, 116*(4). doi:https://doi.org/10.1061/(ASCE)0733-9364(1990)116:4(705)

Tolmer, C.-E., Castaing, C., Diab, Y., & Morand, D. (2017). Adapting lod definition to meet bim uses requirements and data modeling for linear infrastructures projects: Using system and requirement engineering. *Visualization in Engineering, 5(1)*, 1-18. doi:10.1186/s40327-017-0059-9

Toma, S.-V., Chiriţă, & Şarpe, D. (2012). Risk and Uncertainty. *Procedia Economics and Finance, 3*, 975-980. doi:https://doi.org/10.1016/S2212-5671(12)00260-

Tserng, H., Ho, S., & Jan, S. (2013). Developing BIM-assisted as-built schedule management system for general contractors. *Journal of Civil Engineering and Management*, 20(1): 47-58. doi:http://dx.doi.org/10.3846/13923730.2013.851112

Turkan, Y., Bosche, F., Haas, C., & Haas, R. (2012). Automated progress tracking using 4d schedule and 3d sensing technologies. *Automation in Construction, 22*, 41-421. doi:https://doi.org/10.1016/j.autcon.2011.10.003

Vaisman, A., & Zimányi, E. (2014). *Data Warehouse Systems - Design and Implementation.* Springer. doi:10.1007/978-3-642-54655-6

Van der Aalst, W., Schonenberg, M., & Song, M. (2011). Time prediction based on process mining. *Information Systems, 36(2)*, 450-475. doi:10.1016/j.is.2010.09.001

Van der Graaf, C. (2017). *The discovery of bottlenecks within the BIM design process - The concept of process mining in combination with BIM models in the design phase of construction projects (Master's thesis).*

Van Lith, P. (2020). *The application of data-driven decision making based on construction project process data (master's thesis).*

Van Schaijk, S. (2016). *BIM based process mining - Enabling knowledge reassurance and fact-based problem discovery within the AECFM industry (master's thesis).* Retrieved from https://www.ofcoursecme.nl/mdocs-posts/building-information-model-bim-based-process-mining/

Venkatesh, P., & Venkatesan, V. (2017). Delays in Construction Projects: A review of causes, need and scope for further research. *Malaysian Construction Research Journal, 23*(3), 89-109.

Vereen, S., Rasdorf, W., & Hummer, J. (2016). Development and Comparative Analysis of Construction Industry Labor Productivity Metrics. *Journal of Construction Engineering and Management, 142*(7), 04016020–04016020. doi:10.1061/(ASCE)CO.1943-7862.0001112

Wambeke, B. W., Hsiang, S. M., & Liu, M. (2011). Causes of Variation in Construction Project Task Starting Times and Duration. *Journal of Construction Engineering and Management, 137(9)*, 663-677. doi:10.1061/(ASCE)CO.1943-7862.0000342

Wang, H.-W., Lin, J.-R., & Zhang, J.-P. (2019). Work package-based information modeling for resource-constrained scheduling of construction projects. *Automation in Construction, 109, 102958*, 1-20. doi:10.1016/j.autcon.2019.102958

Wang, W.-C., Weng, s.-W., Wang, S.-H., & Chen, C.-Y. (2014). Integrating building information models with construction process simulations for project scheduling support. *Automation in Construction, 37*, 68-80. doi:10.1016/j.autcon.2013.10.009

Wang, Z., & Azar, E. (2019). BIM-based draft schedule generation in reinforced concrete-framed buildings. *Construction Innovation*, 1-16. doi: https://doi.org/10.1108/CI-11-2018-0094

Wiewiora, A., Trigunarsyah, A.-T., Murphy, G., & Liang, C. (2009). Barriers to effective knowledge transfer in project-based organisations. *International Conference on Global Innovation in Construction Proceedings*, (pp. 1-12). Holywell Park, Loughborough University.

Winch, G. M., & Kelsey, J. (2005). What do construction project planners do? *International Journal of Project Management, 23(2)*, 141-149. doi:10.1016/j.ijproman.2004.06.002

Wix, J., & Karlshøj, J. (2010). *Information Delivery Manual Guide to Components and Development Methods.* buildingSMART International. Retrieved from http://iug.buildingsmart.org/idms/development/IDMC_004_1_2.pdf

Yang, J., Park, M.-Y., Vela, P., & Golparvar-Fard, M. (2015). Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future. *Advanced Engineering Informatics 29*, 211-224. doi:dx.doi.org/10.1016/j.aei.2015.01.011 1474-0346/

Yang, J.-B., & Teng, Y.-L. (2017). Theoretical development of stochastic delay analysis and forecast method. *Journal of the Chinese Institute of Engineers*, VOL. 40, NO. 5, 391–400. doi:10.1080/02533839.2017.1336115

Youssef, W., El-Dash, K., & Ramadan, O. (2013). Construction Activities Duration Patterns in the Middle East. *International Journal of Construction Engineering and Management, 2*(4), 122-135. doi:10.5923/j.ijcem.20130204.04

Zhang, L., & Zou, X. (2015). *Repetitive project scheduling : theory and methods.* Elsevier.

Zhou, J., Love, P., Wang, X., Teo, K., & Irani, Z. (2013). A review of methods and algorithms for optimizing construction scheduling. *The Journal of the Operational Research Society, 64(8*, 1091-1105. doi:10.1057/jors.2012.174

Zoelen, B. v. (2019, Juni 19). *200 miljoen extra, maar dan heb je ook wat: zo verloopt de bouw van zeesluis IJmuiden*. Retrieved from Het Parool: https://www.parool.nl/amsterdam/200-miljoen-extra-maar-dan-heb-je-ook-wat-zo-verloopt-de-bouw-van-zeesluis-ijmuiden~bee5d671/

Zomaya, A. Y., & Sakr, S. (2017). *Handbook of big data technologies.* Springer. doi: https://doi.org/10.1007/978-3-319-49340-4

Appendices

Appendix 1 : Most important factors causing delays according to different sources
Appendix 2 : Most important factors affecting productivity
Appendix 3 : ERD of MySQL tables
Appendix 4 : SQL code for the creation of the tables and views
Appendix 5 : Python code to extract, transform and load the data from Excel into MySQL
Appendix 6 : Workflow to create Synchro 4D as-planned and as-built models
Appendix 7 : Workflow to connect MySQL to PowerBI
Appendix 8 : Examples of sections of data in MySQL tables

Appendix 1 - Most important factors causing delays according to different sources

Table 9: Top ten causes of construction delays in developing and developed countries (Venkatesh & Venkatesan, 2017)

Top 10 causes of delays in developing countries	Top 10 causes of delays in developed countries
- Delay in payments by clients - Delay in drawings, changes & errors in designs - Contractor's financial difficulties - Deficiencies in planning & scheduling - Delay in delivery of materials - Change orders / increase in scope of work - Poor site supervision and management - Economy, law & order, inflation, political instability - Slow decision making by owner - Subcontractor & supplier related causes	- Weather / Ground conditions - Delay in drawings, changes & errors in designs - Subcontractor & supplier related causes - Change orders / increase in scope of work - Slow decision making / approvals from client - Delay in approvals / permits from other authorities - Changes in site conditions / poor site conditions - Contractor's financial difficulties - Delay in monthly payments from client - Force majeure / Acts of Gods

Table 10: Top ten causes of delays to public infrastructure projects (Ahmad, Ayoush, & Al-Alwan, 2019)

Rank	Causes of delays
1	Delay in progressing payment by owner
2	Slow decision making by owner
3	Mismatch between the location of infrastructure services and what was approved in the tender and provided in drawings by the owner of the services
4	Changes in order specification by owner
5	Problems in financing project by contractor
6	Delays in performing inspections and tests by consultant
7	Delay in approving changes in the scope of work by consultant
8	Failure or delay in delivering project site clearance by owner
9	Lack of cooperation by contractor with the owner vis-à-vis barriers and utilities
10	Land acquisitions: citizens refuse compensation for their land or demolition of their homes

Table 11: Top 15 factors causing delays (Gündüz, Nielsen, & Özdemir, 2013)

Rank	Causes of delays	Factor group
1	Inadequate contractor experience	Contractor related
2	Ineffective project planning and scheduling	Contractor related
3	Poor site management and supervision	Contractor related
4	Design changes by owner or agent during construction	Design related
5	Late delivery of materials	Material related
6	Unreliable subcontractors	Contractor related
7	Delay in performing inspection and testing	Consultant related
8	Unqualified/inexperienced workers	Labour related
9	Change orders	Owner related
10	Delay in site delivery	Owner related

11	Delay in approving design documents	Owner related
12	Delay in progress payments	Owner related
13	Slowness in decision making	Owner related
14	Poor communication and coordination with other parties	Consultant related
15	Unexpected surface and subsurface conditions (soil, hw t.)	External related

Table 12: Causes of task starting time and duration variation (Wambeke, Hsiang, & Liu, 2011)

Category	Code	Individual cause of variation
1. Prerequisite work	1.1	Obtaining required permits for the work to begin
	1.2	Completion of previous work (i.e., work to be done before yours is not done yet)
	1.3	Rework being required owing to the quality of previous work
	1.4	Poor quality of previous work (though not to a level that requires rework)
	1.5	Inspections for previously completed work
2. Detailed design and work method	2.1	Design constructability
	2.2	Quality of documents (errors in design and/or drawings)
	2.3	Turnaround time from engineers when there is a question with a drawing
	2.4	Strict specification requirements
	2.5	Quality control requirements
	2.6	Work complexity
	2.7	Work sequence or method is not well planned
	2.8	Low degree of repetition (inability to develop efficient system owing to task constantly changing)
	2.9	Inadequate instruction on detailed working method
3. Labour force	3.1	Socializing (talking with fellow workers)
	3.2	Absenteeism
	3.3	People arriving late and/or leaving early because of illness, injury, family or personal reason
	3.4	Low morale and/or lack of motivation
	3.5	Getting moved to another job/task before the one you were working on was completed
	3.6	Crew size is inadequate
	3.7	Inefficiencies associated with personnel turnover (i.e., new employees)
	3.8	Experience on similar tasks (i.e., there is a learning curve associated with nonrepetitive tasks)
	3.9	Worker/crew lack of skills/experience to perform the task(s) being asked of them
	3.10	Language barrier among workers and/or worker-supervisor
4. Tools and equipment	4.1	Personnel lift (unavailable, no operator, not the priority, maintenance)
	4.2	Power tools (not trained, used by someone else, misplaced, maintenance)
	4.3	Crane or forklift (unavailable, no operator, not the priority, maintenance)
	4.4	Hand tools (used by someone else, misplaced, maintenance)
	4.5	Other heavy equipment (e.g., backhoe, loader, dump truck) not available

	4.6	Personal protective equipment (PPE) (not enough, used by someone else, misplaced, unserviceable)
5. Materials and components	5.1	Material needs to be moved to where you need it
	5.2	Material to arrive from distributor or supplier
	5.3	Trying to get consumables
	5.4	Error in material size
	5.5	Error in material type
6. Work/job site	6.1	Overcrowded work area/job site congestion
	6.2	Difficult access to work area
	6.3	Site layout—distance between material storage and where material is required for work is excessive
7. Management/supervision/information flow	7.1	Wait to get answers to questions you have about the design or drawing
	7.2	Need guidance or instruction from supervisor
	7.3	Lack of field manager (foreman) skill/knowledge
	7.4	Coordination between different trades
	7.5	Overcommitment because of a tight work schedule
	7.6	Foreman availability
	7.7	Change in scope of work
	7.8	Foreman communication skills
	7.9	Communication between: owner/engineer and project manager
	7.10	Communication between: project manager and foreman
	7.11	Communication between: foreman and workers
8. Weather/external conditions	8.1	Weather impacts (excessive heat, cold, wind, rain)

Appendix 2 - Most important factors affecting productivity

Table 13: Project environment factors and management efforts (from Park, 2006)

Project environment factors	Management efforts
Weather	pre-project planning
labour skill	change management
labour availability	constructability
materials availability	materials management
site conditions	zero accident techniques (safety)
project complexity	quality management
regulatory requirements	team building
project team experience	automation/integration technology
project team turnover	
detailed engineering design location	
business market conditions	
absenteeism	
technology use	
human factor	

Table 14: Most frequently cited factors affecting productivity (Hasan et al., 2018)

Rank most frequently cited	Factors affecting construction productivity
1	Non-availability of materials
2	Inadequate supervision
3	Skill shortage
4	Lack of proper tools/equipment
5	Incomplete drawing and specifications
6	Poor communication
7	Rework
8	Poor site layout
9	Adverse weather conditions
10	Change orders

Table 15: Variables for productivity improvement (Arditi, 1985)

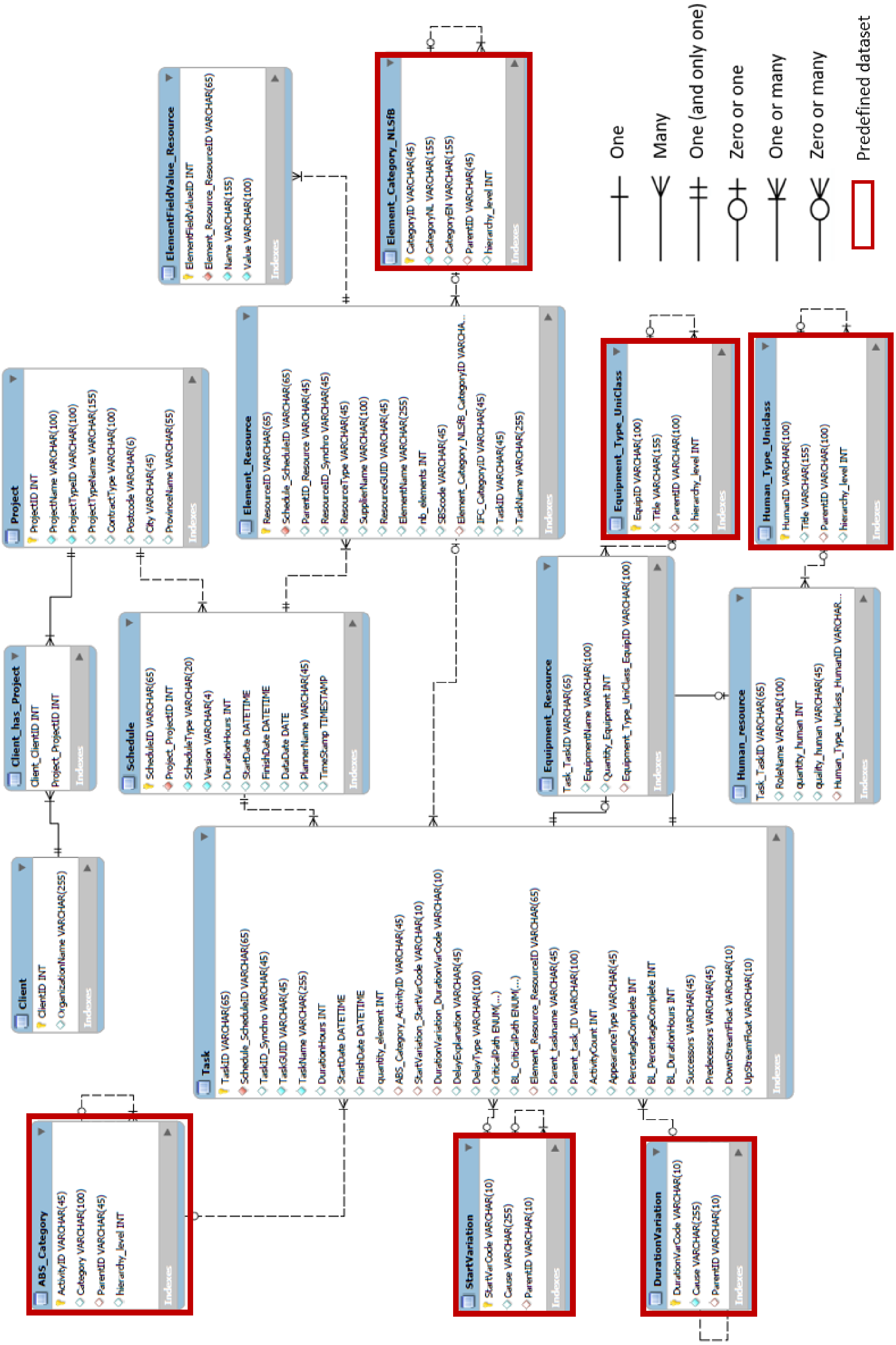
Category	Variables
Management	Labour relations, cost control, supervision
Materials	Delivery, storage, packaging, prefabrication, standardization, product availability, new products
Engineering	Design standards, design improvements, systems engineering, standard specifications
Construction Techniques	Precast elements, preassemble, foreign developments

Regulations	EPA, OSHA, EEO, Local codes
Labour	Contract agreement, training, quality control, turnover, availability
Equipment	Capacity, Simplicity, Maintainability, Utilization

Table 16: Factors Affecting Labor Productivity (Heravi & Eslamdoost, 2015)

Positive factors	Negative factors
Supervision	Poor decision making
Proper coordination	Schedule compression
Effective communication	Frequent change order
Proper planning	Materials, tools and equipment deficiency
Proper HSE program	Unfavourable external condition
Technical excellence	
Suitable site layout	
Labour competence	
Sufficient facilities and accommodation	
Motivation of labour	

Appendix 3 - ERD of MySQL tables



Appendix 4 - SQL code for the creation of the tables and views

Creation of MySQL tables and relationships

```
-- MySQL Workbench Forward Engineering

SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE,
SQL_MODE='ONLY_FULL_GROUP_BY,STRICT_TRANS_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_ZERO,NO
_ENGINE_SUBSTITUTION';

-- -----------------------------------------------------
-- Schema BIM_schedules
-- -----------------------------------------------------

-- -----------------------------------------------------
-- Schema BIM_schedules
-- -----------------------------------------------------
CREATE SCHEMA IF NOT EXISTS `BIM_schedules` DEFAULT CHARACTER SET utf8 ;
USE `BIM_schedules` ;

-- -----------------------------------------------------
-- Table `BIM_schedules`.`DurationVariation`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`DurationVariation` (
  `DurationVarCode` VARCHAR(10) NOT NULL,
  `Cause` VARCHAR(255) NOT NULL,
  `ParentID` VARCHAR(10) NULL,
  PRIMARY KEY (`DurationVarCode`),
  UNIQUE INDEX `DelayCode_UNIQUE` (`DurationVarCode` ASC) VISIBLE,
  INDEX `fk_DurationVariation_DurationVariation1_idx` (`ParentID` ASC) VISIBLE,
  CONSTRAINT `fk_DurationVariation_DurationVariation1`
    FOREIGN KEY (`ParentID`)
    REFERENCES `BIM_schedules`.`DurationVariation` (`DurationVarCode`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Client`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Client` (
  `ClientID` INT NOT NULL AUTO_INCREMENT,
  `OrganizationName` VARCHAR(255) NULL,
  PRIMARY KEY (`ClientID`),
  UNIQUE INDEX `ClientID_UNIQUE` (`ClientID` ASC) VISIBLE)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Project`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Project` (
  `ProjectID` INT NOT NULL AUTO_INCREMENT,
  `ProjectName` VARCHAR(100) NOT NULL,
  `ProjectTypeID` VARCHAR(100) NOT NULL,
  `ProjectTypeName` VARCHAR(155) NULL,
  `ContractType` VARCHAR(100) NULL,
  `Postcode` VARCHAR(6) NULL,
  `City` VARCHAR(45) NULL,
  `ProvinceName` VARCHAR(55) NULL,
  PRIMARY KEY (`ProjectID`))
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Schedule`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Schedule` (
  `ScheduleID` VARCHAR(65) NOT NULL,
  `Project_ProjectID` INT NOT NULL,
  `ScheduleType` VARCHAR(20) NOT NULL,
  `Version` VARCHAR(4) NOT NULL,
  `DurationHours` INT NULL,
  `StartDate` DATETIME NULL,
  `FinishDate` DATETIME NULL,
  `DataDate` DATE NULL,
  `PlannerName` VARCHAR(45) NULL,
  `TimeStamp` TIMESTAMP NULL DEFAULT CURRENT_TIMESTAMP,
  PRIMARY KEY (`ScheduleID`),
  INDEX `fk_Schedule_Project1_idx` (`Project_ProjectID` ASC) VISIBLE,
  CONSTRAINT `fk_Schedule_Project1`
    FOREIGN KEY (`Project_ProjectID`)
```

```
    REFERENCES `BIM_schedules`.`Project` (`ProjectID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`ABS_Category`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`ABS_Category` (
  `ActivityID` VARCHAR(45) NOT NULL,
  `Category` VARCHAR(100) NULL,
  `ParentID` VARCHAR(45) NULL,
  `hierarchy_level` INT NULL,
  PRIMARY KEY (`ActivityID`),
  INDEX `fk_ABS_Category_ABS_Category1_idx` (`ParentID` ASC) VISIBLE,
  CONSTRAINT `fk_ABS_Category_ABS_Category1`
    FOREIGN KEY (`ParentID`)
    REFERENCES `BIM_schedules`.`ABS_Category` (`ActivityID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`StartVariation`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`StartVariation` (
  `StartVarCode` VARCHAR(10) NOT NULL,
  `Cause` VARCHAR(255) NULL,
  `ParentID` VARCHAR(10) NULL,
  PRIMARY KEY (`StartVarCode`),
  INDEX `fk_StartVariation_StartVariation1_idx` (`ParentID` ASC) VISIBLE,
  CONSTRAINT `fk_StartVariation_StartVariation1`
    FOREIGN KEY (`ParentID`)
    REFERENCES `BIM_schedules`.`StartVariation` (`StartVarCode`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Element_Category_NLSfB`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Element_Category_NLSfB` (
  `CategoryID` VARCHAR(45) NOT NULL,
  `CategoryNL` VARCHAR(155) NOT NULL,
  `CategoryEN` VARCHAR(155) NULL,
  `ParentID` VARCHAR(45) NULL,
  `hierarchy_level` INT NULL,
  PRIMARY KEY (`CategoryID`),
  INDEX `fk_Element_Category_NLSfB_Element_Category_NLSfB1_idx` (`ParentID` ASC) VISIBLE,
  CONSTRAINT `fk_Element_Category_NLSfB_Element_Category_NLSfB1`
    FOREIGN KEY (`ParentID`)
    REFERENCES `BIM_schedules`.`Element_Category_NLSfB` (`CategoryID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Element_Resource`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Element_Resource` (
  `ResourceID` VARCHAR(65) NOT NULL,
  `Schedule_ScheduleID` VARCHAR(65) NOT NULL,
  `ParentID_Resource` VARCHAR(45) NULL,
  `ResourceID_Synchro` VARCHAR(45) NULL,
  `ResourceType` VARCHAR(45) NULL,
  `SupplierName` VARCHAR(100) NULL,
  `ResourceGUID` VARCHAR(45) NULL,
  `ElementName` VARCHAR(255) NULL,
  `nb_elements` INT NULL,
  `SBScode` VARCHAR(45) NULL,
  `Element_Category_NLSfB_CategoryID` VARCHAR(45) NULL,
  `IFC_CategoryID` VARCHAR(45) NULL,
  `TaskID` VARCHAR(45) NULL,
  `TaskName` VARCHAR(255) NULL,
  PRIMARY KEY (`ResourceID`),
  INDEX `fk_Element_Resource_Element_Category_NLSfB1_idx` (`Element_Category_NLSfB_CategoryID` ASC)
VISIBLE,
  INDEX `fk_Element_Resource_Schedule1_idx` (`Schedule_ScheduleID` ASC) VISIBLE,
  CONSTRAINT `fk_Element_Resource_Element_Category_NLSfB1`
    FOREIGN KEY (`Element_Category_NLSfB_CategoryID`)
    REFERENCES `BIM_schedules`.`Element_Category_NLSfB` (`CategoryID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `fk_Element_Resource_Schedule1`
```

```
    FOREIGN KEY (`Schedule_ScheduleID`)
    REFERENCES `BIM_schedules`.`Schedule` (`ScheduleID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Task`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Task` (
  `TaskID` VARCHAR(65) NOT NULL,
  `Schedule_ScheduleID` VARCHAR(65) NOT NULL,
  `TaskID_Synchro` VARCHAR(45) NULL,
  `TaskGUID` VARCHAR(45) NOT NULL,
  `TaskName` VARCHAR(255) NOT NULL,
  `DurationHours` INT NULL,
  `StartDate` DATETIME NULL,
  `FinishDate` DATETIME NULL,
  `quantity_element` INT NULL,
  `ABS_Category_ActivityID` VARCHAR(45) NULL,
  `StartVariation_StartVarCode` VARCHAR(10) NULL,
  `DurationVariation_DurationVarCode` VARCHAR(10) NULL,
  `DelayExplanation` VARCHAR(45) NULL,
  `DelayType` VARCHAR(100) NULL,
  `CriticalPath` ENUM('TRUE', 'FALSE') NULL,
  `BL_CriticalPath` ENUM('TRUE', 'FALSE') NULL,
  `Element_Resource_ResourceID` VARCHAR(65) NULL,
  `Parent_taskname` VARCHAR(45) NULL,
  `Parent_task_ID` VARCHAR(100) NULL,
  `ActivityCount` INT NULL,
  `AppearanceType` VARCHAR(45) NULL,
  `PercentageComplete` INT NULL,
  `BL_PercentageComplete` INT NULL,
  `BL_DurationHours` INT NULL,
  `Successors` VARCHAR(45) NULL,
  `Predecessors` VARCHAR(45) NULL,
  `DownStreamFloat` VARCHAR(10) NULL,
  `UpStreamFloat` VARCHAR(10) NULL,
  INDEX `fk_Task_ABS_Category1_idx` (`ABS_Category_ActivityID` ASC) VISIBLE,
  INDEX `fk_Task_DurationVariation1_idx` (`DurationVariation_DurationVarCode` ASC) VISIBLE,
  INDEX `fk_Task_StartVariation1_idx` (`StartVariation_StartVarCode` ASC) VISIBLE,
  PRIMARY KEY (`TaskID`),
  INDEX `fk_Task_Schedule1_idx` (`Schedule_ScheduleID` ASC) VISIBLE,
  INDEX `fk_Task_Element_Resource1_idx` (`Element_Resource_ResourceID` ASC) VISIBLE,
  CONSTRAINT `fk_Task_ABS_Category1`
    FOREIGN KEY (`ABS_Category_ActivityID`)
    REFERENCES `BIM_schedules`.`ABS_Category` (`ActivityID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `fk_Task_DurationVariation1`
    FOREIGN KEY (`DurationVariation_DurationVarCode`)
    REFERENCES `BIM_schedules`.`DurationVariation` (`DurationVarCode`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `fk_Task_StartVariation1`
    FOREIGN KEY (`StartVariation_StartVarCode`)
    REFERENCES `BIM_schedules`.`StartVariation` (`StartVarCode`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `fk_Task_Schedule1`
    FOREIGN KEY (`Schedule_ScheduleID`)
    REFERENCES `BIM_schedules`.`Schedule` (`ScheduleID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `fk_Task_Element_Resource1`
    FOREIGN KEY (`Element_Resource_ResourceID`)
    REFERENCES `BIM_schedules`.`Element_Resource` (`ResourceID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`ElementFieldValue_Resource`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`ElementFieldValue_Resource` (
  `ElementFieldValueID` INT NOT NULL AUTO_INCREMENT,
  `Element_Resource_ResourceID` VARCHAR(65) NOT NULL,
  `Name` VARCHAR(155) NOT NULL,
  `Value` VARCHAR(100) NOT NULL,
  PRIMARY KEY (`ElementFieldValueID`),
  INDEX `fk_ElementFieldValue_Resource_Element_Resource1_idx` (`Element_Resource_ResourceID` ASC) VISIBLE,
  CONSTRAINT `fk_ElementFieldValue_Resource_Element_Resource1`
    FOREIGN KEY (`Element_Resource_ResourceID`)
    REFERENCES `BIM_schedules`.`Element_Resource` (`ResourceID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
```

```sql
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Client_has_Project`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Client_has_Project` (
  `Client_ClientID` INT NOT NULL,
  `Project_ProjectID` INT NOT NULL,
  PRIMARY KEY (`Client_ClientID`, `Project_ProjectID`),
  INDEX `fk_Client_has_Project_Project1_idx` (`Project_ProjectID` ASC) VISIBLE,
  INDEX `fk_Client_has_Project_Client1_idx` (`Client_ClientID` ASC) VISIBLE,
  CONSTRAINT `fk_Client_has_Project_Client1`
    FOREIGN KEY (`Client_ClientID`)
    REFERENCES `BIM_schedules`.`Client` (`ClientID`)
    ON DELETE CASCADE
    ON UPDATE CASCADE,
  CONSTRAINT `fk_Client_has_Project_Project1`
    FOREIGN KEY (`Project_ProjectID`)
    REFERENCES `BIM_schedules`.`Project` (`ProjectID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Equipment_Type_UniClass`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Equipment_Type_UniClass` (
  `EquipID` VARCHAR(100) NOT NULL,
  `Title` VARCHAR(155) NULL,
  `ParentID` VARCHAR(100) NULL,
  `hierarchy_level` INT NULL,
  PRIMARY KEY (`EquipID`),
  INDEX `fk_Equipment_Type_UniClass_Equipment_Type_UniClass1_idx` (`ParentID` ASC) VISIBLE,
  CONSTRAINT `fk_Equipment_Type_UniClass_Equipment_Type_UniClass1`
    FOREIGN KEY (`ParentID`)
    REFERENCES `BIM_schedules`.`Equipment_Type_UniClass` (`EquipID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Equipment_Resource`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Equipment_Resource` (
  `Task_TaskID` VARCHAR(65) NOT NULL,
  `EquipmentName` VARCHAR(100) NULL,
  `Quantity_Equipment` INT NULL,
  `Equipment_Type_UniClass_EquipID` VARCHAR(100) NULL,
  INDEX `fk_Equipment_Resource_Equipment_Type_UniClass1_idx` (`Equipment_Type_UniClass_EquipID` ASC)
VISIBLE,
  PRIMARY KEY (`Task_TaskID`),
  CONSTRAINT `fk_Equipment_Resource_Equipment_Type_UniClass1`
    FOREIGN KEY (`Equipment_Type_UniClass_EquipID`)
    REFERENCES `BIM_schedules`.`Equipment_Type_UniClass` (`EquipID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `fk_Equipment_Resource_Task1`
    FOREIGN KEY (`Task_TaskID`)
    REFERENCES `BIM_schedules`.`Task` (`TaskID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Human_Type_Uniclass`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Human_Type_Uniclass` (
  `HumanID` VARCHAR(100) NOT NULL,
  `Title` VARCHAR(155) NULL,
  `ParentID` VARCHAR(100) NULL,
  `hierarchy_level` INT NULL,
  PRIMARY KEY (`HumanID`),
  INDEX `fk_Human_Type_Uniclass_Human_Type_Uniclass1_idx` (`ParentID` ASC) VISIBLE,
  CONSTRAINT `fk_Human_Type_Uniclass_Human_Type_Uniclass1`
    FOREIGN KEY (`ParentID`)
    REFERENCES `BIM_schedules`.`Human_Type_Uniclass` (`HumanID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `BIM_schedules`.`Human_resource`
-- -----------------------------------------------------
```

```
CREATE TABLE IF NOT EXISTS `BIM_schedules`.`Human_resource` (
  `Task_TaskID` VARCHAR(65) NOT NULL,
  `RoleName` VARCHAR(100) NULL,
  `quantity_human` INT NULL,
  `quality_human` VARCHAR(45) NULL,
  `Human_Type_Uniclass_HumanID` VARCHAR(100) NULL,
  INDEX `fk_Human_resource_Human_Type_Uniclass1_idx` (`Human_Type_Uniclass_HumanID` ASC) VISIBLE,
  PRIMARY KEY (`Task_TaskID`),
  CONSTRAINT `fk_Human_resource_Human_Type_Uniclass1`
    FOREIGN KEY (`Human_Type_Uniclass_HumanID`)
    REFERENCES `BIM_schedules`.`Human_Type_Uniclass` (`HumanID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `fk_Human_resource_Task1`
    FOREIGN KEY (`Task_TaskID`)
    REFERENCES `BIM_schedules`.`Task` (`TaskID`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;


SET SQL_MODE=@OLD_SQL_MODE;
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;
```

Creates MySQL views for data analysis in PowerBI

```
DROP VIEW IF EXISTS DurVarPath, StartVarPath, HumanTypePath, EquipTypePath, ElementTypePath,
ElementNLSfBPath, ABSTypePath;

-- Creates views for the purpose of dimension modelling as part of data analysis
CREATE VIEW  DurVarPath AS
    WITH RECURSIVE DurVarPath AS
    (
      SELECT DurationVarCode, Cause, ParentID, 1 AS Depth, Cause AS Path
        FROM durationvariation
        WHERE ParentID = 'NULL'
      UNION ALL
      SELECT dvar.DurationVarCode, dvar.Cause, dvar.ParentID, DvP.depth + 1, CONCAT(DvP.Path, ' > ',
dvar.Cause)
        FROM DurVarPath AS DvP
          JOIN durationvariation AS dvar ON DvP.DurationVarCode = dvar.ParentID
    )
    SELECT * FROM DurVarPath;

CREATE VIEW  StartVarPath AS
    WITH RECURSIVE StartVarPath AS
    (
      SELECT StartVarCode, Cause, ParentID, 1 AS Depth, Cause AS Path
        FROM startvariation
        WHERE ParentID = 'NULL'
      UNION ALL
      SELECT svar.StartVarCode, svar.Cause, svar.ParentID, svP.depth + 1, CONCAT(svP.Path, ' > ',
svar.Cause)
        FROM StartVarPath AS SvP
          JOIN startvariation AS Svar ON svP.StartVarCode = svar.ParentID
    )
    SELECT * FROM StartVarPath;


CREATE VIEW HumanTypePath AS
WITH RECURSIVE HumanTypePath AS
(
  SELECT HumanID, Title, ParentID, 1 AS Depth, Title AS Path
    FROM human_type_uniclass
    WHERE ParentID = 'NULL'
  UNION ALL
  SELECT Eq.humanID, Eq.Title, Eq.ParentID, EqP.depth + 1, CONCAT(EqP.Path, ' > ', Eq.Title)
    FROM humanTypePath AS EqP
      JOIN human_type_uniclass AS Eq ON EqP.humanID = Eq.ParentID
)
SELECT * FROM humanTypePath;

CREATE VIEW EquipTypePath AS
    WITH RECURSIVE EquipTypePath AS
    (
      SELECT EquipID, Title, ParentID, 1 AS Depth, Title AS Path
        FROM equipment_type_uniclass
        WHERE ParentID = 'NULL'
      UNION ALL
      SELECT Eq.EquipID, Eq.Title, Eq.ParentID, EqP.depth + 1, CONCAT(EqP.Path, ' > ', Eq.Title)
        FROM EquipTypePath AS EqP
          JOIN equipment_type_uniclass AS Eq ON EqP.EquipID = Eq.ParentID
    )
    SELECT * FROM EquipTypePath;

CREATE VIEW  ElementNLSfBPath AS
    WITH RECURSIVE ElementNLSfBPath AS
    (
      SELECT CategoryID, CategoryNL, CategoryEN, ParentID, 1 AS Depth, CategoryNL AS Path
        FROM element_category_nlsfb
        WHERE ParentID = 'NULL'
      UNION ALL
      SELECT ESfB.CategoryID, ESfB.CategoryNL, ESfB.CategoryEN, ESfB.ParentID, ESfBP.depth + 1,
CONCAT(ESfBP.Path, ' > ', ESfB.CategoryNL)
        FROM ElementNLSfBPath AS ESfBP
          JOIN element_category_nlsfb AS ESfB ON ESfBP.CategoryID = ESfB.ParentID
    )
    SELECT * FROM ElementNLSfBPath;

CREATE VIEW  ABStypePath AS
    WITH RECURSIVE ABStypePath AS
    (
      SELECT ActivityID, Category, ParentID, 1 AS Depth, Category AS Path
        FROM abs_category
        WHERE ParentID = 1
      UNION ALL
      SELECT abs.ActivityID, abs.Category, abs.ParentID, absP.depth + 1, CONCAT(absP.Path, ' > ',
abs.Category)
        FROM ABStypePath AS absP
          JOIN abs_category AS abs ON absP.ActivityID = abs.ParentID
```

```
    )
    SELECT * FROM ABStypePath;


-- Creates a view where the entity-attribute-value structure is transposed/transformed into a table
suitable for analysis
DROP VIEW IF EXISTS elementfieldvalue_resource_col;
SET SESSION group_concat_max_len = 10000000000;
SELECT
  GROUP_CONCAT(DISTINCT
    CONCAT(
      'ifnull(MAX(case when name = ''',
      Name,
      ''' then value end),0) AS `',
      Name, '`'
    )
  ) INTO @sql
FROM
  elementfieldvalue_resource;
SET @sql = CONCAT('CREATE VIEW elementfieldvalue_resource_col AS
                SELECT ElementFieldValueID, Element_Resource_ResourceID, ', @sql, '
                FROM elementfieldvalue_resource
                GROUP BY Element_Resource_ResourceID');

PREPARE stmt FROM @sql;
EXECUTE stmt;
DEALLOCATE PREPARE stmt;
```

Appendix 5 - Python code to extract, transform and load the data from Excel into MySQL

Python code to insert predefined classification tables from Excel into MySQL database

```
import openpyxl as xl
import mysql.connector

# Establish a MySQL connection
#db = mysql.connector.connect(user='scott', password='password', host='127.0.0.1', database='schedules')
db = mysql.connector.connect(user = "root",
                                db = 'bim_schedules')  #<--- name of database

#establish connection to workbook
wb = xl.load_workbook('ABStable.xlsx', data_only=True) #<-insert workbook name
ws = wb ['Sheet1']


# Get the cursor, which is used to traverse the database, line by line
cursor = db.cursor()

cursor.execute('SET FOREIGN_KEY_CHECKS=0;')

#maps data from the Excel
data = map(lambda x: {'ActivityID': x[0].value,
                      'Category': x[1].value,
                      'ParentID': x[2].value,
                      'hierarchy_level': x[3].value},
                      ws[5: ws.max_row])

for row in data:
    # executes raw MySQL syntax
    cursor.execute('INSERT INTO ABS_Category '
                   'VALUES ("{ActivityID}", "{Category}", "{ParentID}", "{hierarchy_level}" );'
                   .format(**row))  # construct MySQL syntax through format function

####-------------------------------------------------------

wb = xl.load_workbook('NL-SfB_Table.xlsx', data_only=True)
ws = wb ['Sheet1']

data = map(lambda x: {'CategoryID': x[0].value,
                      'CategoryNL': x[1].value,
                      'CategoryEN': x[2].value,
                      'ParentID': x[3].value,
                      'hierarchy_level': x[4].value},
                      ws[2: ws.max_row])

for row in data:
    cursor.execute('INSERT INTO Element_Category_NL-SFB '
                   'VALUES ("{CategoryID}", "{CategoryNL}", "{CategoryEN}", "{ParentID}",
"{hierarchy_level}" );'
                   .format(**row))

####-------------------------------------------------------

wb = xl.load_workbook('Uniclass_Roles.xlsx', data_only=True)
ws = wb ['Sheet1']

data = map(lambda x: {'HumanID': x[0].value,
                      'Title': x[1].value,
                      'ParentID': x[2].value,
                      'hierarchy_level': x[3].value},
                      ws[2: ws.max_row])

for row in data:
    cursor.execute('INSERT INTO Human_Type_Uniclass '
                   'VALUES ("{HumanID}", "{Title}", "{ParentID}", "{hierarchy_level}" );'
                   .format(**row))

####-------------------------------------------------------

wb = xl.load_workbook('Uniclass_TE_Tools.xlsx', data_only=True)
ws = wb ['Sheet1']

data = map(lambda x: {'EquipID': x[0].value,
                      'Title': x[1].value,
                      'ParentID': x[2].value,
                      'hierarchy_level': x[3].value},
                      ws[2: ws.max_row])

for row in data:
```

```
    cursor.execute('INSERT INTO Equipment_Type_Uniclass '
                   'VALUES ("{EquipID}", "{Title}", "{ParentID}", "{hierarchy_level}" );'
                   .format(**row))

####--------------------------------------------------------

wb = xl.load_workbook('StartDurationVariation.xlsx', data_only=True)
ws = wb ['Sheet1']

data = map(lambda x: {'StartVarCode': x[0].value,
                      'Cause': x[1].value,
                      'ParentID': x[2].value},
                      ws[2: ws.max_row])

for row in data:
    cursor.execute('INSERT INTO StartVariation '
                   'VALUES ("{StartVarCode}", "{Cause}", "{ParentID}" );'
                   .format(**row))

####--------------------------------------------------------

wb = xl.load_workbook('StartDurationVariation.xlsx', data_only=True)
ws = wb ['Sheet1']

data = map(lambda x: {'DurationVarCode': x[0].value,
                      'Cause': x[1].value,
                      'ParentID': x[2].value},
                      ws[2: ws.max_row])

for row in data:
    cursor.execute('INSERT INTO DurationVariation '
                   'VALUES ("{DurationVarCode}", "{Cause}", "{ParentID}" );'
                   .format(**row))

cursor.execute('SET FOREIGN_KEY_CHECKS=1;')

cursor.close()
db.commit()
db.close()
```

Python code to insert single export of Synchro 4D schedule as Excel into MySQL database

```
import openpyxl as xl
import mysql.connector
import re
from datetime import datetime
import uuid

#get and insert path of workbook
filename = ('C:\\Users\\path\\BAM-viaduct_As-Planned.xlsx') #<--- Excel workbook

wb = xl.load_workbook(filename, read_only=True, data_only=True)

# Establish a MySQL connection
#db = mysql.connector.connect(user='scott', password='password', host='127.0.0.1', database='schedules')
db = mysql.connector.connect(user = "root",
                                db = 'bim_schedules')  #<--- name of database

# Get the cursor, which is used to traverse the database, line by line
cursor = db.cursor()

# Creates Unique QUID for every new inserted schedule
ScheduleID = str(uuid.uuid4())

####--------------------------------------------------------

ws = wb ['Codes']

add_client = ("INSERT IGNORE INTO Client "
              "(ClientID, OrganizationName) "
              "VALUES (%s, %s)")

for row in ws.iter_rows(min_row=2):
    for cell in row:
        if cell.value == "ClientID":
            ClientID = ws.cell(row=cell.row, column=6).value
        if cell.value == "OrganizationName":
            OrganizationName = ws.cell(row=cell.row, column=6).value

            data_client = (ClientID, OrganizationName)
            cursor.execute(add_client, data_client)

####--------------------------------------------------------

add_project = ("INSERT IGNORE INTO Project "
              "(ProjectID, ProjectName, ProjectTypeID, ProjectTypeName, ContractType, Postcode, City,
ProvinceName) "
              "VALUES (%s, %s, %s, %s, %s, %s, %s, %s)")

for row in ws.iter_rows(min_row=2):
    for cell in row:
        if cell.value == "ProjectID":
            ProjectID = ws.cell(row=cell.row, column=6).value
        if cell.value == "ProjectName":
            ProjectName= ws.cell(row=cell.row, column=6).value
        if cell.value == "ProjectTypeID":
            ProjectTypeID = ws.cell(row=cell.row, column=6).value
        if cell.value == "ProjectTypeName":
            ProjectTypeName = ws.cell(row=cell.row, column=6).value
        if cell.value == "ContractType":
            ContractType = ws.cell(row=cell.row, column=6).value
        if cell.value == "Postcode":
            Postcode = ws.cell(row=cell.row, column=6).value
        if cell.value == "City":
            City = ws.cell(row=cell.row, column=6).value
        if cell.value == "ProvinceName":
            ProvinceName = ws.cell(row=cell.row, column=6).value

            data_project = (ProjectID, ProjectName, ProjectTypeID, ProjectTypeName, ContractType, Postcode,
City, ProvinceName)
            cursor.execute(add_project, data_project)

####--------------------------------------------------------

add_client_has_project = ("INSERT IGNORE INTO Client_has_project "
                          "(Client_ClientID, Project_ProjectID) "
                          "VALUES (%s, %s)")

data_client_has_project = (ClientID, ProjectID)
cursor.execute(add_client_has_project, data_client_has_project)

####--------------------------------------------------------

add_schedule = ("INSERT IGNORE INTO Schedule "
```

```
                "(ScheduleID, Project_ProjectID, ScheduleType, Version, DurationHours, StartDate,
FinishDate, DataDate, PlannerName) "
                "VALUES (%s, %s, %s, %s, %s, %s, %s, %s, %s)")

ws = wb ['Tasks']

DurationHours = ws['D2'].value
StartDate = ws['E2'].value
StartDate =StartDate.strftime("%Y:%m:%d %H:%M:%S") #converts data.datetime to string
FinishDate = ws['F2'].value
FinishDate =FinishDate.strftime("%Y:%m:%d %H:%M:%S")

ws = wb ['Codes']

for row in ws.iter_rows(min_row=2):
    for cell in row:
        if cell.value == "ScheduleType":
            ScheduleType = ws.cell(row=cell.row, column=6).value
        if cell.value == "Version":
            Version = ws.cell(row=cell.row, column=6).value
        if cell.value == "PlannerName":
            PlannerName = ws.cell(row=cell.row, column=6).value
        if cell.value == "DataDate":
            DateDate = ws.cell(row=cell.row, column=6).value

            data_schedule = (ScheduleID, ProjectID, ScheduleType, Version, DurationHours, StartDate,
FinishDate, DateDate, PlannerName)
            cursor.execute(add_schedule, data_schedule)



add_Element_Resource = ("INSERT IGNORE INTO Element_Resource "
                        "(ResourceID, Schedule_ScheduleID, ParentID_Resource, ResourceID_Synchro,
ResourceType, SupplierName, "
                        "ResourceGUID, ElementName, nb_elements, SBScode,
Element_Category_NLSfB_CategoryID, IFC_CategoryID, TaskID, TaskName) "
                          "VALUES (%s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s)")

####------------------------------------------------------------

ws = wb ['Resources']

for row in ws.iter_rows(min_row=2):
    ParentID_Resource = row[0].value
    ResourceID_Synchro = row[1].value
    ResourceType = row[2].value
    SupplierName = row[3].value
    ResourceGUID = row[4].value
    ElementName = row[5].value
    nb_elements = row[6].value
    SBScode = row[7].value
    Element_Category_NLSfB_CategoryID = row[8].value
    IFC_CategoryID = row[9].value
    TaskID = row[10].value
    TaskName = row[11].value

    data_Element_Resource = (ResourceID_Synchro + ScheduleID, ScheduleID, ParentID_Resource,
ResourceID_Synchro, ResourceType, SupplierName, ResourceGUID,
                             ElementName, nb_elements, SBScode, Element_Category_NLSfB_CategoryID,
IFC_CategoryID, TaskID, TaskName)
    cursor.execute(add_Element_Resource, data_Element_Resource)

####------------------------------------------------------------

add_ElementFieldValue_Resource = ("INSERT IGNORE INTO ElementFieldValue_Resource "
                          "(Element_Resource_ResourceID, Name, Value) "
                          "VALUES (%s, %s, %s)")

AmountColumn = ws.max_column - 12
#sets amount of columns equal to number of variables

d = {}
for i in range(AmountColumn):
    d["variable_name" + str(i)] = ws.cell(row=1, column=13+i).value
#retrieves 'Name' and stores them in a dictionary

for i in range(AmountColumn):
    for row in ws.iter_rows(min_row=2):
        ResourceID_Synchro = row[1].value
        variable_data = row[12+i].value
        VariableNameInput = 'variable_name' + str(i)

        data_ElementFieldValue_Resource= (ResourceID_Synchro + ScheduleID, d.get(VariableNameInput),
variable_data)
        cursor.execute(add_ElementFieldValue_Resource, data_ElementFieldValue_Resource)

#delete rows where the value is null
```

```
cursor.execute("DELETE FROM elementfieldvalue_resource WHERE elementfieldvalue_resource.value IS NULL or
elementfieldvalue_resource.value = '';")
#MySQL automatically converts commas to dots. Expression below again converts dots to commas for numeric
values.
cursor.execute("UPDATE elementfieldvalue_resource SET Value = REPLACE(Value, '.', ',') WHERE value REGEXP
'^[^A-Z]*$';")

####--------------------------------------------------------

ws = wb ['Tasks']
add_Task = ("INSERT IGNORE INTO Task "
            "(TaskID, Schedule_ScheduleID, TaskID_Synchro, TaskGUID, "
            "TaskName, DurationHours, StartDate, FinishDate, quantity_element, ABS_Category_ActivityID, "
            "StartVariation_StartVarCode, DurationVariation_DurationVarCode, DelayExplanation, DelayType,"
            "CriticalPath, BL_CriticalPath, Element_Resource_ResourceID, Parent_taskname, Parent_task_ID,
ActivityCount, AppearanceType, "
            "PercentageComplete, BL_PercentageComplete, BL_DurationHours, Successors, Predecessors,
DownStreamFloat, UpstreamFloat)"
            "VALUES (%s, %s, %s, %s, %s, %s, %s, %s, %s, %s, "
            "%s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s)")

for row in ws.iter_rows(min_row=2):
    TaskID_Synchro = row[0].value
    TaskGUID = row[1].value
    TaskName = row[2].value
    DurationHours = row[3].value
    StartDate = row[4].value
    FinishDate = row[5].value
    quantity_element = row[6].value
    quantity_element = re.sub('\(|\)', '', str(quantity_element))  # removes ()
    ABS_Category_ActivityID = row[7].value
    StartVariation_StartVarCode = row[8].value
    DurationVariation_DurationVarCode = row[9].value
    DelayExplanation = row[10].value
    DelayType = row[11].value
    CriticalPath = row[12].value
    BL_CriticalPath = row[13].value
    Element_Resource_ResourceID = row[14].value
    if Element_Resource_ResourceID:
        Element_Resource_ResourceID = row[14].value + ScheduleID
    if not Element_Resource_ResourceID:
        Element_Resource_ResourceID = row[14].value
    Parent_taskname = row[15].value
    Parent_task_ID = row[16].value
    ActivityCount = row[17].value
    AppearanceType = row[18].value
    PercentageComplete = row[19].value
    BL_PercentageComplete = row[20].value
    BL_DurationHours = row[21].value
    Successors = row[22].value
    Predecessors = row[23].value
    DownStreamFloat = row[24].value
    UpstreamFloat = row[25].value

    data_task = (TaskID_Synchro + ScheduleID, ScheduleID, TaskID_Synchro, TaskGUID,
               TaskName, DurationHours, StartDate, FinishDate, quantity_element, ABS_Category_ActivityID,
               StartVariation_StartVarCode, DurationVariation_DurationVarCode, DelayExplanation, DelayType,
               CriticalPath, BL_CriticalPath, Element_Resource_ResourceID, Parent_taskname, Parent_task_ID,
ActivityCount, AppearanceType,
               PercentageComplete, BL_PercentageComplete, BL_DurationHours, Successors, Predecessors,
DownStreamFloat, UpstreamFloat)
    cursor.execute(add_Task, data_task)

####--------------------------------------------------------

add_Equipment_Resource = ("INSERT IGNORE INTO Equipment_Resource "
                          "(Task_TaskID, EquipmentName, Quantity_Equipment,
Equipment_Type_UniClass_EquipID) "
                          "VALUES (%s, %s, %s, %s)")

for row in ws.iter_rows(min_row=2):
    Task_TaskID = row[0].value
    EquipmentName = row[30].value
    Quantity_Equipment = row[31].value
    Equipment_Type_UniClass_EquipID = row[32].value

    data_Equipment_Resource = (Task_TaskID + ScheduleID, EquipmentName, Quantity_Equipment,
Equipment_Type_UniClass_EquipID)
    cursor.execute(add_Equipment_Resource, data_Equipment_Resource)

cursor.execute("DELETE FROM Equipment_Resource WHERE EquipmentName IS NULL or EquipmentName = '';")

####--------------------------------------------------------

add_Human_Resource = ("INSERT IGNORE INTO Human_Resource "
                      "(Task_TaskID, RoleName, quantity_human, quality_human,
Human_Type_Uniclass_HumanID) "
                      "VALUES (%s, %s, %s, %s, %s)")
```

```
for row in ws.iter_rows(min_row=2):
    Task_TaskID = row[0].value
    RoleName = row[26].value
    quantity_human = row[27].value
    quality_human = row[28].value
    Human_Type_Uniclass_HumanID = row[29].value

    data_Human_Resource = (Task_TaskID + ScheduleID, RoleName, quantity_human, quality_human,
Human_Type_Uniclass_HumanID)
    cursor.execute(add_Human_Resource, data_Human_Resource)

cursor.execute("DELETE FROM Human_Resource WHERE RoleName IS NULL or RoleName = '';")

cursor.close()
db.commit()
db.close()
```

Appendix 6 - Workflow to create Synchro 4D as-planned and as-built models

Creating a Synchro 4D As-planned schedule and insert it into the database

1. Create a new Synchro 4D file and import a Synchro Project export or IFC file.
2. Create a BIM-based as-planned schedule where at the deepest level of the schedule one task is related to a maximum of one 3D Resource (element of the BIM model).

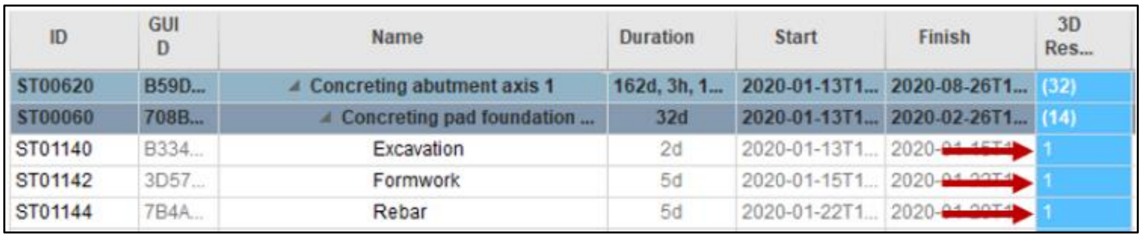


Figure 93: Element versus Task relation at the deepest level of the schedule

3. Create/insert attribute ProjectSBS (userfield) which represents the system/object breakdown structure and the ElementCategoryID (userfield) which represents the NL-SFB element code to which the elements in BIM model are related. Preferably, both codes are set within the designing process, thus are embedded in the BIM model elements.
4. Go to the 'Resource tab --> customize columns..' and set the fixed attributes in the exact order according to the figure below. (This is the same order as the Element_Resource table in SQL) Again, go to 'customize columns..' and select all other element-related attributes (variable attributes) that need to be saved (such as User fields that contain dimensions of the elements). (All these attributes are imported into the ElementFieldValue_Resource table.)

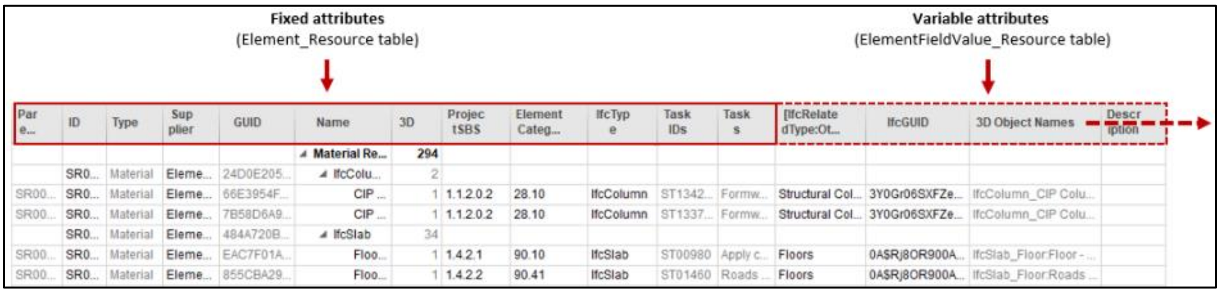


Figure 94: Attributes part of the Resource tab

5. Create activity codes: RoleName, Quality human, Quantity human (see figure below as example). Create resource userfield 'HumanID' and assign the code Roles codes from Uniclass according to the role assigned to a specific task.

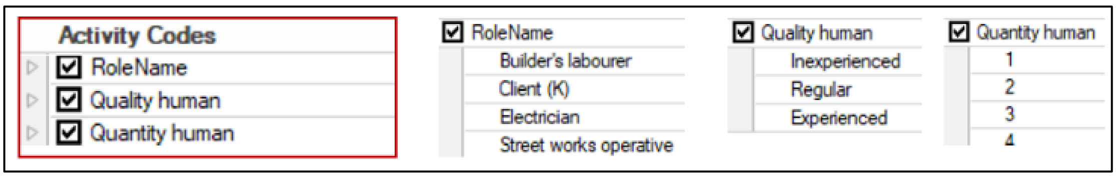


Figure 95: Human_Resource activity codes

6. Create activity codes: Equipmentname, Quantity Equipment (see figure below as example). Create resource userfield 'EquipID' and assign the Tool and Equipment codes from Uniclass according to the equipment assigned to a specific task.

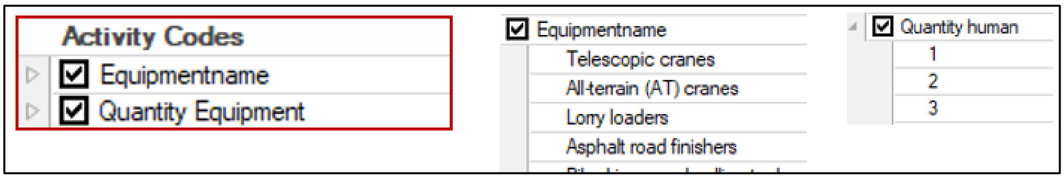


Figure 96: Equipment_Resource activity codes

7. Create task user fields: ABScategoryID allowing for a task/activity code to be assigned to a task at the deepest level according to the standard activity breakdown structure. Also, create userfield 'DelayExp' which allows for additional delay explanation to be written down if necessary.

8. Go to the 'Gantt tab --> customize columns..' and set the attributes in the exact order according to the figure below. These attributes facilitate the data for the Task, Human_Resource and Equipment_Resource table in SQL.

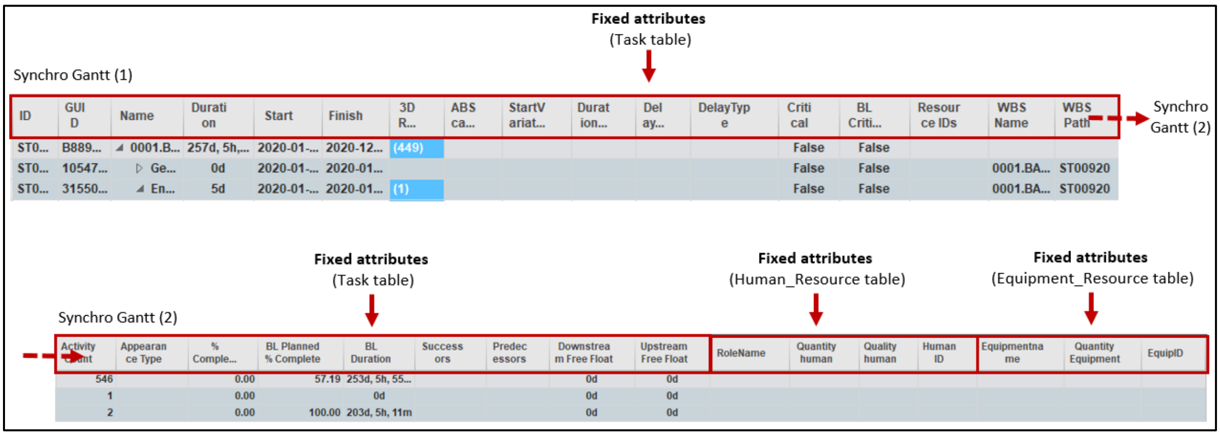


Figure 97: Attributes part of the Gantt tab

9. The fixed attributes of variation/delay remain empty in the as-planned situation. Once the as-built-progress and as-built schedule are made, the variation/delay information can be assigned.

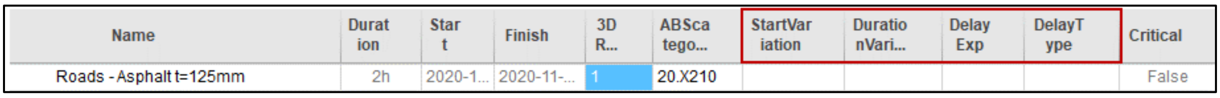


Figure 98: Variation/delay attributes

10. Create activity code 'Projectinfo' and create the exact code value names according to Figure 99. Assign the code value as part of the description ('0001' as an example in Figure 99). The ProjectTypeID and ProjectTypeName should correspond with classifications as listed in the NL-SfB table. The three possible ScheduleType are: 'As-planned', 'As-built-in-progress' and 'As-built'. The ScheduleType should be updated according to the type of export.

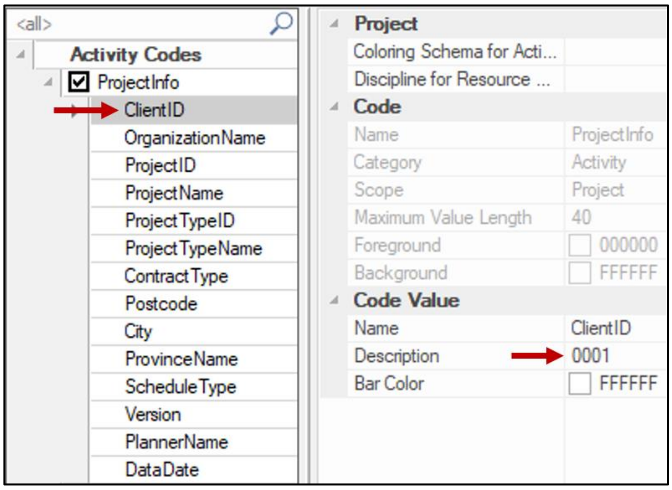


Figure 99: Project info setup

11. Select all tasks --> select 'Baseline Selected Tasks' and create baseline. Then, 'Reschedule' and 'Compute Critical Path'.
12. Export project as Excel. Set table layout to custom and 'Duration Units' to hours.
13. Open the Excel file --> 'file' --> 'info' --> 'check for issues' --> 'inspect document' --> 'inspect' --> Go through the list and hit 'Remove all' for all issues --> resave the Excel as workbook.xlsx.
14. Within the python script, adjust the variable 'filename' to match the path to the newly saved workbook.xlsx and make sure the name of the database is set up correctly. Then, run the script.

Creating a Synchro 4D As-built-progress or As-built schedule and insert it into the database

1. Adjust the schedule according to the progress on-site (dates, duration, input factors and internal environment). When lengthening the duration of a task, enrich the task with a DurationVariaton code and DelayType as in Figure 100. (DelayType is an activity code: 'Excusable.Compensable', 'Excusable.Non compensable' and 'Non excusable'.) Similarly, attach a 'StartVariation' code when a task is starting later than planned (does not apply for tasks that are moved due to the scheduling logic).

Name	Duration	Start	Finish	3D R...	ABScatego...	StartVariation	DurationVari...	Delay Exp	DelayType	Critical
Pile-driving abutment axis 1 - vertical	3d, 6h	2020-0...	2020-01-...	(8)						False
400x400mm - Vertical	2h	2020-0...	2020-01-...	1	10.X2211					False
400x400mm - Vertical	2h	2020-0...	2020-01-...	1	10.X2211					False
400x400mm - Vertical	2h	2020-0...	2020-01-...	1	10.X2211					False
400x400mm - Vertical	2h	2020-0...	2020-01-...	1	10.X2211					False
400x400mm - Vertical	2d	2020-0...	2020-01-...	1	10.X2211		4.5		Non excu...	False

Figure 100: Assign variation/delay information

An example of a section of the standard variation codes that can be applied is presented in Figure 101.

A	B	C
Code	Cause	ParentID
1	Prerequisite work	NULL
1.1	Obtaining required permits for the work to begin	1
1.2	Completion of previous work (i.e., work to be done before yours is not done yet)	1
1.3	Rework being required owing to the quality of previous work	1
1.4	Poor quality of previous work (though not to a level that requires rework)	1
1.5	Inspections for previously completed work	1
2	Detailed design and work method	NULL
2.1	Design constructability	2
2.2	Quality of documents (errors in design and/or drawings)	2
2.3	Turnaround time from engineers when there is a question	2
2.4	Strict specification requirements	2
2.5	Quality control requirements	2
2.6	Work complexity	2
2.7	Work sequence or method is not well planned	2
2.8	Low degree of repetition (inability to develop efficient	2
2.9	Inadequate instruction on detailed working method	2
3	Labor force	NULL
3.1	Socializing (talking with fellow workers)	3
3.2	Absenteeism	3

Figure 101: Example of a section from standard variation codes

2. Click 'Reschedule', then click 'Update progress of the current Project automatically' and select the data date. Then 'Compute Critical Path' and update the project info 'ScheduleType', 'Version' and 'DataDate' according to the export.
3. Select the baseline. Then export and save the project as described for the part of creating an as-planned schedule.

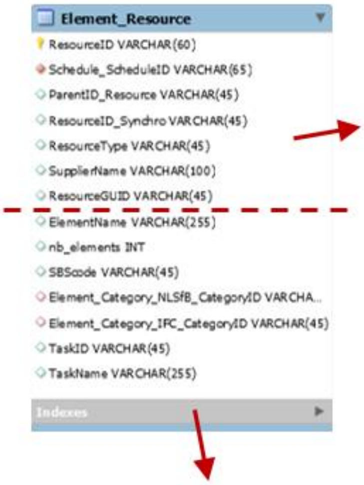
Appendix 7 - Workflow to connect MySQL to PowerBI

1. Create the 'views' for which MySQL queries are written according to Appendix 4.
2. Establish a connection between the MySQL tables and PowerBI within PowerBI.
3. PowerBI automatically establishes the relationships between the tables based on the primary keys and foreign keys. However, it is necessary to check for faulty relationships.
4. Execute 'split columns by delimiter' on the path column for the abstypepath, durvarpath, ElementNLSfBPath, equiptypepath, humantypepath and startvarpath tables. This operation allows to retrieve the data in a hierarchal manner.

Appendix 8 - Examples of sections of data in MySQL tables

ProjectID	ProjectName	ProjectTypeID	ProjectTypeName	ContractType	Postcode	City	ProvinceName
1	BAM Viaduct	1826	Viaducten	UAV-GC	2801SC	Gouda	Zuid-Holland

Figure 102: Project table



ResourceID	Schedule_S	ParentID_Res	ResourceID_Sy	ResourceTy	SupplierName	ResourceGUID
SR000036801b195...	1b195b...	SR00003660	SR00003680	Material	ElementSupplier4	EAC7F01A-E...
SR000036804cded...	4cded4b...	SR00003660	SR00003680	Material	ElementSupplier4	EAC7F01A-E...
SR00003680ab821...	ab8214...	SR00003660	SR00003680	Material	ElementSupplier4	EAC7F01A-E...
SR000037001b195...	1b195b...	SR00003660	SR00003700	Material	ElementSupplier4	855CBA29-E...
SR000037004cded...	4cded4b...	SR00003660	SR00003700	Material	ElementSupplier4	855CBA29-E...
SR00003700ab821...	ab8214...	SR00003660	SR00003700	Material	ElementSupplier4	855CBA29-E...
SR000037101b195...	1b195b...	SR00003660	SR00003710	Material	ElementSupplier4	6A5B8A2B-D...
SR000037104cded...	4cded4b...	SR00003660	SR00003710	Material	ElementSupplier4	6A5B8A2B-D...
SR00003710ab821...	ab8214...	SR00003660	SR00003710	Material	ElementSupplier4	6A5B8A2B-D...
SR000037201b195...	1b195b...	SR00003660	SR00003720	Material	ElementSupplier4	AB4D642F-2...
SR000037204cded...	4cded4b...	SR00003660	SR00003720	Material	ElementSupplier4	AB4D642F-2...
SR00003720ab821...	ab8214...	SR00003660	SR00003720	Material	ElementSupplier4	AB4D642F-2...
SR000037301b195...	1b195b...	SR00003660	SR00003730	Material	ElementSupplier4	BC970D37-9...
SR000037304cded...	4cded4b...	SR00003660	SR00003730	Material	ElementSupplier4	BC970D37-9...
SR00003730ab821...	ab8214...	SR00003660	SR00003730	Material	ElementSupplier4	BC970D37-9...
SR000037401b195...	1b195b...	SR00003660	SR00003740	Material	ElementSupplier4	4DFE3938-E...
SR000037404cded...	4cded4b...	SR00003660	SR00003740	Material	ElementSupplier4	4DFE3938-E...
SR00003740ab821...	ab8214...	SR00003660	SR00003740	Material	ElementSupplier4	4DFE3938-E...
SR000037601b195...	1b195b...	SR00003660	SR00003760	Material	ElementSupplier4	10AA323A-3...
SR000037604cded...	4cded4b...	SR00003660	SR00003760	Material	ElementSupplier4	10AA323A-3...
SR00003760ab821...	ab8214...	SR00003660	SR00003760	Material	ElementSupplier4	10AA323A-3...

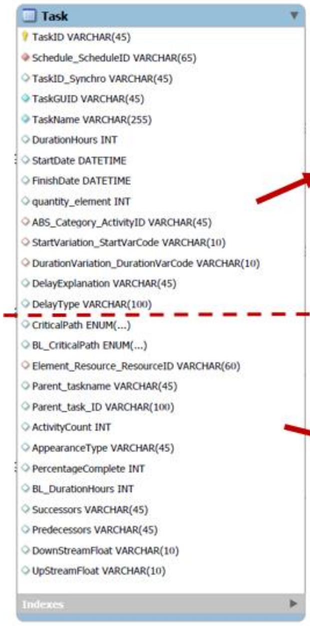
ElementName	nb_e	SBScode	Element_C	Element_(	TaskID	TaskName
Floor:Floor - Concrete t=200mm:613469_(#712972)	1	1.4.2.1	90.10	IfcSlab	ST00980	Apply curbside
Floor:Floor - Concrete t=200mm:613469_(#712972)	1	1.4.2.1	90.10	IfcSlab	ST00980	Apply curbside
Floor:Floor - Concrete t=200mm:613469_(#712972)	1	1.4.2.1	90.10	IfcSlab	ST00980	Apply curbside
Floor:Roads - Asphalt t=125mm:613454_(#712694)	1	1.4.2.2	90.41	IfcSlab	ST01460	Roads - Asphalt t=125mm
Floor:Roads - Asphalt t=125mm:613454_(#712694)	1	1.4.2.2	90.41	IfcSlab	ST01460	Roads - Asphalt t=125mm
Floor:Roads - Asphalt t=125mm:613454_(#712694)	1	1.4.2.2	90.41	IfcSlab	ST01460	Roads - Asphalt t=125mm
Floor:Floor - Concrete t=300mm - VAR:533280_(#40099)	1	1.3.3.3	90.41	IfcSlab	ST13200, ST1322...	Formwork, Rebar, Concr...
Floor:Floor - Concrete t=300mm - VAR:533280_(#40099)	1	1.3.3.3	90.41	IfcSlab	ST13200, ST1322...	Formwork, Rebar, Concr...
Floor:Floor - Concrete t=300mm - VAR:533280_(#40099)	1	1.3.3.3	90.41	IfcSlab	ST13200, ST1322...	Formwork, Rebar, Concr...
Floor:Floor - Concrete cast in place t=075mm:565879_(#599631)	1	1.4.1.1	90.10	IfcSlab	ST00960	Apply curbside
Floor:Floor - Concrete cast in place t=075mm:565879_(#599631)	1	1.4.1.1	90.10	IfcSlab	ST00960	Apply curbside
Floor:Floor - Concrete cast in place t=075mm:565879_(#599631)	1	1.4.1.1	90.10	IfcSlab	ST00960	Apply curbside
Precast Bufferplate:4500x1000x0300:537963_(#41825)	1	1.2.2.1	16.10	IfcSlab	ST05620	Install slab
Precast Bufferplate:4500x1000x0300:537963_(#41825)	1	1.2.2.1	16.10	IfcSlab	ST05620	Install slab
Precast Bufferplate:4500x1000x0300:537963_(#41825)	1	1.2.2.1	16.10	IfcSlab	ST05620	Install slab
Precast Bufferplate:4500x1000x0300:613439_(#712252)	1	1.2.1.1	16.10	IfcSlab	ST01080	Install slab
Precast Bufferplate:4500x1000x0300:613439_(#712252)	1	1.2.1.1	16.10	IfcSlab	ST01080	Install slab
Precast Bufferplate:4500x1000x0300:613439_(#712252)	1	1.2.1.1	16.10	IfcSlab	ST01080	Install slab
Precast Bufferplate:4500x1000x0300:613427_(#711520)	1	1.2.1.1	16.10	IfcSlab	ST01110	Install slab
Precast Bufferplate:4500x1000x0300:613427_(#711520)	1	1.2.1.1	16.10	IfcSlab	ST01110	Install slab
Precast Bufferplate:4500x1000x0300:613427_(#711520)	1	1.2.1.1	16.10	IfcSlab	ST01110	Install slab

Figure 103: Section of the Element_resource table

ElementFieldV	Element_Resource_ResourceID	Name	Value
107339	SR000044104cded4bc-5846-...	RUF:(BaseQuantities(IfcElementQuantity))NetVolume	15,46
107669	SR000044104cded4bc-5846-...	RUF:(BaseQuantities(IfcElementQuantity))OuterSurface...	388613,61
108989	SR000044104cded4bc-5846-...	RUF:(Dimensions)Length	10969,66
111299	SR000044104cded4bc-5846-...	RUF:(Dimensions)Volume	15,46
114269	SR000044104cded4bc-5846-...	RUF:(Dimensions)Elevation at Bottom	5850
115259	SR000044104cded4bc-5846-...	RUF:(Dimensions)Elevation at Top	6700
117569	SR000044104cded4bc-5846-...	RUF:(IfcRelatedType:Dimensions)Breedte	2000
118229	SR000044104cded4bc-5846-...	RUF:(IfcRelatedType:Dimensions)Hoogte	850
120869	SR000044104cded4bc-5846-...	RUF:IfcMaterial	Concrete Cast in Place (default)
121199	SR000044104cded4bc-5846-...	RUF:IfcRelatedType	2000x850mm
121529	SR000044104cded4bc-5846-...	RUF:IfcRelatedTypeClass	IfcBeamType
121859	SR000044104cded4bc-5846-...	RUF:(Other)Type Id	CIP Rectangular Beam: 2000x850mm
122519	SR000044104cded4bc-5846-...	RUF:IfcElementTag	532471
123179	SR000044104cded4bc-5846-...	RUF:(Geometric Position)Join Status	Both joins enabled
124169	SR000044104cded4bc-5846-...	RUF:(Geometric Position)y Justification	Origin
124499	SR000044104cded4bc-5846-...	RUF:(Geometric Position)y Offset Value	0
124829	SR000044104cded4bc-5846-...	RUF:(Geometric Position)yz Justification	Uniform
125159	SR000044104cded4bc-5846-...	RUF:(Geometric Position)z Justification	Top
125489	SR000044104cded4bc-5846-...	RUF:(Geometric Position)z Offset Value	0
125819	SR000044104cded4bc-5846-...	RUF:(Other)Family and Type	CIP Rectangular Beam: 2000x850mm
126149	SR000044104cded4bc-5846-...	RUF:(Other)Family	CIP Rectangular Beam: 2000x850mm
126479	SR000044104cded4bc-5846-...	RUF:(IfcRelatedType:Identity Data)Assembly Code	2.BM.CIP.REC
127469	SR000044104cded4bc-5846-...	RUF:(IfcRelatedType:Identity Data)Description	2000x850mm
127799	SR000044104cded4bc-5846-...	RUF:(IfcRelatedType:Identity Data)Keynote	2.BM
129119	SR000044104cded4bc-5846-...	RUF:(Structural)Cut Length	11980,02
129449	SR000044104cded4bc-5846-...	RUF:(Structural)Enable Analytical Model	True
129779	SR000044104cded4bc-5846-...	RUF:(Structural)Rebar Cover - Bottom Face	Rebar Cover Settings: Rebar Cover 1
130769	SR000044104cded4bc-5846-...	RUF:(Structural)Rebar Cover - Other Faces	Rebar Cover Settings: Rebar Cover 1
131099	SR000044104cded4bc-5846-...	RUF:(Structural)Rebar Cover - Top Face	Rebar Cover Settings: Rebar Cover 1

Figure 104: Section of the Elementfieldvalue_resource table

TaskID	Schedule_ScheduleID	TaskID_Sy	TaskGUID	TaskName	DurationH	StartDate	FinishDate	quanti	ABS_Categ	StartV	Duratic	DelayE	DelayT
ST0114...	d03ac1cd-ee67-401...	ST01142	3D572ED6...	Formwork	40	2020-01...	2020-01-2...	1	10.X3310	NULL	3.6	NULL	Non ...
ST0114...	d9f0356e-38f8-459...	ST01142	3D572ED6...	Formwork	40	2020-01...	2020-01-2...	1	10.X3310	NULL	3.6	NULL	NULL
ST0114...	476caee5-637f-43b...	ST01143	8279A949...	Remove formwork	8	2020-01...	2020-01-2...	1	10.X3319	NULL	NULL	NULL	NULL
ST0114...	d03ac1cd-ee67-401...	ST01143	8279A949...	Remove formwork	8	2020-01...	2020-01-3...	1	10.X3319	NULL	NULL	NULL	NULL
ST0114...	d9f0356e-38f8-459...	ST01143	8279A949...	Remove formwork	8	2020-01...	2020-01-3...	1	10.X3319	NULL	NULL	NULL	NULL
ST0114...	476caee5-637f-43b...	ST01144	7B4A724E...	Rebar	24	2020-01...	2020-01-2...	1	10.X3330	NULL	NULL	NULL	NULL
ST0114...	d03ac1cd-ee67-401...	ST01144	7B4A724E...	Rebar	40	2020-01...	2020-01-2...	1	10.X3330	NULL	2.9	NULL	Non ...
ST0114...	d9f0356e-38f8-459...	ST01144	7B4A724E...	Rebar	40	2020-01...	2020-01-2...	1	10.X3330	NULL	2.9	NULL	NULL
ST0114...	476caee5-637f-43b...	ST01145	D3F60DB9...	Backfilling	8	2020-01...	2020-01-2...	1	10.X104	NULL	NULL	NULL	NULL
ST0114...	d03ac1cd-ee67-401...	ST01145	D3F60DB9...	Backfilling	8	2020-01...	2020-02-0...	1	10.X104	NULL	NULL	NULL	NULL
ST0114...	d9f0356e-38f8-459...	ST01145	D3F60DB9...	Backfilling	8	2020-01...	2020-02-0...	1	10.X104	NULL	NULL	NULL	NULL
ST0115...	476caee5-637f-43b...	ST01150	2A797647...	Remove formwork	8	2020-02...	2020-02-0...	1	10.X3319	NULL	NULL	NULL	NULL
ST0115...	d03ac1cd-ee67-401...	ST01150	2A797647...	Remove formwork	8	2020-02...	2020-02-1...	1	10.X3319	NULL	NULL	NULL	NULL
ST0115...	d9f0356e-38f8-459...	ST01150	2A797647...	Remove formwork	8	2020-02...	2020-02-1...	1	10.X3319	NULL	NULL	NULL	NULL
ST0116...	476caee5-637f-43b...	ST01160	3CBB3265...	Remove formwork	8	2020-02...	2020-02-1...	1	10.X3319	NULL	NULL	NULL	NULL
ST0116...	d03ac1cd-ee67-401...	ST01160	3CBB3265...	Remove formwork	8	2020-02...	2020-02-2...	1	10.X3319	NULL	NULL	NULL	NULL
ST0116...	d9f0356e-38f8-459...	ST01160	3CBB3265...	Remove formwork	8	2020-02...	2020-02-2...	1	10.X3319	NULL	NULL	NULL	NULL
ST0119...	476caee5-637f-43b...	ST01190	1D9E6E99...	400x400mm - Vertical	2	2020-01...	2020-01-0...	1	10.X2211	NULL	NULL	NULL	NULL
ST0119...	d03ac1cd-ee67-401...	ST01190	1D9E6E99...	400x400mm - Vertical	2	2020-01...	2020-01-0...	1	10.X2211	NULL	NULL	NULL	NULL
ST0119...	d9f0356e-38f8-459...	ST01190	1D9E6E99...	400x400mm - Vertical	2	2020-01...	2020-01-0...	1	10.X2211	NULL	NULL	NULL	NULL



Criticall	BL_Crit	Element_Resource_	Parent_taskname	Parent_task_ID	Activit	Appeara	Percer	BL_Perc	BL_Dur	Succes	Predecess	Dow	Up
FALSE	FALSE	SR00006690d03...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	32	NULL	ST0114...	0	0
FALSE	FALSE	SR00006690d9f...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	32	NULL	ST0114...	0	0
FALSE	FALSE	SR00006690476...	Concreting pad ...	ST00920.ST00...	1	Install	0	100	8	NULL	ST1292...	0	0
FALSE	FALSE	SR00006690d03...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	8	NULL	ST1292...	0	0
FALSE	FALSE	SR00006690d9f...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	8	NULL	ST1292...	0	0
FALSE	FALSE	SR00006690476...	Concreting pad ...	ST00920.ST00...	1	Install	0	100	24	NULL	ST0114...	0	0
FALSE	FALSE	SR00006690d03...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	24	NULL	ST0114...	0	0
FALSE	FALSE	SR00006690d9f...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	24	NULL	ST0114...	0	0
FALSE	FALSE	SR00006690476...	Concreting pad ...	ST00920.ST00...	1	Install	0	100	8	NULL	ST0114...	0	0
FALSE	FALSE	SR00006690d03...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	8	NULL	ST0114...	0	0
FALSE	FALSE	SR00006690d9f...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	8	NULL	ST0114...	0	0
FALSE	FALSE	SR00005010476...	Concreting pad ...	ST00920.ST00...	1	Install	0	100	8	NULL	ST0114...	0	0
FALSE	FALSE	SR00005010d03...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	8	NULL	ST0114...	0	0
FALSE	FALSE	SR00005010d9f...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	8	NULL	ST0114...	0	0
FALSE	FALSE	SR00004270476...	Concreting pad ...	ST00920.ST00...	1	Install	0	100	8	NULL	ST0115...	0	0
FALSE	FALSE	SR00004270d03...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	8	NULL	ST0115...	0	0
FALSE	FALSE	SR00004270d9f...	Concreting pad ...	ST00920.ST00...	1	Install	100	100	8	NULL	ST0115...	0	0
FALSE	FALSE	SR00006950476...	Pile-driving abu...	ST00920.ST00...	1	Install	0	100	2	NULL	NULL	0	0
FALSE	FALSE	SR00006950d03...	Pile-driving abu...	ST00920.ST00...	1	Install	100	100	2	NULL	NULL	0	0
FALSE	FALSE	SR00006950d9f...	Pile-driving abu...	ST00920.ST00...	1	Install	100	100	2	NULL	NULL	0	0

Figure 105: Section of the task table

Task_TaskID	EquipmentName	Quantity_Equipment	Equipment_Type_UniClass_EquipID
ST00180476caee5-637f-43b0-8956-d5fa7d3eb...	Wire tensioners	1	TE_70_50_45_97
ST00180d03ac1cd-ee67-401b-af59-e7e8d6910...	Wire tensioners	1	TE_70_50_45_97
ST00180d9f0356e-38f8-459e-9084-6dc879e6e...	Wire tensioners	1	TE_70_50_45_97
ST00380476caee5-637f-43b0-8956-d5fa7d3eb...	Telescopic cranes	1	TE_30_10_10_88
ST00380d03ac1cd-ee67-401b-af59-e7e8d6910...	Telescopic cranes	1	TE_30_10_10_88
ST00380d9f0356e-38f8-459e-9084-6dc879e6e...	Telescopic cranes	1	TE_30_10_10_88
ST00940476caee5-637f-43b0-8956-d5fa7d3eb...	Excavators, loaders and dredgers	1	TE_40_70
ST00940d03ac1cd-ee67-401b-af59-e7e8d6910...	Excavators, loaders and dredgers	1	TE_40_70
ST00940d9f0356e-38f8-459e-9084-6dc879e6e...	Excavators, loaders and dredgers	1	TE_40_70
ST00950476caee5-637f-43b0-8956-d5fa7d3eb...	Excavators, loaders and dredgers	1	TE_40_70
ST00950d03ac1cd-ee67-401b-af59-e7e8d6910...	Excavators, loaders and dredgers	1	TE_40_70
ST00950d9f0356e-38f8-459e-9084-6dc879e6e...	Excavators, loaders and dredgers	1	TE_40_70
ST00960476caee5-637f-43b0-8956-d5fa7d3eb...	Excavators, loaders and dredgers	1	TE_40_70
ST00960d03ac1cd-ee67-401b-af59-e7e8d6910...	Excavators, loaders and dredgers	1	TE_40_70
ST00960d9f0356e-38f8-459e-9084-6dc879e6e...	Excavators, loaders and dredgers	1	TE_40_70
ST00970476caee5-637f-43b0-8956-d5fa7d3eb...	Excavators, loaders and dredgers	1	TE_40_70
ST00970d03ac1cd-ee67-401b-af59-e7e8d6910...	Excavators, loaders and dredgers	1	TE_40_70
ST00970d9f0356e-38f8-459e-9084-6dc879e6e...	Excavators, loaders and dredgers	1	TE_40_70

Figure 106: Section of the Equipment_Resource table

CategoryID	CategoryNL	CategoryEN	ParentID	hierarchy_level
0-	PROJECT TOTAAL	Sites, Project, Building systems	NULL	1
0-.0	indirecte projectvoorzieningen	Indirect project fittings	0-	2
0-.1	werkterreininrichtingen	Site facilities	0-	2
0-.10	indirecte projectvoorzieningen; werkterreininrichting, algemeen (verzamelniveau)	Indirect project fittings, site fittings, general	0-.1	3
0-.11	indirecte projectvoorzieningen; werkterreininrichting, bijkomende werken algemeen	Indirect project fittings, site fittings, additional works	0-.1	3
0-.12	indirecte projectvoorzieningen; werkterreininrichting, personen/materiaalvoorzieningen	Indirect project fittings, site fittings, people/plant	0-.1	3
0-.13	indirecte projectvoorzieningen; werkterreininrichting, energievoorzieningen	Indirect project fittings, site fittings, energy	0-.1	3
0-.14	indirecte projectvoorzieningen; werkterreininrichting, beveiligingsvoorzieningen	Indirect project fittings, site fittings, security	0-.1	3
0-.15	indirecte projectvoorzieningen; werkterreininrichting, doorwerkvoorzieningen	Indirect project fittings, site fittings, continuation	0-.1	3
0-.16	indirecte projectvoorzieningen; werkterreininrichting, voorzieningen belendende percelen	Indirect project fittings, site fittings, neighboring sites	0-.1	3
0-.17	indirecte projectvoorzieningen; werkterreininrichting, onderhoudsvoorzieningen	Indirect project fittings, site fittings, maintenance	0-.1	3
0-.2	materieelvoorzieningen	Plant fittings	0-	2
0-.20	indirecte projectvoorzieningen; materieelvoorzieningen, algemeen (verzamelniveau)	Indirect project fittings, plant fittings, general	0-.2	3
0-.21	indirecte projectvoorzieningen; materieelvoorzieningen, transport	Indirect project fittings, plant fittings, transport	0-.2	3
0-.22	indirecte projectvoorzieningen; materieelvoorzieningen, gereedschappen (algemeen)	Indirect project fittings, plant fittings, tools	0-.2	3
0-.3	risicodekking	Riskcoverage	0-	2
0-.30	indirecte projectvoorzieningen; risicodekking, algemeen (verzamelniveau)	Indirect project fittings, risk coverage, general	0-.3	3
0-.31	indirecte projectvoorzieningen; risicodekking, verzekeringen	Indirect project fittings, risk coverage, insurance	0-.30	3
0-.32	indirecte projectvoorzieningen; risicodekking, waarborgen	Indirect project fittings, risk coverage, safeguards	0-.31	3
0-.33	indirecte projectvoorzieningen; risicodekking, prijsstijgingen	Indirect project fittings, risk coverage, inflation	0-.32	3
0-.4	projectorganisatie	Project organisation	0-	2
0-.40	indirecte projectvoorzieningen; projectorganisatie, algemeen (verzamelniveau)	Indirect project fittings, project organisation, general	0-.4	3
0-.41	indirecte projectvoorzieningen; projectorganisatie, administratie	Indirect project fittings, project organisation, administration	0-.4	3
0-.42	indirecte projectvoorzieningen; projectorganisatie, uitvoering	Indirect project fittings, project organisation, execution	0-.4	3

Figure 107: Section of the Element_Category_NLSfB Table

+ X number of variables →

ElementFieldValueID	Element_Resource_ResourceID	3DObjectNames	RUF:(Base(	RUF:(Bas	RUF:	RUF:(B	RUF:(B	RUF:(Bas	RUF:(E	RUF:(B	RUF:(Bas	RUF:(Ba	RUF:	RUF:(E	RUF:(I	RUF	RUF:(C Level	RUF:((Offset	RUF:(Constraints)Level
92989	SR00003660476caee5-637f-43b0-8956-d5fa7d3eb19b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62299	SR00003660d03ac1cd-ee67-401b-af59-e7e8d6910bd9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30619	SR00003660d9f0356e-38f8-459e-9084-6dc879e6e66f	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62960	SR00003680476caee5-637f-43b0-8956-d5fa7d3eb19b	IfcSlab_Floor:Floor - ...	0	11,61	0	0	2322...	0	0	28,8	0	0	2,32	0	278...	200	0	0	0
31280	SR00003680d03ac1cd-ee67-401b-af59-e7e8d6910bd9	IfcSlab_Floor:Floor - ...	0	11,61	0	0	2322...	0	0	28,8	0	0	2,32	0	278...	200	0	0	0
260	SR00003680d9f0356e-38f8-459e-9084-6dc879e6e66f	IfcSlab_Floor:Floor - ...	0	11,61	0	0	2322...	0	0	28,8	0	0	2,32	0	278...	200	0	0	Level: +8.00 Top of Deck
62961	SR00003700476caee5-637f-43b0-8956-d5fa7d3eb19b	IfcSlab_Floor:Roads ...	0	52,27	0	0	6533...	0	0	108,94	0	0	6,53	0	351...	125	0	0	0
31281	SR00003700d03ac1cd-ee67-401b-af59-e7e8d6910bd9	IfcSlab_Floor:Roads ...	0	52,27	0	0	6533...	0	0	108,94	0	0	6,53	0	351...	125	0	0	0
261	SR00003700d9f0356e-38f8-459e-9084-6dc879e6e66f	IfcSlab_Floor:Roads ...	0	52,27	0	0	6533...	0	0	108,94	0	0	6,53	0	351...	125	0	0	Level: +8.00 Top of Deck
62962	SR00003710476caee5-637f-43b0-8956-d5fa7d3eb19b	IfcSlab_Floor:Floor - ...	0	0	0	0	0	0	0	338,32	0	0	43...	0	570...	0	0	0	0
31282	SR00003710d03ac1cd-ee67-401b-af59-e7e8d6910bd9	IfcSlab_Floor:Floor - ...	0	0	0	0	0	0	0	338,32	0	0	43...	0	570...	0	0	0	0
262	SR00003710d9f0356e-38f8-459e-9084-6dc879e6e66f	IfcSlab_Floor:Floor - ...	0	0	0	0	0	0	0	338,32	0	0	43...	0	570...	0	0	0	Level: +8.00 Top of Deck
62963	SR00003720476caee5-637f-43b0-8956-d5fa7d3eb19b	IfcSlab_Floor:Floor - ...	0	17,99	0	0	1349...	0	0	39,13	0	0	1,35	0	420...	75	0	0	0
31283	SR00003720d03ac1cd-ee67-401b-af59-e7e8d6910bd9	IfcSlab_Floor:Floor - ...	0	17,99	0	0	1349...	0	0	39,13	0	0	1,35	0	420...	75	0	0	0
263	SR00003720d9f0356e-38f8-459e-9084-6dc879e6e66f	IfcSlab_Floor:Floor - ...	0	17,99	0	0	1349...	0	0	39,13	0	0	1,35	0	420...	75	0	0	Level: +8.00 Top of Deck
62964	SR00003730476caee5-637f-43b0-8956-d5fa7d3eb19b	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	0
31284	SR00003730d03ac1cd-ee67-401b-af59-e7e8d6910bd9	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	0
264	SR00003730d9f0356e-38f8-459e-9084-6dc879e6e66f	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	Level: +8.00 Top of Deck
62965	SR00003740476caee5-637f-43b0-8956-d5fa7d3eb19b	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	0
31285	SR00003740d03ac1cd-ee67-401b-af59-e7e8d6910bd9	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	0
265	SR00003740d9f0356e-38f8-459e-9084-6dc879e6e66f	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	Level: +8.00 Top of Deck
62966	SR00003760476caee5-637f-43b0-8956-d5fa7d3eb19b	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	0
31286	SR00003760d03ac1cd-ee67-401b-af59-e7e8d6910bd9	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	0
266	SR00003760d9f0356e-38f8-459e-9084-6dc879e6e66f	IfcSlab_Precast Buff...	0	0,3	0	0	1350	0	0	12,3	0	0	1,35	0	2600	1...	0	0	Level: +8.00 Top of Deck
62967	SR00003800476caee5-637f-43b0-8956-d5fa7d3eb19b	IfcSlab_Floor:Roads ...	0	52,27	0	0	6533...	0	0	108,94	0	0	6,53	0	351...	125	0	0	0

Figure 108: Section of the ElementFieldValue_Resource_Col view