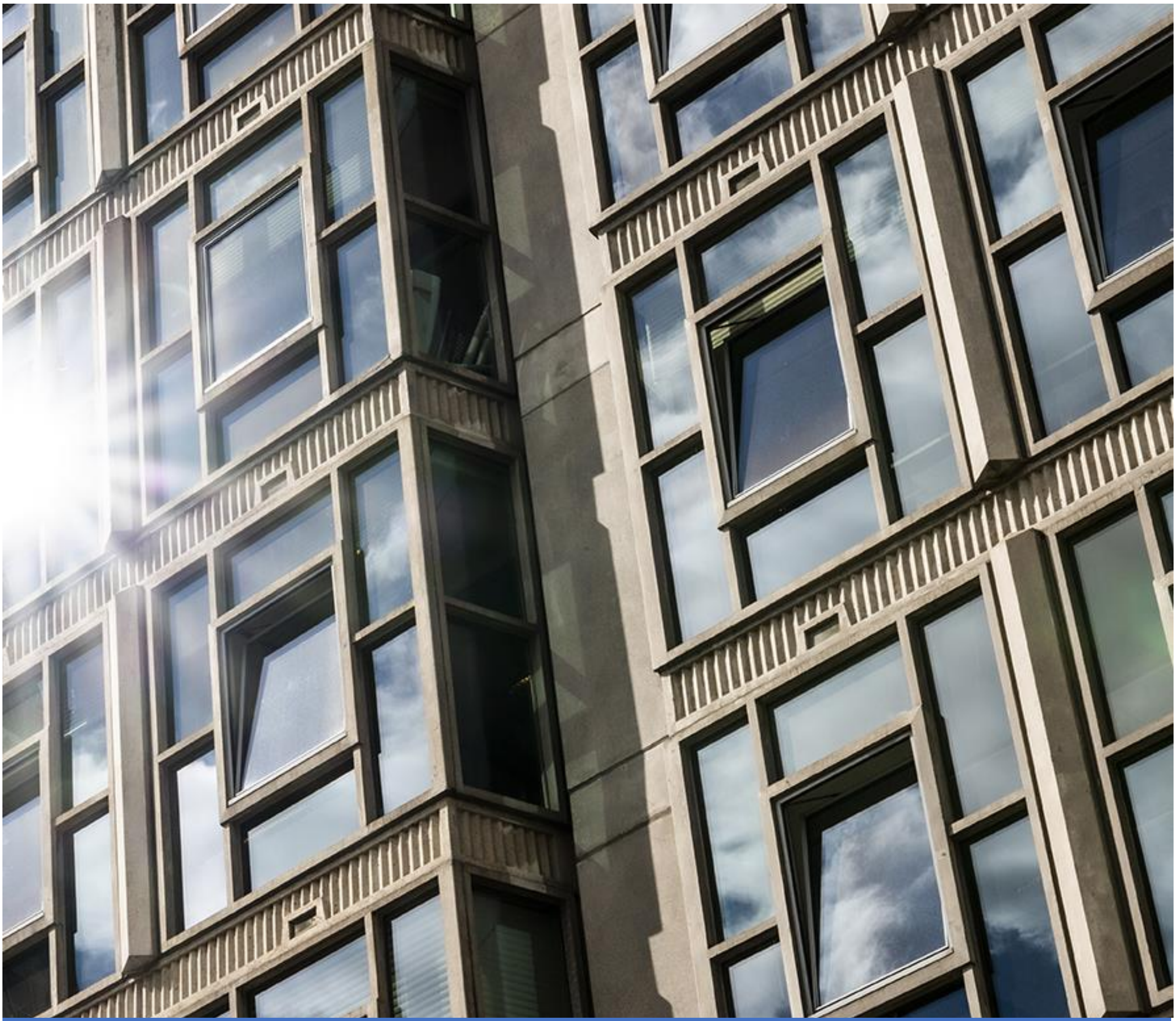
Graph-based data integration for ensuring FAIR project management information

Dusseldorp, Niels

Award date:
2022



Graph-based data integration for ensuring FAIR project management information

Construction Management & Engineering – 40 ects

Chair:	*Bauke de Vries*
1st supervisor:	*Pieter Pauwels*
2nd supervisor:	*Ekaterina Petrova*
Company Advisor:	*Dirk Bakker*

Dusseldorp, Niels 30/11/2022 Graduation Project

Table of contents

Preface

Dear reader,

Before you lies the result of my graduation research, which is my final deliverable of my educational career. A little over 6 years ago, my journey into the world of the Built Environment started with the intention of becoming an architect or urban planner, only to find out that the practice of project management was a way more suitable fit. This thesis thus focusses on the practice of project management, and especially the role of technology in facilitating such practices. It has always fascinated me how new innovations can be applied to making work easier, faster, or simply more fun. Therefore, I am happy to present you this graduation's work, in which I have aimed at making our work a bit more fun!

To achieve this goal, I have had to develop not only an understanding of the theoretical frameworks of project management, but also about how it is actually applied in real-life construction projects. For gaining such insights, I have to thank Count&Cooper, the company facilitating my graduation internship. They allowed me to take a look at the way-of-working of one of the Netherlands' largest infrastructure projects. I'm grateful for their efforts in teaching me how a project is done, but also how the underlying technological infrastructure works. Without their knowledge and experience, I would not have learned so much as I have now. In specific, I want to mention Dirk Bakker, and the Digital Project Management Team, as they have supported me throughout the entire graduation research. They have taught me to work agile, I which I was encouraged to critically look at all of the work to be done, and approach my progress in a structured manner.

Furthermore, I would like to thank Pieter Pauwels and Ekaterina Petrova, as the creation of this thesis would not have been possible without them. You have helped me in scoping my research, structuring my thesis and telling the right story. I have enjoyed our discussions a lot, and have learned a great deal about critically developing a working method, with its corresponding tools. You have helped me learn about software assessment, user environment design, and the wide range of possibilities graph databases possess. Of this my initial knowledge was just limited half a year ago, but now I am proud to provide you with this thesis, filled with everything I have learned.

I sincerely hope you enjoy reading this thesis, and hope it inspires you to think of your own ways to make working more easy, fast and fun through the clever use of data!

Niels Dusseldorp

Eindhoven, November 2022

Summary

Due to the digitalization of construction project management, an increasing amount of data is generated, which is incredibly valuable for the completion of a project within time and budget. However, this data associated to a construction project is not of a common structure, nor is it consistent over the different construction phases in a project. This makes the amount of information that is created, stored and exchanged within a construction project unfathomable. This can be attributed to the temporary nature of construction projects, and its corresponding group of contractors, sub-contractors and other associates. Each participant in a construction project uses their own applications, each generating their own respective pieces of information, which in turn might differ over the duration of a project. This complexity of different data types, different participants and the amount of produced data, only increases if project scale increases. This raises the issue of relevant information becoming lost in the large amount of data associated to a project. It therefore becomes increasingly difficult for a person associated to a project to stay up-to-date on progress, planned work and responsibilities. To still be able to manage a project, one becomes increasingly dependent on the use of technology. This raises difficulties, as certain parts of information might be stored in one system, but it should be compared to data in another system. Yet, the systems are not designed to work together, nor does their exported data structure match. This raises the question on how to maintain project management information findability, and accessibility, independent of project scale.

To answer such a question, an initial literature review is conducted on the origin of project management practices, and how these have developed up until now. What can clearly be seen is that construction project management has two distinct practices, both integral to project success. On the one hand, the construction modelling aspect of project management is identified. On the other hand, all associated documentation on management practices is identified, like planning, risk management and administrative tasks. Both of these groups have a striking difference in how its underlying standardization and data landscape are arranged. For construction modelling, recent years has seen the introduction of Building Information Modelling (BIM). This method is based on the principle of moving away from 2D drawings, towards a common, decentralized 3D model, which combines all different types of information. This is made possible due to the introduction of not only naming convention standards, but also data exchange standards. Through the introduction of IFC, a standard for 3D model representation, one is able to combine the models of multiple disciplines associated to a construction project, in a so-called IFC viewer. This allows knowledge gain, as data becomes integrated. The advantages of the adoption of such a standard is acknowledged by government bodies, as they have included the standards in their own policy making on construction. Contrary to this, other information associated to a construction project, only has standards on naming conventions, but not on data exchange formats. This limits the development of a platform which provides multi-source insights, comparable to IFC-viewers. Yet, there is a demand to gain insight in combined data, coming from multiple sources. This stresses the need for a standardized working method in combining multi-source project management information, to continue to provide accessible and findable information.

To identify this standardized working method, this thesis makes use of the Contextual Design Process. Within this process the current way-of-working is analyzed, and improved with the creation of a revised working method. This revised working method then is visualized in an user interaction design. To identify this process, two process coordinators are consulted, one part of a small project, one part of a big project. They were tasked to identify a process present in both of their projects, to ensure a working method which works independent of scale. Furthermore, they were asked to think about a case is which applies to the demand of multi-

source project management information insights. The resulting process deals with generating more insight in the progress of activities, and the documents associated to such activities. For this specific case, a user interaction design is presented, in the form of mock-up concepts.

These mock-up concepts are then translated to a prototype implementation. This implementation however should have an underlying data structure capable of integrating data from multiple sources. With the ISO 19650 standard a proposal is given on the integration of data associated with the delivery of a project, through the initiation of a Common Data Environment (CDE). It proposes a framework in which all data created, is send towards a common environment, where it is to be stored and used in further stages of the project. Setting up such a system requires the structuring of a data landscape, capable of housing such a variety of data. As the ISO standard only proposes a framework, one has to find a suitable storage technology. The application of graph technology is highlighted as a potential solution, as its way of storage is focused more on the relation between bits of data, instead of solely on the data itself. Since project management data is semantically rich, and highly dependent on its relations, graph methodology thus is a suitable fit.

Graph databases come in two different types, the labelled property graph (LPG), and the RDF-based graph. They are similar in the methodology they employ, which structures bits of data in nodes, which have relations to other bits of data, stored on other nodes. While both are able to combine semantically rich data, they differ quite a lot from each other. LPG's consist out of nodes, which are labelled. Each node can hold a certain set of properties, hence the labelled property graph name. RDF-based graphs have each bit of data separated in an individual node. Each property thus is a node related to another node. Furthermore, in RDF each node is identified by an Uniform Resource Identifier (URI), which can hold additional semantic information on the node itself. In this thesis three graph applications are reviewed, Neo4j, which is an LPG, GraphDB, which is RDF-based and Weaver, which is a hybrid of the two types.

For the proposed implementation, Neo4j is chosen. Within this tool the data associated with generating insight on activity progress, and its associated documents is loaded. This data originates from Primavera, used for project planning, and Relatics, a model-based Systems Engineering tool. The implementation includes the mapping of relations between these two data sets, to ensure data integration. Once the data is loaded into Neo4j, it can be visualized within a dashboard. The layout of the dashboard is based on the findings of the contextual design method. The dashboard is created in the NeoDash software. This dashboard shows three environments, one providing insight in all activities starting in the upcoming two weeks, one providing insight in the progress on preceding activities of specifically selected activity, and finally one which allows the comparison of activity data from Relatics to activity data of Primavera.

This prototype has shown that with the aid of a graph database, it is possible to integrate multi-source data into a singular tool. This is achieved through data mapping, which creates a fixed way of combining data. If data is exported to standard formats like CSV before being loaded into a graph, one is able to create one fixed scheme for the combination of data, which can be re-used over the spread of different projects. This ensures that data is findable and accessible. Furthermore, since the process identified by the process coordinators is present both in a small project, as well as a big project, and the mapping model is also usable for both, the scalability is achieved as well. A standard working method for findable and accessible information can thus be achieved by applying reusable data mapping structures, which is possible with the aid of a graph database.

Summary - Dutch

Vanwege de digitalisatie van project management binnen de bouw genereren we alsmaar meer data, welke van essentieel belang is bij het afronden van een project binnen tijd en budget. Echter is het zo dat deze data vaak niet van een constante structuur is, noch is het type of de soort data consistent hetzelfde gedurende de verschillende fases van een project. Dat maakt de hoeveelheid data die wordt gegenereerd, opgeslagen en uitgewisseld niet te bevatten. Die inconsistentie is te wijten aan de tijdelijke aard van een bouwproject, waarbij betrokken partijen vaak door wisselen. Elke betrokken partij heeft zijn eigen manier van werken, en de daarbij horende manier van dataopslag, welke ook weer kan variëren over de duur van een project. Deze complexiteit van data types, betrokken partijen en data hoeveelheid neemt alleen maar toe wanneer de project schaal toeneemt. Hierdoor raakt relevante informatie verloren in de overvloed aan data, waardoor het alsmaar moeilijker wordt om het juiste stukje informatie te vinden. Hierdoor wordt men steeds afhankelijker van het gebruiken van technologie, om up-to-date te blijven met de voortgang van een project. Echter wordt project informatie niet opgeslagen in een enkel systeem, maar is men afhankelijk van vele systemen, elk met hun eigen doel. Dit kan zorgen voor problemen, als data uit het ene systeem vergeleken moet worden met data uit een ander systeem, maar de systemen niet gemaakt zijn om samen te werken. Om toch projecten overzichtelijk te kunnen houden, beantwoord deze thesis de vraag hoe informatie vindbaar en toegankelijk gehouden kan worden, ongeacht de schaal van een project.

Om die vraag te beantwoorden is eerst een literatuur studie gedaan over de verschillende principes die allemaal komen kijken bij project management, en hoe deze zich hebben ontwikkeld tot nu. Hier kan het onderscheid gemaakt worden tussen twee groepen, beide belangrijk voor project succes. Het modelleren van het project, en het vastleggen van alle management gerelateerde processen omtrent het project. Op het gebied van standaardisatie van data stromen verschillende deze groepen significant. Voor het modelleren is in de laatste jaren Building Information Modelling (BIM) geintroduceerd. Deze methodiek heeft als doel van 2D tekeningen naar een centraal opgeslagen 3D model te gaan, waarbij verschillende modellen van sub onderdelen gecombineerd kunnen worden tot een model. Dit is niet alleen mogelijk door afspraken voor naamgeving, maar ook door de introductie van IFC, een datastandaard voor de 3D representatie van een model. Door modellen te combineren in een zogenaamde IFC viewer, kan men extra kennis vergaren over wat men gaat bouwen. Gezien de waarde die die kennis oplevert, wordt deze standaard ook opgenomen in het beleid van overheden.

Dit is echter niet zo voor de data met betrekking tot project management, waar enkel standaarden zijn gepubliceerd over naamgeving, maar niet over data uitwisseling. Dit limiteert de ontwikkeling van een platform voor management data integratie, vergelijkbaar met IFC viewers. Echter is er weldegelijk vraag naar inzicht op basis van gecombineerde data, wat de urgentie onderstreept om ook voor management data, een gestandaardiseerde manier te vinden om data te integreren, om ook hier informatie toegankelijk en vindbaar te houden.

Om deze gestandaardiseerde manier te vinden, maakt deze thesis gebruik van de contextual design process. In deze methode wordt de huidige manier van werken geanalyseerd en verbeterd in een herziene methode. Deze herziene methode wordt vervolgens visueel gemaakt in een interactie ontwerp, opgesteld door de gebruiker. Om deze herziene manier van werken op te stellen zijn twee proces coördinatoren geconsulteerd. Een onderdeel van een klein project, de ander van een groot project, om zo een proces te identificeren welke aanwezig is ongeacht project schaal. Zij zijn gevraagd een process te identificeren welke op beide projecten aanwezig was, en zich bezig hield met een informatie vraagstuk met data uit

meerdere bronnen. Zij kozen het genereren van inzicht in activiteiten voortgang, en de bijbehorende voortgang van gekoppelde documenten, als proces voor de herziene werkwijze. Van deze specifieke case is een interactie ontwerp gemaakt.

Dit interactie ontwerp dient als de basis voor de realisatie van een prototype applicatie, welke de standaardwerkwijze bevat. De achterliggende data opslag moet echter ook de data daadwerkelijk kunnen verbinden, omdat die uit meerdere bronnen komt. De ISO norm 19650 stelt het opzetten van een Common Data Environment (CDE) voor als platform voor data integratie. Echter omvat dit enkel een kader, welke een omgeving voorstelt waar alle data naartoe wordt gestuurd, opgeslagen wordt en gebruikt kan worden op latere momenten. Dit vergt het creëren van een data structuur waar meerdere formats in opgeslagen kunnen worden. Echter, omdat de standaard enkel een kader voorstelt, moet men zelf nog de juiste technologie vinden. In literatuur wordt de toepassing van een graph database voorgesteld als een oplossing met potentieel. Binnen zo'n database wordt de relatie tussen stukjes informatie geprioriteerd, in plaats van de data zelf. Gezien de semantiek van management data, en de vele relaties, lijkt de toepassing van een graph database is logische zet.

Een graph database kan van twee types zijn, een labeled property graph (LPG), of een op RDF gebaseerde graph. Ze gebruiken dezelfde methodiek, waarbij informatie wordt opgeslagen op nodes, welke weer een relatie hebben met andere stukjes informatie, opgeslagen op een andere node. Beide zijn in staat zo data te combineren, maar ze hebben ook hun verschillen. Bij een LPG heeft iedere node een label, en kan een node meerdere eigenschappen dragen. Bij RDF graph is dat niet zo, hier is iedere eigenschap opgeslagen op zijn eigen node. Verder maakt een RDF graph gebruik van Uniform Resource Identifiers (URI), om de nodes aan te duiden. Deze soort links kunnen extra informatie in zich opslaan. In deze thesis zijn drie graph applicaties geanalyseerd. Neo4j, een LPG, GraphDB, een RDF graph, en Weaver, welke een soort hybride graph aanbiedt.

Voor de voorgestelde nieuwe werkwijze is Neo4j gebruikt. De nieuwe werkwijze wil inzicht creëren in activiteit voortgang, en voortgang van bijbehorende activiteiten. De data om dat inzicht te genereren komt uit Primavera, gebruikt voor planning, en Relatics, gebruikt voor werkuiteenzetting. Om de data te combineren, moeten de relaties tussen de datasets in kaart worden gebracht in een datatemplate. Wanneer dit is gedaan, kan data in de graph worden geladen en worden gevisualiseerd. De visualisaties zijn gebaseerd op de bevindingen van het contextual design process. Dit resulteert in een dashboard, gebouwd in NeoDash. Dit dashboard toont 3 visualisaties, eentje die inzicht geeft over alle activiteiten die starten in de komende twee weken. Eentje die inzicht geeft in voorliggende activiteiten, en de daarbij behorende activiteiten, en als laatste eentje die het mogelijk maakt Primavera activiteiten te vergelijken met Relatics activiteiten.

Dit prototype laat zien dat met de hulp van een graph database, het mogelijk is data uit verschillende bronnen te integreren in een applicatie. Dit wordt bereikt door het in kaart brengen van de data in een datatemplate, welke data relaties vastzet. Omdat data eerst geëxporteerd moet worden naar een standaardformat zoals CSV, is het gecreeërde datatemplate herbruikbaar voor meerdere projecten. Dit template structureerd de data zo, dat deze vindbaar en toegankelijk blijft. Omdat de case een proces heeft gekozen welke zowel op een klein als groot project plaatsvind, en het template toepasbaar is op beide, is de schaalbaarheid van de methode ook bereikt. Een standaard methode voor vindbare en toegankelijke informatie kan dus bewerkstelligd worden door de datastructuren te standaardiseren, wat mogelijk is in een graph database toepassing.

Abstract

Due to an increasingly complex information landscape within construction projects, individual participants of project teams tend to struggle more and more in finding the correct information. This struggle increases when project scale increases, due to the many stakeholders involved. Within the scope of project management information, one can distinguish two information types: construction modelling and project management information. While modelling has seen the development of data integration standards like IFC, these standards lack for project management information. To ensure information findability and accessibility of project management information, a standard working method is needed.

This standard is proposed through the application of the contextual design process, which identifies and suggests improved business processes, with a user interface design as a result. To identify this process, two process coordinators are consulted, both of a big and small project. To make the application of this standard possible, a technology is needed to facilitate project management data integration. For this a graph database is selected. This technology allows the mapping of data, which can be done in a standardized and reusable format. The findings of the contextual design process provide the input for the needed graph visualizations, which are translated to a dashboard. Together this provides a standardized and reusable method for project management data integration.

Keywords: Project Management Information Systems, Graph databases, Data integration, Contextual Design Process

List of Abbreviations

AEC	Architecture, Engineering & Construction
API	Application Programming Interface
BIM	Building Information Management
BRR	Business readiness rating
CDE	Common Data Environment
EIR	Exchange Information Requirements
FAIR	Findable, Accessible, Interoperable, Re-usable
HTTP	HyperText Transfer Protocol
ICT	Information Communication Technology
IFC	Industry Foundation Classes
ISO	International organization for standardization
LPG	Labelled Property Graph
PIM	Project Information Model
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier

List of figures

List of tables & Listings

1. Introduction

Within this introduction, the background and reasoning of the research is introduced. Furthermore, the problem statement is given, and questions on how to solve that problem are defined. The objectives that are set for the answering of these questions are given. Lastly, this section includes a reading guide for the remainder of this document.

1.1 Background

The construction industry, and its broader ecosystem, erects buildings, infrastructure, and industrial structures that are the foundation of our economies and are essential to our daily lives. More importantly, the construction industry is going through a disruptive phase. The McKinsey report states that in the coming years, fundamental change within the AEC sector is to be catalyzed through changing market characteristics, among which scarcity of skilled labor, persisting cost pressure from both infrastructure and affordable housing. This is accompanied by stricter regulations on work-site sustainability and safety, while also the sophistication and needs of customers and owners evolve (McKinsey & Company, 2020). The market characteristics can thus be seen as catalysts for change within the construction sector, which in turn poses challenges upon information management. Information management within construction applies to many aspects of the process, as the process of building is complex in its own nature, as interactions between the project disciplines, size of the project, interactions between stakeholders, strategic importance of the project and multiple critical paths, are the most significant contributors to project complexity (Erol et al., 2020). Walker et al. argues that this diverse set of characteristics cause uncertainty, and therefore increase the complexity of construction. To mitigate this uncertainty, more efficient collaborative practices are proposed as a potential solution (Walker et al., 2017).

Collaborative practices based on project information within construction rely heavily on technological software applications, however, due to the temporary nature of construction projects and its respective project consortia, innovation in the field of digital collaborative practices is limited. *“The built environment has a very wide span of data. With various application domains, diverse stakeholders and uses each with several levels of details—the amount of data communicated and exchanged is unfathomable. These robust set of interactions lead to the existence of diverse data about the built environment being exchanged in a large diversity of ways”* (Pauwels et al., 2022). This lack of unity in data created in a project results in a lack of integration of the data, increasing the challenge of findability and thus useability of the project information. This is an issue also raised by a large group of stakeholders, who have joined forces to create the FAIR Data Framework (Wilkinson et al., 2016). They argue that good data management is not a goal in itself, but a means leading to knowledge discovery and innovation. Within the FAIR Data Framework guidelines are set up on ensuring data findability, accessibility, interoperability and reusability.

To be able to innovate within the digitalization of collaborative practices a shift in the way data is handled is thus needed. For this, the FAIR Data Framework provides guidelines, not a specific method or software package. To facilitate such a shift, these guidelines are required to be applied to the information management practices in construction projects. The nature of information management within construction projects is inherently complex, partially due to the temporary nature of construction projects. As each project is executed by a consortium, which is formed by a set of companies, who in turn employ a large set of sub-contractors. Furthermore, each participant of the construction project uses their own modelling methodologies and corresponding information storage.

Therefore, within such a project, information management covers many different aspects of the building process, of which data is stored in a variety of datatypes and storage systems.

Within those systems, the types and content of the data alternates over the various building phases. Having a smoothly running project, defined by increasing data integration, findability and accessibility of construction project information, has several beneficial effects. Projects with effective data management in place are more likely to finish in time, with the budget limitations, which is the main goal of project management (Martínez-Rojas et al., 2015). Creating such conditions to facilitate effective data management thus is a relevant field of further research.

1.2 Problem statement

As described, large scale projects tend to struggle with their information management. This poses several challenges, both in the technological domain and in the procedural domain. Due to an increase in amount of data associated to a project, data infrastructure, storage and data access mechanisms work less efficiently (Martínez-Rojas et al., 2015). This results in information becoming less accessible and findable to the employee, causing delay, increased risk and failure. Due to incorrect data handling the data and its structure becomes increasingly unstructured. With an increase in project scale, it is less easy to understand the entirety of the project, and how personal tasks relate to the whole. This can partially be attributed to the way we organize our information, but it also finds ground in information management procedures. Not only the scale of construction projects contributes to this issue, but also the temporal and fragmented nature of construction projects in general. With each construction project being unique concerning its constituent elements, each requiring its own particular form of expertise and management. It is observed that gained knowledge derived from a construction project is often lost due to the short time between capturing the knowledge, the turnover of technical staff and the dismantling of teams engaged for a specific project at the project's end (Vaz-Serra & Edwards, 2021).

Here, an important distinction should be made, as within project management one needs to deal with a large variety of data. This data comes from different systems, which each serve their specific purpose. These flows of data can be categorized in two groups, construction modelling, and project management information. Construction modelling has seen its fair share of innovation with standardized BIM adoption, allowing for model integration for knowledge gain. Project management information however has seen a more limited shift, as solely standards on naming conventions have been adopted. Yet, a data standard is still lacking, increasing difficulty on data integration and knowledge gain.

To accommodate for the aforementioned problems with regards to project management knowledge, a solution can be to streamline information management processes digitally, and thus more sustainably store knowledge. If the FAIR principles are taken into account, data should become findable, accessible, interoperable and reusable through the effective use of metadata, stored in a common location (Wilkinson et al., 2016). This need for data integration has also been acknowledged through the creation of ISO 19650 - Organization and digitization of information about buildings and civil engineering works, including building information modelling. This standard provides guidelines on possible solutions for data integration, specifically for the AEC sector. Among these practices, the initiation of 'Common Data Environments' shows potential (ISO, 2019).

A CDE can be initiated in many forms, however, due to the semantically rich nature of project management data, a graph database is proposed as a tooling with potential (J. Werbrouck et al., 2019). This methodology shows great promise in the ability of multi-source data linkage. Ideally, an environment in which information is able to flow freely from system to system, without losing its meaning or integrity, which is a concept called interoperability, is created. However, due to non-compatible software, proprietary information and legal issues this process has proven to be difficult (Pauwels et al., 2022). While this intent still requires a lot of development and research, the potential of graph technology in aiding the integration of data from different sources is stated. The application of graphs is also seen as a step which facilitates further Building Information Modelling (BIM) innovation, as it is aims to move away from the information carrier level of files to the data and information level desired (Simeone et al., 2020).

Based on its focus on relations between data, the graph database can be considered as tooling with potential in making data integration advancements. However, graph databases also have their weaknesses, as it is less strong in supporting data formats like documents or building geometry (Pauwels et al., 2022b). Therefore, it is important to asses where graph technology can provide value for the storage and combination of project management data.

To be able to gain knowledge on project management information storage, integration and presentation, this thesis aims to answer the following research question and sub-questions:

"How to maintain construction project information findability and accessibility independent of project scale, with the application of a graph database?"

This question can only be answered if the following sub-questions are answered. These questions will serve as a guide in the set-up of the thesis, as they will form individual steps in the research.

1. *"What types of data are associated to a construction project?"*
2. *"How can the development of a graph database be made useable and thus accessible to the business?"*
3. *"What is the current state of art with regards to graph databases?"*
4. *"What practices are important in data preparation for graph databases?"*
5. *"What role should existing implemented project management systems fulfill in combination with the graph database?"*

1.3 Thesis Objectives

As mentioned previously, construction projects suffer heavily from the non-effective functioning of their information structures, which cause delay and cost increase. By assessing current processes present in the AEC sector, investigating the corresponding data flows and understanding how they correlate, steps can be taken to improve the effectiveness of the information structure. It implements these procedural findings in the creation of a revised working method for data integration, with a graph database as underlying data structure. While much research has gone into the development of BIM, and its associated modelling processes, research in the creation of standards and enterprise platforms for construction management programs has stayed behind, this thesis aims to contribute to this identified research gap. For this, the following objectives have been set:

- Business processes struggling with information findability and accessibility are identified
- An assessment is done of present project management data structures
- Explore the noted graph database potential for data integration
- Create a standardized way of data integration for project management information.

Through this approach, several objectives are to be achieved. First, the thesis aims to provide insight in business processes as they are currently, through these insights potential improvements can be identified. This identification process is conducted in cooperation with experts from multiple projects, differing in scale. This allows us to also gain insight in the effects of project scale increase to processes part of a construction project.

Secondly, the data associated to the current way of working is analyzed. Creating insight in the source formats of different systems applied in the business, and how this data can be transformed from its native format, to a standardized format. When this data is identified, it is assessed which data is suitable for combination, to be able to retrieve valuable additional or new knowledge.

As the thesis specifically focusses on the application of graph databases, since they are marked as a tools with potential for the AEC sector, a thorough assessment is done on the state-of-art of graph databases. As graphs come in many shapes and sizes, this thesis aims to provide a structured view of the unique characteristics of multiple types of graphs. Discussing its advantages, but also its drawbacks. It does not aim to pinpoint the 'better' graph, but merely provide a broad review of the different types.

Finally, the thesis aims to improve upon existing information structures, to increase information findability and accessibility. To achieve this, the creation of a standardized way of working for data integration is proposed. To validate if such a standard way of working actually ensures information findability and accessibility, the contextual design process provides a highly iterative process. This is all done based on input gained through expert consultation.

1.4 Thesis Outline

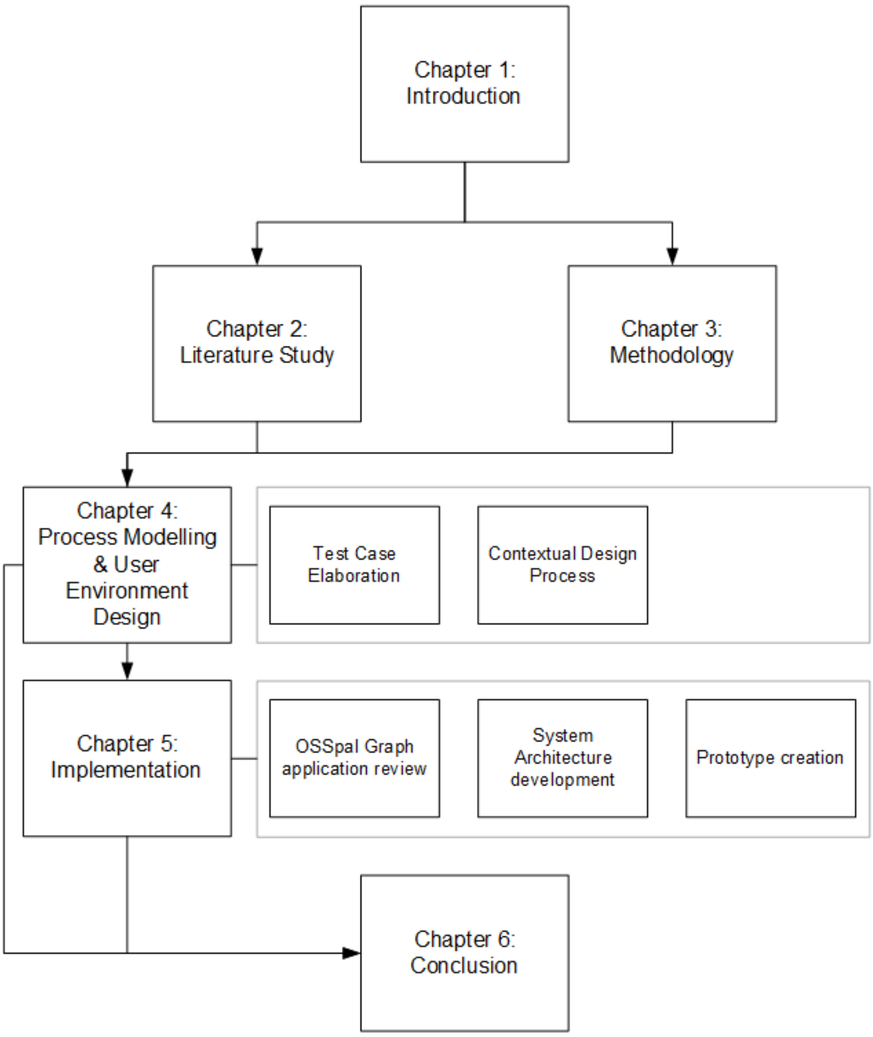
In order to find the answers to the research questions defined in Section 1.2, this master thesis encompasses the following elements. Initially, a literature review is conducted in chapter 2. The methodology of the literature review is discussed at the start of the chapter. Throughout this literature review the information landscape of the AEC sector is explored. It furthermore argues current construction management practices and its corresponding methods, while also discussing the state-of-art with regards to graph databases. In doing so, the literature review thus answers sub-question 1 and 2 of the research questions.

In chapter 3 the methodology is explained. This section consists of 3 sub-sections, which explain the process of identifying a solution for the given research problem of a lack of project management data integration. First, a reference is made to the setup of the literature review. Second, a detailed description is given of the conceptual design process. This process consists out of 8 individual steps, spread over 2 phases. Each step is elaborated separately, with further elaboration on how it is applied within the expert interviews if necessary. Third and last, the implementation of the knowledge gained is discussed. Here all the steps taken to create the proposed new working method, with its supporting technology, are discussed. This starts with the assessment framework used to provide a structured review of the different graph database systems. Second, the process of data collection and preparation is discussed. Then, the knowledge gained is used for the development of a prototype application. Finally, the validation is discussed.

In Section 4 the results of the process modelling & user interaction design is discussed, within this section the cases used for the process identification are described. Then, based on the input gained throughout the contextual design phase, the situation as-is is described. Then the to-be situation which was imagined throughout the contextual design process. Finally, a look ahead is given on the proposed standardized way of working.

In the Section 5 the implementation of the knowledge gained out of the process modelling and user environment design, as well as the contextual design process is translated to a proposed standardized way of working, illustrated by a working prototype. This section includes the review of 3 different graph applications. One of these applications is used to facilitate data integration in the proposed standardized working method. This requires the mapping of data, and the creation of a dashboarding interface. The chapter then includes steps on validating the created product and some concluding remarks.

In the conclusion the initial aim and research question is reflected upon, with a discussion on the choices which were made and results that were found. The structure of the thesis is summarized in Figure 1.



Chapter 1: Introduction
Chapter 2: Literature Study
Chapter 3: Methodology
Chapter 4: Process Modelling & User Environment Design
Test Case Elaboration
Contextual Design Process
Chapter 5: Implementation
OSSpal Graph application review
System Architecture development
Prototype creation
Chapter 6: Conclusion

Figure 1: Thesis structure

2. Literature Review

A literature study is conducted to review the current state-of-art with regards to the information landscape in the AEC sector. Within this literature review, the origin and development of project management practices are discussed. Furthermore, it reviews current methods and corresponding data. Of these methods, standards and policies are assessed, which together present the current standing and attitude towards project management data.

2.1 Literature review methodology

To be able to systematically review literature about project management information and its corresponding systems and methods, an unbiased and structured approach is needed. To do so, a systematic literature review is initiated according to the PRISMA 2020 method, short for 'Preferred Reporting Items for Systematic reviews and Meta-Analyses.'(Page et al., 2021) This is important due to the rapidly changing nature of digitalization within construction. Therefore it is essential to provide conclusions on existing methods' strengths and weaknesses, as well as the potentials and threats of new innovations. The systematic review aims to provide a broad scope of available methods associated with information management of construction projects. Within this scope data flows, data formats and system dependencies should be elaborated. It thus aims to provide a scope of the most commonly used systems with regards to information management within construction projects.

To be able to specify if the found literature is eligible or ineligible for the research, inclusion and exclusion criteria have to be drawn up. For the literature to be included, it has to adhere to the following criteria;

- Literature deals with infrastructural, utility or housing construction project data
    - Other fields of engineering deviate from construction due to their different composition of project participants.
- Literature assesses existing systems
    - Proposed frameworks within existing methods are taken into account, but newly proposed systems, without actually being developed and validated, are not taken into account.
- Literature is not older than 5 years
    - Due to the rapidly developing nature of technology in general, the scope is limited to only 5 years. This is especially important for literature concerning the development and application of software systems, since these iterate and improve themselves as well. Issues, drawbacks or other hinderances described in older literature might have been resolved in later stages.

Furthermore, some exclusion criteria have also been drawn up, to facilitate the scope of the research. These are as follows;

- Literature focusses on solely 3D modelling software
- Literature reviews non-western construction practices
- The literature deals with data other than planning and work preparation

The literature to be reviewed is collected through online databases. To be able to do this in a systematic way, the keywords have to be chosen carefully. The keywords used are collected through the review of several seminal works within the field of construction information management. These keywords are structured as follows:

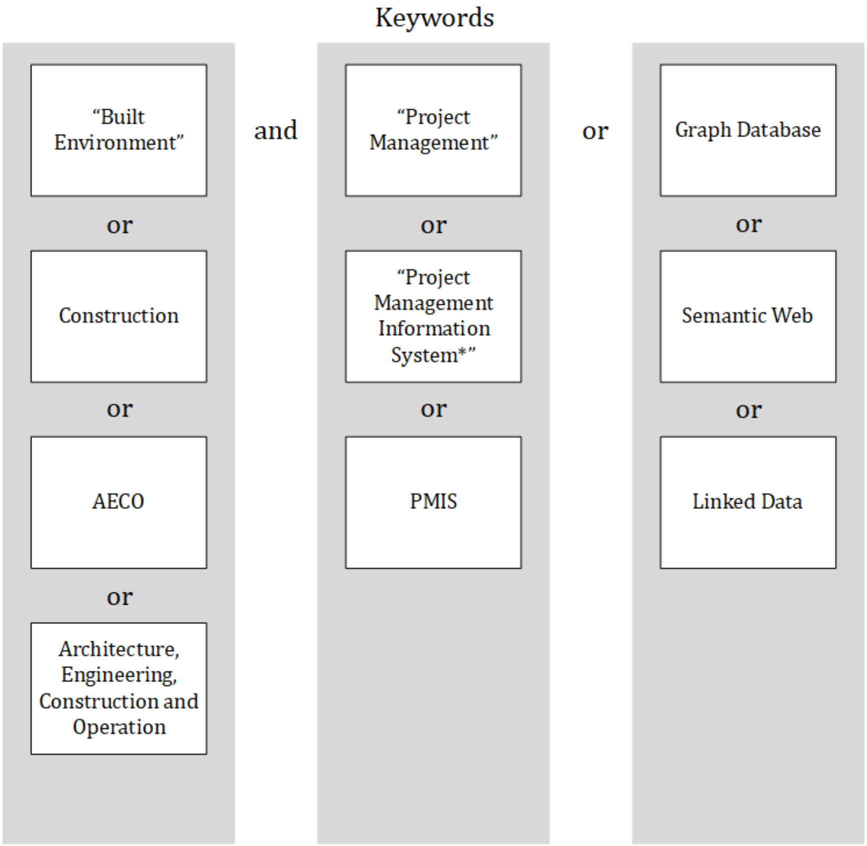


Figure 2: Literature Review Keywords

This yields the following search query to be inserted into the databases: (TITLE-ABS-KEY (graph AND database OR "semantic web" OR "linked data") OR TITLE-ABS-KEY ("project management information system" OR pmis OR "project management") AND TITLE-ABS-KEY (construction OR "built environment" OR aeco OR architecture, AND engineering, AND construction AND 7rganization)) AND (LIMIT-TO (SUBJAREA , "ENGI")) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018))

To find the literature needed for the systematic review, 3 data sources will be used; Scopus, Web of Science and ABI complete. On 20/05/2022 Scopus was conducted, which yielded 142 possible entries. On 25/05/2022 Web of Science was conducted, which yielded 283 possible entries. On 17/05/2022 ABI Complete was conducted, which yielded 13 possible entries. These entries were checked according to the set criteria, yielding 40 papers with relevant themes to the search. These 40 papers serve as the base for the literature review, and are complemented with additional literature found through the waterfall method, and further specific search queries on specific topics.

To be able to provide insight in project management information, it is important to understand the current practices of information management, and its corresponding methods and tools. Complex project management environments are characterized by interrelated sub-systems, the involvement of various stakeholders and disciplines, while having overlapping phases (Ershadi et al., 2022).

2.2 Project management complexity

As stated by the McKinsey report, one of the AEC sector's disrupting factors is the digitalization of products and processes, as "*Digital technologies can enable better collaboration, greater control of the value chain, and a shift toward more data-driven decision making."* (McKinsey & Company, 2020). The way digital tools play a role in the management of a construction project is thus considered of great relevance to project success. To be able to discuss the management of a 'construction project' it is important to agree on a common understanding of the term, which is described by Kania et al. as follows; "*Construction projects can be characterized as temporary organizations that are interdisciplinary in character and that are created to achieve specific goals in a set amount of time."* (Kania et al., 2021). It further states that the number of parties involved in the various processes during construction causes communications to become increasingly complex, which further highlights the large dynamic of a construction project.

This shift in digitalization offers several advantages within construction, however, this can only be achieved if the use of new innovations is adopted within the working practices. This often requires a tailored approach, as each construction project is unique concerning its constituent elements, requiring particular forms of professional inputs and management (Vaz-Serra & Edwards, 2021). The key to understanding the specificity of communication and, as a consequence, managing its course, is to understand the formal organizational structure of the project, analyzing its schedule, its realization, and identifying the necessary knowledge resources, the exchange of which determines the character of this communication (Kania et al., 2021). This differs for each project, due to its temporary nature and its temporary organizations. Furthermore, one is able to gain a competitive advantage by applying the gained knowledge and experience within a company as a strategic asset within a company. Applying this gained knowledge proves to be difficult due to a construction project's temporary nature, since partners of the initial project might not be the same as the next, this thus also requires inter-organizational innovation (Vaz-Serra & Edwards, 2021). Being able to communicate efficiently within a project proves to be of great importance, as argued by Luo et al. project complexity decreases if communication is good, while a high project complexity is bad for project performance (Luo et al., 2017). The main challenge according to Luo can be summarized in two categories; organizational complexity and technological complexity.

Within this technological complexity, one argues that not only organizations part of construction projects can have a discrepancy in working methods, but the tools they use and thus the data they produce as well. This poses an issue of interoperability, which alludes to the struggle of information management, where data is not stored centrally, but in its respective systems. These so-called data islands are not communicating, nor do they have a standardized way of combining them, as they differ in file format and scheme.

These systems and their corresponding files differ quite a lot, therefore it is important to understand the control mechanisms affecting a construction project. Construction project management makes use of many tools and techniques, such as the Work Breakdown Structure, Gantt Charts, PERT/CPM networks, Trade-off analysis, etc. (Love et al., 2001). Together these methods provide insight in planning, progress and work to be done. This in turn deals with the progress of the different phases of a construction project, which can generally be divided in three main phases; design, execution and maintenance. While each phase tends to be independently managed, information is still shared across phases. Therefore, effective data management is integral to project success (Martínez-Rojas et al., 2015).

2.3 History of construction project management related data

Within the AEC industry, multiple parties collaborate in order to complete projects, both of buildings as well as infrastructure. The composition of the parties involved in such projects often differs from project to project, but also between project phases. The data associated with a construction project has made a development of its own in the past decades, having had a shift from a mostly document centric way of working, towards a more digitized and modelled way. With the emergence of CAD digital drawing and the initial efforts in creating other document-oriented solutions for storing and sharing information around the 1970s, the shift began (Isikdag et al., 2007). This was taken a step further in 1984, when the International Standards Organisation (ISO) commissioned the creation of the ISO standard 10303: *Industrial Automation Systems – Product Data Representation and Exchange*. The creation of this standard introduced one of the first large scale standardization efforts within construction file exchange in 1994. The standardization which came in the form of the STEP file format for file exchange, which makes use of the EXPRESS language for its data definition and specification (ISO, 1994).

Apart from the drawings, other data is also associated with the realization of a construction project, as the project is to be completed within time and scope, while maintaining quality and cost. This can be considered 'construction management ', which first and foremost is the challenge to deliver the project within defined constraints. The second is integration of input and optimization of allocating resources to meet the primary objectives. A project is a set of objectives (people, money, materials, space, energy, communication, provisions, etc.) which needs to be handled properly to meet the need of predefined objectives (Kumar et al., 2019). Before the 1950s this was managed by master builders and creative architects, based on personal experience. The 1950s saw the initial adoption of systematic techniques and tools to complete complex projects. Construction management knows two forefathers of the discipline. One is Henry Gantt, known for his creation of planning and control techniques, most notably the famous Gantt chart. The other is Henri Fayol, who stated that the 5 most important aspects of leading within general management are planning, organizing, commanding, coordinating and controlling. Some of these Fayolian principles, beliefs and views still influence contemporary management theories (Kumar et al., 2019). Based on these principles and views, the 1950's saw the adoption of two mathematical planning models.

1. The "Program Evaluation and Review Technique or PERT", developed by Booz-Allen & Hamilton as part of the United States Navy's (in conjunction with the Lockheed Corporation) Polaris missile submarine program to be able to find the shortest possible project duration.
2. The "Critical Path Method" (CPM) developed in a joint venture by both DuPont Corporation and Remington Rand Corporation for managing plant maintenance projects. This method deals with the critical path, which is a set of consecutive activities which define the project duration, if one of these activities delays, the entire project delays.

Both types of planning models still form the base of many planning practices in construction projects, PERT is usually applied if uncertainty is high, while CPM is applied to more routinely conducted practices. The two models are shown in Figure 3, here it can be seen that in a PERT scheme activities are modelled as a process between milestones. In the CPM an activity is modelled as a node, which possesses a duration.

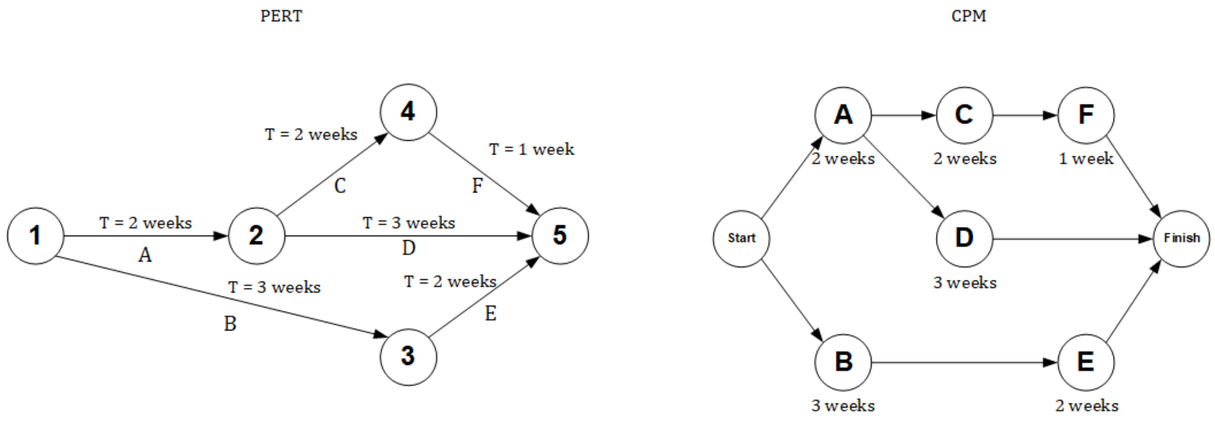


Figure 3: PERT and CPM diagrams

Furthermore, to have better insight in the work to be done, the methodology of the Work Breakdown Structure was developed by the US Department of Defense and NASA in 1963. The Project Management Body of Knowledge (PMBOK) defines the work breakdown structure as 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." (Brotherton, 2008). To further standardize these working methods the Project Management Institute (PMI) was formed in 1969. This institute created the aforementioned PMBOK as a guide to project management standards and guidelines. These methods both create data on the work to be done, either it being in the form of a planning, or the dissection of all tasks related to a project. To be able to efficiently plan and monitor project progress, software packages were developed around the 1980s, with Primavera launching in 1983 and Microsoft Project launching in 1987 (Kumar et al., 2019). The further digitalization of project management data came with the invention of the internet, where several web-based applications are offered. One of these platforms is Relatics, which was launched in 2003. This platform brings systems engineering practices, combined with the aforementioned common practices to the construction sector (Relatics, 2022). Relatics is widely adopted in Dutch construction practices, but is seeing international adoption as well.

2.4 Present solutions to information management complexity

The McKinsey report argues that this technological complexity can be solved by digital innovation, thus applying new practices and methods to the construction industry. To be able to understand the current state-of-art in construction project digitalization, several themes will be further elaborated.

2.4.1 Efforts in standardization of construction data

Efforts have been made to achieve higher levels of standardization within the AEC sector. BuildingSMART has committed to creating and disseminating common, open data standards for the built asset industry. Since 1995 buildingSMART has focused on solving common industry problems by collaboratively developing open, neutral and international digital data sharing solutions and standards (buildingSMART, 2022). One of those efforts in the introduction of the 'Industry Foundation Classes', more commonly referred to as IFC. This standard allows a standardized description of objects in the built environment, including buildings and civil infrastructure. It is an open and international standard, documented in ISO 16739-1, which is meant as a vendor neutral and widely usable standard data model. Within this data model characteristics like material or color are documented, but also its connections and location, geometry and type. It is used as a format for information exchange form one party to another party, while also serving as a archiving format for long-term preservation, as well as operational purposes. The IFC data can be encoded in various supported formats, such as XML, JSON, RDF and most commonly STEP, which is then transmitted over web services, imported or exported in files, or managed in centralized or linked databases. As it is a standard, software vendors of building authoring tools should provide the means to export, import, transmit and view the IFC formatted files (buildingSMART, 2022).

Not only the design documentation is standardized by ISO standards, but also all managerial tasks involved with the construction process have standardizations. The recently revised ISO 21500: 2021 – Project, programme and portfolio management – Context & Concepts, specifies organizational context and underlying concepts for undertaking project, program, and portfolio management. In doing so, ISO 21500:2021 addresses the organizational environment, external environment, strategy implementation, and integrated governance and management approaches. This standard is closely related to the ISO 21502:2020 – Project, Programme And Portfolio Management – Guidance On Project Management, which in turn provides guidelines for project management. It is applicable to any organization, including public, private and charitable, as well as to any type of project, regardless of purpose, delivery approaches, life cycle model used, complexity, size, cost or duration (ISO, 2021). This document provides high-level descriptions of practices that are considered to work well and produce good results within the context of project management (ISO, 2020a). However, both standards refer to best practices within the field of construction management, not to the digital storage of managerial data.

2.4.2 Building Information Modelling

Another one of the proposed solutions to mitigating the complexity of all data associated to a construction project, is the implementation of Building Information Modelling (BIM). Since the inception of BIM, the shift towards model-based working practices has gained huge momentum around the world in recent years and the AEC industry across the globe is undergoing a fundamental transition from conventional paper-based workflows to digitized ones. This process is catalyzed by a growing amount of government initiatives and laws, facilitating a more widespread adoption of the technology (Borrmann et al., 2018).

The industry is undergoing a shift from document based working to model based working, which implies a shift form CAD-based drawing methods to 3D-based modelling with BIM. Overall, the BIM domain entails a set of information technology (IT) tools for generating, managing, and sharing building information among project actors, involving more digital functionalities than three-dimensional (3D) modeling. It thus serves as a tool to increase the interoperability of information in building construction, as it makes use of industry standards with regards to model exchange. It thus has a significant impact on the AEC industry, with specific advantages like cost reduction, documentation efficiency, time reduction, increased information findability, increased employee productivity and an increase in financial control (Mesároš & Mandičák, 2017). The origin of BIM within the construction sector lies in 1962, when Douglas C. Englebart gave his vision on the the way of working of the future architect, in his paper Augmenting Human Intellect. *“the architect next begins to enter a series of specifications and data – a six-inch slab floor, twelve-inch concrete walls eight feet high within the excavation, and so on. When he has finished, the revised scene appears on the screen. A structure is taking shape. He examines it, adjusts it… These lists grow into an ever more detailed, interlinked structure, which represents the maturing thought behind the actual design.”* This idea of working digitally was made more specific by Eastman in 1975, when he wrote his paper on the use of computers instead of drawings in building design (Eastman, 1975).

BIM implementation can happen in multiple ways, which in turn causes several models to assist the implementation of the method within a project. One of the most commonly used frameworks for such an assessment is the BIM Maturity Index. This index is created by Bew & Richards in 2008, showing the different levels of BIM implementation.

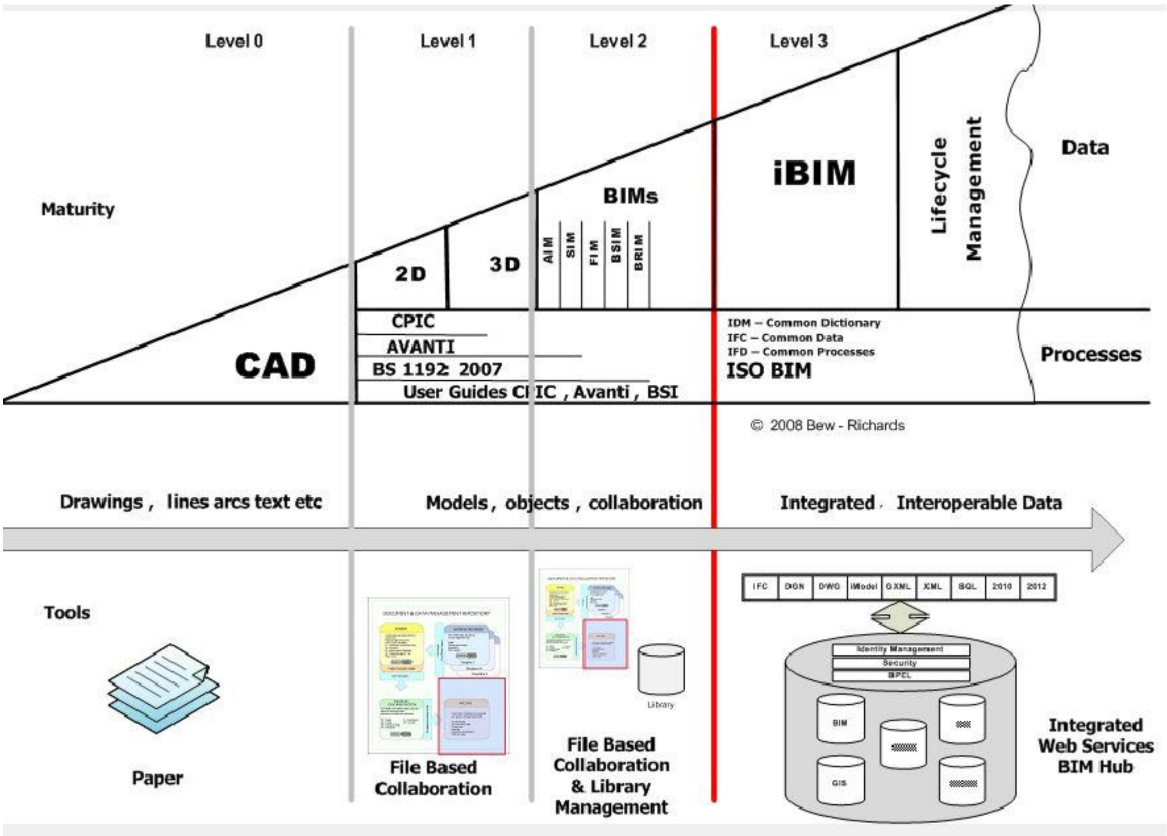


Figure 4: BIM Maturity level model (Bew & Richards, 2008)

These levels will be elaborated separately, with 0 being the least implemented version of BIM, and 3 the deepest and most integrated version of BIM. The levels define the technological progress which a project has achieved, according to the degree of collaboration and information sharing between the different stakeholders involved in a project. BIM level 0 implies a low collaboration, where all of the information involved in the construction project is stored on paper. This limits data to two dimensional drawings and printed data. This data can be considered unmanaged and can only be interpreted by humans, therefore not serving any purpose in interoperability or collaboration. Level 1 of the BIM maturity index introduces file based collaboration, where a construction project is described by two- and three dimensional objects, stored in local but digital files. While this does improve modelling possibilities, the local storage still does not allow for any advantageous collaborative purposes. BIM level 2 introduces the use of file sharing of models created in level 1. Each discipline holding their own federated digital model, which they can combine with models of designers of other disciplines. In this combined model, additional information can be added, which may include construction sequencing, often referred to as BIM 4D, and cost information, referred to as BIM 5D. The highest level, level 3, is achieved when all aforementioned modelling moves towards a cloud based environment, implying that one collaborative server provides access to all project stakeholders. This also implies that instead of different models to be combined, one works on one common model, which is stored in a common data environment. This model is constantly updated, which means one is not providing file based updates anymore, but one is shifted towards object based model updates. In such a system one can say that all objects and information with regards to the project are linked to one another, thus also planning and cost. However, in BIM level 3 one also argues that the incorporation of lifecycle data is needed, and that the model used for construction is transferred to the operational party after completion of construction, to be used for operational modelling. While the ambition to adopt BIM level 3 is large, reality shows that the adoption of BIM level 2 in current construction practices proves to already be rather troublesome. Attrill and Mickovski analyzed the adoption of BIM in UK construction, as the UK government specified the legal obligation for publicly funded projects to make use of collaborative working practices, which are specified as BIM level 2. They argue that adoption is lacking as there is confusion about the specific definition of the levels, since specific requirements are missing. Furthermore, they argue that while BIM level 2 is widespread amongst the industry, it is poorly exploited. This is mostly attributed to a lack of motivation to change, as well as the effort it takes to define and incorporate processes and standards (Attrill & Mickovski, 2020). Current practice for model-based collaboration is documented in international standards such as ISO 19650 – Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM), which relies on the concept of federating disciplinary models in a common data environment (CDE) based on information containers (ISO, 2018). These information containers can be seen as a collection of files, which makes currently implemented mechanisms for model-based collaboration rely merely on file management (Borrmann et al., 2021).

With the adoption of BIM level 2 in the works, the application of BIM practices is becoming more common in the construction sector. However, the introduction of BIM in information management only provides a tool for interoperability of construction models, with the possible inclusion of planning and cost. Next to these models, as argued earlier, other types of information like work breakdown structures or planning of non-construction related tasks are also key to project success.

2.4.3 Project Management Information Systems

Practices like risk management, cost estimation, work breakdown structures, planning and logistics also influence the construction process, while having to rely on accurate and accessible data. These tasks should all be conducted over the course of the construction period, which consists of different phases. A project lifecycle can generally be divided in three main phases; design, execution and maintenance. While each phase tends to be independently managed, information is still shared across phases. Therefore, effective data management is integral to project success (Martínez-Rojas et al., 2015). To facilitate effective data management, information and communication technologies (ICT) play a fundamental role. The use of these tools is becoming increasingly necessary to bolster the competitiveness of construction companies, many efforts are made that focus on the design, development, and practice of techniques for the management of construction information (Benjaoran, 2009). To be able to fulfill said tasks, project management information systems (PMIS) are used. Such systems allow for the structuring of information as well as the sharing of data and documents in a customizable environment. As such systems combine many services into a single solution, they can be regarded as enterprise environments, as they are able to constitute core components of work processes and the digital infrastructure in companies (Rolland & Hanseth, 2021).

With the invention of the internet, not only BIM, but also the PMIS domain gained momentum in its development, as a whole new domain of possible innovations opened up. However, the concept of project management has a far wider scope than the concept of Building Modelling. Present PMIS tools have developed diverging into specific directions, to be able to deal with specific problems. They have developed far beyond their initial purpose as document management system (Chen et al., 2021). This divergence was enlarged by the advent of the smartphone, as well as the gaining of momentum of services like cloud computing and online software packages. Chen et al. argue that while both BIM and PMIS have made substantial innovations over the past decade, the two services offer little direct connection. While being inherently different services, both BIM and PMIS system adoption face familiar issues. Oesterreich and Teuteberg clustered the barriers into four dimensions: structure, people, technology and task. They find that social barriers relating to "people" and "structure" dimensions are the most important barriers such as people's resistance to change, compared to which, the pure technical issues are less critical. This resistance to adopt new technology lies within cost, but also a lack of knowledge, willingness to change and collaborative maturity (Oesterreich & Teuteberg, 2019).

In a systematic review of the current market leading PMIS systems in the construction industry, Chen et al. show this wide diversification of functionalities of the systems provided by the market, as each has followed its own diverging development route. Within the paper it is argued that these functionalities can be ordered according to categories, covering information management, communication and coordination, planning, monitoring and control among others. Although the functionalities of the systems differ, each tries to offer an all-solution package, touching at least upon each of the set categories once. This together with the ever emerging new technologies for system integration make it difficult to make a comparison (Chen et al., 2021).

With the segmentation between BIM and PMIS, it can be argued that they are still considered as two different services, each serving the purpose to their own respective issues. Contrary to BIM, PMIS lacks industry standards in data storage and exchange. Thus, the way in which information is stored in a PMIS environment does not follow the principles of the aforementioned FAIR Data Framework. This discrepancy is twofold, firstly since PMIS facilitates the sharing of documents, which poses challenges in itself. Often documents are not linked, implying poor information management, which is estimated to be the primary cause of time delays in construction (Senthilvel et al., 2020). Information stored in documents is considered to be unstructured, as often the required piece of information is surrounded by a plethora of non-required information. While this might be the most common form of information exchange, it does not contribute to the data integration between the systems, as system operability and data exchange between different actors and companies is only available at the file level, which is mainly based on proprietary formats with poor semantics (Martínez-Rojas et al., 2015). Specifically with regards to document-based systems, Martínez-Rojas argues that when document-based systems are in use, the relevant information is often buried in irrelevant information, as both is stored in the same document. This decreases the findability of the relevant information. Secondly, the set-up of PMIS has a lot of freedom, which is also reflected in the offerings on the market, each having their own respective set of functionalities. This freedom and the inherent malleability of such digital technologies opens up a larger design space for organizations to develop novel solutions and innovate new products and services – even based on previously failed projects and solutions (Arvidsson & Mønsted, 2018). However, such malleability also has its pitfalls, as such freedom could constitute inertia or make it harder to change over time. An extensive literature review on digital transformation by Vial states that inertia is one of the most significant barriers of digital transformation. In the context of digital transformation, lack of organizational capabilities and resources, as well as path dependency across supply chains in specific industries are identified as typical sources of inertia. (Vial, 2019). This idea is backed by Rolland and Hanseth, who also state that digital technologies are not merely enabling for digital transformation processes, but also a source of inertia, especially due to a knowledge dept and switching costs (Rolland & Hanseth, 2021). A shared understanding of interoperability and communication would greatly improve the handling of complex construction projects (Chen et al., 2021).

While PMIS potentially can solve many of the aspects of project management, often additional services are used for project completion. As apart from the sharing of general knowledge and files within a project, one also needs to stay involved in the ongoing processes during project execution. To be able to communicate clearly about project requirements, verifications, tests, deviations, risks and other tasks associated with the project, a structured framework is needed. Within construction, systems engineering has become paramount to project success. Systems Engineering is an engineering discipline whose responsibility is creating and executing an interdisciplinary process to ensure that the customer and stakeholder's needs are satisfied in a high quality, trustworthy, cost efficient and schedule compliant manner throughout a system's entire life cycle (Hala et al., 2020). The concept of systems engineering is not specific to the construction sector, yet it is considered as the standard way of working, and a common language being used both by contractor and client. The Dutch Ministry of Infrastructure and Water management is using the framework and intentions of Systems Engineering as a guideline for their projects (Rijkswaterstaat, 2022).

Another important system to ensure project success is the planning tool. Project planning presents a wide variety of decision problems, which can be classified in four categories: project representation, project scheduling, resource allocation and risk analysis (Pellerin & Perrier, 2019). The project representation refers to the aforementioned Work Breakdown Structure (WBS). Project scheduling evidently is one of the most important aspects of the planning process, as it considers the networking of activity sequences and durations, simultaneously with resource trade-offs, often based on assumptions made in the early stage stages of the planning process. As the practice of project scheduling has to deal with uncertainty and lack of knowledge throughout the initial planning process, one has to take measures to still be able to ensure reliability of the planning forecasts. This can be done by incorporating buffering strategies to allow the planning to be less sensitive to disruptions happening during construction. As the level of uncertainty is high, resource allocation also needs to take this into account. For this, two approaches exist, reactive and proactive resource allocation. Reactive methods allocate all available resources already during planning, which requires reallocating if disruptions occur. Proactive methods, in contrast, focus at incorporating a safety buffer in the allocation of resources, to be able to safely encounter variations in the execution phase (Pellerin & Perrier, 2019). Finally, one also has to deal with risk assessment. It is considered a crucial step during project planning, as it analyses the effect of deviating from planned activities. It assesses the risk itself, assesses the weight of the risk, possibilities of mitigation and selection of proper response to the risk. There is no clear systematic methodology of risk assessment, therefore it is heavily reliant on lessons learned from previous similar projects.

While a large number of project management software packages have been developed to assist the project manager, only a limited number of project planning and control software packages is commonly cited in experimental and comparative studies, which are Microsoft Project and Oracle's Primavera (Pellerin & Perrier, 2019). While often being labeled as project management tools, they are in practice planning and control information systems, which cannot be compared to other PMIS, which support a wider variety of project management functions. The tools are being extended to be integrated in BIM 4D and 5D practices, respectively incorporating planning and cost into construction modelling. However, as Pellerin and Perrier also argue, integration with other disciplines like quality, sustainability and safety are interesting for future research, as there are only a few papers actually addressing this.

With these findings a striking difference can be found within the development of information systems for the AEC sector. The BIM development initiated not only standards, but also a framework for further innovation, in the form of the BIM maturity index. This guideline steers towards Common Data Environments where all standardized information of the modelling process is gathered, combined and presented to provide additional insight in the modelling process. Furthermore, it has the potential to be enriched with additional information of planning and cost. This is made possible due to efforts in standardization, and subsequently legislation, allowing the adoption of the standards by software providers. Contrary to this development, services attached to a building process outside of the design and modelling, lack internal data standardization. While IFC does allow the integration of practices like planning and cost into construction modelling, it cannot be used solely for multi-source integration within the project management information scope. PMIS, Systems Engineering and Planning packages all have their own respective standards, but these standards solely deal with a common understanding of the practices, not necessarily with the data that is produced in the practicing of these respective disciplines of construction management. This allows the clear distinction of two disciplines in the project management information landscape, construction modelling, focused on the object, and project management information, focused on the work to be done. For construction management practices to also increase in their data integration capabilities, one can argue that an additional top layer system and standard is required to fully be able to combine all information.

2.5 Common Data Environment

If one wants to make use of their data for their decision making process, throughout the project lifecycle, one should arrange appropriate data storage, making data organized and exploitable (Martínez-Rojas et al., 2015). To ensure this goal, while making use of the FAIR Data Principles, a proposed solution is the Common Data Environment (CDE). This is a virtual storage location for collecting and managing documentation of building projects, mostly offered as a cloud service. Because all project information is managed in this common environment, the chance of misunderstanding and information loss is strongly reduced (J. Werbrouck et al., 2019).

2.5.1 ISO 19650 implications on Common Data Environments

In ISO 19650 – Organization and digitization of information about buildings and civil engineering works, including building information modelling, the term CDE is given standardized meaning, which is part of a solution in providing project participants with a delivery of right information, at the right time. Within the ISO 19650, several stages of the

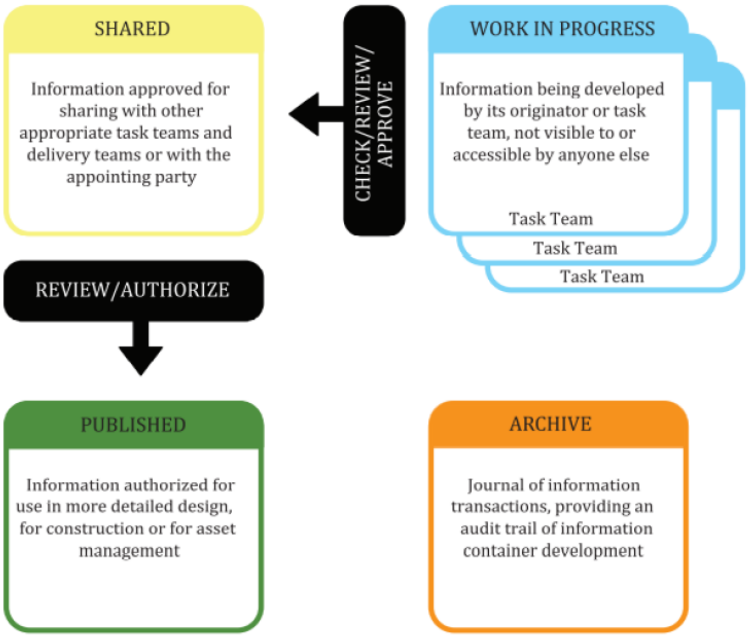
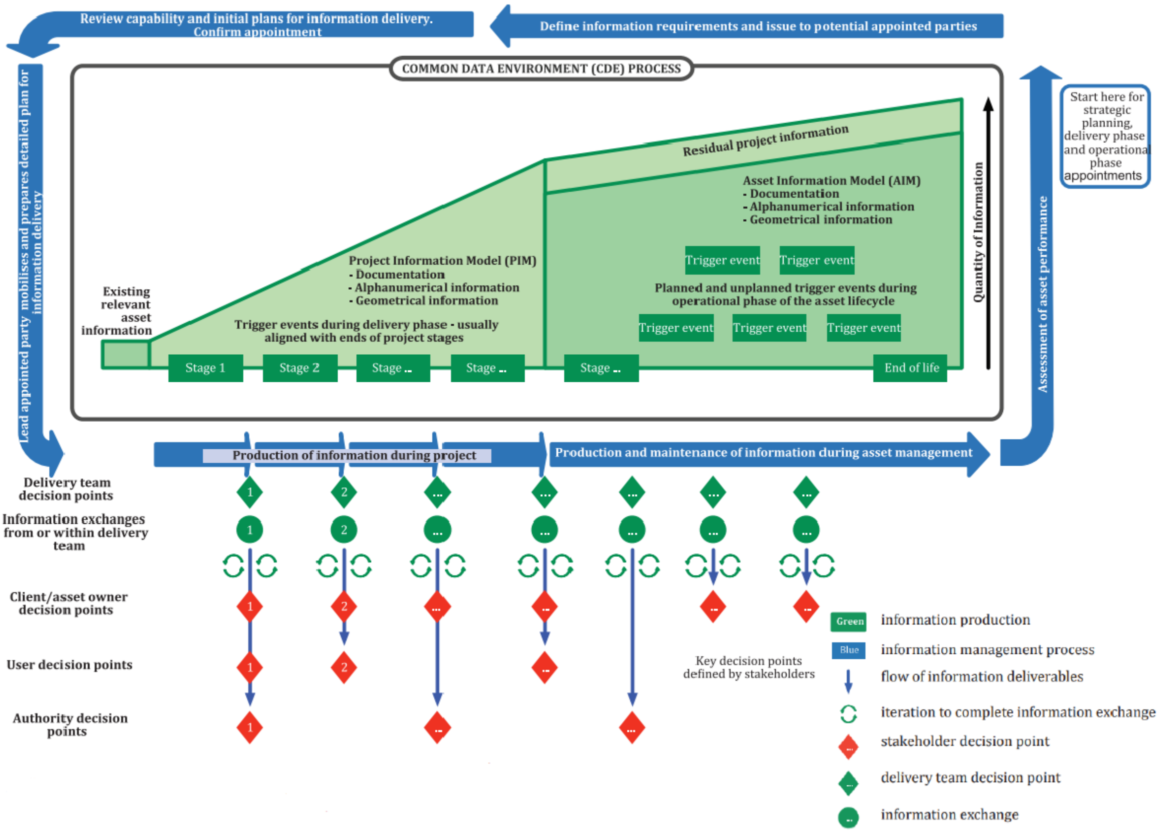

SHARED
Information approved for sharing with other appropriate task teams and delivery teams or with the appointing party
CHECK/REVIEW/ APPROVE
WORK IN PROGRESS
Information being developed by its originator or task team, not visible to or accessible by anyone else
Task Team
Task Team
Task Team
REVIEW/AUTHORIZE
PUBLISHED
Information authorized for use in more detailed design, for construction or for asset management
ARCHIVE
Journal of information transactions, providing an audit trail of information container development

Figure 5: CDE data statuses (ISO, 2019)

construction lifecycle are discussed. Each stage has its own respective data flows. These data flows are visualized in Figure 5, where these four different stages of data are discussed.

It is important to highlight that a CDE intents to work with information containers, which are different from files or documents. It is intended that these information containers are standardized pieces of data, that can be linked to data coming from other systems. This data moves through 4 stages, Work in Progress, Shared, Published and Archived. Transferring between these statuses is possible, after a check or review of the data. How this translates to a project wide application, is showed in Figure 6.



Review capability and initial plans for information delivery. Confirm appointment
Define information requirements and issue to potential appointed parties
COMMON DATA ENVIRONMENT (CDE) PROCESS
Lead appointed party mobilises and prepares detailed plan for information delivery
Start here for strategic planning, delivery phase and operational phase appointments
Assessment of asset performance
Residual project information
Asset Information Model (AIM)
- Documentation
- Alphanumerical information
- Geometrical information
Quantity of information
Project Information Model (PIM)
- Documentation
- Alphanumerical information
- Geometrical information
Existing relevant asset information
Trigger events during delivery phase - usually aligned with ends of project stages
Trigger event
Trigger event
Planned and unplanned trigger events during operational phase of the asset lifecycle
Trigger event
Trigger event
Trigger event
Stage 1
Stage 2
Stage ..
Stage ..
Stage ..
End of life
Production of information during project
Production and maintenance of information during asset management
Delivery team decision points
Information exchanges from or within delivery team
Client/asset owner decision points
User decision points
Authority decision points
1
2
...
Key decision points defined by stakeholders
Green information production
Blue information management process
flow of information deliverables
iteration to complete information exchange
stakeholder decision point
delivery team decision point
information exchange

Figure 6: Common Data Environment process according to ISO 19650 (ISO,2019)

Within such an application, information flows originate from different delivery teams. Here it is shown that the interaction with the project itself is shown as iterative, where the input has to meet the demand. Within the model, all of this information contributes to the creation of a singular Project Information Model (PIM). This model is subject to predefined Exchange Information Requirements (EIR), which contains the information standards and production methods and procedures for how project teams should deliver data.

This information landscape is to be created by the lead party in the construction project. This party also has to bear the responsibility of the effective management of the information, which includes the creation of the aforementioned data standards, and the validation if stakeholders adhere to the set standards. In this process, one thus has to deal with the integration of many pieces of information, each created by different delivery teams. This shows the required scale of correct data management, if a CDE is to be applied correctly. While the framework proposed by ISO 19650 is clear, it merely serves as a guideline on practices. It does not propose specific tooling to facilitate its proposed framework.

2.5.2 Data integration platform

To initiate such a CDE, different methods can be applied. Cloud based platforms offered as complete software packages also present themselves as CDE systems, of which Autodesk BIM 360 and Trimble Connect are among the largest in the market for BIM oriented CDE's. However, as argued by Simeone et al. it is often the case that currently used versions of CDE systems instead of making clever use of data, act more like shared document repositories, with their full potential being left unexploited. Also, it is argued that beyond the BIM modelling, other sources of managerial data are left out, while still remaining relevant to the construction project (Simeone et al., 2020). They propose to apply a graph-based methodology to improve coherence, reliability and accessibility of information stored and shared, as well as improving quality of the integration of models, documents and other information carriers that contribute to the central core of data acting as a reference. Thereby such an implementation focusses on the addition of an extra information level compared to current CDEs. Graph methodology is closely related to the 'semantic web' framework. Tim Berners-Lee coined the term 'semantic web' in 2001, which offers technology that contributes towards solving the noted interoperability problem at all three of the layers: physical, syntactic, and semantic interoperability (Berners-Lee et al., 2001).

As described in the ISO 19650, projects possess a wide variety of semantically rich data, which should be standardized to be used effectively in a CDE. The combination of non-complementary data formats which are semantically rich, is where graph methodology shows potential, as it prioritizes the relationship between parts of data. These relationships can be intuitively visualized using graph databases, making them useful for heavily inter-connected data (Yoon et al., 2017).

2.6 Graph technology

The graph databases refer to a storage system technology making use of graph structures, with nodes and edges, used to represent and store data. It is based around the principle of relationships between objects, which makes them specifically focused on the processing of highly connected data and flexibility in the usage of data models (Pokorný, 2015). Pokorný distinguished 4 basic components of graph database technology: graph storage, graph querying, scalability and transaction processing.

Due to the structuring of the data based on relationships, graph databases possess a property called *index-free adjacency,* implying that each node is related to its neighbor node. Thus each nodes serves as an index for its adjacent nodes, which is more efficient than having global indexing tables, which have to be fully traversed when querying. A common example of a large scale application of a graph database is the social networking site Facebook, which makes use of graphs to store their massive amounts of data. These types of 'Big Graphs' have over 1 billion nodes, and 140 billion edges. It is relevant to use a graph if the relationship between elements matters, contrary to other database types, where information is stored in a tabular format, which focusses less on the relationships. The data storage of a graph database is often referred to a NoSQL, which contrary to SQL does not make use of a tabular structure. Graph technology includes two different types, the labelled property graph and the RDF-based graph. The labelled property graph (LPG) employs graph methodology in either a local or centrally stored database, while the RDF-based graph employs the tools of the semantic stack to store large parts of data online.

This difference between SQL and graph applications is also apparent in the querying of data. SQL has a standardized and structured query language, while graph applications do not have 1 uniform language, as there are multiple types of graphs. While query languages differ due to this lack of standardization, the setup is similar. The languages make use of the index-free adjacency, which results in the search for a node, and its adjacent nodes, which are then filtered based on specified properties or its identifier, this is called point querying. One of the most common graph query languages is Cypher, provided by Neo4j, which is loosely based on the SQL syntax (Pokorný, 2015). SPARQL is another graph query language, which is the standardized query language for RDF-based graphs. The output of a graph query can both be a table or another graph, which is a simplification or transformation of the existing graph data. This implies that the querying of a graph is heavily correlated to graph visualization.

To be able to accommodate for the large amounts of data, it can be necessary to partition a graph, yet distributing a graph across multiple machines can be rather difficult in the case of an LPG. It should be avoided to have relationships spanning multiple machines, this is called the *minimum point-cut problem.* For most data, the number of links or relationships is too large to efficiently compute an optimal partition; therefore most databases use random partitioning (Lluís Larriba-Pey et al., 2014). While this seems problematic, reality shows that current systems can handle such large quantities of data, that only in the case of Big Graphs one should take action in partitioning (Pokorný, 2015). In the case of an RDF-based graph, the online storage mitigates this issue.

According to Pokorný there are three generic use cases for graphs: Create, read, update, delete (CRUD) applications, query processing and batch mode analytics or data discovery. Often a graph application is optimized for one of such purposes.

As there are many graph databases on offer, it is important to highlight their inherent differences. For the application of the semantic web, the Resource Description Framework (RDF) is the standard model, as defined by W3C. However, recently the labeled property graph (LPG) has seen a rise of attention. This model has been used in multiple AEC-related use-cases and has shown that typical characteristics of LPGs could be beneficial to linked data models for the AEC sector, such as using relationship labels, fast and easy graph search, scalability and performing complex graph algorithms (Donkers et al., 2020). These two types of graph databases will be elaborated separately.

2.6.1 RDF-based Graph

A way for a CDE to ensure a high level of data integration is the correct application of linked data technologies in a RDF-based graph, with its corresponding data handling processes. This technology relies on the use of the Resource Description Framework (RDF) which is a data model that has been standardized as part of the Semantic Web technology stack (Berners-Lee et al., 2001). RDF allows individual data objects to be linked to one another, with the application of a relation, a concept which is defined as a 'triple'. These triples connect an object to a subject with a predicate, stating the exact relationship between the two. Each subject and property is uniquely identified by a Uniform Resource Identifier (URI) which ensures that the data can be identified on the web. Objects can either be identified by a URI or by a literal. In the case of the last, the object is a string which cannot be uniquely referenced. This data representation comes in the form of a web-wide graph, of which digital agents are semantically capable of interpreting this data, and using it for specific purposes with minimal human intervention (J. Werbrouck et al., 2019). RDF-triples can be noted in different syntaxes, of which the most common are N-triples, XML, JSON-LD and Turtle.

For these triples to be machine readable and interpretable in a linked data application, the data elements have to obtain their semantic meaning by schemas called 'ontologies' or 'vocabularies'. These schemas define the type of a data object, and how it should be structured. They are a collection of classes and relations between those classes, that can be used to describe data. This is particularly important as the meaning of each concept has to be defined beforehand, in a structured manner with a shared understanding of the concepts, for it to be machine readable. These schemas comprise of classes, attributes and relations, each named and valued. If one adds data, this data should become an instance of a predefined class. To be able to create such ontologies, the RDF schema (RDFS) and Web Ontology Language (OWL) are used. OWL has the broader set of functionalities, as contrary to RDFS it supports subclasses, disjunctive relations, cardinality constraints, transitivity, uniqueness and invasiveness.

SPARQL is adopted as the standard query language for RDF, based on triples patterns. The query language retrieves information from a graph based on pattern matching. SPARQL is designed and endorsed by W3C, not only as a query language, but also as a HTTP-based transport protocol, allowing it to access all web-based SPARQL endpoints.

Related to RDF-based graphs, it is important to highlight the Information Container for Document Delivery (ICDD), which applies to creating a container structure for project information delivery. ICDD is a final stage of standardization of ISO 21597, which has been developed '*in response to the need of the construction industry to handle multiple documents as one information delivery or data drop.*' (J. Werbrouck et al., 2019). The ISO 21597 comprises of a document, which defines an open and stable container format to exchange files of a heterogeneous nature to deliver, store and archive documents that describe an asset throughout its entire lifecycle (ISO, 2020b). The standard makes it suitable for all parties dealing with information concerning the built environment, where there is a need to exchange multiple documents related to one another, to be able to share them with such relations attached. It is possible to refer to both documents added in the container, as well as to external web references. This container is created based on two ontologies, the container ontology, defining the description of metadata of the documents, and the linkset ontology, defining the semantic links between documents. This method is developed to make the application of linked data more suitable for document management.

2.6.2 Labelled Property Graph CDE

Contrary to the RDF-based graph, the labelled property graph does not natively make use of ontologies for its storage functionalities. Another notable difference is the ability of labeled property graphs to be able to carry properties directly within their nodes and relationships. While a RDF-based graph has to add another node with either an URI or a literal, depicting the property, the labeled property graph is able to add this property directly into the relationship or object. LPG's thus don't make use of URI's. If one compares the visual representation of such models, this causes a more comprehensible view, as not all information is shown as separate nodes, only the ones intended to be shown as nodes. The internal structure of nodes and relationships is described by key-value pairs. Labeled property graphs are very node-centric, which is contrary to the more edge-centric RDF-based triples (Donkers et al., 2020).

The most commonly used LPG graph database is the native graph database Neo4j. This database makes use of the Cypher query language. As a labeled property graph does not make use of specific schema's, prefixes are not used in the query language. Cypher merely makes use of a combination of Cypher keywords. Neo4j has committed to the effort of the creation of a new standardized query language for property graph systems, together with ISO/IEC's Joint Technical Committee 1, which assesses IT standards. This committee voted positively in June 2019 for the development of a new standardized query language. This language, called GQL, draws heavily from existing query languages, mainly being inspired by Neo4j's Cypher, Oracle's PGQL and SQL itself . The language is still in development, yet it provides great perspectives in standardization of property graphs in the near future, and is stated to release a standard in April 2024 (ISO/IEC JTC1 SC32 Working Group, 2022). Contrary to SPARQL, which is standardized and developed by the W3C organization, the conception of GQL is more commercially oriented. GQL was first proposed by Neo4j to other software vendors in July 2016, only later to be taken into consideration by the ISO/IEC JTC1.

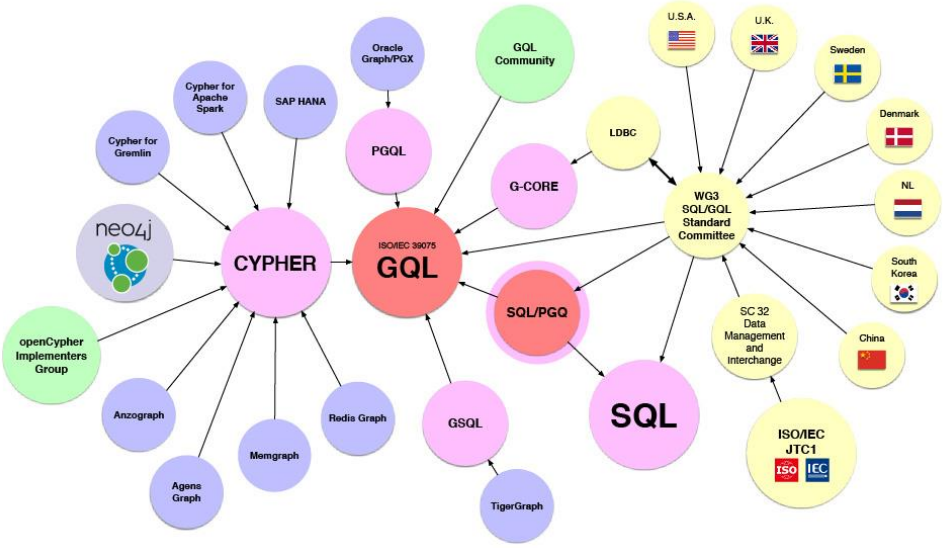


Figure 7: GQL positioning of query language (ISO/IEC JTC1 SC32 Working Group, 2022)

Labeled property graphs have been developed with the purpose of using the data, storing data and querying data as efficiently as possible (Donkers et al., 2020). The structure of data in a labeled property graph is considerably more simple than the structure of the same dataset in an RDF-based graph. Donkers et al. argue that due to this difference in graph density, when datasets increase in size, a labeled property graph performs faster compared to a RDF-based graph. However, it should be mentioned that graph structure does also have an effect on speed.

2.6.3 Other graph applications in the AEC sector

The literature study has also provided additional insights in the application of graph technology in the AEC sector, with other use cases being proposed, developed and presented. These use cases highlight the further potential of graph technology aiding in AEC data integration, but also with regards to modelling and monitoring.

As argued before, the Common Data Environment is one of the proposals in which graph databases can fulfill a useful role. These CDEs are proposed as central environments for the storage and exchange of construction data during design and construction of a project. As the built environment consists of a wide share of built assets, with a wide variety of asset properties, each with their own 3D shape, standardization to some extent is needed. This standardization comes in the form of the ifcOWL ontology, which is a reinterpretation of native IFC models to RDF, which allows building models to be queried with the use of SPARQL (Krijnen & Beetz, 2018). This is taken one step further with the Linked Building Data ecosystem, in which focus has shifted from the development of one big building-oriented ontology, towards smaller modular ontologies. Cornerstone of this development is the Building Topology Ontology (BOT) which contains the definitions to describe topological relationships within a building. Other ontologies can then be used to further enrich the data, depending on the use case (Werbrouck et al., 2022).

This development corresponds with the application of graphs in the creation of digital twins. A digital twin allows the virtual representation of physical asset's condition as data, which can be integrated bi-directionally at any point in time (Ozturk, 2021). It thus is able to provide insight in the current state of a physical asset, through a constantly updated virtual copy. If one is in possession of a digital twin, stored in a graph-based environment, one is able to connect real-time sensor data to such a model. This offers significant benefits for the buildings lifecycle management, as the graph-based model allows to overcome barriers of data integration (Farghaly et al., 2019). This allows the integration of data like room bookings, human resource data, but also building energy performance. The ability to link such datasets and act upon the knowledge gained allows for, for example, more efficient energy use and heating of a building. Concepts like preventive maintenance based on building use and smart lighting are other use cases proposed within the research subject of lifecycle assessment, also coming back in the ambition to achieve BIM level 3 on the BIM maturity scale of Bew & Richards, which also expresses the desire to incorporate lifecycle management in BIM adoption. However to achieve BIM level 3, several challenges also have to be overcome, for which Borrmann et al. also propose the application of graph technology. As the application of this technology supposedly would aid in the shift from file-level updates to object-level updates. This would greatly advance the capabilities of a CDE, as only the altered elements would be updated, through the use of update patches, instead of uploading a completely new full model (Borrmann et al., 2021).

2.7 Literature conclusion

Throughout the course of time, from the first notions of digital innovation aiding the construction process of buildings by Eastman and Englebart, towards the aim to apply new technologies to approach BIM Maturity level 3, the information landscape around construction projects has seen quite a development.

Within the literature review, the key developments of two disciplines within the management of a construction process are discussed, the first being the modeling of the to-be-built asset. The second is the management of all construction management-related tasks. Between those two disciplines a discrepancy arises. While the modelling element of construction management has seen the introduction of Building Information Modelling, which not only introduces a standardized file exchange format for 3D representations of an asset, but also introduces a roadmap towards common, yet decentralized, data storage. It does so through the use of enterprise-like systems, including a shift from file-based information container to free flowing data from system to system. This is further catalyzed by the initiation of multiple standards, like ISO 16739-1 and ISO 19650, documenting both naming conventions, as well as data standards. The introduction of those standards also motivates governments to implement policies in BIM inclusion in building practices.

Contrary to the development of BIM for modelling, the development of an enterprise solution to all construction project management related data is very limited. IFC offers the integration of several project management aspects like cost and planning to be integrated in construction modelling, but cannot act as a standard for integration of multiple sources of project management information. While standards on naming conventions do increase the common understanding of construction management, standards on data exchange formats are missing. This also causes innovation on the creation of such enterprise platforms to be lagging. As there are no data exchange standards, it thus is the challenge to create another standard way of integrating project management data. With the creation of this standard, the principles of the FAIR framework should be applied. In the literature it is argued why graph technology has potential in aiding this search for a data integration methodology. How the technology can aid, should thus be researched further.

3. Methodology

Within this chapter, it is discussed which steps are taken to answer the research question on how to ensure project management information findability and accessibility, independent of project scale. In the literature review it was identified that within the information of a construction project, one can find many different forms of data. This data can be separated in two categories, construction modelling and project management information. While within the modelling discipline of the AEC, advancements have been made on data integration through the use of data standards, like IFC, such advancements are lacking for project management information. Due to a lack of data exchange standards, innovation on an enterprise platform actually capable of data integration is lacking. As concluded in the literature review, a standard method on the integration of project management data is to be created, to ensure the requested findable and accessible project management information. For such a method, the application of graph technology has been identified as a potential aiding technology, as it allows for the combination of different datatypes.

To do so, several steps have to be taken. These steps will be individually elaborated in this chapter. To be able to create a standardized working method, one first has to get a more detailed sense the information management processes present in the business. To identify such processes the contextual design method is used. Within this method, expert opinion is used to identify processes currently struggling, due to an inefficient information landscape. The most relevant process is identified, and used as a case for the further development process, with eventually proposed an user interaction environment, with a standard way of working.

This standard way of working will be used as the case for the prototyping of a working tool. In the implementation this development process is discussed. This working tool will use graph technology to facilitate data integration. Within this section the different implementations of 3 graph applications is discussed, of which one is used to create the prototype. This concludes with the prototype creation, based on the input gather in the contextual design method.

3.1 Methodological approach and research design

In light of the goal of this thesis, it is important to find out which information management processes the AEC sector is using, in terms of data storage and data visualization. If one is able to understand such processes, one would get a good overview of the subject of applied information management, and thus be able to start discussion on the current state-of-art, potential improvements and the human factor. As the thesis deals with findability and accessibility, one cannot neglect the human aspect of technology interaction. As this constitutes the set goal of the thesis, a methodology which concerns the end-user is particularly interesting. A suitable approach to apply a human-centered development approach comes in the form of the contextual design process, described by Holtzblatt & Beyer. *“Contextual Design is a structured, well-defined user-centered design process that provides methods to collect data about users in the field, interpret and consolidate that data in a structured way, use the data to create and prototype product and service concepts, and iteratively test and refine those concepts with users.”* (Holtzblatt & Beyer, 2014). This identification of business process, and the creation of a corresponding way of working, including a user interaction environment, can be found in Section 3.3.

The findings of the contextual design method are used to create a prototype tool. This process is described in Section 3.4. To be able to provide the process of contextual design with a working backend environment, the OSSpal framework will be used. This framework provides a structured review of openly accessible systems for business application (Wasserman et al., 2017). While it also provides a grading systematic, this will solely be used to describe the inherent differences of the system, not to compare and select the better system.

Of the applications assessed, one will be chosen based on gathered business criteria. Within this application the project management data will be loaded. This ETL procedure is described in Section 3.4.2 The development of the visualization is described in Section 3.4.3. The validation of the prototype is described in Section 3.4.4.

3.2 Literature Review

A literature review is conducted to provide insight in the development and current practices within the field of project management, and its related data within the AEC sector. The study provides a structured review on both existing practices, as well as the application of graph technology, as this has been highlighted as a technology with a high potential. The literature is conducted according to the PRISMA 2020 methodology. This methodology provides a framework allowing the identification of future research priorities (Page et al., 2021), making it suitable for the verification of the described potential of graph technology application, as well as elaborating on the current state-of-art of project management practices, data associated to construction projects and information systems innovations. Further literature study is conducted based on the waterfall principle, which makes use of papers identified as references in papers found through the application of the PRISMA method. The literature review can be found in chapter 2.

3.3 Contextual Design

The contextual design process consists of two phases, the first being the ‘Requirements & Solutions’ phase, the second being the ‘Define & Validate concepts’ phase. Figure 8 shows the individual steps involved on the process, spread over the two phases. Of each phase an assessment is made if all steps will be required to be taken, as the phases tend to have some overlapping features. As the end-users participating in this research have the Dutch Nationality, all interviewing sessions will be conducted in Dutch. The findings will be

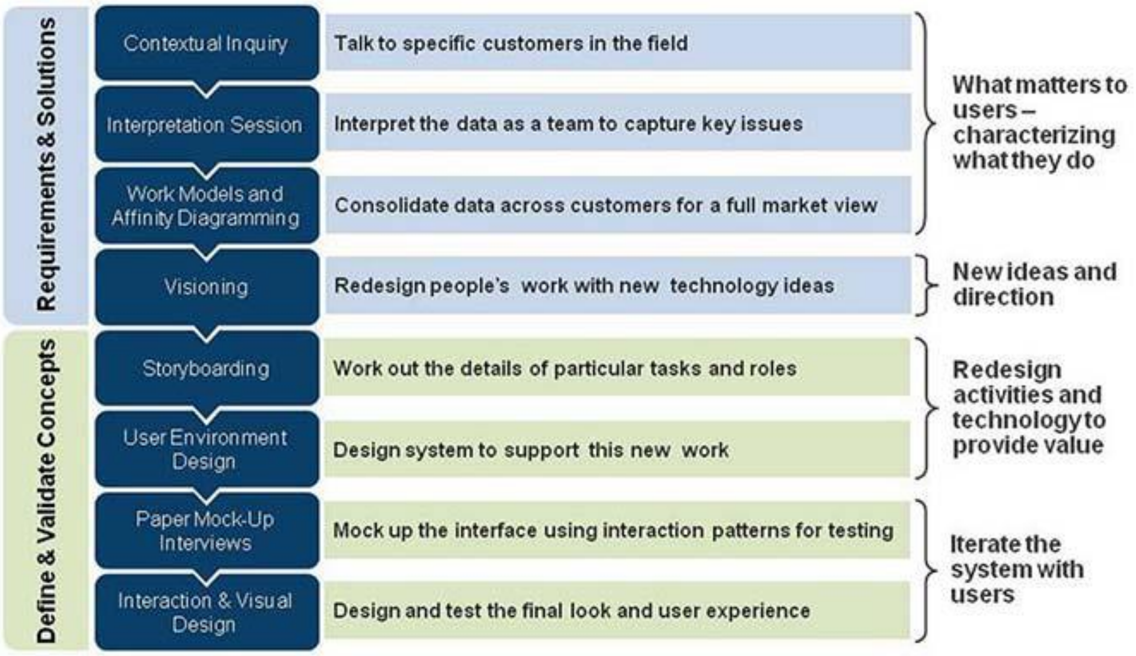

Requirements & Solutions
Contextual Inquiry
Talk to specific customers in the field
Interpretation Session
Interpret the data as a team to capture key issues
Work Models and Affinity Diagramming
Consolidate data across customers for a full market view
What matters to users – characterizing what they do
Visioning
Redesign people's work with new technology ideas
New ideas and direction
Define & Validate Concepts
Storyboarding
Work out the details of particular tasks and roles
User Environment Design
Design system to support this new work
Redesign activities and technology to provide value
Paper Mock-Up Interviews
Mock up the interface using interaction patterns for testing
Interaction & Visual Design
Design and test the final look and user experience
Iterate the system with users

Figure 8: Stages of Contextual Design Process (Holtzblatt & Beyer, 2014)

interpreted and translated to English to be used in the thesis. Original Dutch input will be added to the appendix.

3.3.1 Contextual Inquiry

The goal of the contextual inquiry is to understand the end-user, who directly interact with the proposed solution, the indirect user, whose information provided helps the end-user finishing their tasks, the manager of both users, responsible for workflow supervision, and thus project success. The end-user gets priority in the contextual inquiry, as the actions of the end-user will define the success of the solution design, however, other should not be neglected.

The contextual inquiry is an explicit step in understanding who the user actually is and how she works on a day-to-day basis. The difficulty here lies with the fact that work tends to become rather habitual for the end-user. Therefore traditional forms of interviewing are set aside for observation and inquiry about the actions of the user as they unfold. By doing so, one is able to understand the motivation and strategy of certain actions. Then, the interviewer and user, through discussion, decide on a shared understanding of the work.

Due to the stated research problem, which argues that an increase of project scale makes it more difficult to find and access relevant project information, it is decided to focus on two projects, one of a large scale, with 400-500 people working on a daily basis, and one of a smaller scale, with 40-50 people working. The first step then is to identify the end-users within the project or enterprise. In the case of this research, the focus will lie on two process coordinators. Each of them involved in their own respective infrastructural construction project. It is these people who are held responsible for the quality, consistency and continuation of the project. This person coordinates the starting and finishing of construction activities, and is often asked to share knowledge on progress with regards to construction activities, together with the according documentation. While other stakeholders work together closely with the process managers, it is decided to choose the process manager as the main end-user, as others often have a dependency on them, instead of vice-versa.

To be able to conduct the contextual inquiry, a session is set up with the two process managers. It is intended during this session to gain insight in the way of working of both experts. Since both of them work on their own respective project, each with a different scale, it is expected that their way-of-working is significantly different, while the questions and provided answers can be of the same kind. To be able to offer a way of working which works for both of the experts, the aim is to isolate a process which is present in both projects. This will be used as a use case for the development of a new tool. The goal is to identify a process which becomes increasingly problematic once the project scale increases. This case should be made very specific, as it will serve as the proof of concept. The session potentially yields other cases which could be interesting for further research.

The experts both have the function of process coordinator. This implies that they are responsible for maintaining project quality, as well as keeping track if the project remains within schedule. In both projects, an information structure of Relatics, Primavera and Sharepoint is used to keep track of planning, work and progress. Observation of the work showed that the process coordinator often is bombarded with questions about certain sets of information present in one of these three systems, but even more often about combined sets of information, requiring the combination of systems and its respective data. It is shown that these requests tend to cause a lot of time waste, and the desire for a more efficient way to gain cross-platform knowledge is desired. Therefore the research is focused on the process of answering to an information request.

To be able to discuss the potential cases in a structured manner, an interview session was organized. This session consisted of 3 parts, an introduction, a brainstorm and a discussion. Within the introduction the research problem was briefly discussed, together with the possibilities of applying graph technology. During the brainstorm, a poster, which can be found in appendix 1, was used to structure the information. Here the experts were allowed to put as many ideas on the board as they could think of.

To facilitate a structured brainstorm session, a poster was created as a guideline for the requested information. This poster consists out of 4 elements, a column for input, process and output, as well as a requirements box. Within the poster the information request is abstracted to a 3-step process, which is input-process-output. These elements will be elaborated individually.

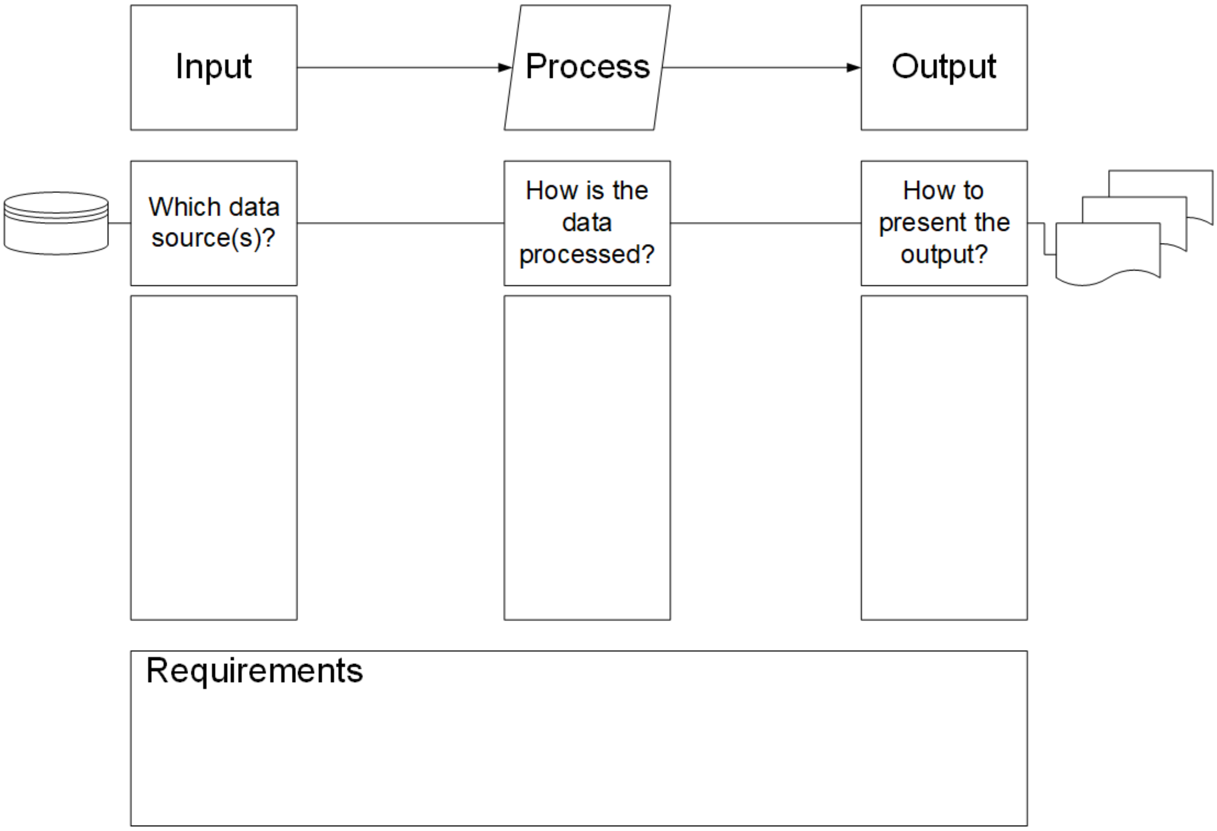


Figure 9: Interview poster format for contextual inquiry

- Input
    - Each information request has a data source. As data can be stored in many different ways, it is important to know how specific data is stored for particular information requests. Relatics, Primavera and Sharepoint are the main data sources, but both projects also use other secondary systems to keep track of their progress.
- Process
    - To be able to answer a question, sometimes data should be processed in a certain way. This could imply for example the cleaning of a dataset, or the combination of multiple datasets.
- Output
    - The output of the information request essentially encompasses the answer to the initial request itself. However, there are many ways to represent such an answer, especially if it is a recurring request. Examples of possible

presentations are generated reports, dashboards or web pages. Initially, the participants will be tasked to think about possible outputs of information requests, as these outcomes are the most important to the business. After thinking of these outputs, they are challenged to reflect on the required inputs and processes to achieve such output.

- Requirements
    - To be able to facilitate the selection criteria for the 'Functionality' assessment of the OSSpal method, the experts are consulted for their functionality requirements in using the graph database environment. These requirements discuss the functionality of the system itself, and how it should and could be used.

3.3.2 Interpretation session

During the interpretation session, which is done with a collection of people with different functions, these insights are reviewed in a discussion. This session lets everyone bring their own unique perspective to the collected data, sharing findings on process, underlying technology and preferred interaction implications. Throughout such a discussion, a shared understanding is achieved of the day-to-day working, and the corresponding process of the process coordinators. With this shared understanding present, the experts were tasked to identify common information requests, present in both projects. With the aid of the poster many different requests will be identified. These requests will then be filtered on the base of being present in both projects, and then prioritized. This eventually leaves one process, which will serve as the case for the proof of concept developed in the conceptual design process. This isolated process will form the subject of the change making process, which will return in the work models and visioning.

3.3.3 Work models and affinity diagramming

Visual representation of how current work processes are done is a great tool to create a summary of actions performed by different stakeholders not only on individual level but also on interdisciplinary level. To be able to help guide the development of a standard way-of-working, it is important to gain insight in the different perspectives one has on the intentions and motivation of a process, to prevent misconception on the purpose of the said process. There are different ways to show these various perspectives. As described by Holtzblatt and Beyer, there are 5 different contextual work models.

The first is the Flow model, which emphasizes the relationship between different actors, both formal and informal. Next to this, the relationships between actors is described, with their role division and interaction captured in the workflow. The flow model is intended to show how work is divided into these roles and responsibilities.

Second is the cultural model, which captures the culture, but also the constraints to which work has to adhere. It describes how people involved in the work work around those constraints to make sure that the work is done. In the case of the AEC sector this can be a broad scope, as one both has to deal with company policy, as well as a wide set of standards, regulations and project requirements. The model allows the highlighting of possible conflicts in the development of a new working standard.

Third, they describe the sequence model. This model describes the detailed steps one has to take to accomplish tasks involved in the set work. The steps are subjective to each person involved in the work, as each deploys their own personal strategy, with its respective intents and goals. It focusses on the end user, providing a more in-depth understanding of what they do, what he is obliged to do, what he prefers to do and how flexible he can be.

Fourth, one describes the physical model. Within this model the ‘physical journey’ of the end-user is described, to be able to fulfill their goal. It allows you to see all the actual steps being taken, either in the real-world, or through software systems, one deems necessary to complete a task. This often brings forth potential inefficiencies, unnecessary actions and obstacles which the user has to work around.

Finally, Holtzblatt & Beyer describe the artifact model, which shows all artifacts that are involved in the process of completing the work, either as input, means or result. This can be the transfer of files, but also physical goods like reports. It is interesting to visualize if artifacts transform from one type to another, providing insight in the handling of information involved in the process.

Of these five models, the relevant models are to be selected for the development to be done, as not all models tend to be relevant for a development. Thus, dependent on the selected case, a selection of these 5 proposed models will be worked out.

As described, each person is able to go through the process of completing work differently, thus consolidation is the final step to be taken. This brings together common findings of the entire customer population, sketching a single picture. This prioritization of elements of the flows shows what really matters in the work, and guides in the structuring of a coherent response.

3.3.4 Visioning

Up to this point, all of the steps taken are focused on understanding the users as they are. From this point on, the invention of a design solution comes into play. This phase concludes the ‘requirements & solutions’ phase. Within this phase, one is challenged to analyze how the technology and overall design solution will be improving the existing workflow of the end-users. This focuses the conversation on how to improve people's lives with technology, rather than on what could be done with technology without considering the impact on peoples' real lives. This provides us with the vision of the standardized working method. This vision is intentionally rough and high-level, setting the possible design direction without fleshing out every detail. This results in a vision of a system, its delivery, and support structures to make the new work practice successful.

3.3.5 Storyboarding

The vision defines the high-level rough design in response to the users’ needs. However, for this vision to become actionable, one has to define more detailed functions, behavior and structures, all in proposed new system. One is to make storyboards of each step of a task that a user should do to accomplish a certain activity. This next level of design must take the users’ tasks into account and ensure the right function is defined in the right system places for a smooth workflow. The storyboards describe how a task may be handed off between users, and how it may be supported by several systems operating together.

3.3.6 User environment design

Within the storyboards, the coherence of individual tasks is given. The tasks provided should however be combined in one coherent system. In the user environment design step each outlined task is given a space to be performed. Holtzblatt and Beyer draw a parallel between the user environment design and the creation of a floor plan of a building. Each showing the function of each part, how each part relates to another, and how one flows between parts. The drawing of such structures does not take into account the creation of an user interface, but purely focusses on functionalities. Within each part of the system, each functionality is described with context on how it is supposed to work, resulting is a diagrammatic view of all functionalities. Using a diagram which focuses on keeping the system coherent for the user counterbalances other forces that would sacrifice coherence for ease of implementation or delivery.

3.3.7 Paper mock-up interviews

To make sure that resources in the development process are used as efficient as possible, one initiates with testing and iterating designs, before one invests in the final design and code writing for a prototype. This reduces the amount of iterations when the project has already assumed a more labor intensive status, as having that application already implies to involve software engineers for the backend and frontend to write a code and sure it is working. That option is not only time-consuming but also cost consuming. The simpler the testing process, the more time is available for multiple iterations to work out the detailed design with users.

To facilitate that process, paper prototyping provides rough mockups of the system, with the use of notes and drawings. These notes and drawings are used to represent windows, dialog boxes, buttons and menus. To be able to validate the resulting paper prototype, the design is tested with other users, which are then given the possibility to provide feedback. This iterates several times, where the paper prototype is adjusted accordingly. With the implementation of such a method, it is ensured that the final result will be true to the user's needs, resulting in a customer-centered way to resolve disagreements and define very specific system requirements. Throughout the iteration, larger system structures stabilize, which allow for the development of more specific user interface criteria. Once that is achieved, the development of a working prototype can commence.

3.3.8 Interaction & visual design

If the paper prototyping is starting to show a more stabilizing structure for the proposed tool, one is able to commence with the actual development of a working prototype. This can either be in a temporary online environment, or by actually starting to develop the system and its corresponding code. The creation of such a prototype should provide ample ground for conversation about visualization details, like color, font and text sizes, as functionality is already decided upon. Furthermore, it can be used as a stage in which one is able to validate earlier set requirements. Does the navigation work as expected? Does the tool provide the required answers and is the way its visualized according to preferences? To finalize this stage, all ambiguity should be cleared, making it crucial to very carefully listen to the user's feedback.

3.4 Implementation

To be able to put the findings of both the literature review, as well as the contextual design process to use in a working proof-of-concept, a method for the implementation is drawn up. As argued in the literature study, the application of graph technology is considered to have potential in connecting multiple information sources, serving as the backend standardized data integration method. However, as graph technology comes in multiple forms, the decision of which system to use is of importance for the further development process. To be able to provide a structured and supported review of different systems, the OSSpal methodology is applied, which is elaborated in Section 3.4.1. After this review, one of the systems will be chosen, based on the prioritization of several software characteristics, provided by expert opinion. The selected system will then receive project-specific data, this data is to be collected and prepared before being loaded into the graph database. This process is described in Section 3.4.2. Once the data is prepared, the development of the proposed working method can commence. The data will be loaded into one of the graph systems, this requires data mapping to ensure a proper translation from tabular to graph data. Once the data load is completed, it should be disclosed towards the user, which is a process which has seen the prototyping in the contextual design process. The development process of the data visualization is described in Section 3.4.3. Finally, to reflect on the set goals and requirements from the contextual design method, a validation is conducted, which is described in Section 3.4.4.

3.4.1 Open Source Software assessment

The backend of the development of the data integration method will consist of an application of a graph database, as literature has shown that this technology has potential in combining different datasets. As discussed in the literature study, linked data can serve many purposes in the AEC sector, in the meanwhile applications of graph databases in construction project management are still rather absent. There is a lack of empirical comparison of graph models for linking data in the AEC industry (Donkers et al., 2020). This stresses the relevance of a well-structured review of database systems.

Since the application of graph methodology does not come in a standardized matter, as both a labelled property graph as a RDF-based graph is among the possibilities, both having their own respective advantages and disadvantages, it is important to review multiple tools, and discuss their strengths and weaknesses. To ensure replicability of this thesis, thereby allowing for further research and continuation of this development, the choice is made to only assess graph database platforms which have a free entry level tier. Through this approach it is possible for others to easily replicate the described steps, without having the burden of costly licensing.

For the assessment of the graph database platform, the OSSpal method is used. OSSpal, short for Open Source Software pal, with ‘pal’ referring to ‘friend’, provides a framework to review and assess open source software packages according to a structured manner. The use of open source software however comes with its own set of challenges. While some open source software is provided through vendors who offer regular releases, technical support, and professional services, other software providers might not have established a source of commercial support, while the software itself might be of good quality from a technical perspective. The method finds its origin in the Business Readiness Rating (BRR) method, which provides a calculated score attributed to each software package. This method was further improved to better suit the needs of the ‘business’, as the initial method proved to be invaluable to users. The initial method hid lower levels of assessment details due to its calculated score. Next to this, the developers of the method improved the structuring of assessments by grouping them according to project activity and software categories.

While the OSSpal method provides a means for both qualitative and quantitative assessment of software, with the possibility of calculated scoring, the goal of this assessment is not to point towards the ‘better’ system. This is done as the requirements for the graph application are very case specific, having to take into account specific business preferences. Therefore the OSSpal methodology is purely used as a structure to qualitatively asses each system. The quantitative assessment of the method is left outside of this thesis’ scope, as each system potentially is able to fulfill the desired function in the creation of an information management tool. The assessment merely functions as a review of existing methods, and their inherent differences. The final decision on use for the to-be-developed tool is based on additional priorities set by a group of software developers.

The OSSpal method consists of 7 areas of evaluation, which are as follows:

- **Functionality**
  How well will the software meet the average user’s requirements?
- **Operational Software Characteristics**
  How secure is the software? How well does the software perform? How well does the software scale to a large environment? How good is the UI? How easy to use is the

software for end-users? How easy is the software to install, configure, deploy, and maintain?

- **Support and Service**
  How well is the software component supported? Is there commercial and/or community support? Are there people and organizations that can provide training and consulting services?
- **Documentation**
  Is there adequate tutorial and reference documentation for the software?
- **Software Technology Attributes**
  How well is the software architected? How modular, portable, flexible, extensible, open, and easy to integrate is it? Are the design, the code, and the tests of high quality? How complete and error-free are they?
- **Community and Adoption**
  How well is the component adopted by the community, market, and industry? How active and lively is the community for the software?
- **Development** Process
  What is the level of professionalism of the development process and of the project organization as a whole?

The implementation of this methodology normally consists out of four phases, however, as only the qualitative aspect is described, weighting the requirements is left out. Each of the remaining steps will be shortly described. Firstly, to be able to assess the functionality aspect of the 3 software packages, expert opinion will be conducted in the interview sessions, these interviewing sessions have been part of the contextual design process. These session also provide significant insight in the functional requirements of such a system. They will provide all features and components that will be the scope of the functionality assessment.

Second, testing of the systems will take place, which can be used to define a classification value for each criterion and sub-criteria. This testing is done with a simplified dataset, which should be implemented in the system successfully. This assessment is done on a five-point scale, defined as follows: 1: Unacceptable, 2: Poor, 3: Acceptable, 4: Very good, 5: Excellent, however, supported by qualitative arguments. Since the criteria are not weighted, as each application, or developer, might prioritize other software characteristics, it is not decided upon which system is the better, rather, a structured view of each systems' qualities and drawbacks is given.

However, as there is a standardized method to be developed, one application is to be chosen. To be able to choose one of the systems to use for the creation of a proof-of-concept of the standardized working method, experts coming from the AEC business sector will be consulted. This consultation process is structured in two interview iterations, combined with findings of the literature study. First, input is gained through expert opinion interviewing sessions, secondly through the consultation of a group of expert developers from the AEC sector. As is in line with the TU/e code of scientific integrity additional documents regarding the interviewed people are signed.

The initial expert interviews conducted for the contextual design process make use of the poster shown in Figure 9. This poster includes a requirements section, which serves as input for the functional requirements for the application of the OSSpal method. The second round of requirement selection happens through the consultation of several expert development experts working in the AEC sector. These experts will be given a demonstration of the to be analyzed graph platforms. With this presentation, the key characteristics of the different systems will be elaborated upon. With this presentation in mind, the developer is asked to provide their top 3 assessment criteria for the selection of such a system, as if they were the one making the decision. This motivates the expert to prioritize between possible assessment criteria. The collection of this data is done anonymously, through the use of an online form, ensuring that one expert does not influence another with their answers. Finally, the findings of both sessions will be combined into a set of functionality criteria. These criteria will be compared to findings from the literature review, and might be extended accordingly. These criteria and the following assessment of the graph database tools will be discussed in the development chapter of this thesis.

3.4.2 Project management data collection

To be able to fill the chosen graph, one is in need of data. As argued in the literature review, there is a significant gap of research is project management related data, coming from the AEC sector. This data covers the scope of planning, work preparation and document management. Within this scope, ISO has created standards with regards to naming conventions, but not within file exchange formats, contrary to the Building Information Modelling development. As this is identified as the research gap, the retrieval of specific data will lie within the project management data field. The specifically needed data is identified based on the identification of the specific case in the contextual design process, which is decided upon within the 'Requirements & Solutions' phase.

The process which describes the gathering of data, its preparation and its application in the graph is called the Extract, Transform, Load – procedure, ETL in short. Throughout the ETL procedure several steps have to be taken to guide the process of moving data from a source system into a so-called data warehouse, which is a central place of data storage. Within the scope of this thesis, such a data warehouse can be considered similar to the aforementioned Common Data Environments. Before it is able to be loaded into such a warehouse, it is often needed that it is translated from raw data to a data format which is suitable for the enterprise data warehouse. This transformation happens in the data landing. Here the data is gathered, and through dedicated parsers, it is transformed to a standardized format. This process is visualized in Figure 10.

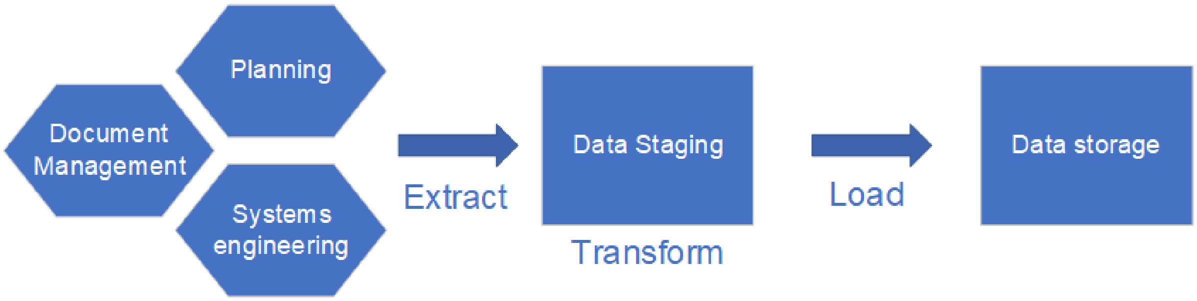


Figure 10: ETL procedure framework

This process is to be done carefully, as data quality defines the ability to acquire accurate insights and facilitates better decision-making (Biplob et al., 2018). Here, the common saying 'garbage in, garbage out' applies. To ensure one does not provide 'garbage' as input, the ETL procedure takes an intermediate step, prior to loading. After data extraction from the source systems, data first is to be cleaned, before being transformed to a standardized format. In this step all non-relevant data, duplicates and other inconsistencies are to be filtered out. Only then the data will be transformed and loaded into the data warehouse.

Within the data staging one is to translate the source systems' data towards a standardized format. One of the most supported data formats is .csv, short for comma-separated values, which is a standardized format for tabular data. As in the case of this thesis, the data warehouse is instanced in the form of a graph database, a discrepancy of data formats arises. As shown in the literature review, graph methodology deviates from the common tabular data structures of other databases. Therefore, between the transform step and the load step of the ETL procedure an additional step has to be taken, as one does not only need to translate native data formats to common data formats, but one also needs to map tabular data towards a graph structure. The mapping makes sure that relations between parts of data within a dataset are registered, but also allows for the creation of cross-source relations to be made. This is done through the creation of nodes and relationships, where each node represents a type of data element, either with or without properties, and relationships are linking the nodes together. These nodes define the different datatypes, not to be confused with file formats. These datatypes serve as a form of template, defining which properties are attached to a certain form of data. They will serve as a collection of instances with one respective type, each possessing the same set of properties, and having the same kind of relations. Often in tabular systems, tables are linked through the use of key values or identifiers, referencing to corresponding identifiers in other tables. These values make it possible to map the relationships between nodes. Throughout the mapping phase one thus designs the architecture of the graph, defining how elements of data are related to one another. An example is given in Figure 11, where we see the node 'room' and the node 'building', which have a relation which is called 'isPartOf', here 'room' and 'building' are considered types, each having their own list of instances. These instances then can be a 'Living Room', 'Bedroom' and 'Kitchen' all part of the building 'Home'. Once the design of this mapping is completed, the datasets prepared in the transformation phase are to be attached to the mapping. In the case of the example, attaching a column of a csv file called 'roomName' to the property 'name' of the type 'room'. An important note should be made on the mapping process, as due to the different setups of graph database platforms, the steps to be taken to achieve correct mapping deviates per platform. These differences are explained in Section 5.1 which highlights the to-be tested systems. Once the data is mapped, it can be imported into the data warehouse. Now, the data is ready to be used for various analytical purposes.

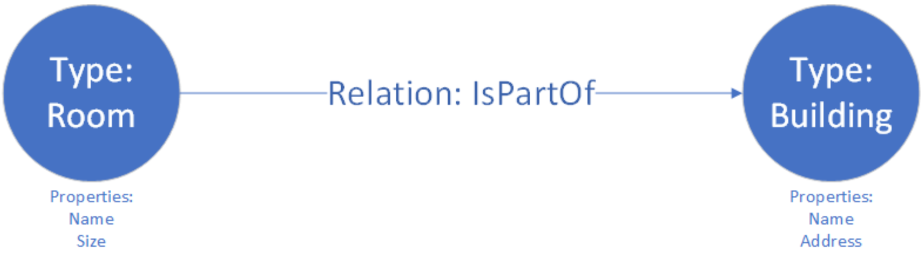


Figure 11: Graph principle example

3.4.3 Development of standardized data integration method

For the development of a standard way-of-working, which finds its purpose based on the input gathered throughout the contextual design process, several types of technology are to be discussed. Initially, the options for the different graph tools, to be assessed according to the OSSpal methodology, are described. Secondly, the mapping process is guided by the creation of a UML class diagram, which facilitates a structured view of the intended connected data structure. Finally, the concept of dashboarding is discussed, which will be applied to visualize the connected data present in the graph system.

For the comparison, three systems are selected, where a RDF-based graph system, a Labeled Property Graph system (LPG), and a hybrid system, respectively; GraphDB, Neo4j, and Weaver will be compared. This section highlights the characteristics of each of these systems. These systems will be tested with a simplified dataset. In Section 5.1 the assessment of the systems according to the OSSpal framework can be found. These three systems will be discussed separately.

GraphDB

GraphDB is a Semantic Graph Database, compliant with W3C Standards. The database can also be called a RDF triplestore, and is thus described as an RDF-based graph. It is regarded as one of the most used RDF databases in use, supporting all RDF serialization formats, as it makes use of the RDF4J framework (Ontotext, 2022). The platform offers both a free version, a standard version, and an enterprise version. The free version and the standard version offer the same functionalities, however, the computational power of the free version is less compared to the standard version. For the assessment of this tool, the free version is used. The enterprise version adds the functionality of semantic inferencing, which implies the possibility to derive new semantic facts from existing facts. The database can run both locally and online. For the testing, a local database is used.

Weaver

Weaver can be considered a hybrid graph database system, positioned between RDF-based graphs and Labeled property graphs. The system finds its origin as a graph-technology based collaboration platform applied to the Shipbuilding industry. This evolved in a data integration platform used as a Configuration Management Database for the AEC sector, used to exchange, validate and manage data from a large variety of applications used in complex projects (Weaver, 2022).

Neo4j

Neo4j, an abbreviation for Network Exploration and Optimization for Java, is a Java-based software package serving as a Graph Database management system. Neo4j is the prime example of a labeled property graph (LPG), as it possesses the largest market share of LPG's. Neo4J offers a wide suite of products, with Neo4j Aura and Neo4j DBMS being the largest products. While Neo4j is a locally stored system, Neo4j Aura uses cloud servers to store the data, making it accessible from everywhere. For the test a local database is initiated.

As argued before, the change from a tabular data structure to a graph data structure requires a mapping process. To make sure this mapping process runs as smoothly as possible, a UML class diagram is made of the case's data structure. UML, short for Unified Modelling Language, offers a standard visual modelling language intended to be used for the modelling of business process, and the analysis, design and implementation of software-based systems. This makes UML the de-facto standard formalism for software design and analysis (Berardi et al., 2005). Within a UML class diagram the structure of a designed system is shown, at the level of classes and interfaces. Furthermore it is able to show each classes' features, constraints and relationships (UML-Diagrams, 2022). Classes are defined as a set of objects with common features, these features are described by the attributes present in a class. Two classes can have two attributes of the same name, while the multiplicity or type might differ. Attributes can have a multiplicity or can be considered singular. An association is a relation between instances of two or more classes, this association can have a multiplicity as well (Berardi et al., 2005). It is intended that the UML class diagram assists in classifying the data structure of the case-specific data. Once the structure of the data is known in the UML Diagram, the process of mapping can happen accordingly.

If the case-specific data is loaded into the graph, with the correct mapping applied, it is ready for analysis. As stated in the description of the ETL procedure, the final goal of the ETL procedure is to present data that is ready for analytical purposes. When applied to graph databases, it is argued that data visualization is among the challenges of graph database application. Improvement of human-data interaction is fundamental, especially in the visualization of large-scale graph data, and of query and analysis results (Pokorný, 2015).

The aforementioned contextual design process takes up this challenge with a user-centered approach, where the business process and way of working of the future user are taken into account to design the human-data interaction. As for these processes, insights for decision-making are needed, in which information visualization through dashboarding plays a fundamental role (Frazao et al., 2021). Dashboards are proposed as a potential remedy to the information overload problem, which they aim to solve through the presentation of multi-source information (Yigitbasioglu & Velcu, 2012). Dashboarding proves to be efficient tooling for short-time decision-making, and when designed correctly, allows managers to learn about the variables involved to make more bold predictions of the future (Frazao et al., 2021). This also stresses the importance of the contextual design process, as it focuses on understanding the processes happening in the working field before the tooling is developed. The exact visualization to be used within the dashboard is thus based on the identified process, which will iteratively be finetuned to ensure correct design. This dashboard will be the final result of this thesis.

3.4.4 Validation

To ensure the aforementioned ‘good design’, but also if the proposed way-of-working meets the initial goals of the thesis, the following validation steps are taken. This proposed way-of-working deals with the absence of a standardized platform for project management information, which is identified as a research gap. To validate if the further research and development contributes to the existing state-of-art, and achieves set research goals, several steps have been taken. To assess whether the findings of the contextual design process and the implementation of the proof-of-concept have contributed towards achieving these goals, testing and validation has to be done.

Within the contextual design process, a case study is conducted. Within this case study the literature review serves as a starting point for a new development, which is tested in a business oriented environment. The case study initially considers two projects in which the company ‘Count&Cooper’ is involved as project management party. The projects that are considered are ‘Kademakers’, a quay renovation project in the Amsterdam canals, and ‘VeenIX’, a highway extension and renovation close to Amstelveen. These projects differ in scale substantially, which creates an environment to test scale-independent information management. Succeeding in providing value for both projects with the same tooling validates the scalability of the applied technology. Within the contextual design process, multiple experts from these projects are consulted for design iterations, which will be used to develop the final proof-of-concept. each design iteration validates if progress is according to the current business processes, and adds to achieving the set research goals.

As the thesis deals with a proof-of-concept, validation through extensive testing does not fall within the timescale of the project. Therefore, a scenario testing is proposed. The contextual design process will propose a revised way of working, which is an improvement based on the as-is situation. Based on this intended working method, several scenarios will be drawn up.

The prototype tool will be tested according to these scenario’s, to validate if it fulfills its intended purpose.

4. Process modelling & user environment design

In this chapter the previously suggested methods are used in the context of a use case coming from a construction project. This analysis is commenced to be able to investigate how an application of a graph database can aid in increasing the findability and accessibility of project management based information, and thus collect the relevant data. Throughout this analysis, the framework of the contextual design process will be used. This process provides a framework for systematic user-centered design, based on the identification and improvement of business processes. Through the collection of data about users in the field, which is interpreted and consolidated in a structured way, it can be used to create and prototype product and service concepts. To do so, several iterations of end-user interaction are proposed, in which the interviewer merely guides the interviewees, and lets them discuss the relevant topics and draw conclusions. First, the case itself is described in Section 4.1, then the applied contextual inquiry is discussed in Section 4.2. These results are then translated to Business Process Model and Notation (BPMN) schemes and other work models about the current status in Section 4.3. After consolidation the desired revised processes are drawn in Section 4.4, which also presents a vision with its corresponding storyboards. In Section 4.5 the user environment design is discussed, which shows the dependencies between tasks and actions identified by the end-user. Then, in Section 4.6, the input of the end-user on how to present and visualize their required data is discussed, with the use of paper mock-ups. Finally, in Section 4.7 a conclusion is provided on the ideal process, and its corresponding data structure.

4.1 Case description

The case which is used in the contextual design process is the ‘VeenIX’ project in Amstelveen, the Netherlands. This project consists of a highway renovation over the course of a 11 kilometer stretch. Throughout this renovation the highway will be widened and put partially underground. Experts participating in the realization of this project are consulted in the several design steps. The entire project has around 500 people involved, ranging from project management and contractor, to all the sub-contractors and designers. Next to this, also an expert is consulted from another project, ‘Kademakers’ in Amsterdam, the Netherlands. This project deals with the inner-city quay renovation in the Amsterdam canals. This project is considerably smaller, with around 40 people working on a day-to-day basis. It is decided to gather input from two projects that differ greatly with regards to scale, to be able to choose a business process which is present both at a small scale project, as well as a big scale project. What can be seen clearly is that while the projects differ in the amount of people involved, data stored and cost, the amount of systems in use is similar. This causes an increase in information exchanges happening in the same amount of systems, stressing the importance of solid information provisioning. With this in mind, the further steps will be applied with data coming from the VeenIX project. Further information on the projects is shown in Table 1.

Table 1: Data on case projects

	Kademakers	VeenIX
Project duration	6 years, 1 year per quay	6 years
Estimated project cost	€25 million	€1 billion
People involved on a daily basis	40	500
Documents in use	+-500	+-80.000
Information systems in place	8	8

Picture	 *Figure 12: Kademakers project (Beens Groep, 2022)*	 *Figure 13: VeenIX project (Count&Cooper, 2022)*

Of both projects, the process coordinators are asked to participate. Through this, the aim is to identify processes present both in small and large scale construction projects. While the end result of the projects might differ quite a lot, as it is a highway renovation project compared to a quay renovation project, the working methods with regards to project management are very similar. Both projects make use of the Relatics system for their Systems Engineering, which makes use of a fixed format for data export. Furthermore, both projects also make use of Oracle Primavera, a software package used for planning. In Primavera the planner is able to add additional metadata to planning objects, this could cause small discrepancies in the comparison of the datasets. Finally, in both projects, Sharepoint is used as the document management service.

4.2 Contextual Inquiry

The contextual inquiry described in the contextual design process includes three phases – user identification, interviews, and interpretation session. The initial step defines the end-user of the to-be designed data visualization tool. With this end-user, the interviews and feedback iterations will be conducted. The answers of the interviews are interpreted and prepared to be used in further steps of the contextual design process.

4.2.1 User identification

As mentioned in the methodology's Contextual Design Chapter, the identification of the end-user is of great importance for the further development of the contextual design process. In the context of this thesis, the end-user is chosen to be a process coordinator. The process coordinator is chosen due to their responsibility to guard project quality and continuity. They make use of the entire information landscape present in the construction project. They are familiar with requests on specific data and the tools that are used to fulfill those requests as-is. Furthermore, they bear the responsibility in signing off tasks, making them very important in the maintaining of project progress.

4.2.2. Interview sessions

The interview sessions happened according to a structured process, which is elaborated in the methodology chapter of this thesis. The session was held with two process coordinators, one of the 'VeenIX' project, one of the 'Kademakers' project. The structure of the session is shown in Figure 14.

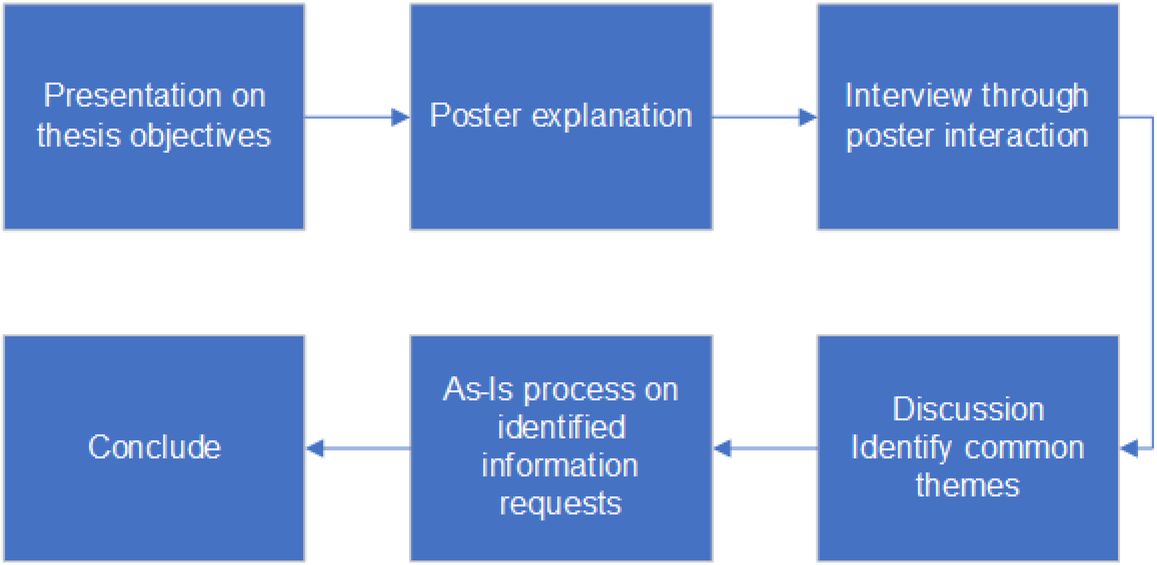

Presentation on thesis objectives
Poster explanation
Interview through poster interaction
Discussion Identify common themes
As-Is process on identified information requests
Conclude

Figure 14: Interview session structure

Within this process it is intended to prepare and guide the interviewee, allowing them to understand why they are participating, and letting them focus solely on providing insight in their way-of-working, instead of having to worry about understanding specific terminology. Initially, the problem statement of the thesis is provided. This is done to provide the context of the inquiry. Interviewees are introduced to the terminology of the elements to be researched, to prevent ambiguity throughout the rest of the interview session. Then, the poster used to gather the required information is explained. Once everything is clear for the interviewee, the interview itself is conducted. The session has a very interactive nature where the interviewer guides the interviewees when they fill the posters with input. The interviewer will ask additional explanatory questions based on the input the interviewee provides, and in the end guides the discussion where the interviewees identify and prioritize their common processes.

After the introduction, the interviewees are given the following question: 'What kind of information requests do you get on a regular basis with regards to your project management data?' Based on this question, they are asked to fill in the poster. Once the poster is filled, the answers given by each respective process coordinator will be discussed. Through this discussion common themes might be identified, which will be prioritized in the interpretation phase. This concludes the gathering of individual input by the process coordinators, which serves as input for the interpretation session.

Within the poster used for the interview, which can be found in appendix 1, three categories can be identified, input, process, and output. These posters were filled in in Dutch, the following tables show the English interpretation of the given input. The original answers can be found in appendix 2 & 3. These were filled in in the following way:

Table 2: Kademakers Interview Input

Input	Process	Output
Metacom – Final Budget (pdf) Ground Water Works specification administration	*Find the most recent version of the budget*	*The budgeted cost for specificationpost X are €…?*
Gapples & E-mail	*PDF report on SharePoint (manual action by workpreparation) status of*	*Examination X is validated and finished.*

	examination is checked and validated (excel)	
Primavera planning	*Find the correct activity*	*The planned startdate of activity X is?*
VISI	*Fill in data in Relatics*	*Dashboard - Deviation X is approved, but not yet insured.*

Table 3: VeenIX interview input

Input	**Process**	**Output**
Contract -> Design note -> implementation plan -> tasks for design and workpreparation	*Depends on individual -> Input via SharePoint + process/V&V/Configuration person filled in in Relatics.*	*- Is the project architecture created and complete? (SBS, WBS, A, RBS, DBS)* *- What is the progress on workpackage level, are we on schedule?* *- What is the current status on the 10 top risks, is this up-to-date?* *- idem with deviations* *- idem with quality control* *- idem with Verification & Validation*
Primavera Planning	*Planner retrieves progress*	*What is happening at the construction site, how do those actions reflect in the planning?* *- Planning is complete with regards to the main line* *- Devation impact is added to task level*
SharePoint document library	*Add SharePoint filters or use document libraries*	*Where can I find my most recent documents, instead of all documents?*

4.2.3. Interpretation of individual interviews

After the individual answering of the question on information requests, the end-users were asked to elaborate on their answers, and react to each other. After the short pitch of their answers, the end-users was given the task to identify the common themes in their information requests. These common themes were the following:

- Where can I find the relevant and most recent version of a document?
  - Which documents are relevant for me?
- What is the effect of a delaying activity?
  - Do critical paths change?

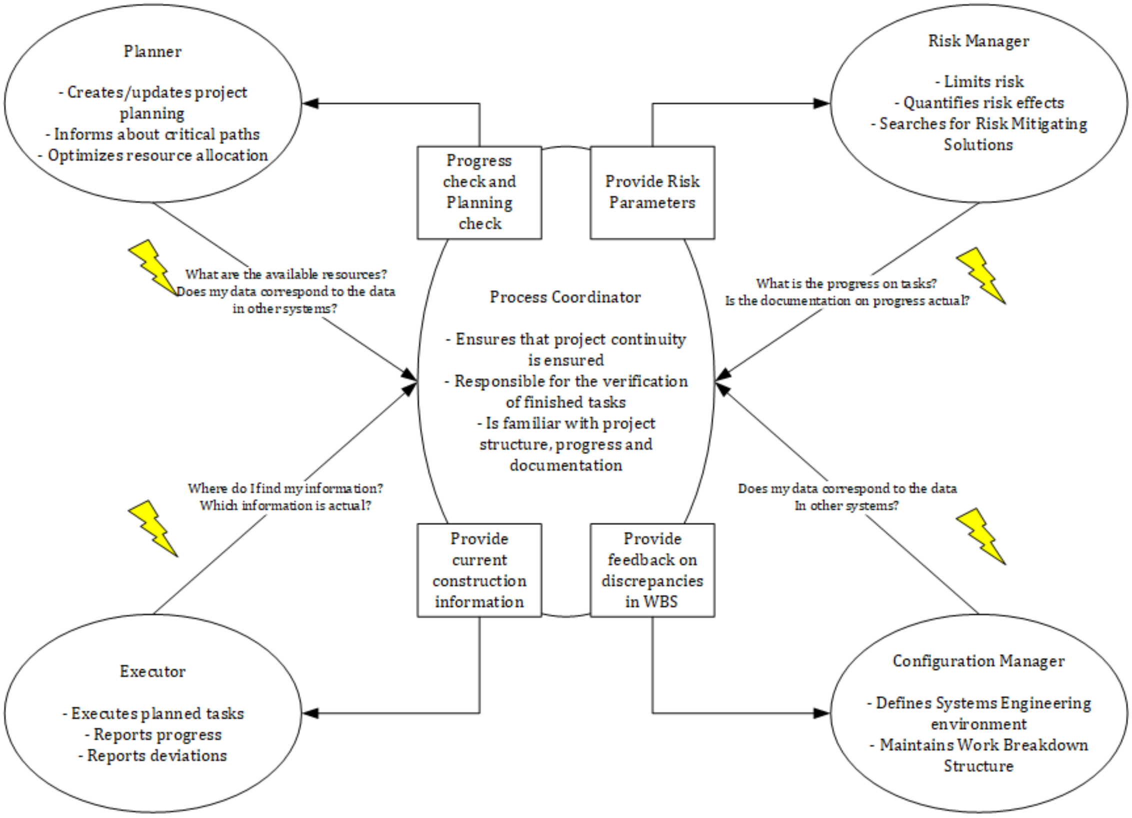
These two themes were then put up for discussion, where the end-users were tasked to prioritize the theme where they both saw the most value to be gained, if the information related to that particular theme was more findable and accessible. Through this process the process of finding relevant documents, which are the most recent and relevant for the respective user has been identified as the case in which most value can be created. Here, the process coordinators are asked to reflect on their own working methods. This is done through the creation of a Business Process Model and Notation diagram, where they model their current processes with regards to finding relevant documents to a particular task. These models will be elaborated in Section 4.3.

4.3 As-is Business Process Model of Case

With the identified case of 'Where can I find the relevant and most recent version of a document?' in mind. The following step of the contextual design process is commenced. Within this step, the results of work modelling and visioning described in methodology Section 3.3.3 and 3.3.4 are given.

4.3.1 Work Modelling

The flow model of the identified case has the process coordinator, the end-user of the tool, at its center. Surrounding the process coordinator one can find all other participants and their corresponding methods with which they collaborate to exchange information of the project, to complete their individual tasks. The squares depict which information is requested from which project participant, the lightning bolt highlights doubts and questions raised due to poor understanding of the exchanged data.



Planner
- Creates/updates project planning
- Informs about critical paths
- Optimizes resource allocation
Risk Manager
- Limits risk
- Quantifies risk effects
- Searches for Risk Mitigating Solutions
Progress check and Planning check
Provide Risk Parameters
What are the available resources?
Does my data correspond to the data in other systems?
What is the progress on tasks?
Is the documentation on progress actual?
Process Coordinator
- Ensures that project continuity is ensured
- Responsible for the verification of finished tasks
- Is familiar with project structure, progress and documentation
Where do I find my information?
Which information is actual?
Does my data correspond to the data In other systems?
Provide current construction information
Provide feedback on discrepancies in WBS
Executor
- Executes planned tasks
- Reports progress
- Reports deviations
Configuration Manager
- Defines Systems Engineering environment
- Maintains Work Breakdown Structure

Figure 15: Flow model

As the input is gathered through the participation of two process coordinators, each participating in their own respective project. The sequence of going through the process of retrieving the relevant and most recent version of documents related to specific tasks is different. Therefore, two different sequence models are elaborated upon. These sequence models are modelled as Business Process Model and Notation Diagrams (BPMN). Within these models the process of providing progress on a document associated to an activity is shown. Within this process it is possible that multiple people are involved, these people are shown in separate lanes.

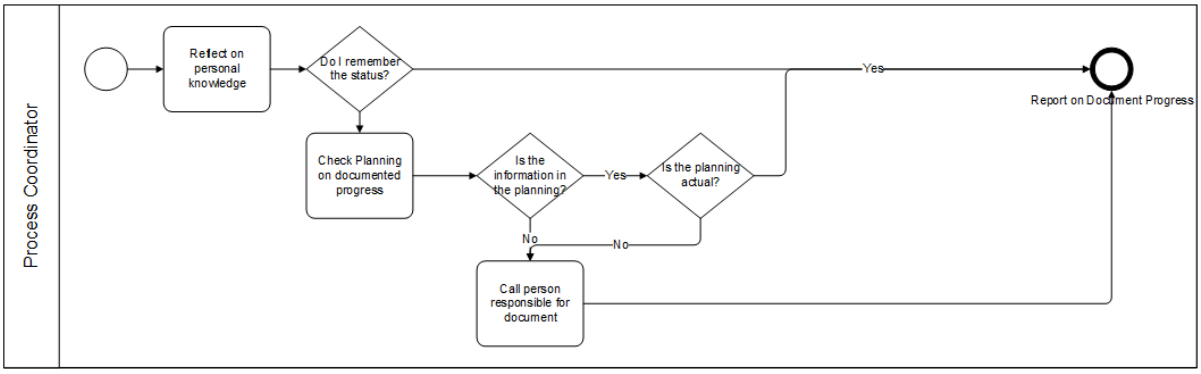

Process Coordinator
Refect on personal knowledge
Do I remember the status?
Yes
Report on Document Progress
Check Planning on documented progress
Is the information in the planning?
Yes
Is the planning actual?
No
No
Call person responsible for document

Figure 16: BPMN of Kademakers

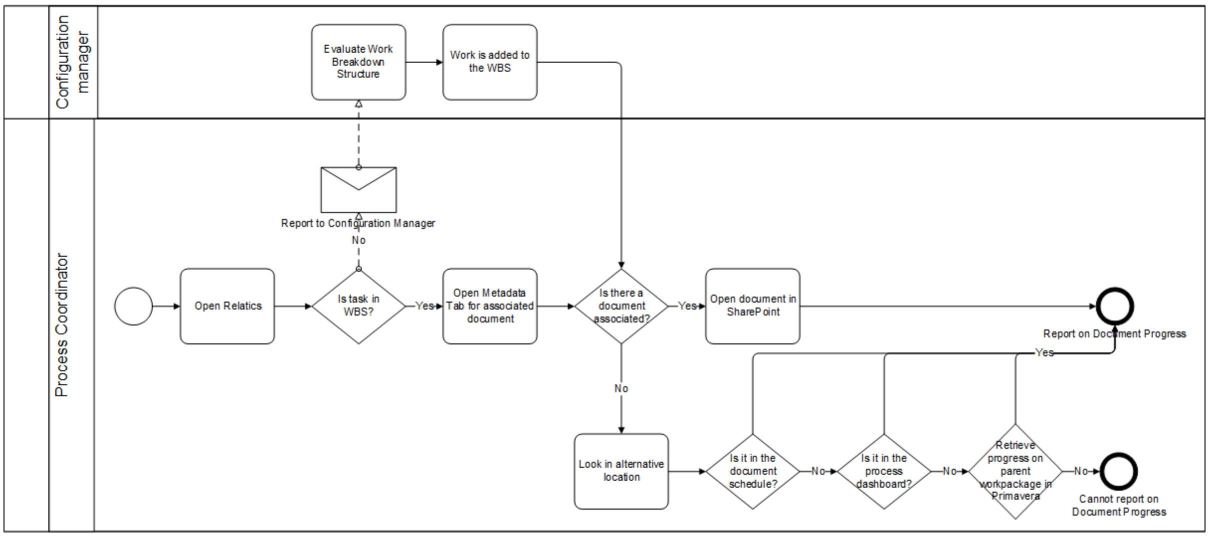

Configuration manager
Process Coordinator
Evaluate Work Breakdown Structure
Work is added to the WBS
Report to Configuration Manager
No
Open Relatics
Is task in WBS?
Yes
Open Metadata Tab for associated document
Is there a document associated?
Yes
Open document in SharePoint
Report on Document Progress
Yes
No
Look in alternative location
Is it in the document schedule?
No
Is it in the process dashboard?
No
Retrieve progress on parent workpackage in Primavera
No
Cannot report on Document Progress

Figure 17: BPMN of VeenIX

Within the physical model it is shown how the interaction between ‘places’ happens when one goes through the process. Here we talk about the digital domain and the physical domain. It

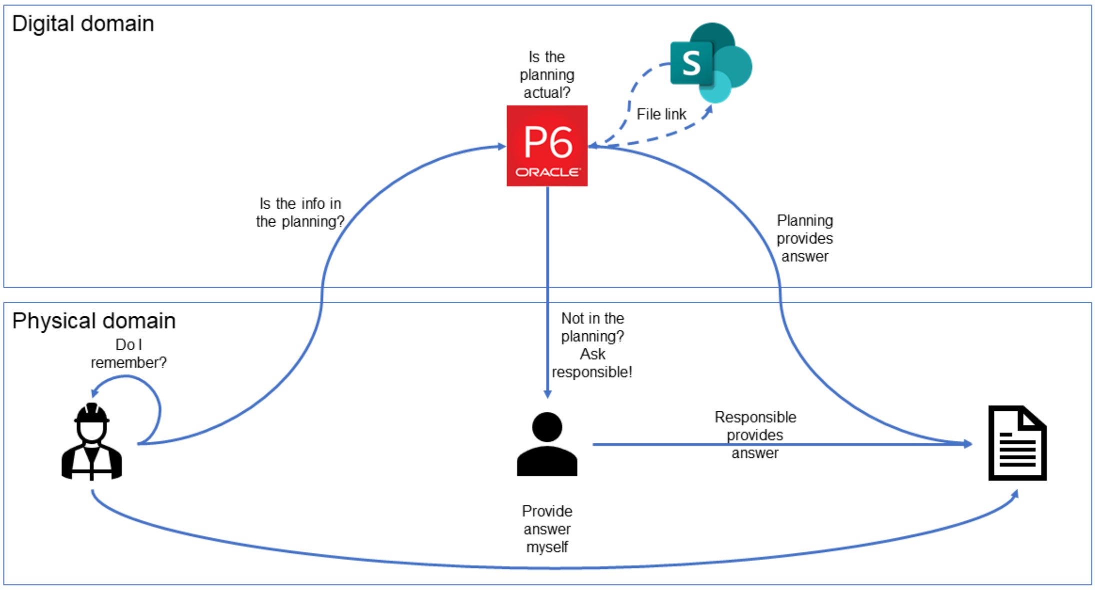


Figure 19: Physical Model of Kademakers

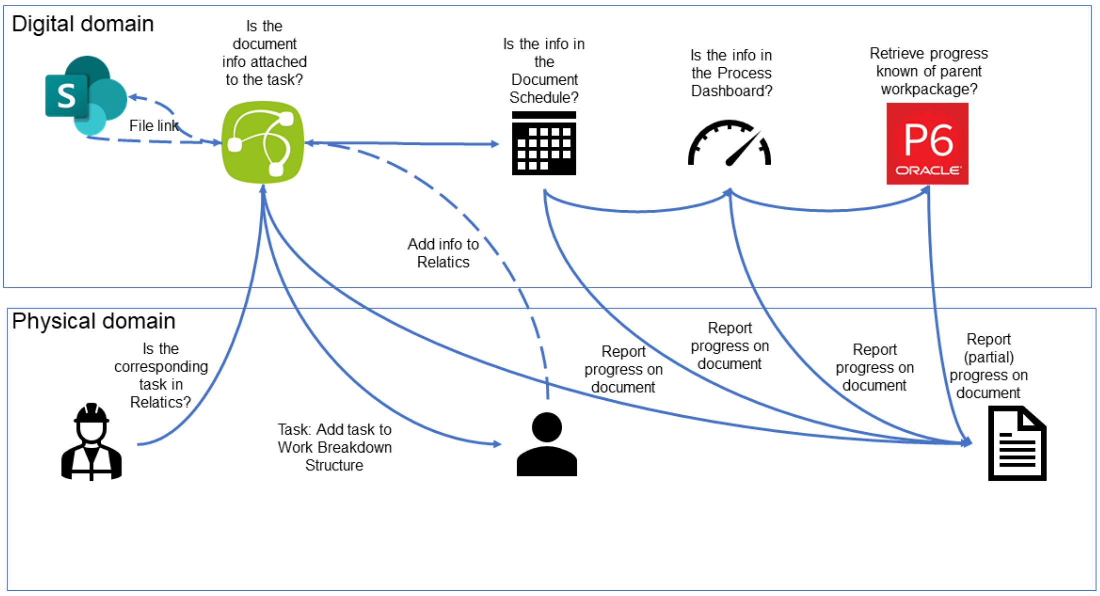

Digital domain
Is the document info attached to the task?
File link
Is the info in the Document Schedule?
Is the info in the Process Dashboard?
Retrieve progress known of parent workpackage?
P6 ORACLE
Add info to Relatics
Physical domain
Is the corresponding task in Relatics?
Task: Add task to Work Breakdown Structure
Report progress on document
Report progress on document
Report progress on document
Report (partial) progress on document

Figure 18: Physical Model of VeenIX

shows the steps taken by the end-users to collect the required data, together with the respective routes they take past several systems and methods, both digital as well as physical.

4.4 Process formation and consolidation

As the contextual design process described, it starts with understanding the user and their working methods as is. This is concluded in Section 4.3. In this section this input is used to envision how one wants to alter their working methods to consolidate their as-is processes. This is done through the creation of a vision, illustrating the newly proposed way of working. Specific tasks identified in this vision are then further elaborated as individual storyboards.

4.4.1 Visioning

Information gathered and interpreted in the contextual inquiry is used for the aforementioned modelling. in these models all of the dependencies and relations between stakeholders are mapped. Next to this, the current processes and individual actions are described. With this current state in mind, the visioning phase is commenced. Here, the end-user is tasked to analyze the current state and think of solutions to improve the existing workflow. In visioning, the team uses the data to drive conversations about how to improve users' work by using technology to transform the work practice. This focuses the conversation on how to improve people's lives with technology, rather than on what could be done with technology without considering the impact on peoples' real lives (Holtzblatt & Beyer, 2014).

As the contextual inquiry is conducted with two process coordinators, each from their own projects, there are differences in the way one works. However, the question asked is the same, as well as the involved stakeholders in the respective process. In both the sequence model and the physical model the difference can be seen clearly. When reviewing the sequence models, one sees that the element of scale in a project provides a significant difference. In the 'Kademakers' project the retrieval of progress is significantly easier, as often the process coordinator can rely on their own memory. Furthermore, if the systems fail to provide the desired information to the process coordinator, one simply knows who to approach to gain the right information. This is in stark contrast with the process on the 'VeenIX' project. Where one first of all is dependent solely on the digital documentation on progress. Also, due to the large amount of work packages, with its corresponding activities, one is not sure if the documentation in the Work Breakdown Structure is present at all. This doubt is little, to non-existent for the process coordinator of 'Kademakers'. In the case of these two process coordinators, the element of scale thus defines if one is able to maintain personal knowledge on progress and responsibilities, or if one needs to rely on tools to gain that knowledge.

Another difference that was found in the comparison of the two projects, is the structuring of the information landscape. In the case of the small project, doubt on information accuracy is limited, as the scale of the project is easy to grasp. The process coordinator in this case is up-to-date on the tasks to be done, and its respective progress. This makes the retrieval of the information easier, as one simply knows better where to look. Yet this is not the only advantage, as also the amount of systems which hold the information are limited. In the case of the larger project, this information reliability is lacking, first of all due to the amount of information, with can be overwhelming, but also due to the many sources which might hold the answer to the question. Especially since each source is maintained by a different person, inconsistencies can arise. The process of the smaller project can be traversed with more certainty to finding the required information, compared to the process of the large project.

For the modelling of the vision, the process coordinator is regarded as the end-user. However, this does not mean that the to be developed tool is not of use for other stakeholders mentioned in the Flow model. In this case, for example, a planner might not want to have to request a meeting, or ask for a report each time he needs the progress on certain documentation, this person might just also want to be able to get such insights out of the present systems. Also, for stakeholders providing documentation of progress, reports and risks, the system can be of value. It can show them an overview of the current status of the work they are responsible for. The visioning includes rough ideas for these functionalities.

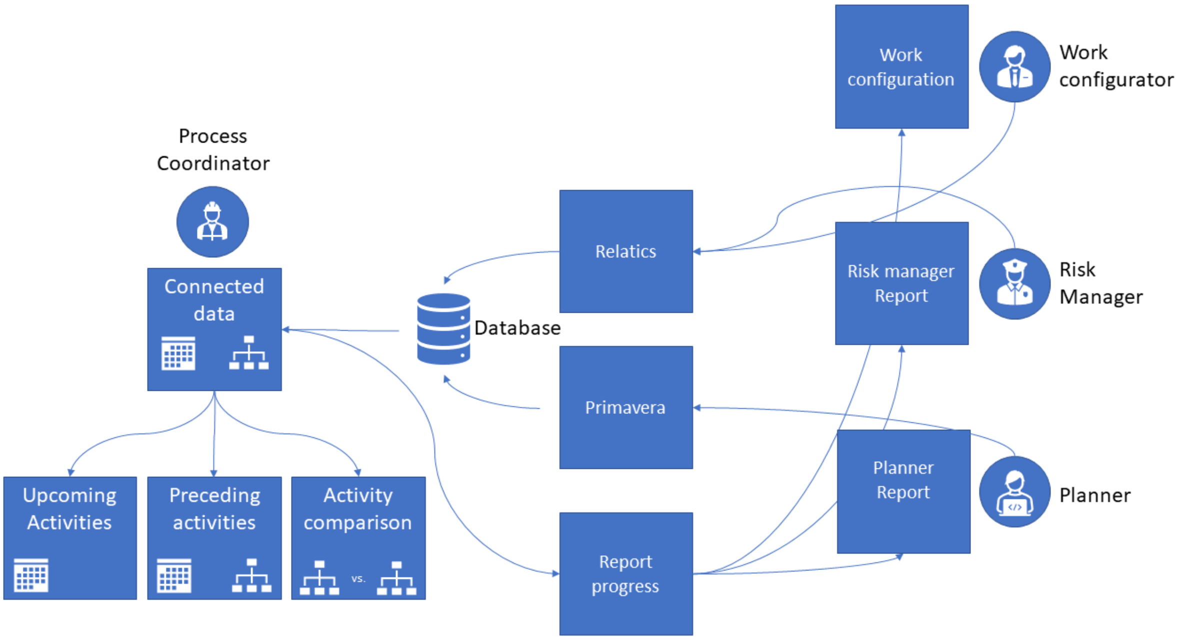


Figure 20: Simplified Visioning Diagram

Within the vision, of which a larger and more extensive variant can be found in appendix 3, the proposed new way of working is visualized. The focus user here is the process coordinator, for this user several interfaces are proposed. However, as mentioned, other stakeholders are also part of the vision, as they provide the data relevant for the decision making process of the process coordinator. These other stakeholders are influenced by the newly developed working method through the inclusion of a feedback loop, consisting out of function specific reports. Since the vision proposes three interfaces and a reporting feature, the way of working will change significantly. To model this new way of working, the contextual design process suggests the making of storyboards. These storyboards will be based on the individual concepts created to aid to new working method.

4.4.2 Storyboards

In the visioning diagram, 3 ideas have arisen to facilitate the revised way of working. These ideas are as follows: Upcoming activities, preceding activities and activity comparing dashboards, with a reporting functionality attached. All of these supposed functionalities aim to provide the user with an overview of the situation in the project, without having to go into the specifics of the data. All of these supposed features serve as a starting point for conversation and if needed, improvement. It aims to direct attention to deviating elements, limiting delay and risk. These individual concepts will be elaborated in the form of storyboards, showing how one is supposed to interact with the information for said goals. These storyboards thus help to visualize a possible layout for the particular concept, as they show the required features in a combined scheme.

In Figure 21 one can find the storyboard of the ‘upcoming activities’ concept. Within this dashboard, the user is able to have a clear overview of the activities which are about to start in the upcoming period. The standard period for the look ahead is set at two weeks, while this can also be altered by setting specific parameters. The tables used in the dashboard allow for the filtering of the activities according to its respective metadata.

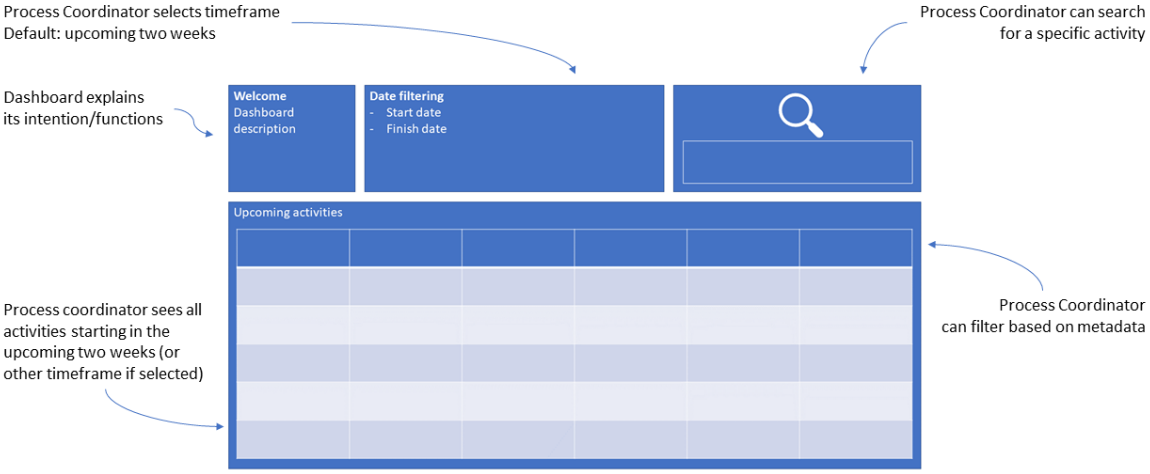

Process Coordinator selects timeframe
Default: upcoming two weeks
Process Coordinator can search for a specific activity
Dashboard explains its intention/functions
Welcome
Dashboard description
Date filtering
- Start date
- Finish date
Upcoming activities
Process coordinator sees all activities starting in the upcoming two weeks (or other timeframe if selected)
Process Coordinator can filter based on metadata

Figure 21: Upcoming Activities storyboard

This dashboard also serves as a starting point for the use of the further proposed dashboards, as further dashboards provide additional insights in the data associated to specific activities, and its cross-platform relations.

The second dashboard is intended to provide insight if a specific activity is able to start or not. To be able to draw that conclusion, one has to check if previous tasks are finished, and if all the needed documentation of those previous tasks is present and validated. This specific activity is to be selected in a search bar. After which one is presented with all of the activities preceding the chosen activity in the left table, and all of their associated documents in the right table. Both tables should support filtering based on metadata.

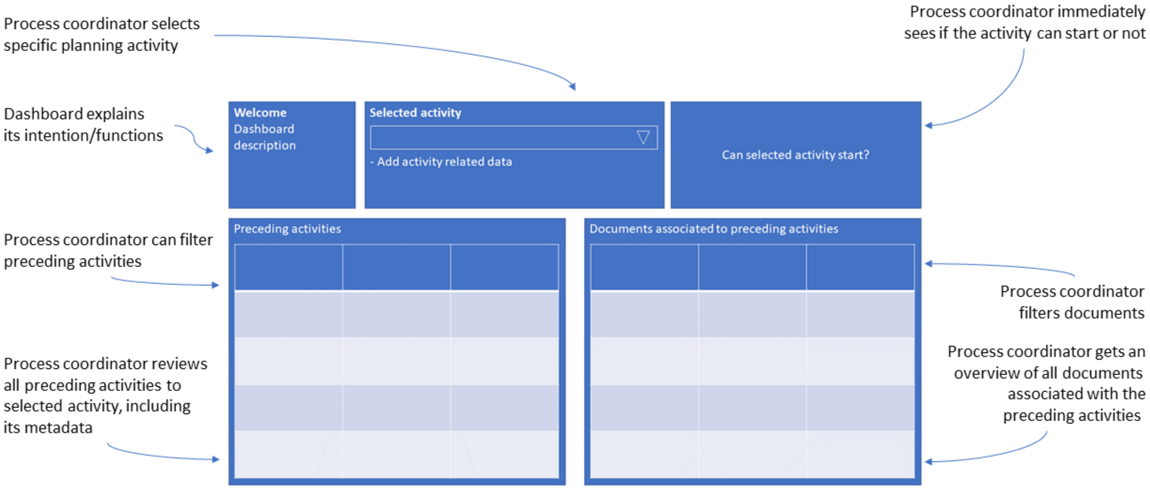

Process coordinator selects specific planning activity
Process coordinator immediately sees if the activity can start or not
Dashboard explains its intention/functions
Welcome
Dashboard description
Selected activity
- Add activity related data
Can selected activity start?
Preceding activities
Documents associated to preceding activities
Process coordinator can filter preceding activities
Process coordinator filters documents
Process coordinator reviews all preceding activities to selected activity, including its metadata
Process coordinator gets an overview of all documents associated with the preceding activities

Figure 22: Preceding activities storyboard

Finally, the third dashboard shows the comparison between the data on activities in the planning, and the data on activities in the Relatics environment. Here, again two tables are shown, where one is able to make the comparison between the data in the two systems, based on tables. Both of these tables again support filtering based on metadata. Furthermore, the dashboard shows the link between the two datasets, and reports an error if there is a discrepancy between the two datasets.

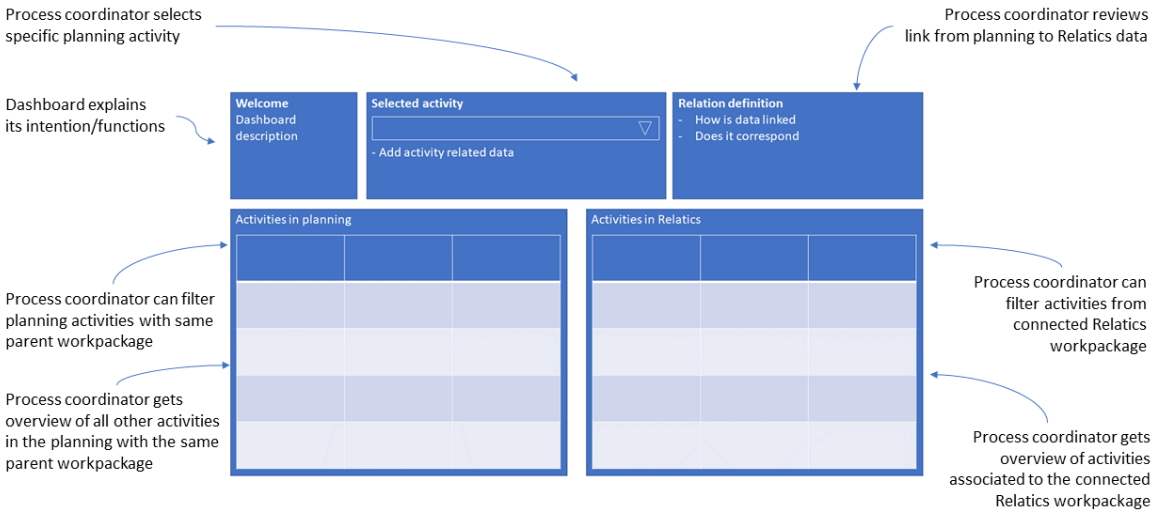

Process coordinator selects specific planning activity
Process coordinator reviews link from planning to Relatics data
Dashboard explains its intention/functions
Welcome
Dashboard description
Selected activity
- Add activity related data
Relation definition
- How is data linked
- Does it correspond
Activities in planning
Activities in Relatics
Process coordinator can filter planning activities with same parent workpackage
Process coordinator gets overview of all other activities in the planning with the same parent workpackage
Process coordinator can filter activities from connected Relatics workpackage
Process coordinator gets overview of activities associated to the connected Relatics workpackage

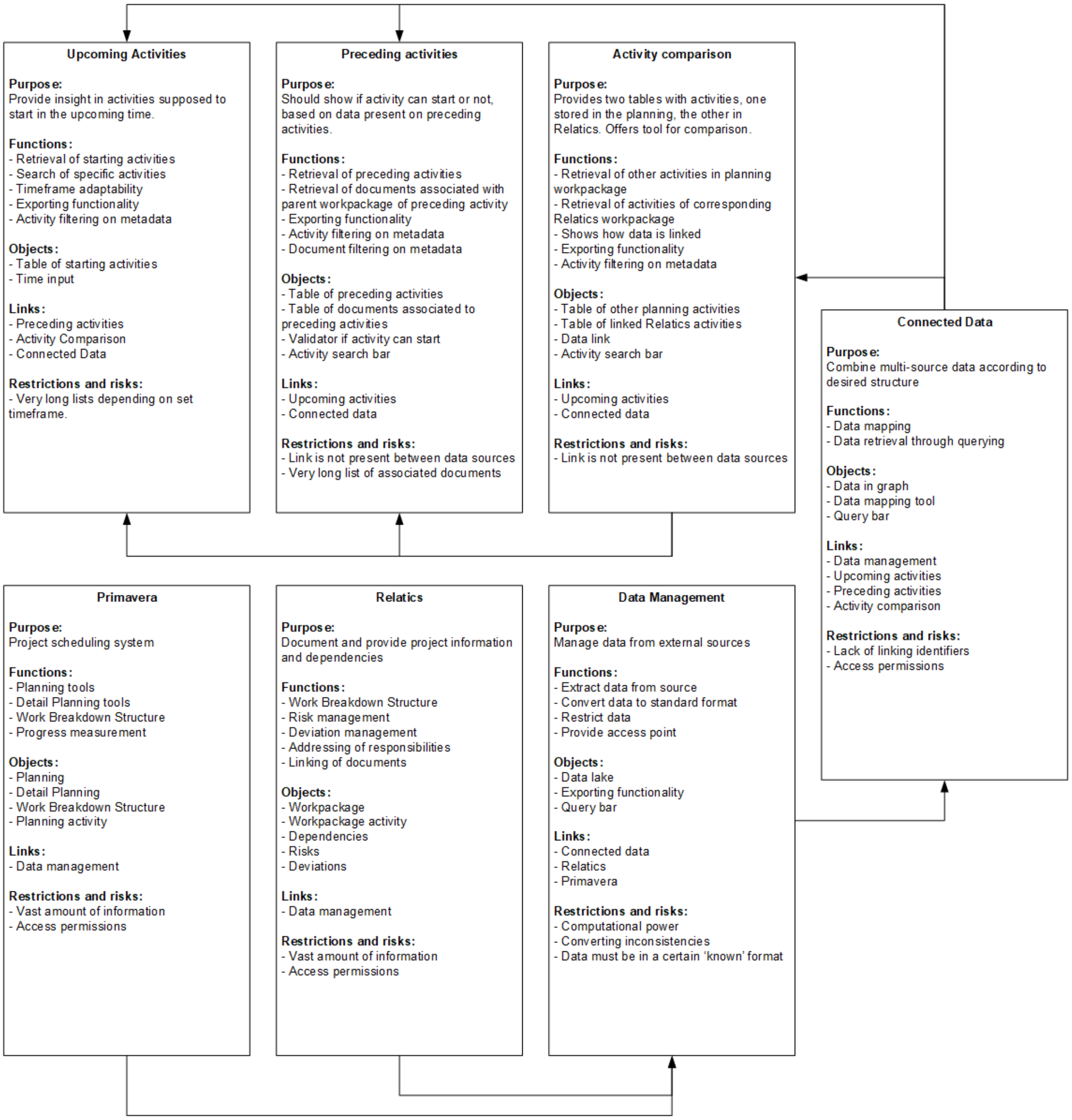
Figure 23: Activity comparison storyboard

4.5 User environment design

The following step of the contextual design process is the user-centered design, which is the translation of a high-level vision towards a more structured, but not yet low-level workspace (Holtzblatt & Beyer, 2014). As the storyboards described the individual functioning of the tasks, the user environment design focusses on the appropriate structure to support a natural flow from task to task. Each task is shown as a part of the system, showing how it supports the user's work, which functionalities are present, and how the user navigates between parts. These parts each can be seen as individual developments or increments to the finalization of the tool. Next to the created tasks, corresponding platforms and data sources are integrated in the model. These affect the proposed tooling, but not without restrictions and risks, and should therefore be taken into account. The tasks and related actions are each visualized in their own square in Figure 24, each consisting out of the following structure:

- Purpose – generalized intention of an activity. It should consist of one distinctive cognitive activity, i.e., a goal, supporting the performance.
- Functions – multiple tasks (sub-category of an activity) helping to achieve purpose.
- Objects – list of elements which would allow to make use of function (fulfil a task) and navigate within the focus area or transition to another.
- Links – relationship with other tasks or actions.
- Restrictions and risks – limitations, concerns of area which affect functionalities.

The tasks are the aforementioned 'upcoming activities', 'preceding activities' and 'activity comparison'. Additional actions include the creation of connected data, the data management itself and the two sources of data, Relatics and Primavera. What can be seen is that the connected data and data management together form the gateway from source system to the to-be-developed system. Furthermore, in the flow from raw data to visualized data, the 'upcoming activities' dashboard functions as a landing, as it provides suggestions on which activities should be looked into further, either in 'preceding activities' or 'activity comparison'.



Upcoming Activities

Purpose:
Provide insight in activities supposed to start in the upcoming time.

Functions:
- Retrieval of starting activities
- Search of specific activities
- Timeframe adaptability
- Exporting functionality
- Activity filtering on metadata

Objects:
- Table of starting activities
- Time input

Links:
- Preceding activities
- Activity Comparison
- Connected Data

Restrictions and risks:
- Very long lists depending on set timeframe.

Preceding activities

Purpose:
Should show if activity can start or not, based on data present on preceding activities.

Functions:
- Retrieval of preceding activities
- Retrieval of documents associated with parent workpackage of preceding activity
- Exporting functionality
- Activity filtering on metadata
- Document filtering on metadata

Objects:
- Table of preceding activities
- Table of documents associated to preceding activities
- Validator if activity can start
- Activity search bar

Links:
- Upcoming activities
- Connected data

Restrictions and risks:
- Link is not present between data sources
- Very long list of associated documents

Activity comparison

Purpose:
Provides two tables with activities, one stored in the planning, the other in Relatics. Offers tool for comparison.

Functions:
- Retrieval of other activities in planning workpackage
- Retrieval of activities of corresponding Relatics workpackage
- Shows how data is linked
- Exporting functionality
- Activity filtering on metadata

Objects:
- Table of other planning activities
- Table of linked Relatics activities
- Data link
- Activity search bar

Links:
- Upcoming activities
- Connected data

Restrictions and risks:
- Link is not present between data sources

Connected Data

Purpose:
Combine multi-source data according to desired structure

Functions:
- Data mapping
- Data retrieval through querying

Objects:
- Data in graph
- Data mapping tool
- Query bar

Links:
- Data management
- Upcoming activities
- Preceding activities
- Activity comparison

Restrictions and risks:
- Lack of linking identifiers
- Access permissions

Primavera

Purpose:
Project scheduling system

Functions:
- Planning tools
- Detail Planning tools
- Work Breakdown Structure
- Progress measurement

Objects:
- Planning
- Detail Planning
- Work Breakdown Structure
- Planning activity

Links:
- Data management

Restrictions and risks:
- Vast amount of information
- Access permissions

Relatics

Purpose:
Document and provide project information and dependencies

Functions:
- Work Breakdown Structure
- Risk management
- Deviation management
- Addressing of responsibilities
- Linking of documents

Objects:
- Workpackage
- Workpackage activity
- Dependencies
- Risks
- Deviations

Links:
- Data management

Restrictions and risks:
- Vast amount of information
- Access permissions

Data Management

Purpose:
Manage data from external sources

Functions:
- Extract data from source
- Convert data to standard format
- Restrict data
- Provide access point

Objects:
- Data lake
- Exporting functionality
- Query bar

Links:
- Connected data
- Relatics
- Primavera

Restrictions and risks:
- Computational power
- Converting inconsistencies
- Data must be in a certain 'known' format

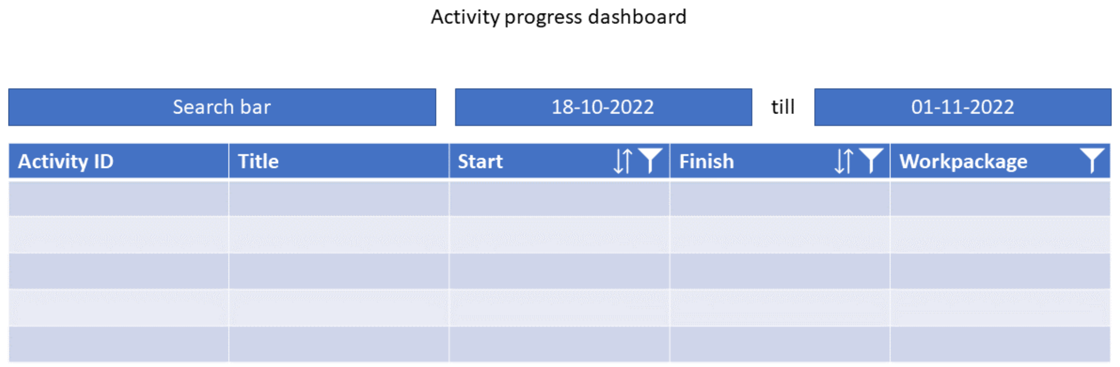
Figure 24: User Environment Design model

4.6 Paper Mock-up Interviews

To prevent too many iterations within the development stage, making it time ineffective and costly, the contextual design process proposes the method of paper mock-up interviews. The process commences with the discussing of all findings up and until the user environment design, which in itself is a result of an iterative process. Based on this, the end user is given the opportunity to provide input for a rough mockup of the system, by using notes and quick sketches. These quick sketches then serve as the basis for a wireframing prototype, resulting in a mock-up of a desired user interface. This takes the created storyboards and the flows of tasks in the user environment design to a more concrete entity. While many of the functionalities correspond to one another, the preference was given to split up all tasks in individual interfaces. It was requested that each interface had its own specific goal, instead of multiple goals. This was done to prevent clutter, and keep interaction simple and clean. Furthermore, while the end-users were familiar with graph methodology and visualization, they preferred data visualization in tables. The tabular structure should also support the earlier mentioned filtering and sorting based on file metadata. This resulted in the following three schemes, corresponding to the aforementioned 3 tasks.

1. Landing scene – Upcoming activities

Within the landing scene the user is welcomed with all activities supposed to start in the upcoming two weeks. This timeframe is given as a preferred scope for the process coordinator. However, if the need arises, the timeframe can be adopted through the use of a custom date selection. Furthermore, the activities can be sorted based on start date and finish date, and filtered based on which workpackage it is associated to. Also, the feature to search for a specific activity was requested.



Activity progress dashboard

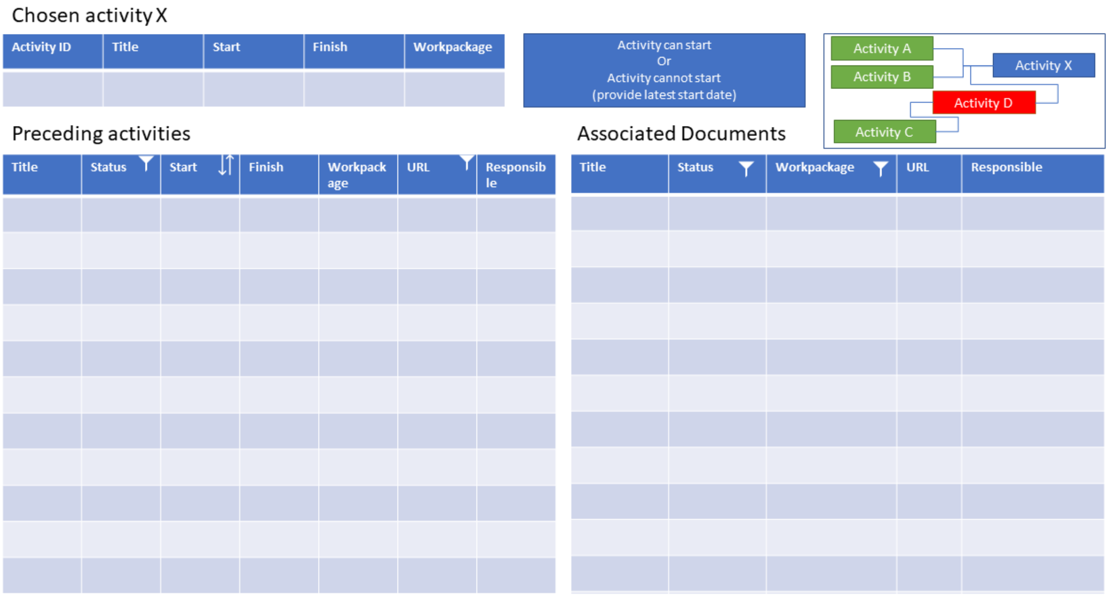
Search bar 18-10-2022 till 01-11-2022

Activity ID	Title	Start	Finish	Workpackage

Figure 25: Activity Progress dashboard mock-up

2. Progress scene – Preceding activities

Within the progress scene, the user is provided with all the present information on project progress, related to a particular activity. This particular chosen activity is shown at the top of the dashboard. Next to this, a conditional box is proposed, which immediately shows if the chosen activity is able to start. Also, a graphical representation of preceding activities is requested. Most importantly, the majority of this dashboard consists of two tables, one showing the preceding activities and their corresponding metadata, the other showing the documents that are associated with the preceding activities' parent workpackages. Again with the option to filter based on metadata associated both the activities and documents.



Chosen activity X

Activity ID	Title	Start	Finish	Workpackage

Preceding activities

Title	Status	Start	Finish	Workpackage	URL	Responsible

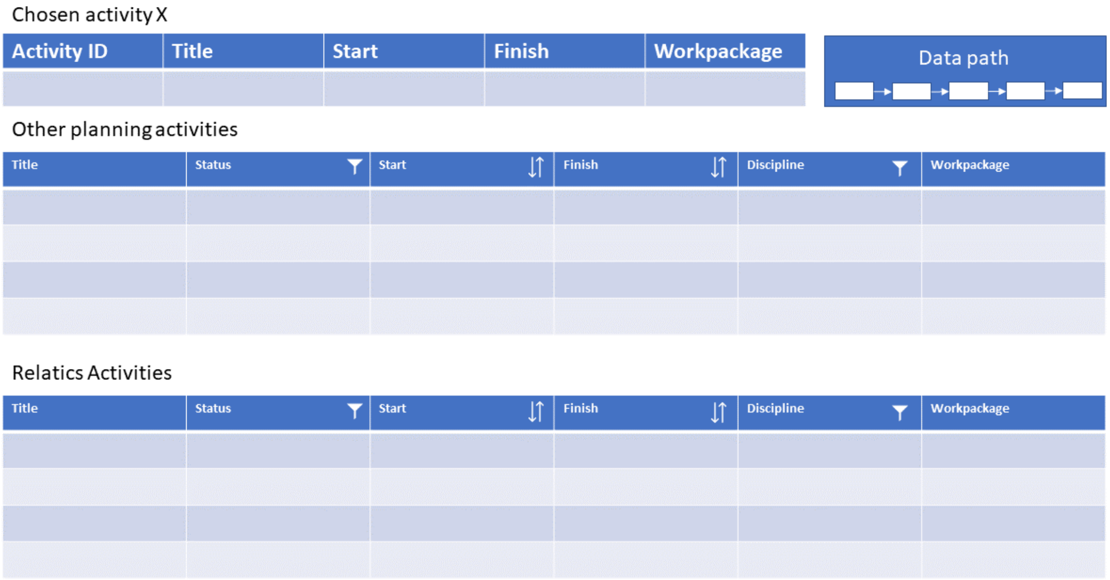
Associated Documents

Title	Status	Workpackage	URL	Responsible

Figure 26: Preceding activities dashboard mock-up

3. Comparison scene – Activity comparison

Finally, the proposed comparison dashboard is visualized. This dashboard is intended to show the comparison between activities documented in the planning and activities documented in the Relatics environment. The presentation should facilitate the comparison between the two data sources, and present this in the form of two tables, structured similarly. In the top table the chosen activity is presented. Then, right of that the data path is shown, providing insight in how the datasets are actually linked. Then, the dashboards aims to compare all activities associated to a workpackage in a planning, to the activities associated to a workpackage stored in Relatics. These tables present the same metadata allowing for quick and clear comparison. Again, filtering and sorting based on metadata is made possible in these tables.



Chosen activity X

Activity ID	Title	Start	Finish	Workpackage

Other planning activities

Title	Status	Start	Finish	Discipline	Workpackage

Relatics Activities

Title	Status	Start	Finish	Discipline	Workpackage

Figure 27: Activity comparison dashboard mock-up

These paper mock-ups are the result of an interactive input session, which was then translated to a more structured digital copy, which is shown in the figures above. These digital copies were validated with both the process coordinators, and adopted according to their feedback. These wireframes for interfaces serve as the guideline for the creation of the actual data visualization tool. The development of which is described in chapter 5.

4.7 To-be Business Process Model

The case study has shown that within an existing information landscape, the design of a tool to combine sources of information does not go without its hiccups. Especially once the project scale increases, the amount of data becomes unfathomable for the process coordinator. The process quickly becomes unclear if the initial search for information does not yield a satisfying answer, as one needs to divert to alternative sources only providing partial information, or need to contact people for their answer. To mitigate those uncertainties, a new preferred business process is proposed. This process should serve as a standard in requesting progress on activities and corresponding documentation for projects of all scales. In Section 4.7.1 the BPMN model of this process is elaborated. In Section 4.7.2 the conditions are described, to which a project and its respective data structure should comply for this method to succeed.

4.7.1 Proposed Business Process Model

Based on the input gathered in the first phase of the contextual design process, the ideal process can be modelled as seen in Figure 28. This BPMN diagram shows a proposed framework consisting of two steps to provide a satisfactory answer to activity progress. It deals with the case whether an upcoming activity is able to start, or not, based on data attached to its preceding activities.

Within the model, the planning is considered to be the single source of truth, as only the planning allows for the detailed scheduling at the activity level. The proposed system then compares its planning information with the information retrieved from Relatics. The first step of the model validates if the cross-platform link of data is existing, and thus validates if the tool is able to provide a satisfactory answer. This gateway either progresses if everything is connected properly, or reports an error if it is not. If the information is not present, it suggests to contact the work configurator. Here a manual action is needed, where either the Relatics environment is updated, or the planning is updated according to conversation between planner and work configurator.

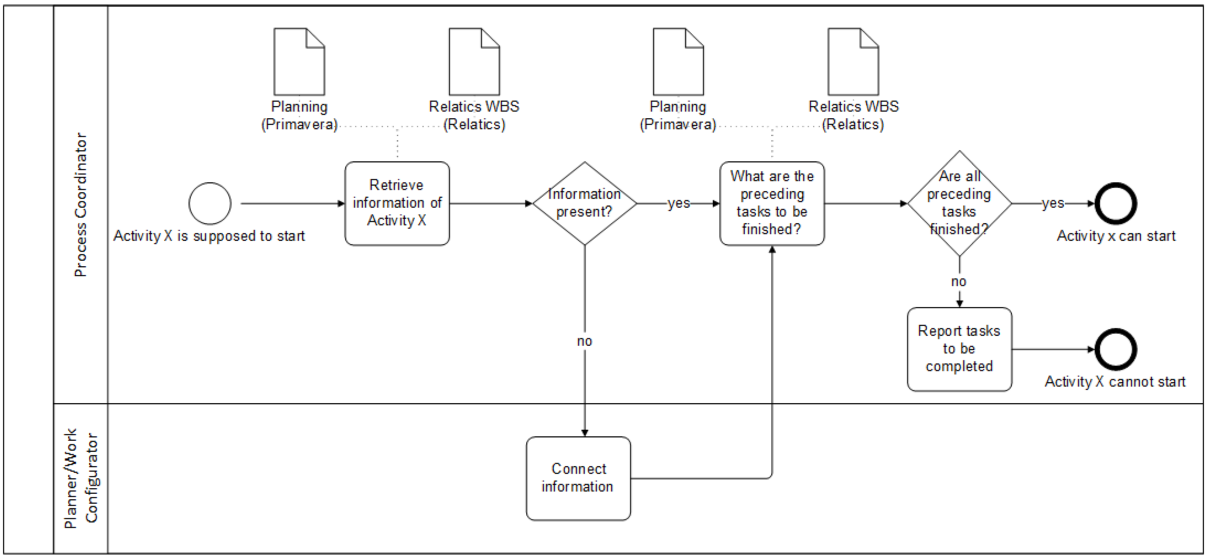


Figure 28: Proposed BPMN of new working method

If the first gateway is passed, and therefore the link between information sources is present, one is able to continue to the retrieval of the meta data of the activities. Here the second gateway presents itself, where it validates if all preceding tasks are finished. Within this gateway the first check is done whether all preceding activities have the status 'finished' attached to them. If so, one is able to continue, and activity X is able to start. If not all preceding activities have the 'finished' status, one retrieves the data associated to those non-finished activities. As the data is linked, one is able to retrieve progress on documentation for the work package of the respective activity, where one might find the reason the activity is not labelled as 'finished' yet. Once this reason is discovered, one is able to report on the tasks to be done before activity x is able to start.

4.7.2 Conditions for preferred system

For this conceptual framework to be successful, the project and its corresponding information landscape have to adhere to several conditions. First of all, for the planning to be considered the single source of truth, the planner should be made aware of each addition in tasks in other software platforms, preferably automated to mitigate human error as much as possible. Furthermore, it is important that metadata within the planning is up-to-date, meaning that it can be trusted that each provided status of activities is accurate. The same can be argued for the status of documents attached to work packages, as these statuses might be the cause for an activity not being considered finished. These conditions in turn all apply to the correct and consistent provision and exchange of information and metadata, which is highly affected by the consistency of the person providing the data. As argued in this chapter, the benefits of being able to efficiently combine data are vast. It mitigates delay in searching for the right location of data storage, or delay due to being dependent on contacting a person. However, these advantages will only be achieved if people consistently provide the correct data.

4.8 Conclusions on process modelling & user environment design

The contextual design process was used to be able to identify an important business process, within the scope of project management data integration. The process identified should apply to the combination of data of multiple sources, and be present both in a small, as well as a big scale project. To ensure these criteria are met, the end-users were identified as the process coordinators of the 'VeenIX' and 'Kademakers' projects, a large and smaller project respectively. This process is intended to serve as an example subject of a standardized work method development, to integrate project management data.

Throughout the contextual design process, the creation of insight in activity progress, and progress on its associated documents is highlighted as the process to be improved. In both of the projects Primavera and Relatics hold the data required for providing this insight. Yet, within this process it became clear that even though it is present in both projects, the methods used to provide such insight are very different. Once project scale increases, the amount of systems, as well as the data in these systems increases as well. One becomes increasingly reliant on the technology used to monitor project progress, and is not able to maintain a personal overview.

Therefore, the contextual design process presents a vision on a data integration method, which centralizes the required data, combines it as desired and presents it in three visualizations. Aimed at providing insight in upcoming activities, preceding activities and document status, and the comparison of activities in Relatics and Primavera. This method is validated as preferable for both a small, as well as a large scale project. To implement this newly proposed way of working, the creation of a working prototype is needed, which is done in chapter 5.

5. Implementation

As the case study has been conducted, valuable input has been gathered on the initiation of a new way of working. Of the contextual design process, the initial 7 steps have been taken, only leaving the interaction and visual design, which involves the actual development of the tool co-created with the end-users. Within chapter 5 this development process is elaborated. As stated in the literature review, graph applications tend to differ from each other. In Section 5.1 the OSSpal methodology is implemented to review three existing graph applications, and highlight their strengths and weaknesses. In Section 5.2, Neo4j, one of the graph applications, is initiated for the prototyping of the proposed tool. In Section 5.3 the ETL procedure of the project data towards the graph is discussed. Section 5.4 discusses how the input gathered in the contextual design process is translated towards a dashboard prototype. The chapter is concluded with a description of the validation of the created proof-of-concept, in Section 5.5.

5.1 Graph application assessment

As described in methodology Section 3.4.1, the Open Source Software assessment is used for the structured review of the different graph database applications. The literature review has argued that graph databases show potential in the linking of data, however, not all graph databases are the same. As described, one can talk about labelled property graphs and RDF-based graphs. To be able to review if these different types of graphs are suitable for an application in the AEC sector, this assessment takes both of the types into account, and also includes a hybrid graph database, consisting of elements of both labeled property graphs and RDF-based graphs. These systems, as argued in Section 3.4.3, will be GraphDB, Weaver and Neo4j. They will be assessed based on the 7 set criteria of the OSSpal method, which can be found in Table 4. The functionality criteria are based upon expert input, as functionality is specific for the application which is assessed.

Table 4: OSSpal criteria

Functionality	**How well will the software meet the average user's requirements?**
Operational Software Characteristics	How secure is the software? How well does the software perform? How well does the software scale to a large environment? How good is the UI? How easy to use is the software for end-users? How easy is the software to install, configure, deploy, and maintain?
Support and Service	How well is the software component supported? Is there commercial and/or community support? Are there people and organizations that can provide training and consulting services?
Documentation	Is there adequate tutorial and reference documentation for the software?
Software Technology Attributes	How well is the software architected? How modular, portable, flexible, extensible, open, and easy to integrate is it? Are the design, the code, and the tests of high quality? How complete and error-free are they?
Community and Adoption	How well is the component adopted by the community, market, and industry? How active and lively is the community for the software?
Development Process	What is the level of professionalism of the development process and of the project organization as a whole?

The functionality criteria are based upon data collection in two iterations, combined with findings in the literature review. First, the initial interview sessions held with the process coordinators concluded with a discussion on requirements for the to-be developed tool. Second, a group of developers within the AEC were asked to provide their 3 most important assessment criteria. Based on this input, the following functionality criteria have been drawn up for the assessment of the graph databases.

The additional requirements can be divided into five themes, each of which will be elaborated on shortly. The themes are Networking, Querying, ETL-Procedure and Development environment were based on the input provided by experts. Ontology compatibility was added as a result of findings on structuring benefits in the literature review.

- Networking
    - To be able to provide a functional database, a broad scope of accessibility is desired. The database is cloud-based and can be accessed from multiple systems, independent of system location. Local storage is considered unwanted.
- Querying
    - For the data in a database to make sense, it should be able to retrieve desired data easily. This is commonly done by writing a query line. As the graph database methodology does not have a standardized query language, each system often provides its own query language. Efforts have been made to standardize querying over graph data through the introduction of SPARQL, which is not only a query language, but also an HTTP-based transport protocol (Ontotext, n.d.). Compatibility with SPARQL is considered beneficial. Also, information provision on using the query language should be assessed.
- ETL-Procedure
    - While the ability to store data is key to a database, the way data is embedded in a database can differ quite a lot. Within an ETL procedure, data is extracted from its origin system, transformed to provide a fit with the graph database tool and finally loaded into the new database system. This assessment also takes into account the mapping of the data towards a new data structure, as a graph structure differs from a tabular data structure.
- Development environment
    - To be able to fill and maintain the graph database system, the development environment should be considered. This environment preferably requires a limited amount of coding for it to work. Next to this the user interface should be intuitive and easy-to-use, decreasing learning times.
- Ontology compatibility
    - As mentioned in the literature review, the inclusion of ontologies in a graph database has significant advantages with regards to standardization and accessibility. The support of such industry standards is thus regarded as an additional bonus.

The systems will be tested in full, which implies that the software will be downloaded and installed. Furthermore, for the testing of the ETL procedure, querying and data visualization, a simplified dataset will be used. This dataset is provided in CSV format, and consists out of data on a project Work Breakdown Structure. As each system has its respective mapping procedure and querying language, the corresponding documentation will be used as well. No prior experience is present before the testing of all three systems, which ensures an unbiased assessment of the testing.

5.1.1. GraphDB

GraphDB is a Semantic Graph Database, compliant with W3C Standards. The database can also be called a RDF triplestore, and is thus described as an RDF-based graph. It is regarded as one of the most used RDF databases in use, supporting all RDF serialization formats, as it makes use of the RDF4J framework (Ontotext, 2022). The platform offers both a free version, a standard version, and an enterprise version. The free version and the standard version offer the same functionalities, however, the computational power of the free version is less compared to the standard version. For the assessment of this tool, the free version is used. The enterprise version adds the functionality of semantic inferencing, which implies the possibility to derive new semantic facts from existing facts.

GraphDB supports the W3C SPARQL Protocol specification for querying, which is a set of specifications that provide languages and protocols to query and manipulate RDF graph content on stored online, or in a local RDF store. SPARQL is considered the standard query language and protocol for Linked Open Data and RDF databases.

To ensure that the data is mapped correctly, allowing it to be loaded into GraphDB, several mapping steps have to be taken. Ontotext, the supplier of GraphDB offers a mapping tool based on the opensource project OpenRefine. This tool, called Ontotext Refine allows the mapping of tabular data to an RDF-based knowledge graph. As GraphDB is a RDF-based graph, the mapping process should adhere to RDF schemas. However these schema’s, also known as ontologies, should already be created. This functionality is not supported by GraphDB, nor by Ontotext Refine. To create such a RDF schema, an external tool is therefore needed.

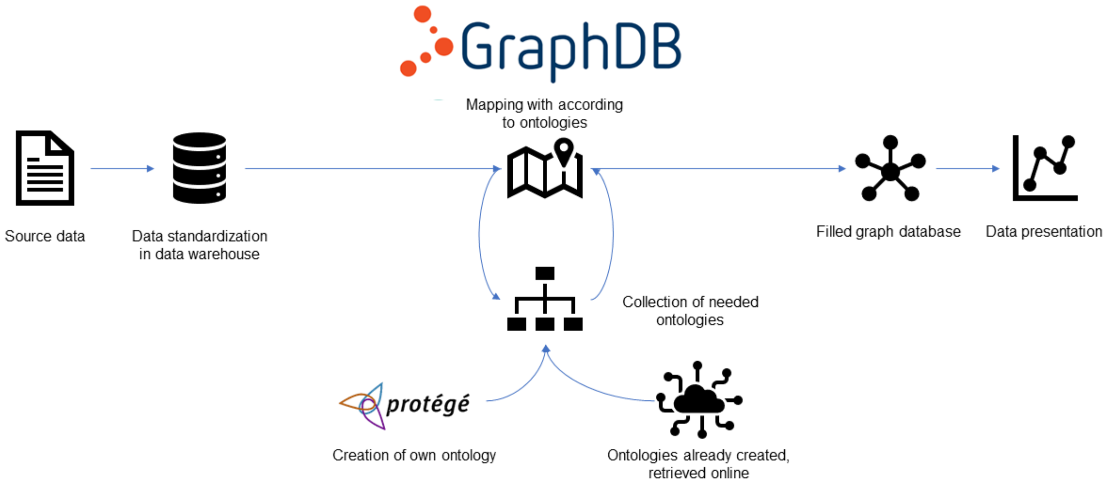

GraphDB
Mapping with according to ontologies
Source data
Data standardization in data warehouse
Filled graph database
Data presentation
Collection of needed ontologies
protégé
Creation of own ontology
Ontologies already created, retrieved online

Figure 29: GraphDB dataflow

While many ontologies, also related to the AEC sector, already exist, it often is the case that it does not suffice in fully mapping the required data. Therefore, it is often the case that an own ontology is to be created, either as an extension to an existing ontology, or as an entirely new one. For the creation of ontologies the Stanford university defined a 7 step process, which makes use of the Protégé platform to describe the ontology. These steps are structured as follows;

1. Determine the domain and scope of the ontology
2. Consider reusing existing ontologies
3. Enumerate important terms in the ontology
4. Define the classes and the class hierarchy

5. Define the properties of classes-slots
6. Define the facets of the slots
7. Create instances

While it also states that there is no way to correctly develop an ontology, the methodology provides an iterative approach which serves as a possible process for ontology development (Noy & Mcguinness, 2001).

Table 5: GraphDB assessment

Functionality:	
1. Networking	GraphDB makes use of both locally and cloud stored databases. As it is a RDF-based graph, GraphDB makes use of URI, which are web-based identifiers, which can serve as easy online access points to parts of data.
2. Querying	GraphDB makes use of the SPARQL query language. This language is standardized by W3C.
3. ETL	The use of GraphDB requires the creation of ontologies for the mapping of data. This process is reliant on a clear ontology development strategy, as most existing ontologies are fit for purpose. Furthermore, ontology creation requires external tooling. Existing ontologies can be used, however chances are that these do not serve the full purpose of the intended graph.
4. Development environment	The development environment of GraphDB is mostly low-code oriented. However, due to the use of URI's, the data in the system is difficult to read for a person.
5. Ontology compatibility	GraphDB does not function without the use of ontologies. It is thus highly dependent on the correct creation of ontologies. However, when done correctly, the database can easily profit from the advantages of the added semantics.
Operational Software Characteristics	The installation of GraphDB is very easy. The interface of the desktop tool runs locally, but in the browser window. The software layout is simple, and makes use of clear categories. For the correct use of the software, one also should have the correct ontologies. If these are not already made, one should make one themselves with an external tool (e.g. Protégé). This process is not guided within GraphDB. It thus requires some advanced knowledge on the subject of ontologies and RDF.
Support and Service	GraphDB has an active developer hub and actively aids developers through stack overflow.

Documentation	The documentation on the use of GraphDB is extensive, however, it focusses purely on the functionalities of the database itself. References to ontology creation would have been of added value. OntoText Refine documentation at time of review was not fully up-to-date. Also, the documentation does not make a distinction between different tiers of GraphDB, while functionalities and interface slightly differ.
Software Technology Attributes	GraphDB offers are large variety of possibilities for incorporation in external services, as it has support for a REST API, as wel as a RDF4J API.
Community and Adoption	Within the GraphDB stack overflow other users, as well as GraphDB developers share experiences and try to help others with questions they have. GraphDB is widely adopted as the go-to RDF Triplestore, used by a wide variety of enterprises, like Fujitsu, the BBC and Elsevier.
Development Process	GraphDB is provided by Ontotext, offering a proven and acclaimed solution for RDF-based graphs.

5.1.2 Weaver

Weaver can be considered a hybrid graph database system, positioned between RDF-based graphs and Labeled property graphs. The system finds its origin as an graph-technology based collaboration platform applied to the Shipbuilding industry. This evolved in a data integration platform used as a Configuration Management Database for the AEC sector, used to exchange, validate and manage data from a large variety of applications used in complex projects (Weaver, 2022).

The system relies on a PostgreSQL backend. This diverts from the typical NoSQL standard, as PostgreSQL typically is denoted as a Relational Database Management System. However, a self-developed layer is laid on top of this backend, creating graph functionalities.

This allows data to be visualized both in a graph, as well as in tables. The Weaver software package is fully cloud-based and can be accessed from anywhere throughout the online portal. Weaver describes itself as a SaaS provider, short for Software as a Service. It therefore offers no native functionalities. The service comes in a free tier, as well as a paid enterprise tier. Weaver supports the use of ontologies, while it also supports the use of self-made, non-machine-readable naming conventions. This allows for added flexibility in the system.

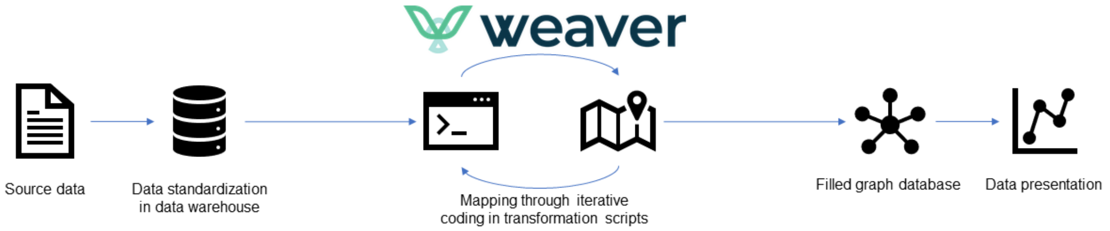

weaver
Source data
Data standardization in data warehouse
Mapping through iterative coding in transformation scripts
Filled graph database
Data presentation

Figure 30: Weaver dataflow

For the importing of data Weaver supports most standards, like XML, JSON, CSV, Excel, Turtle, and TriG formats. Each of which has its own dedicated parser. An unique feature of Weaver is that it first proposes a data structure based on the provided file, which is then to be altered. It is therefore seen as a tool with potential to explore data relations and dependencies. Once data is imported into the Weaver system, the data can be adopted through the use of so-called transform scripts. These scripts make use of Weaver Query, a self-developed language. Through such commands relationships can be formed, adopted and given semantic meaning through the inclusion of ontologies. Here the same applies as for the application of ontologies in GraphDB, if one wants to make use of an ontology, it is often the case that one has to extend an existing one, or create a new one.

Table 6: Weaver assessment

Functionality:	
1. Networking	Weaver positions itself as a Software as a Service provider, implying they offer their services only online. This allows for easy access from everywhere, as one simply has to log-in online.
2. Querying	Weaver supports both SPARQL and Weaver Query. Weaver Query is a more simplified language, able to fulfill in most basic needs. If one needs to make a more challenging query, one needs to resort to using SPARQL. This lack of unison in querying is not considered as positive, as it causes some ambiguity.
3. ETL	Weaver supports a large set of different file formats which it can parse immediately. It also has a dedicated Relatics parser, as the company focusses on AEC applications. Weaver makes relations based on the relations in the given file format. If one want’s to adopt these relations, this can be done with transform scripts.
4. Development environment	The environment of Weaver works intuitive, but lacks speed due to its sole online form. Transformations of datasets has to be done through coding, instead of providing a mapping tool. A mapping interface is not present.
5. Ontology compatibility	Weaver supports the inclusion of ontologies. By default each type relations is set to

	contains, however this can be altered to include URI's through the use of transform scripts.
Operational Software Characteristics	As Weaver consists out of a SaaS model, it requires no installation. This however limits the speed and computational power, often freezing the site if large datasets are opened.
Support and Service	Weaver has an active helpdesk, furthermore if offers a developers community through discord. As it is a relatively young company, they are very open to input on future developments. Direct support is given very personally.
Documentation	Weaver offers guides on its basic functions. The documentation is not that extensive, and focus mainly on the use of Weaver Query.
Software Technology Attributes	Weaver has a self-developed API which support create, read, update, delete (CRUD) functionalities.
Community and Adoption	Weaver is a relatively young provider of graph technology, which is still developing and adding new features. It has its first business application in the 'Pallas' project in the Netherlands, where it aids in systems engineering.
Development Process	Weaver still feels as an in development tool, due to its lacking speed and absent actions through coding.

5.1.3. Neo4j

Neo4j, an abbreviation for Network Exploration and Optimization for Java, is a Java-based software package serving as a Graph Database management system. Neo4j is the prime example of a labeled property graph (LPG), as it possesses the largest market share of LPG's.

This model has shown great promise in other AEC-sector applications, especially in the field of smart homes and cities (Donkers et al., 2020). Neo4J offers a wide suite of products, with Neo4j Aura and Neo4j DBMS being the largest products. While Neo4j is a locally stored system, Neo4j Aura uses cloud servers to store the data, making it accessible from everywhere. An additional service provided is Neo4j Bloom, which is mainly aimed at data interaction. As Neo4j uses both native and online graph storage, there is a lot of freedom with regards to choices in data handling. Contrary to GraphDB, Neo4j does not rely on the use of RDF technology, instead, it uses labeled properties that can be attached to nodes and relations, without the standardized practices of RDF. If one is interested in the inclusion of semantics in Neo4j, a plugin called Neosemantics allows for the inclusion of RDF and its associated vocabularies. Neo4j has its own query language, Cypher, which is widely documented in Neo4j's own educational environment, but which also serves as a large influence to the creation of GQL, which is a standardized query language in development for property graphs.

Neo4j supports the import of CSV files through its embedded data importer tool. This tool allows for the quick and easy importing of data. The tool offers a visual interface in which the user is guided in creating the desired data model, to which they then can attach data coming from large sets of CSV files. The created data models can be exported as a JSON file and used for automation in later stages.

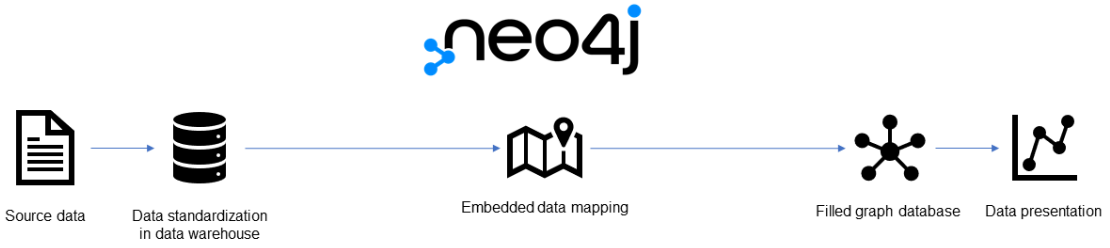


Figure 31: Neo4j dataflow

Queries can be conducted in the Neo4j console, which is able to both show the data in graph view, as well as tabular view, allowing for the quick export of specific queries. However, Neo4j offers an additional tool for viewing data, Neo4j Bloom. This platform makes use of no-code querying for simple requests of data visualization, but also the use of more complex predefined Cypher queries.

Table 7: Neo4j assessment

Functionality:	
1. Networking	Neo4j offers both online and locally running databases. They can be accessed through the use of logging in to an online portal.
2. Querying	Neo4j makes use of Cypher as a query language. Furthermore it is currently actively participating in the creation of GQL, which is a standard query language for labelled property graphs.
3. ETL	While Neo4j only supports the CSV format for direct uploading, however it also has a wide variety of plugins to facilitate the migration of data from RDBMS systems towards Neo4j. It is also possible to import data from API's, as long as it comes in JSON format.
4. Development environment	Neo4j offers a limited coding environment, with intuitive UI. Mapping happens in a no-code environment. Basic visualization is also possible without the use of code. This makes the service very user friendly, and thus easier to implement.
5. Ontology compatibility	Neo4j offers a Neosemantics plugin, which allows for the inclusion of RDF in the database. This plugin however is optional and is not needed for the database to operate smoothly.
Operational Software Characteristics	Installation of the Neo4j Local database is very easy, with a step-by-step guide on your first database initialization. The online version works almost the same, but does not require any downloads. Neo4j supports a

	wide variety of security protocols for its services.
Support and Service	Neo4j offers a developers community, with an active developments forum and hosted events. Furthermore it offers active support through stack overflow.
Documentation	The Neo4j documentation library is very elaborate. It includes an entire educational platform to become familiar with the tooling and query language.
Software Technology Attributes	Neo4j offers a wide variety of API's, drivers and connectors, for integration in one's own software, either for data presentation, as well as data management.
Community and Adoption	Neo4j is regarded as market leader in the labelled property graph market. It is used by a wide variety of enterprises, like NASA, AstraZeneca and Lyft.
Development Process	Neo4j offers a mature product, with a large variety of tools, extensions and plug-ins. The company actively communicates on its developments through release blogs.

5.1.4 Assessment conclusion

Through the use of the OSSpal framework, each of the tools have been reviewed. As argued in the literature review as well, both labelled property graphs and RDF-based graphs show potential in AEC-sector oriented applications. Each of the systems above has been tested with a simplified dataset, which resulted in mapping, querying and visualizing this respective dataset three times. Each system has its strengths and weaknesses, but all are able to fulfill the requested role of data combination. Neo4j excels in ease of use, as it makes use of labelled nodes, which have associated properties. GraphDB is the strongest when it comes to semantic enrichment of the data. Weaver supports the widest variety of file formats for importing, as well as having a dedicated Relatics parser, while also providing the most personal support, furthermore it proposes data models based on its input. Within this review it is not intended to appoint a 'best' provider of graph technology, merely to review and highlight the systems core characteristics.

5.2 Graph initiation with Neo4j

To be able to create the revised working method which was proposed in chapter 4, one of the graph applications should be implemented. As argued, each of the systems is able to facilitate the combination of data, with each system having its advantages and disadvantages. To choose one of the systems for the proposed development, the criteria set by the developers are used once again, as this input tells us the most about requirements coming from the business. The given input is grouped according to common themes, this is shown in Table 8. The original word cloud of the input can be found in appendix 4.

Table 8: Business criteria assessment

Answer	# of times given	Theme
Speed	4	Speed of system
Sufficient specifications for purposed use	1	Speed of system
Ease of data entry	1	Easy to use

Ease of use	1	Easy to use
Ease of mapping	1	Easy to use
Easy to learn	1	Easy to use
Adaptability	2	Ensure business fit
Can be integrated with other systems	1	Ensure business fit
Can be tailored to business needs	1	Ensure business fit

If these criteria are taken into account, Neo4j has the most common advantages. Therefore this system is chosen to serve for the development of the proposed application.

5.2.1 System architecture

The proposed system architecture serves the visualization and data contextualization needs presented in chapter 4. The system architecture consists of two parts, the data architecture and the system architecture. The data architecture considers two data sources, of which data from each source will be standardized and loaded into Neo4j.

The data architecture is modelled in an UML diagram, which shows the main classes and properties of the two datasets used to provide the requested information. Within this model, the distinction is made between data coming from Relatics, and data coming from Primavera. All of the links relevant for the proposed application are shown. Each class shows individual concepts present in the datasets, represented as an individual square. Each including the properties associated to each respective class.

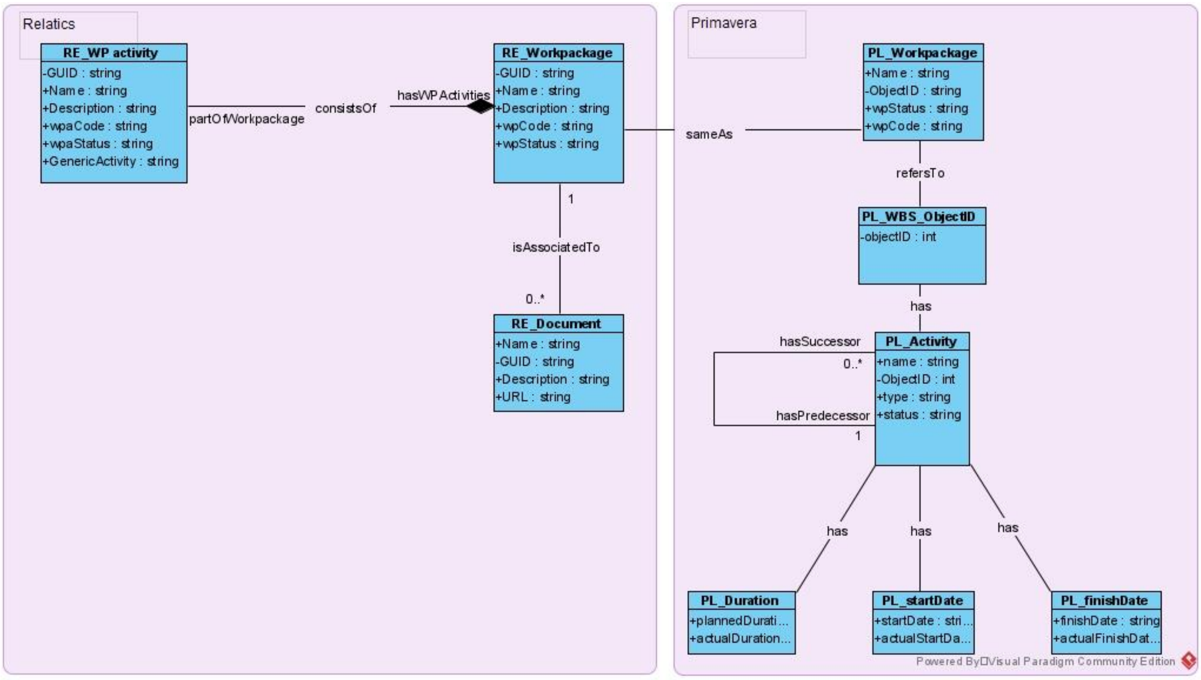


Figure 32: UML diagram of data structure

Within the UML diagram the structure for the data needed to answer the given questions is given. Here it becomes clear that the only common aspect of both datasets is their registration of workpackages. These workpackages are supposed to be equal in both Primavera and Relatics, however, this not always is the case. Within Relatics each class is linked to another class through the use of GUID's, short for Globally Unique Identifier. This however is not the case within the planning data, where each class is linked to another through the use of

ObjectID's. Furthermore, there is no global identifier present in both datasets, which allows us to robustly link the datasets. Therefore, to make the link between the two datasets, matching based on either workpackage name or workpackage code needs to happen. This method of linking is not impossible, however it is prone to human error like spelling mistakes, accidental additional spaces or the addition of dashes or slashes. Activities within the Relatics environment can be of several generic types, which are recurring common activities. If this is not the case, the property will remain empty. The data architecture serves as a guide to the mapping of the data, when uploaded into Neo4j.

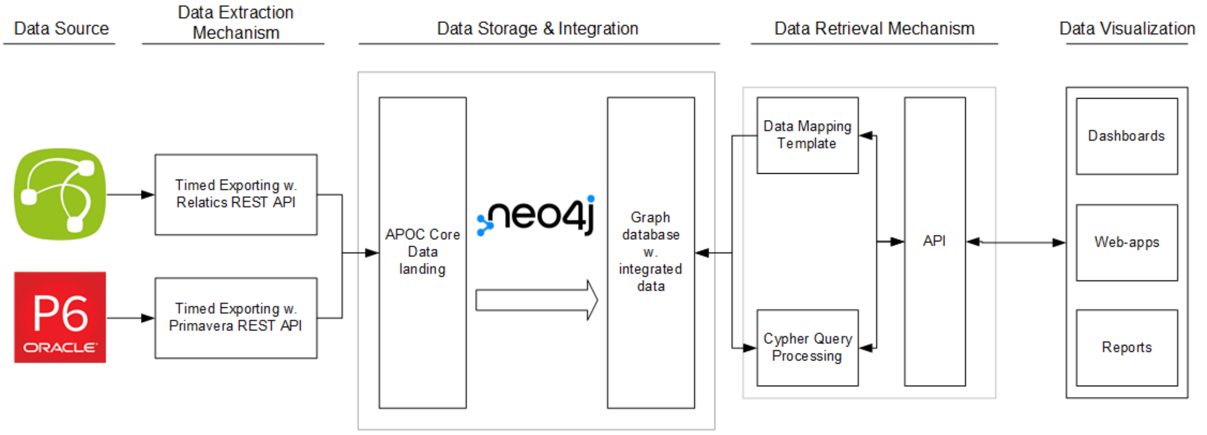


Figure 33: Graph-based system architecture

In Figure 33 a graph-based system architecture is shown. Within this system, data is retrieved from its source systems and imported in a graph database. Here, the raw data is mapped with the use of a data mapping template, and the database is filled. This data can then be retrieved with the use of a Cypher query. This Cypher query is delivered from the visualization platform to the graph though the use of an API. Neo4j does have some native services which can omit the use of this API, but this is not the case for most tools. This system employs Neo4j both as data storage system, as well as integration system, therefore working as the single source of truth.

However, as argued in the literature review, graph methodology does not excel in integration and storing all types of data, for example in the storage of documents, or ordered geometric information. Depending on the intended use of an application employed by the AEC-sector, systems might prefer the retrieval of tabular SQL data. Therefore Figure 34 proposes a system architecture in which both graph methodology, as well as an SQL database are deployed. By doing this, data provisioning for data visualization can be tailored more easily, depending on the applications intentions. Within this model several steps can be identified, which are taken to allow the visualization of source data. Within this proposed architecture, the source data is transferred directly to the data warehouse of azure. Here the data is transformed from its export XML format towards standardized SQL tables through the application of transformation scripts stored externally. This standardizes the data, which in turn is stored again in the Azure database. From this data warehouse it becomes possible to request specific SQL data needed for an external tool. However, within the Azure environment, it is also possible to host Neo4j, which in turn is directly plugged into the SQL data of the Azure database. This data is then combined with a data mapping created in Neo4j. Throughout this data mapping process the multi-source data is integrated into one graph model. This common graph model is then to be used for data visualization.

To access the information stored both in SQL, as well as the graph, the information can be queried through an API call, part of a data retrieval mechanism. This retrieval mechanism

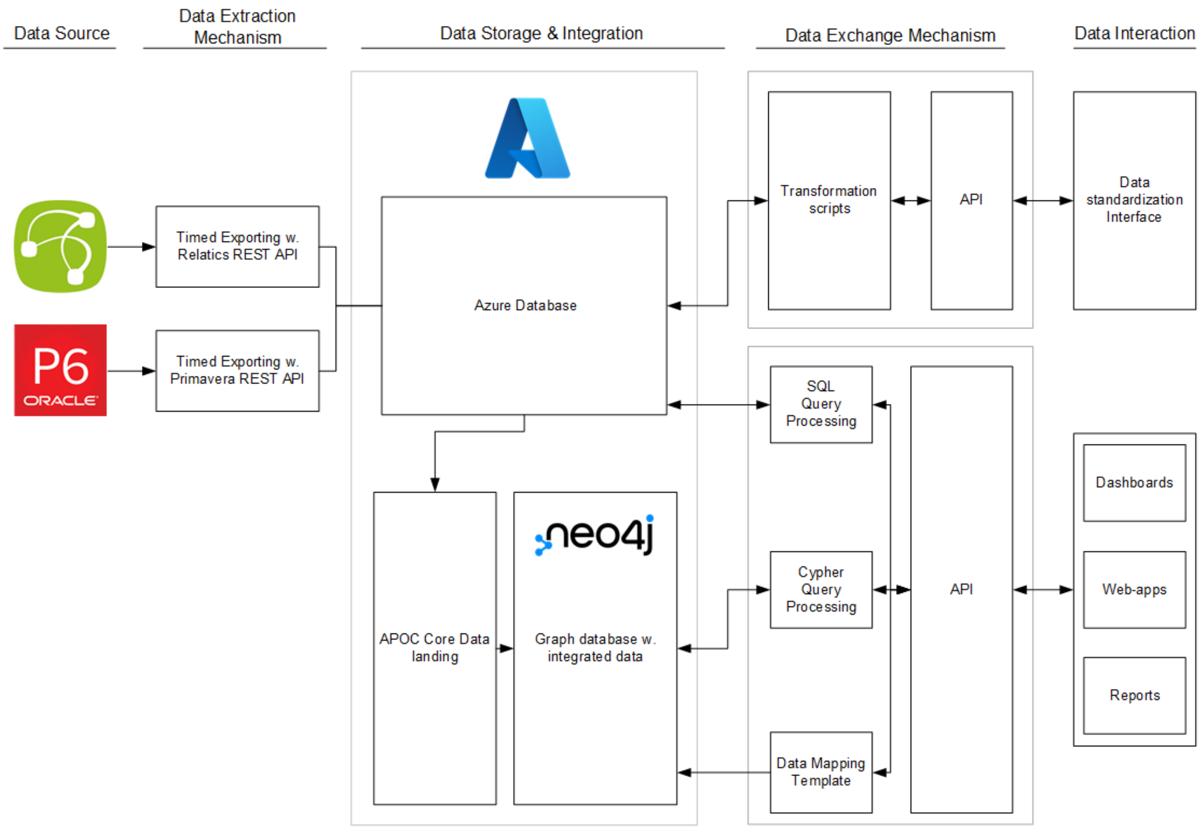


Figure 34: Extended system architecture

processes either SQL or Cypher queries and combines them in a API call towards the respective database. These queries are based on actions taken in the data visualization environment. This environment can present the graph data in dashboards, web-apps or reports. This environment can be created in a variety of software packages. Dashboading and reporting can be achieved with the use of platforms like Microsoft PowerBI, but also with Neo4j's own recently released NeoDash service. Web integration can be achieved with frameworks like Django and React, which allow for HTML integration of user interfaces, through the application of Python and Javascript respectively. This system however is heavily reliant on the use of the API functionalities of the source systems. As both Primavera and Relatics are commercially oriented providers of software, using their API's requires additional licensing cost. This architecture has the advantage that it can provide both data in SQL format, as well as linked data coming from a graph, without having to run two separate databases. This avoids data duplicates, as Neo4j retrieves it information directly from Azure, making Azure the single source of truth in this architecture.

5.2.2 Prototype Architecture

The general proposed architecture of all the systems used is shown in Figure 34. For the development of the prototype this architecture is not followed. As the revised working method is created by the aid of two process coordinators and is based on their experience in ongoing projects, the prototyping should also make use of data coming from these projects Figure 35 shows the process that is currently in place to retrieve the data from Relatics and Primavera, within both the 'Veenix' and the 'Kademakers' projects. This figure shows three phases, the data production, staging and integration. The passing along of the required data happens through actions, either triggered by timers, or manually. This Exchange of information therefore happens both through automated flows, which are highlighted by the gears, as well

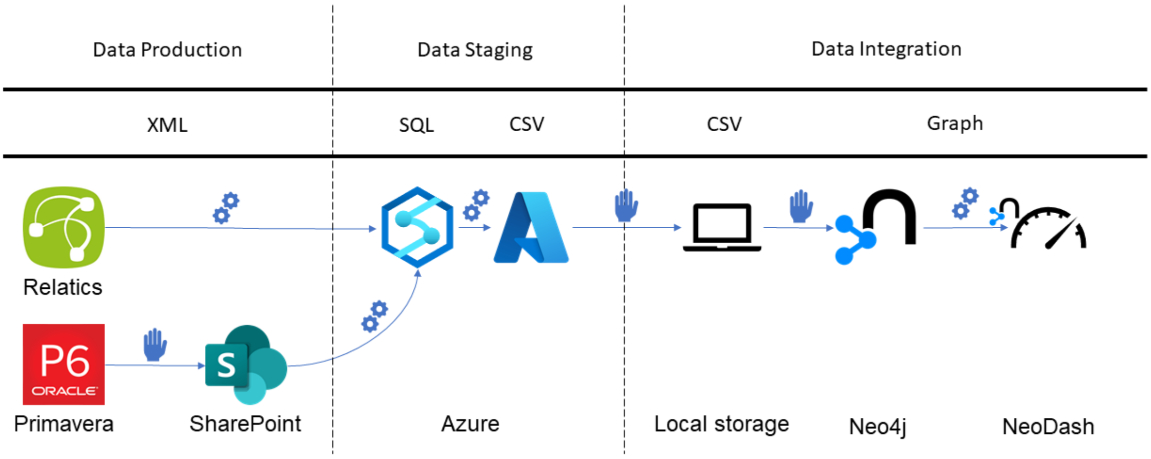


Figure 35: Prototype data structure

as manual actions, highlighted by a hand. In this model, the manual steps taken in the data integration phase can be transformed to automated steps. However, since the to-be developed application only considers a prototype, no live data is used, as this allows for easier tweaking of datasets for testing. If wanted in the future, further automation potentially also eliminates the need for the intermediate exporting to CSV, as Neo4j supports plugging in to SQL databases. Within the data integration phase different datasets will be connected, to provide a ground for cross-platform data visualization. This allows for decision making to be done based on combined multi-source information. To do so, as also discussed in the mock-up sessions of the contextual design process, the data should be visualized in dashboards. Neo4j has its own native dashboarding service, which can directly be connected to an active graph database. This service is called NeoDash. This has the advantage over other dashboarding services, that it does not require additional data structuring, as it is purpose built for Neo4j. To visualize the data, first the data has to be mapped according to a mapping schema. Then, a dashboard can be prototyped within the Neodash environment. This is discussed in Section 5.3 and 5.4 respectively.

5.3 ETL Procedure

As shown in Figure 35, the data delivered to the graph database is provided in CSV format. However, to be able to provide the data in such a format, several initial actions have to be taken. Both Relatics, as well as Primavera, support the exporting of their data in XML format. While the Relatics data can be imported into the Azure environment automatically, Primavera has to be exported to XML manually. When uploaded to Sharepoint, it is automatically retrieved by Azure. Within Azure the Synapse service is deployed, which allows the transformation of XML data to SQL data. Once the data is present in the SQL database, it can be made available for export in CSV. This CSV is manually provided and stored locally.

To import this data into Neo4j, the data should be mapped to a graph data structure. This process is elaborated in Figure 31, of the application review. This mapping process is done in the native data importer of Neo4j, which is a web-based application which can be linked to

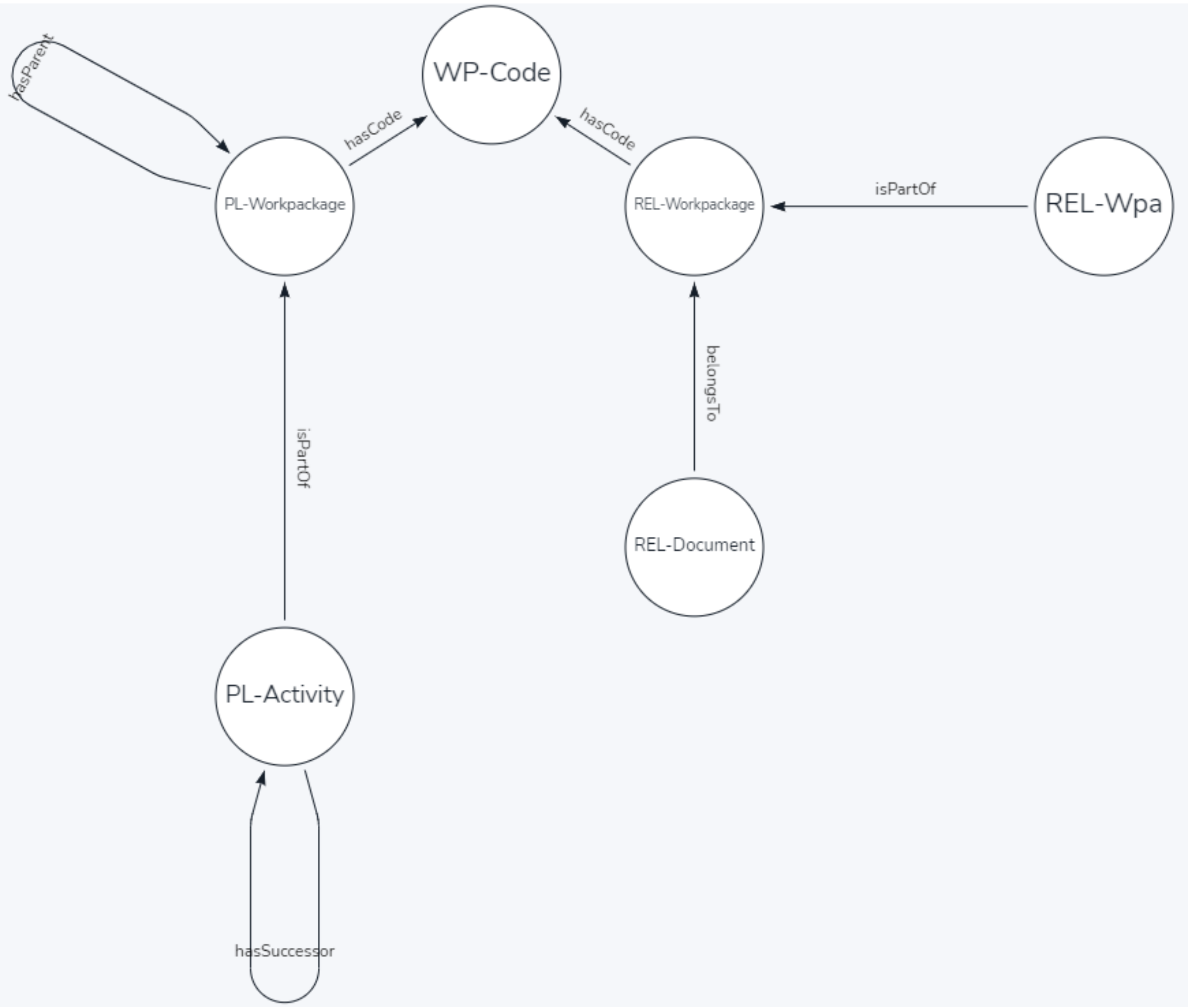


Figure 36: Graph data structure mapping

both locally stored and online graph databases. The structuring of the data in de data importer is done according to the UML schema provided in Figure 32.

Within this scheme, each class is represented, with their respective relationships. Nodes with data coming from Primavera are denoted with PL, short for planning. Relatics data is denoted with the REL abbreviation. These nodes therefore each have a label which corresponds to the classes mentioned in the UML diagram. Each node is linked to a CSV file, each containing all the relevant data for a particular class. Within the importer, one is able to select all the desired properties of a certain class, which will then be added to that particular node. One of these properties is regarded as the node ID, this is an unique value describing a particular instance of such a node. These node ID’s are used to refer to other nodes in relationships.

As shown in the data architecture, the data should either be connected based on workpackage code, or workpackage name. It is chosen to connect based on the code, as this code is provided in a fixed format. However, each node in Neo4j has a fixed type of identifier, all other characteristics are used as properties. To be able to thus reference both the identifier of a planning node, as wel as the identifier of a Relatics node, one either needs to create a mapping table outside of Neo4j, or add a mapping node within the model to link the datasets. This is necessary as the identifiers of Primavera and Relatics do not occur in each other’s datasets. All Primavera nodes have their relations linked on referencing ObjectID’s, which is the Primavera standard. Relatics however works with referencing GUID’s, to link its different classes of data. In this instance, the mapping node is chosen, as this does not require the creation of an additional file to be created and uploaded. Within the mapping node, the identifier is set to the common element of the two datasets. In this case, this is the workpackage code. Once the data model is created, it can be exported to a JSON format for reuse elsewhere. The data importer itself can facilitate the data import in the Neo4j database.

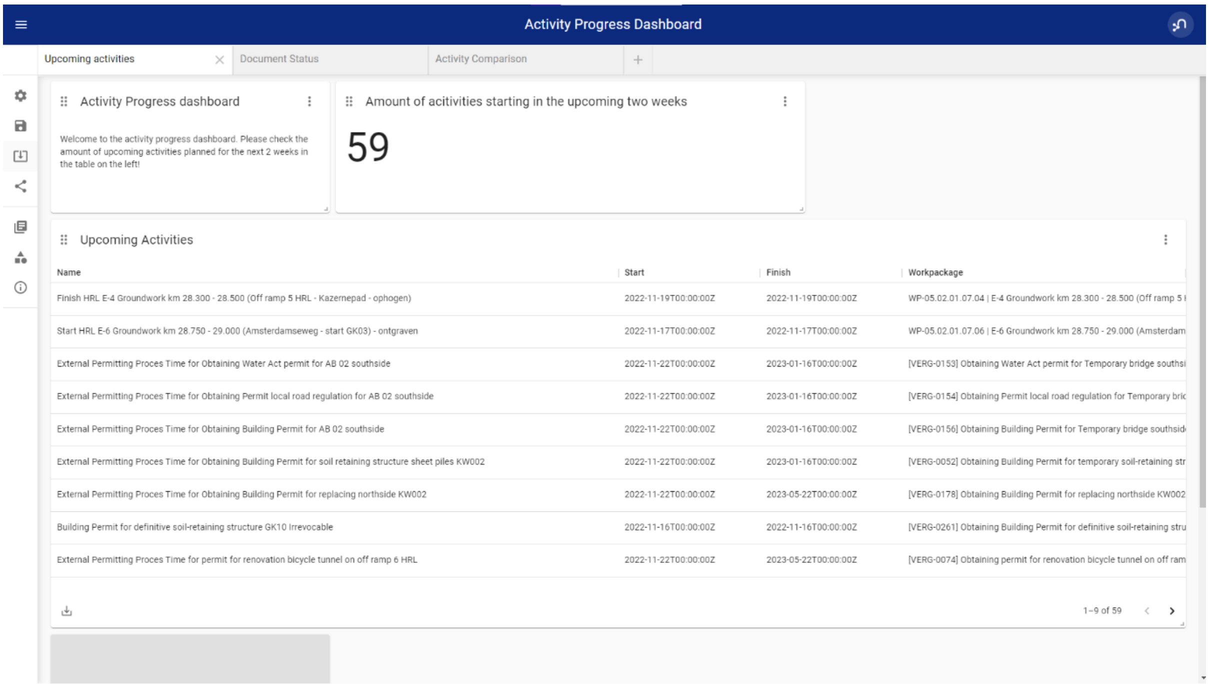
5.4 Dashboarding

During the mock-up phase of the contextual design process, three interfaces have been proposed to visualize the data involved in providing activity and document based progress. These interfaces come in the form of dashboards, each intended to answer a specific part of this progress. To facilitate the actual prototyping and interaction design, which is the last phase of the contextual design process, a dashboarding tool should be chosen. Neo4j has its own native dashboarding tool called NeoDash, which directly connects to the database. Within this service, it is possible to convert graph data to the most common data visualization templates, through the use of cypher querying. Furthermore it supports the use free-to-choose variables, as well as filtering based on metadata, which are functionalities requested in the mock-up sessions.

For the creation of the Neodash dashboards, three interfaces are proposed. Each requires a certain set of functionalities. The functionalities and the corresponding querying of the data will be elaborated per dashboard. As the intended functioning of each dashboard is already described in Section 4.6, the will only be quickly summarized.

1. Upcoming activities

The upcoming activities dashboard is intended to create a quick overview of all activities starting within two weeks, or within another chosen timeframe. This thus requires a query of all planned activities that are starting either today, or in the 14 days following. Furthermore, the addition of an activity counter was added, to be able to provide a quick overview of the amount of activities starting in the upcoming time. This results in the creation of the dashboard in Figure 37.



Activity Progress dashboard

Welcome to the activity progress dashboard. Please check the amount of upcoming activities planned for the next 2 weeks in the table on the left!

Amount of acitivities starting in the upcoming two weeks

59

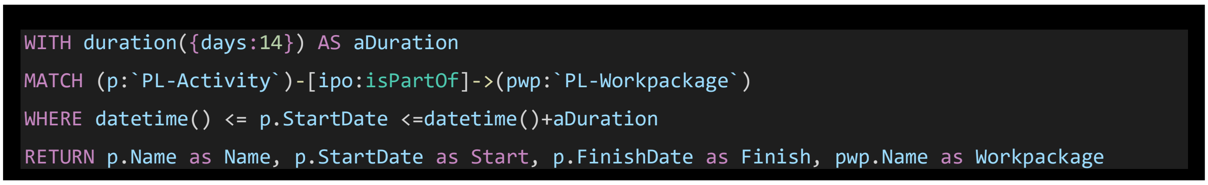
Upcoming Activities

Name	Start	Finish	Workpackage
Finish HRL E-4 Groundwork km 28.300 - 28.500 (Off ramp 5 HRL - Kazernepad - ophogen)	2022-11-19T00:00:00Z	2022-11-19T00:00:00Z	WP-05.02.01.07.04 | E-4 Groundwork km 28.300 - 28.500 (Off ramp 5
Start HRL E-6 Groundwork km 28.750 - 29.000 (Amsterdamseweg - start GK03) - ontgraven	2022-11-17T00:00:00Z	2022-11-17T00:00:00Z	WP-05.02.01.07.06 | E-6 Groundwork km 28.750 - 29.000 (Amsterdam
External Permitting Proces Time for Obtaining Water Act permit for AB 02 southside	2022-11-22T00:00:00Z	2023-01-16T00:00:00Z	[VERG-0153] Obtaining Water Act permit for Temporary bridge southsi
External Permitting Proces Time for Obtaining Permit local road regulation for AB 02 southside	2022-11-22T00:00:00Z	2023-01-16T00:00:00Z	[VERG-0154] Obtaining Permit local road regulation for Temporary bric
External Permitting Proces Time for Obtaining Building Permit for AB 02 southside	2022-11-22T00:00:00Z	2023-01-16T00:00:00Z	[VERG-0156] Obtaining Building Permit for Temporary bridge southsid
External Permitting Proces Time for Obtaining Building Permit for soil retaining structure sheet piles KW002	2022-11-22T00:00:00Z	2023-01-16T00:00:00Z	[VERG-0052] Obtaining Building Permit for temporary soil-retaining str
External Permitting Proces Time for Obtaining Building Permit for replacing northside KW002	2022-11-22T00:00:00Z	2023-05-22T00:00:00Z	[VERG-0178] Obtaining Building Permit for replacing northside KW002
Building Permit for definitive soil-retaining structure GK10 Irrevocable	2022-11-16T00:00:00Z	2022-11-16T00:00:00Z	[VERG-0261] Obtaining Building Permit for definitive soil-retaining stru
External Permitting Proces Time for permit for renovation bicycle tunnel on off ramp 6 HRL	2022-11-22T00:00:00Z	2023-05-22T00:00:00Z	[VERG-0074] Obtaining permit for renovation bicycle tunnel on off ram

1–9 of 59

Figure 37: Activity progress dashboard prototype

There are two remarks to be made on the prototype dashboard, compared to the requested functionalities of the mock-up sessions. First, the search bar for a specific activity is integrated in the filtering of the table itself. One is able to search based on the name of the activity, simply by clicking the header of the table. Second, the alteration of the timeframe is not possible due to limitations in the dashboarding software. While NeoDash does support the use of variables, it is not possible to set them to a default value. Therefore it is chosen to stick to the predetermined timeframe of two weeks. The query needed, with the resulting graph, is shown below.



```
WITH duration({days:14}) AS aDuration
MATCH (p:`PL-Activity`)-[ipo:isPartOf]->(pwp:`PL-Workpackage`)
WHERE datetime() <= p.StartDate <=datetime()+aDuration
RETURN p.Name as Name, p.StartDate as Start, p.FinishDate as Finish, pwp.Name as Workpackage
```

Listing 1: Query for upcoming activities

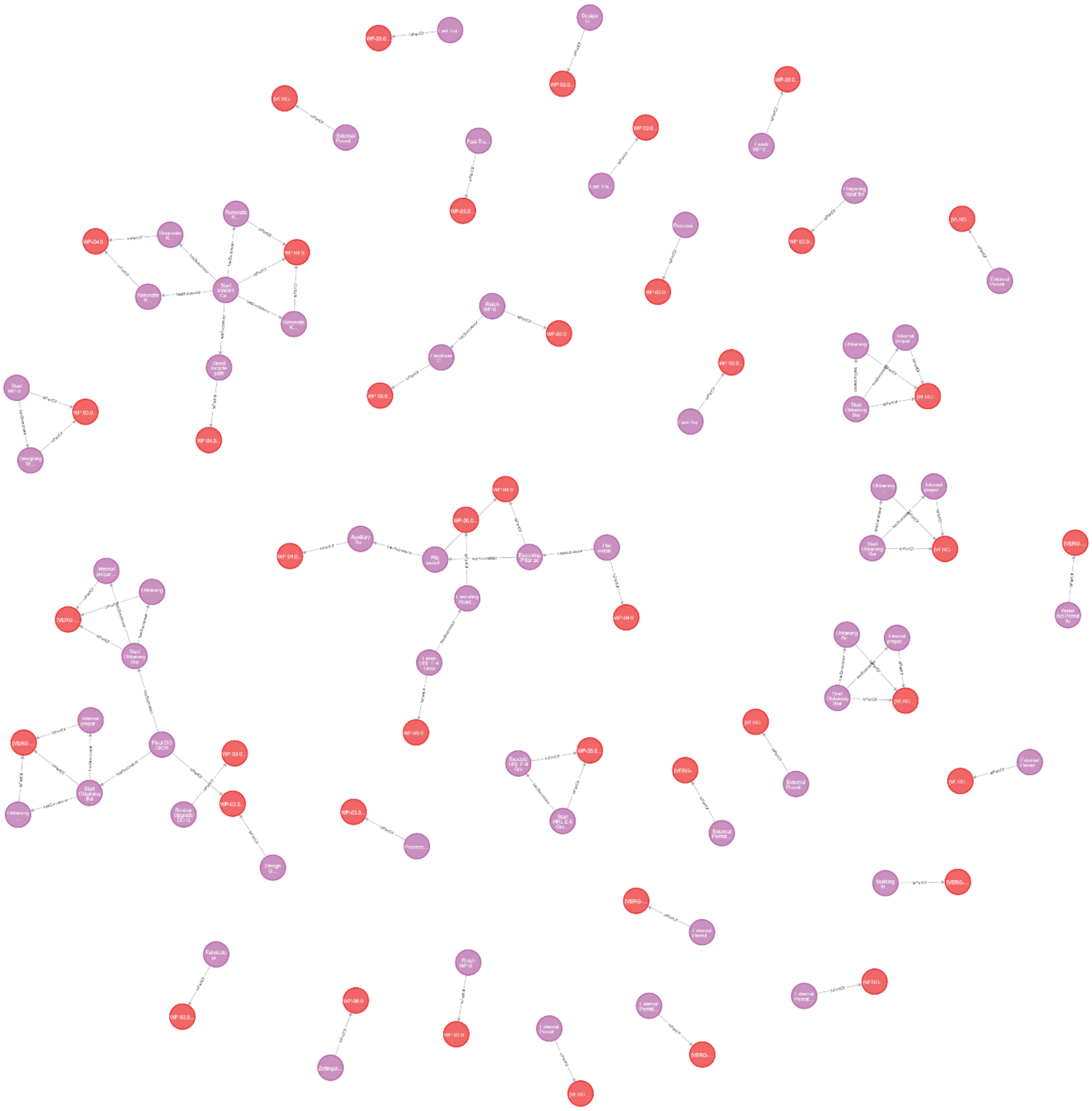
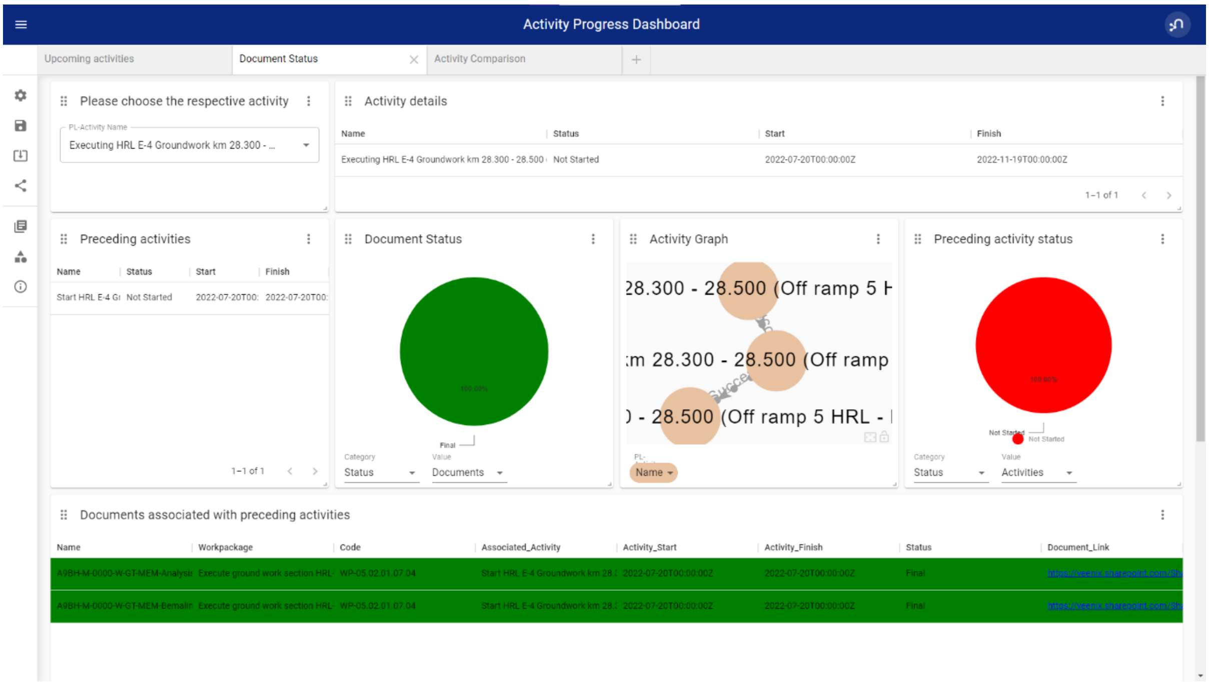


Figure 38: Graph view of upcoming activities

2. Preceding activities

The preceding activities dashboard is intended to provide insight in the status of activities which precede an activity which is about to start. This dashboard requires the ability to select a specific activity. Furthermore, it should show details about the selected activity, its preceding activities, and documents associated with their parent workpackage. Finally, a visualization on preceding activities is requested, as well as a view if the activity can start or not. An activity can only start if all preceding activities are finished, these activity statuses are therefore visualized in a pie chart. An additional feature comes in the showing of a chart which shows the different statuses of the associated documents.



Please choose the respective activity

PL-Activity Name: Executing HRL E-4 Groundwork km 28.300 - ...

Activity details

Name	Status	Start	Finish
Executing HRL E-4 Groundwork km 28.300 - 28.500 (	Not Started	2022-07-20T00:00:00Z	2022-11-19T00:00:00Z

1–1 of 1

Preceding activities

Name	Status	Start	Finish
Start HRL E-4 Gr	Not Started	2022-07-20T00:	2022-07-20T00:

1–1 of 1

Documents associated with preceding activities

Name	Workpackage	Code	Associated_Activity	Activity_Start	Activity_Finish	Status	Document_Link
A9BH-M-0000-W-GT-MEM-Analysis	Execute ground work section HRL-	WP-05.02.01.07.04	Start HRL E-4 Groundwork km 28.3	2022-07-20T00:00:00Z	2022-07-20T00:00:00Z	Final	https://veenix.sharepoint.com/Sh
A9BH-M-0000-W-GT-MEM-Bemalin	Execute ground work section HRL-	WP-05.02.01.07.04	Start HRL E-4 Groundwork km 28.3	2022-07-20T00:00:00Z	2022-07-20T00:00:00Z	Final	https://veenix.sharepoint.com/Sh

Figure 39: Preceding activities and Documentation Dashboard

As this dashboard actively deals with the statuses of both activities, as well as documents, the dashboard is color-coded. If a status is set to final, it appears as green, if it is not started yet, it appears in red. This allows for a quicker understanding of the data. Furthermore, the requested feature of preceding and succeeding activities is given in a graph visualization. Here, one is able to retrieve all information stored in the graph about both predecessors and successors. The queries needed to provide the preceding activities and the associated documents are shown below.

```
MATCH (pla:`PL-Activity`{Name: $neodash_pl_activity_name})<-[hs:hasSuccessor]-(plapred:`PL-
Activity`)
RETURN plapred.Name as Name, plapred.Status as Status, plapred.StartDate as Start,
plapred.FinishDate as Finish
```

Listing 2: Query for preceding activities

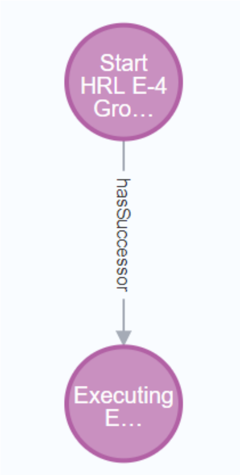
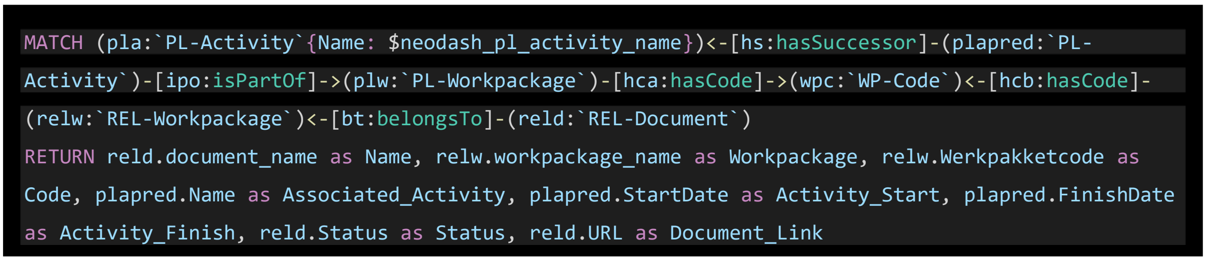


Figure 40: Graph view of preceding activities



```
MATCH (pla:`PL-Activity`{Name: $neodash_pl_activity_name})<-[hs:hasSuccessor]-(plapred:`PL-
Activity`)-[ipo:isPartOf]->(plw:`PL-Workpackage`)-[hca:hasCode]->(wpc:`WP-Code`)<-[hcb:hasCode]-
(relw:`REL-Workpackage`)<-[bt:belongsTo]-(reld:`REL-Document`)
RETURN reld.document_name as Name, relw.workpackage_name as Workpackage, relw.Werkpakketcode as
Code, plapred.Name as Associated_Activity, plapred.StartDate as Activity_Start, plapred.FinishDate
as Activity_Finish, reld.Status as Status, reld.URL as Document_Link
```

Listing 3: Query for associated documents

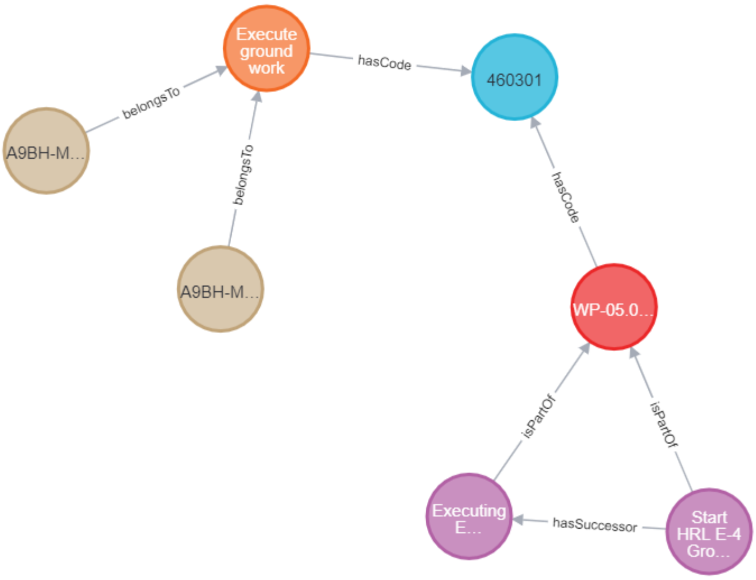
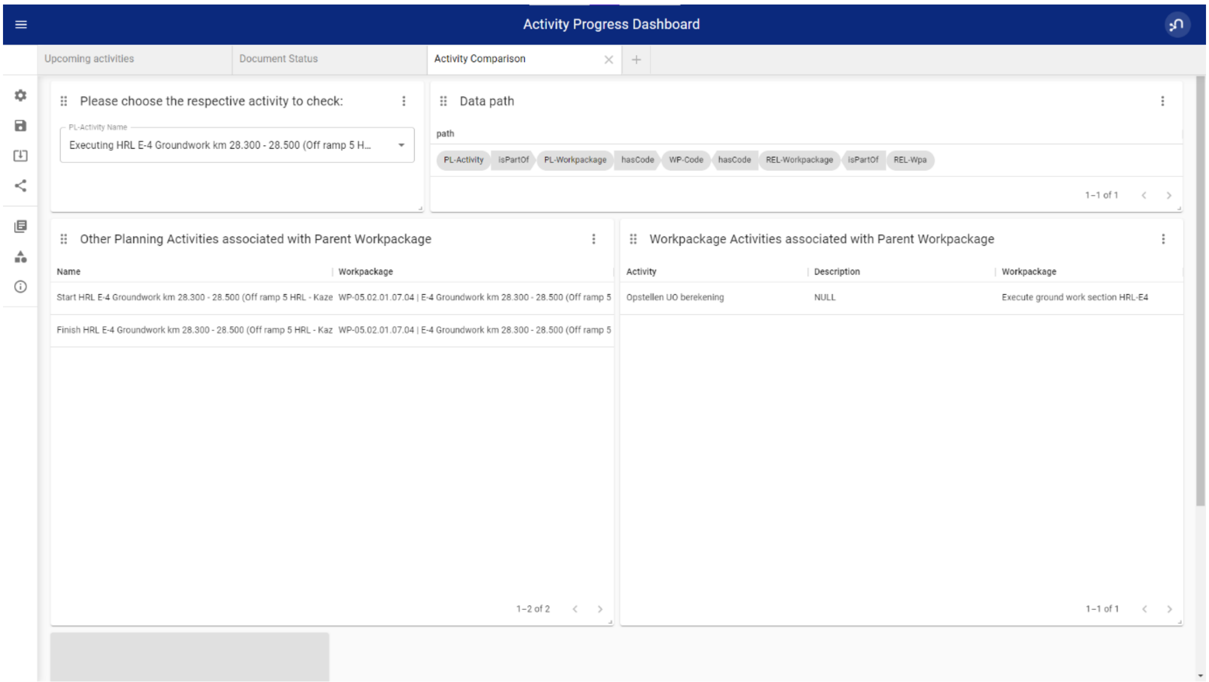


Figure 41: Graph view of associated documents

3. Activity comparison

Finally, in the activity comparison dashboard it is intended to provide a tool for the comparison of activities in the planning and activities in Relatics. Therefore, the dashboard requires the functionality to search and select a specific activity, and show this respective activity's details. Furthermore, it retrieves the additional activities associated to the activity's parent workpackage, just like the activities associated to the corresponding workpackage in Relatics. Both of these lists are shown side by side. Finally, the dashboard shows how the data is linked through a visualization of the data path. Within this data path visualization, it is possible to click each part of the link for additional information. The queries and graph views of the data are shown below.



Please choose the respective activity to check:

PL-Activity Name: Executing HRL E-4 Groundwork km 28.300 - 28.500 (Off ramp 5 H...

Data path

path
PL-Activity isPartOf PL-Workpackage hasCode WP-Code hasCode REL-Workpackage isPartOf REL-Wpa

1–1 of 1

Other Planning Activities associated with Parent Workpackage

Name	Workpackage
Start HRL E-4 Groundwork km 28.300 - 28.500 (Off ramp 5 HRL - Kaze	WP-05.02.01.07.04 | E-4 Groundwork km 28.300 - 28.500 (Off ramp 5
Finish HRL E-4 Groundwork km 28.300 - 28.500 (Off ramp 5 HRL - Kaz	WP-05.02.01.07.04 | E-4 Groundwork km 28.300 - 28.500 (Off ramp 5

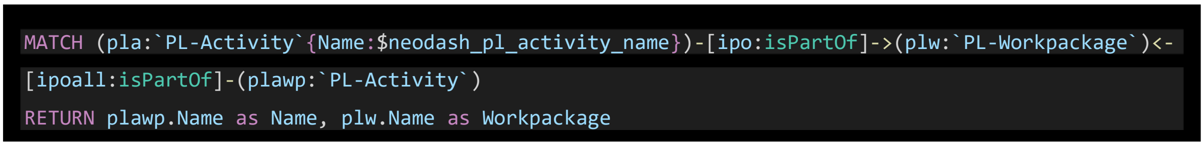
1–2 of 2

Workpackage Activities associated with Parent Workpackage

Activity	Description	Workpackage
Opstellen UO berekening	NULL	Execute ground work section HRL-E4

1–1 of 1

Figure 42: Activity comparison dashboard



```
MATCH (pla:`PL-Activity`{Name:$neodash_pl_activity_name})-[ipo:isPartOf]->(plw:`PL-Workpackage`)<-
[ipoall:isPartOf]-(plawp:`PL-Activity`)
RETURN plawp.Name as Name, plw.Name as Workpackage
```

Listing 4: Query on other activities in workpackage

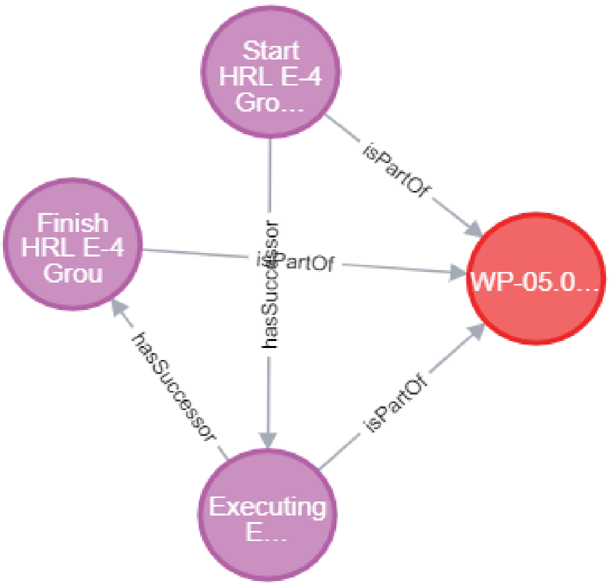
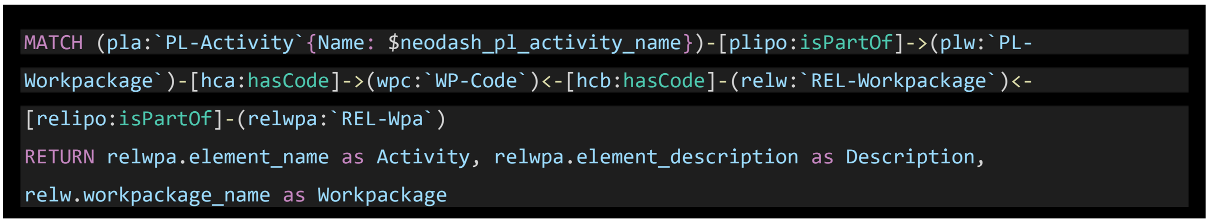


Figure 43: Graph view of other activities in workpackage

Activities in Relatics:



```
MATCH (pla:`PL-Activity`{Name: $neodash_pl_activity_name})-[plipo:isPartOf]->(plw:`PL-Workpackage`)-[hca:hasCode]->(wpc:`WP-Code`)<-[hcb:hasCode]-(relw:`REL-Workpackage`)<-[relipo:isPartOf]-(relwpa:`REL-Wpa`)
RETURN relwpa.element_name as Activity, relwpa.element_description as Description, relw.workpackage_name as Workpackage
```

Listing 5: Query on activities in Relatics

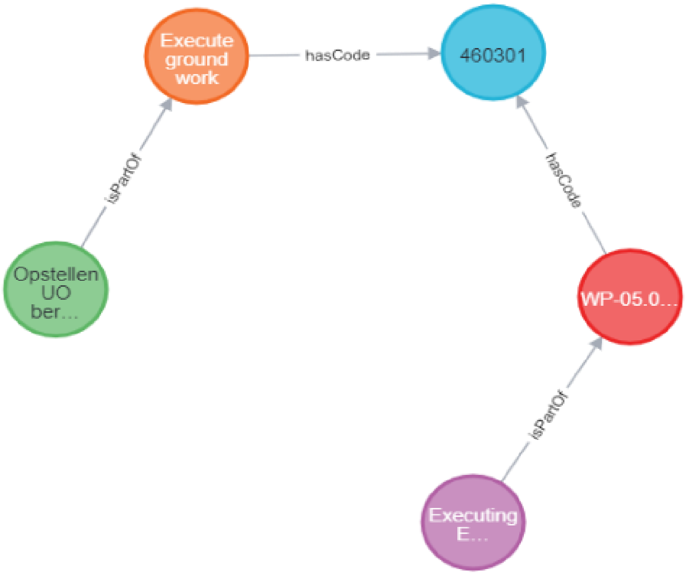


Figure 44: Graph view of Primavera and Relatics activity link

5.5 Validation

Within this section, the developed activity progress dashboard is tested against a set of criteria. These criteria are based upon the goals set in the contextual design process. These goals, as seen in the visioning phase, all support the main goal of providing insight in activity progress. The created dashboard therefore should be tested against this intended purpose of the tool. This validation is done in cooperation with the process coordinators.

5.5.1 Testing Dashboard against user requirements

Throughout the contextual design process, a business process is identified to be revised, with the creation of a new user interface. This proposed user interface is intended to provide insight in the progress of activities and their corresponding documents. Throughout the interactive sessions and iterations many iterations have taken place on functionality and visualization. These conversations yielded several specific use cases, to which the platform should apply. These use cases are stated in Table 9, together with an analysis whether or not this requirement is met.

Table 9: Validation of proposed tool

Requirement	Analysis
Dashboard shows data from multiple sources in a comparable structure	In the created workflow, shown in the system architecture, one can see that the data coming from the source systems is

	standardized first to CSV format. Throughout the mapping procedure a data structure is created, within this data structure it can be seen that naming conventions within the two datasets differ, while the information is comparable. Within the dashboard one common naming convention is chosen, to which both datasets are mapped. This ensures that both datasets can be visualized in comparable structures.
Data is accurate, data loss is minimized	Since the graph database is relying on the provision of standardized data, which is fit for purpose in the graph application, data loss from provision to presentation is non-existent. With the exporting of data from Primavera and Relatics some visualization capabilities of those respective platforms however becomes lost, as NeoDash does not support certain types of data visualizations, like hierarchy trees or Gantt charts.
Dashboard can be accessed by everyone	The NeoDash environment is able to plug in to both locally stored systems, as well as online systems. However, as the prototype is made with a local database, it can only operate on the computer hosting that respective database, sharing it with others would need the transfer of the database to an online server.
Dashboard allows the exporting of (filtered) data	Within the dashboard the functionality is added to export queried data directly to CSV. This therefore does allow exporting, however, depending on the intended use, CSV might not be the most user-friendly format, due to its poor human readability.
Transverse to related data is possible	One of the intention was to be able to transvere easily between related data. While the data representations within NeoDash do provide the possibility to gather information on, for example, preceding activities. These can not directly be selected to renew the view based on that activity. One needs to copy the activity title and paste it in the search bar for the dashboard to refresh. This is a limitation of the NeoDash software. While transversal of data is possible, it is thus not the most userfriendly.

5.5.2 Validation interview with process coordinator

Apart from the criteria mentioned in the earlier iterations of the contextual design process, the visioning also yielded three interfaces, each with their own respective purpose. To validate if this purpose has been achieved, the dashboard is again tested with the process coordinator.

The results of this analysis, whether the dashboards fulfill their respective intentions, is given in Table 10.

Table 10: Expert validation

Dashboard intention	Expert validation	Conclusion
Dashboard is able to show upcoming activities	The upcoming activities are clearly shown. And it suffices the expectations after the paper mock-up. Some feedback for another iteration would be the addition of activity duration would be a nice addition. Furthermore, the interface does not show which part of the metadata is used for the ordering of the data. Text fields do not slide, so sometimes information is information is not visible. It would be nice to make the data in the dashboard clickable, to find further associated data.	Dashboard shows upcoming activities. Most additional feedback is due to software limitations.
Dashboard is able to provide insight in preceding activities and associated documents	Interface is clear and clearly shows the progress on activities. As intended, the dashboard shows immediately if an activity is able to start or not. The graph visualization of preceding and succeeding activities is regarded as valuable in showing data relations. It however can improve on readability. Furthermore, it is not preferable that one has to copy activity titles into the search bar. Interaction across the interfaces would be beneficial, where one can interactively click everything to adapt the dashboard content.	Dashboard provides requested insight. Alternative visualisations compared to paper mock-up are successful. Additional feedback can again not be implemented due to software limitations.
Dashboard is able to provide a view for Primavera and Relatics activity comparison	This dashboard is simple in its intention, just like the visualization. With regards to functionality it therefore ticks the boxes. The added functionality of the data path is regarded as valuable, but does require additional explanation. The dashboard could be made a bit more colorful, and the aforementioned clickable data is also requested. It is considered very valuable that data inconsistencies can be identified easily with the use of this tool.	Dashboard provides requested insight. The layout could have been more colorful, but it does have all the requested functionalities.

Furthermore, the system was tested with data from the ‘Kademakers’ project. From testing with this data the disadvantages of the linking through workpackage codes became evident, as there were very limited successful links, due to inconsistent workpackage code use. This again shows the importance of consistent data storage, and agreements on naming conventions. The dashboards however were able to show progress on activities, as well as provide the insight in preceding work.

These findings conclude the validation of the created dashboard. As shown in the assessment, the prototype is able to achieve all of the set goals. However, the completed dashboards were not able to suffice in all of the required features, mainly due to limitations of the NeoDash software. Most importantly, the functionality of clickable data in tables, to investigate further associated data, is a feature which should be prioritized in a next iteration. Furthermore, testing with data from another project showed the importance of consistent working in source systems. The system thus suffices in demonstrating the application and visualization of multi-source data in a single system, and thus acknowledges the potential of graph databases in data integration for the AEC sector.

5.6 Conclusions on implementation

Within the implementation section an application was developed to aid in a standardized working method for data integration of project management information. This application makes use of Neo4j for its data storage, and visualizes the required data through a NeoDash dashboard. For the development of this standardized work method, a conceptual data architecture was proposed. This architecture shows the application if the full potential of all tools involved was to be used. However, in the prototype application development, this architecture was not fully implemented for testing reasons, as well as financial constraints.

The implementation has shown that the creation of a standardized mapping for data structures creates a reusable template for data integration. This mapping makes use of standardized data, which is combined in a desired structure. This ensures the reusability over multiple projects, as long as those project also make use of the same source systems. Furthermore, the data mapping allows the identification of deviations or missing links, as a result of data being not compliant to the created mapping. In the case of a graph visualization, these parts of connected data will simply not be connected to the rest, as there is an inconsistency in the linking mechanism.

The implementation has thus presented a working prototype, which combined with the findings of the contextual design process forms a proposed standardized working method. It provides insight in the progress of activities, and its corresponding documents associated, which is a process which requires the integration of multi-source data.

6. Conclusion

Due to an increasingly complex information landscape within construction projects, individual participants of project teams tend to struggle more and more in finding the correct information. The aim of this graduation thesis is to find a solution to this lack of findability and accessibility of correct project management information in complex construction projects. This problem has been researched according to the following research question:

“How to maintain construction project information findability and accessibility independent of project scale, with the application of a graph database?”

To answer this research question, the following 5 sub-questions were asked:

1. “What types of data are associated to a construction project?”
2. “How can the development of a graph database be made useable and thus accessible to the business?”
3. “What is the current state of art with regards to graph databases?”
4. “What practices are important in data preparation for graph databases?”
5. “What role should existing implemented project management systems fulfill in combination with the graph database?”

These questions find their ground in the large amount of data which is associated with a construction project. Construction projects consists of many stakeholders, each using their own systems. This often causes a scattered data landscape, with semantically rich data spread over multiple systems. This causes ambiguity on data location and status. Which makes the retrieval of correct, actual information within a reasonable amount of time increasingly difficult, especially once the scale of a project increases. At a certain point, one cannot longer rely on the use of their own knowledge on where to find specific data, or who to contact for a certain question, as this amount of information has become too much to handle for a single person. If one is unable to grasp the entirety of such a project, one has to resort to its digital systems and documentation. Yet due to the inherently scattered and temporal nature of the AEC sector, the existing systems also fail to provide the desired clarity.

Before the processes can be improved, they first have to be understood. Literature research has been conducted on the practices involved in the management of a large scale construction project. This is done to provide insight in the data structures employed in project management. While one often refers to BIM practices when talking about the flows of information within a construction project, this modelling aspect only considers a part of the entire information exchange happening during construction. In literature, if it considers current BIM innovation, one mostly talks about the developments being made to standardize the modelling of construction to such an extent, that one only works with one common construction model, hosted in an online environment. In this shift, one changes from working in a file based manner, to an object based manner. This ambition, referred to as BIM maturity level 3, is facilitated through years of effort in creating the IFC standard and implementing it in the various software offerings used to model construction projects, and its sub-components. This ambition is also registered in the recent ISO 19650 norm, which deals with the organization and digitization of information about buildings and civil engineering works.

This is in stark contrast with the efforts on standardization of other information flows that affect construction projects. Contrary to those ambitions, it is identified that of other information flows affecting a construction project do not have the same amount of standardization efforts taken. Project management related practices like planning, documentation of work, risk and quality management and so on, are documented in ISO 21502:2020. This document provides high-level descriptions of practices that are considered to work well and produce good results within the context of project management. However, contrary to construction modelling practices, this standard solely considers naming conventions and best practices, not data standards. With the creation of such data standards, which come with their corresponding advantages, government incentives to more broadly adopt such standards in building legislation have arisen as well. This urges the business to innovate their modelling practices, and adopt practices allowing for the integration of multiple forms of construction modelling, bringing the AEC sector closer to achieving BIM Maturity level 3.

To achieve both BIM Maturity Level 3, as well as decentral storage of project management related data, the concept of a Common Data Environment is proposed. Within such a system, data is gathered in one location. The application of such a CDE however differs, as software vendors offer commercial CDE's, while it is also possible to see the CDE as a data warehouse, serving as a standardized storage of project related data. The intentions of such a system are documented in ISO 19650, however it does not propose a concrete system for implementation. In the case of the data involved in project management, which has a highly semantic, and the difficulty that provides in integrating it with other sources, the application of graph technology is suggested as a form of CDE.

Within the thesis, thus a gap is identified for the integration of project management data. To achieve the provision of knowledge based on multi-source data integration, yet without the luxury of a widely adopted and developed data standard, a new standard is proposed. This standard comes in the form of a revised working method, based on reviewing current business practices. This proposed standard is created through the application of the contextual design process, in which graph methodology is used for the underlying technology. This process resulted in a user interaction design for a data visualization tool, which ensures findable and accessible project information. To identify this process, two process coordinators, of a small and a big project were consulted. They were are asked to identify a common process, which becomes increasingly complex once project scale increases. This process serves as the main goal for the user environment development.

The selected process deals with the retrieval of information on activity progress, and progress on the documentation associated to such activities. Interaction with the process coordinators has shown once the size of the project increases, it becomes impossible to locate each specific part of information, or to simply know who to contact to retrieve the desired information. The case study identified 3 interfaces closely related to each other, based on the working method and preferences of both process coordinators. These interfaces make use of data visualization of multiple sources, that do not natively work together. Especially once the scale of a project increases, inquiring such insights is considered very difficult. In this process, it becomes clear that due to that these systems are not cooperating natively, data of both sources isn't necessarily equal, while it is supposed to be. Therefore the process suggested includes a validation interface, to be able to compare data from different sources.

For the actual development of a working prototype, a graph application had to be chosen. Here one has to make an principle choice between systems, as graph methodology does not employ a single sort of database. In this thesis three different offerings of graph database software have been tested, GraphDB, Weaver and Neo4j. GraphDB being a RDF-based graph, Neo4j being a labelled property graph and Weaver being a hybrid. Each of which is able to host the data needed, but systems each have their own specific unique selling points. GraphDB is able to add additional semantics to its data, through the application of ontologies. Weaver supports a wide variety of file formats for importing, and is able to recreate data structures by itself. Neo4j employs the use of labels, and can hold properties to each part of its data. In the case of this development, Neo4j was chosen due to its embedded and quick mapping, which can be attributed to its labelling and holding of properties functionality, as this was prioritized through business input.

For the use of project management data in a graph, it was found that prior standardization of the data greatly facilitates a quick mapping and loading procedure. This furthermore eases the understanding of data structures present in existing systems. What becomes clear once the data is loaded, is that a large amount of inconsistencies can be found, partially due to human error. If the source software is not used properly, these inconsistencies become visible in the loading of the data. In this case, Relatics and Primavera are independently operating systems, therefore the allocation of 1 universal GUID spanning the platforms is nearly impossible after initiation. Therefore, the matching of data should happen according to other unique characteristics of the data, in this case the workpackage name, or workpackage code. As this considers a string which is once filled in by hand, this is prone to human error. This stresses the fact that either from the start on, each common element should receive an unique global identifier, or if already implemented, source systems have to be filled in with utmost care. These inconsistencies however do cause the difficulty of finding the right information as well, if applied to the as-is working methods. Through graph visualization these parts of information will end up as loose islands not connected to the rest of the data. Early identification of such flaws might aid in keeping information landscapes structured and up-to-date, as irregularities show up easily.

The result of the prototyping showed that end-users don’t prefer graphs over tables for data visualization. However, this partially changes when one desires to transverse through the data, in this case, the graph visualization, especially with relations visible, is considered of added value. The dashboard fulfills all of the goals set in the contextual design process, and does this by combining multi-source data in one system successfully. Therefore it can be said that is has successfully provided access to project management information. Next to this, as the process identified in the contextual design process is present both in a small, as wel as a big project, it can be stated that the dashboard provides a useful working method independent of project scale. The prototype has been tested with data coming from the large scale ‘VeenIX’ project, as well as data from the small scale ‘Kademakers’ project. The data structure was reused 1-to-1, however, due to the inconsistent use of workpackage codes in the ‘Kademakers’ project the system struggled with connecting the data. This shows the important of consistent data creation at the source systems as well.

As identified, there lies a gap in the lack of data standards for project management information. Within this thesis a standard way of working is proposed, which employs a fixed strategy on data integration. Initially this is applied only to two data sources, while within the AEC sector there are many more on offer. Data integration can be achieved with the use of graph modelling, and can be reused for application in projects elsewhere by reusing the mapped data structures. However, this method of integration and reuse will most likely differ if another graph application would have been used to model the database. Labelled property graphs have proven to be suitable to house project management related data, as their structure resembles the structure of elements in planning and work breakdown structure. However, the same can potentially be achieved by applying an ontology to an RDF-based graph, which describes the structure of a planning element, or a workpackage. This could potentially offer additional advantages of derivation and constraint checking, and therefore has potential for further research.

Furthermore, the prototyping of the dashboard has been done in NeoDash, which is a system offered by Neo4j. While the connection of the system works flawless, the system itself shows its limitations. Within the prototyping it was already concluded that not all requested features could be implemented, or could only be implemented in a non-intuitive manner. It is therefore advised to review data visualization of graph databases in other platforms with more flexibility, in the future. The contextual design process was conducted only with the cooperation of two process coordinators, yet the development influences other stakeholders in the project management environment. It would be interesting to go through the steps involved in the contextual design process with those stakeholders to see the differences in preferences. This would also allow another improvement iteration of the developed application.

All in all, the actions taken in this thesis have contributed to providing data findability and accessibility independent of project scale. It does so by implementing a dashboard supported by graph technology, which is part of a standard working method for data integration of project management information.

7. References

Arvidsson, V., & Mønsted, T. (2018). Generating innovation potential: How digital entrepreneurs conceal, sequence, anchor, and propagate new technology. *Journal of Strategic Information Systems*, *27*(4), 369–383. https://doi.org/10.1016/j.jsis.2018.10.001

Attrill, R., & Mickovski, S. B. (2020). *Issues to be addressed with current BIM adoption prior to the implementation of BIM level 3*. https://edshare.gcu.ac.uk/id/eprint/5179

Benjaoran, V. (2009). A cost control system development: A collaborative approach for small and medium-sized contractors. *International Journal of Project Management*, *27*(3), 270–277. https://doi.org/10.1016/j.ijproman.2008.02.004

Berardi, D., Calvanese, D., & de Giacomo, G. (2005). Reasoning on UML class diagrams. *Artificial Intelligence*, *168*(1–2), 70–118. https://doi.org/10.1016/j.artint.2005.05.003

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). *The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities*. http://www.sciam.com/print_version.cfm?articleID=00048144-10D2...

Biplob, B., Sheraji, G. A., & Khan, S. I. (2018). *Comparison of Different Extraction Transformation and Loading Tools for Data Warehousing*.

Borrmann, A., Esser, S., Vilgertshofer, S., & Borrmann, A. (2021). *Graph-based version control for asynchronous BIM level 3 collaboration*. https://www.researchgate.net/publication/352835924

Borrmann, A., Köning, M., Koch, C., & Beetz, J. (2018). *Building Information Modeling Technology Foundations and Industry Practice*.

Brotherton, S. A. , F. R. T. , & N. E. S. (2008). *Applying the work breakdown structure to the project management lifecycle*.

buildingSMART. (2022). *BuildingSmart Solutions and Standards*. Https://Www.Buildingsmart.Org/Standards/.

Chen, W., Leon, M., & Benton, P. (2021). A systematic review of project management information systems for heavy civil construction projects. In *ECPPM 2021 – eWork and eBusiness in Architecture, Engineering and Construction* (pp. 544–550). CRC Press. https://doi.org/10.1201/9781003191476-73

Donkers, A., Yang, D., & Baken, N. (2020). *Linked Data for Smart Homes: Comparing RDF and Labeled Property Graphs*.

Eastman, C. M. (1975). *The Use of Computers Instead of Drawings in Building Design*. https://www.researchgate.net/publication/234643558

Erol, H., Dikmen, I., Atasoy, G., & Birgonul, M. T. (2020). Exploring the Relationship between Complexity and Risk in Megaconstruction Projects. *Journal of Construction Engineering and Management*, *146*(12), 04020138. https://doi.org/10.1061/(asce)co.1943-7862.0001946

Ershadi, M., Jefferies, M., Davis, P. R., & Mojtahedi, M. (2022). The contribution of project management offices to addressing complexities in principal construction contracting. *Engineering, Construction and Architectural Management*, *29*(1), 287–306. https://doi.org/10.1108/ECAM-04-2020-0244

Farghaly, K., Abanda, F. H., Vidalakis, C., & Wood, G. (2019). BIM-linked data integration for asset management. *Built Environment Project and Asset Management*, *9*(4), 489–502. https://doi.org/10.1108/BEPAM-11-2018-0136

Frazao, D. A. G., Costa, T. S. A. da, Araujo, T. D. O. de, Meiguins, B. S., & Santos, C. G. R. dos. (2021). A brief review of dashboard visualizations employed to support management or business decisions. *Proceedings of the International Conference on Information Visualisation*, *2021-July*, 100–107. https://doi.org/10.1109/IV53921.2021.00025

Hala, N., Mahmoud, E. J., & Melanie, P. (2020). *Transforming the AEC Industry: A Model-Centric Approach*. 13–18. https://doi.org/10.3311/ccc2020-076

Holtzblatt, K., & Beyer, H. (2014). *Contextual Design*. Https://Www.Interaction-Design.Org/Literature/Book/the-Encyclopedia-of-Human-Computer-Interaction-2nd-Ed/Contextual-Design.

Isikdag, U., Aouad, G., Underwood, J., & Wu, S. (2007). *BUILDING INFORMATION MODELS: A REVIEW ON STORAGE AND EXCHANGE MECHANISMS*.

ISO. (1994). *ISO 10303-1:1994*. Https://Www.Iso.Org/Standard/20579.Html.

ISO. (2018). *ISO 19650*. Https://Www.Iso.Org/Standard/68078.Html.

ISO. (2019). *NEN-EN-ISO 19650-1*.

ISO. (2020a). *21502:2020*. Https://Www.Iso.Org/Standard/74947.Html.

ISO. (2020b). *ISO 21597*. Https://Www.Iso.Org/Standard/74389.Html.

ISO. (2021). *ISO 21500:2021*. Https://Www.Iso.Org/Standard/75704.Html.

ISO/IEC JTC1 SC32 Working Group. (2022). *GQL Standard*. Https://Www.Gqlstandards.Org.

J. Werbrouck, P. Pauwels, J. Beetz, & L. van Berlo. (2019). *Towards a decentralised common data environment using linked building data and the solid ecosystem*. http://hdl.handle.net/1854/LU-8633673

Kania, E., Śladowski, G., Radziszewska-Zielina, E., Sroka, B., & Szewczyk, B. (2021). Planning and monitoring communication between construction project participants. *Archives of Civil Engineering*, *67*(2), 455–473. https://doi.org/10.24425/ace.2021.137179

Krijnen, T., & Beetz, J. (2018). A SPARQL query engine for binary-formatted IFC building models. *Automation in Construction*, *95*, 46–63. https://doi.org/10.1016/j.autcon.2018.07.014

Kumar, R., Manoj, E., & Rajak, K. (2019). Changing Trends in Construction Project Management: a Review from the History to Present Day Construction Project Management Practices. *International Journal of Civil Engineering and Technology (IJCIET)*, *10*(3), 288–293.

Lluís Larriba-Pey, J., Martínez-Bazán, N., & Domínguez-Sal, D. (2014). *LNCS 8714 - Introduction to Graph Databases*.

Love, P. E. D., Holt, G. D., Shen, L. Y., Li, H., & Irani, Z. (2001). *Using systems dynamics to better understand change and rework in construction project management systems*. www.elsevier.com/locate/ijproman

Luo, L., He, Q., Jaselskis, E. J., & Xie, J. (2017). Construction Project Complexity: Research Trends and Implications. *Journal of Construction Engineering and Management*, *143*(7). https://doi.org/10.1061/(asce)co.1943-7862.0001306

Martínez-Rojas, M., Nicolás Marín, ;, & Vila, M. A. (2015). *The Role of Information Technologies to Address Data Handling in Construction Project Management*. https://doi.org/10.1061/(ASCE)CP.1943-5487

McKinsey & Company. (2020). *The next normal in construction*.

Mesároš, P., & Mandičák, T. (2017). Exploitation and Benefits of BIM in Construction Project Management. *IOP Conference Series: Materials Science and Engineering*, *245*(6). https://doi.org/10.1088/1757-899X/245/6/062056

Noy, N. F., & Mcguinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. www.unspsc.org

Oesterreich, T. D., & Teuteberg, F. (2019). Behind the scenes: Understanding the socio-technical barriers to BIM adoption through the theoretical lens of information systems research. *Technological Forecasting and Social Change*, *146*, 413–431. https://doi.org/10.1016/j.techfore.2019.01.003

Ontotext. (n.d.). *SPARQL vs SQL*. Https://Www.Ontotext.Com/Knowledgehub/Fundamentals/What-Is-Sparql/#:~:Text=SPARQL%2C%20pronounced%20’sparkle’%2C,Can%20be%20mapped%20to%20RDF.

Ontotext. (2022). *About GraphDB*. Https://Graphdb.Ontotext.Com/Documentation/10.0/about-Graphdb.Html.

Ozturk, G. B. (2021). Digital Twin Research in the AECO-FM Industry. *Journal of Building Engineering*, *40*. https://doi.org/10.1016/j.jobe.2021.102730

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. In *The BMJ* (Vol. 372). BMJ Publishing Group. https://doi.org/10.1136/bmj.n71

Pauwels, P., Costin, A., & Rasmussen, M. H. (2022). Knowledge Graphs and Linked Data for the Built Environment. In *Structural Integrity* (Vol. 20, pp. 157–183). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-82430-3_7

Pellerin, R., & Perrier, N. (2019). A review of methods, techniques and tools for project planning and control. In *International Journal of Production Research* (Vol. 57, Issue 7, pp. 2160–2178). Taylor and Francis Ltd. https://doi.org/10.1080/00207543.2018.1524168

Pokorný, J. (2015). Graph databases: Their power and limitations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9339*, 58–69. https://doi.org/10.1007/978-3-319-24369-6_5

Relatics. (2022). *Systems Engineering*. Https://Www.Relatics.Com/Systems-Engineering/.

Rijkswaterstaat. (2022). *Systems Engineering*. Https://Www.Rijkswaterstaat.Nl/Zakelijk/Zakendoen-Met-Rijkswaterstaat/Werkwijzen/Werkwijze-in-Gww/Systems-Engineering.

Rolland, K. H., & Hanseth, O. (2021). Managing path dependency in digital transformation processes: A Longitudinal Case study of an Enterprise Document Management Platform. *Procedia Computer Science*, *181*, 765–774. https://doi.org/10.1016/j.procs.2021.01.229

Senthilvel, M., Oraskari, J., & Beetz, J. (2020). *Common Data Environments for the Information Container for linked Document Delivery*. https://www.w3.org/community/lbd/

Simeone, D., Cursi, S., Coraglia, U. M., & Fioravanti, A. (2020). *Reasoning in Common Data Environments Re-thinking CDEs to enhance collaboration in BIM processes*.

UML-Diagrams. (2022). *UML Class and Object Diagrams Overview*. Https://Www.Uml-Diagrams.Org/Class-Diagrams-Overview.Html.

Vaz-Serra, P., & Edwards, P. (2021). Addressing the knowledge management “nightmare” for construction companies. *Construction Innovation*, *21*(2), 300–320. https://doi.org/10.1108/CI-02-2019-0013

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. In *Journal of Strategic Information Systems* (Vol. 28, Issue 2, pp. 118–144). Elsevier B.V. https://doi.org/10.1016/j.jsis.2019.01.003

Walker, D. H. T., Davis, P. R., & Stevenson, A. (2017). Coping with uncertainty and ambiguity through team collaboration in infrastructure projects. *International Journal of Project Management*, *35*(2), 180–190. https://doi.org/10.1016/j.ijproman.2016.11.001

Wasserman, A. I., Guo, X., McMillian, B., Qian, K., Wei, M. Y., & Xu, Q. (2017). OSSpal: Finding and evaluating open source software. *IFIP Advances in Information and Communication Technology*, *496*, 193–203. https://doi.org/10.1007/978-3-319-57735-7_18

Weaver. (2022). *About us*. Https://Www.Wvr.Io/about-Us.

Werbrouck, J., Pauwels, P., Beetz, J., & Mannens, E. (2022). *LBDserver-a Federated Ecosystem for Heterogeneous Linked Building Data*. https://www.w3.org/RDF/

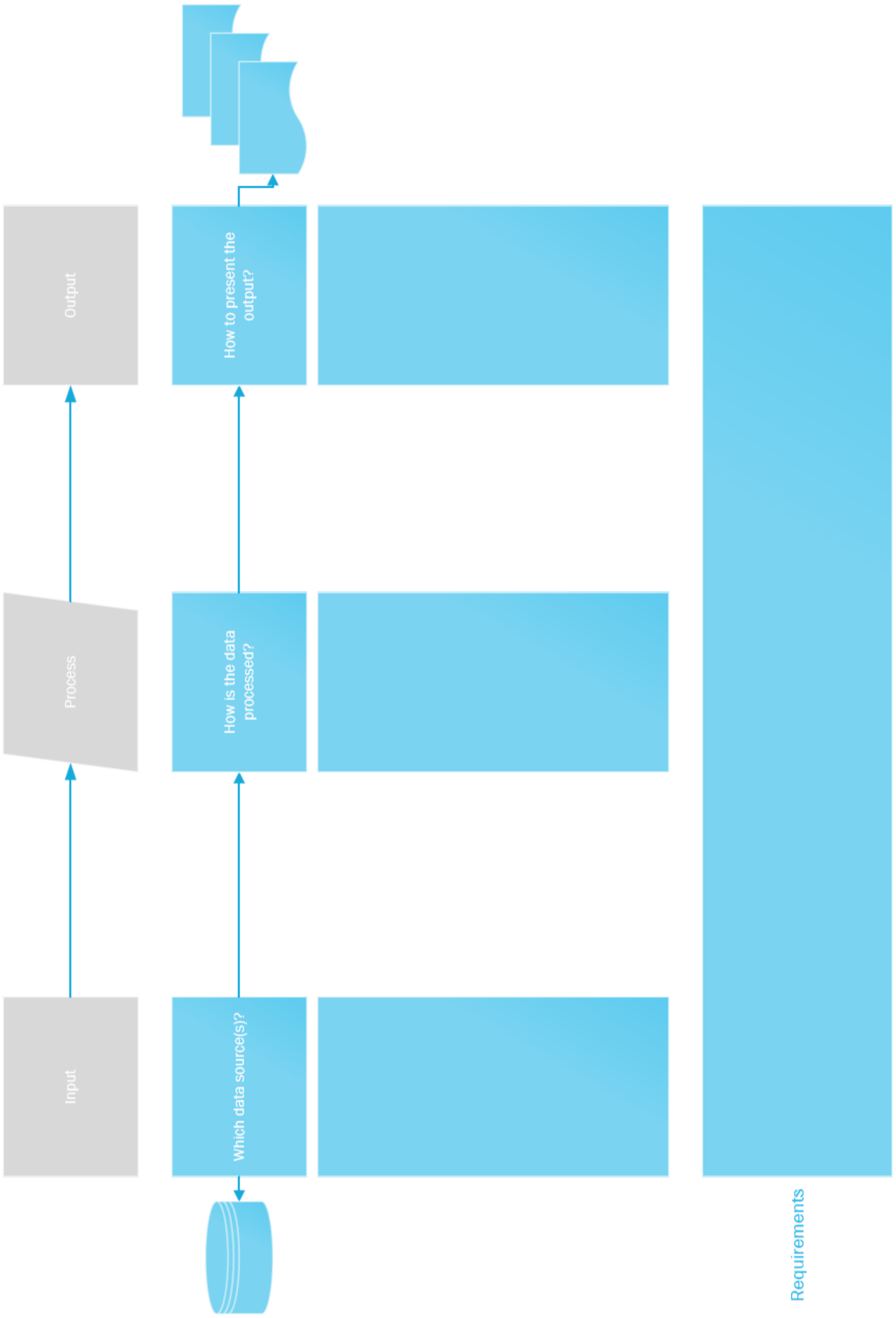
Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, *13*(1), 41–59. https://doi.org/10.1016/j.accinf.2011.08.002

Yoon, B.-H., Kim, S.-K., & Kim, S.-Y. (2017). Use of Graph Database for the Integration of Heterogeneous Biological Data. *Genomics & Informatics*, *15*(1), 19. https://doi.org/10.5808/gi.2017.15.1.19

8. Appendices

Appendix 1: Poster used for interview sessions



Input

Process

Output

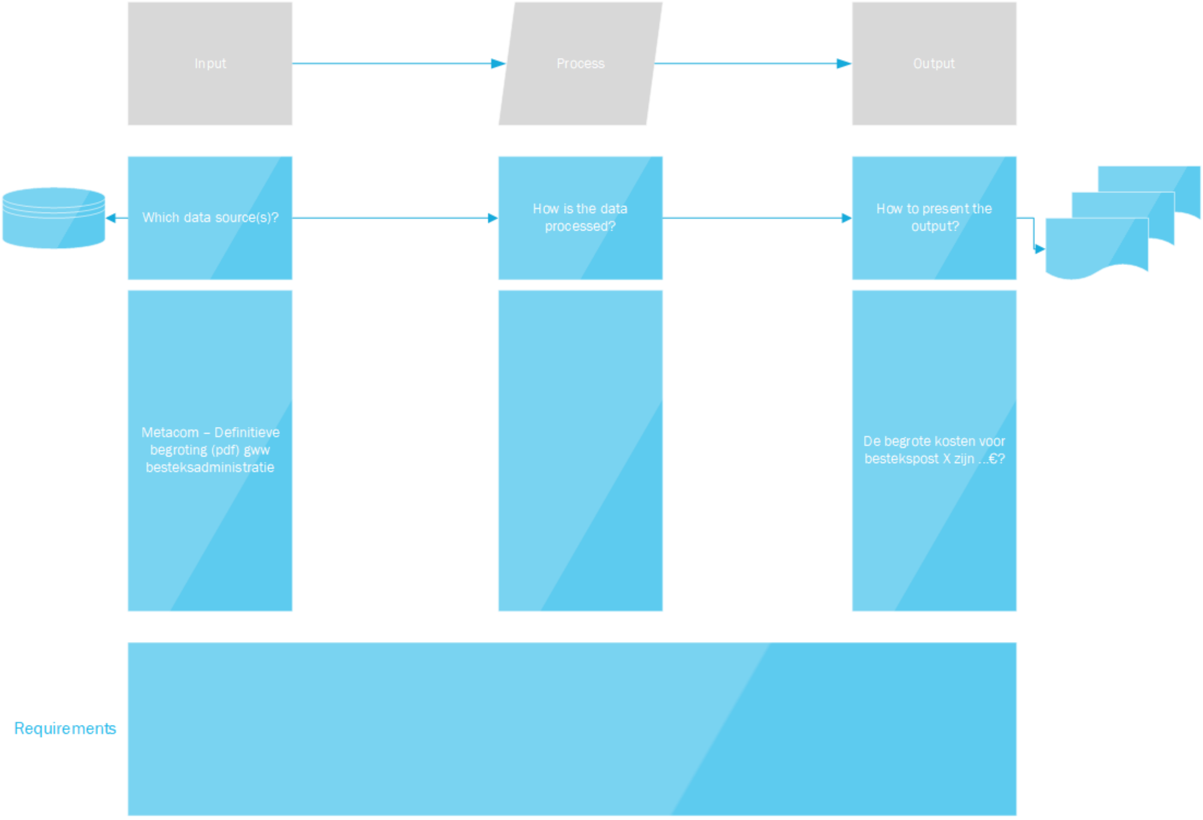
Which data source(s)?

How is the data processed?

How to present the output?

Requirements

Appendix 2: Poster interview session Results of Kademakers project



Input

Process

Output

Which data source(s)?

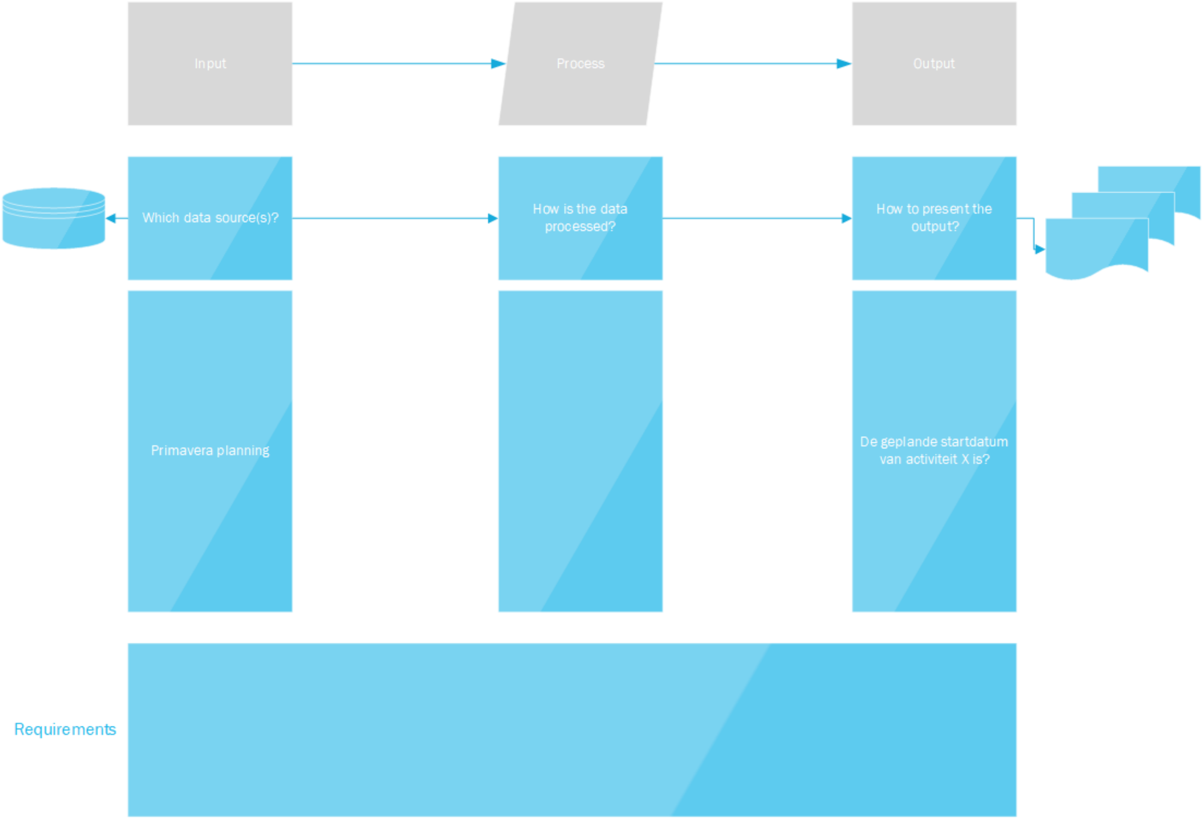
How is the data processed?

How to present the output?

Metacom – Definitieve begroting (pdf) gww besteksadministratie

De begrote kosten voor bestekspost X zijn ...€?

Requirements



Input

Process

Output

Which data source(s)?

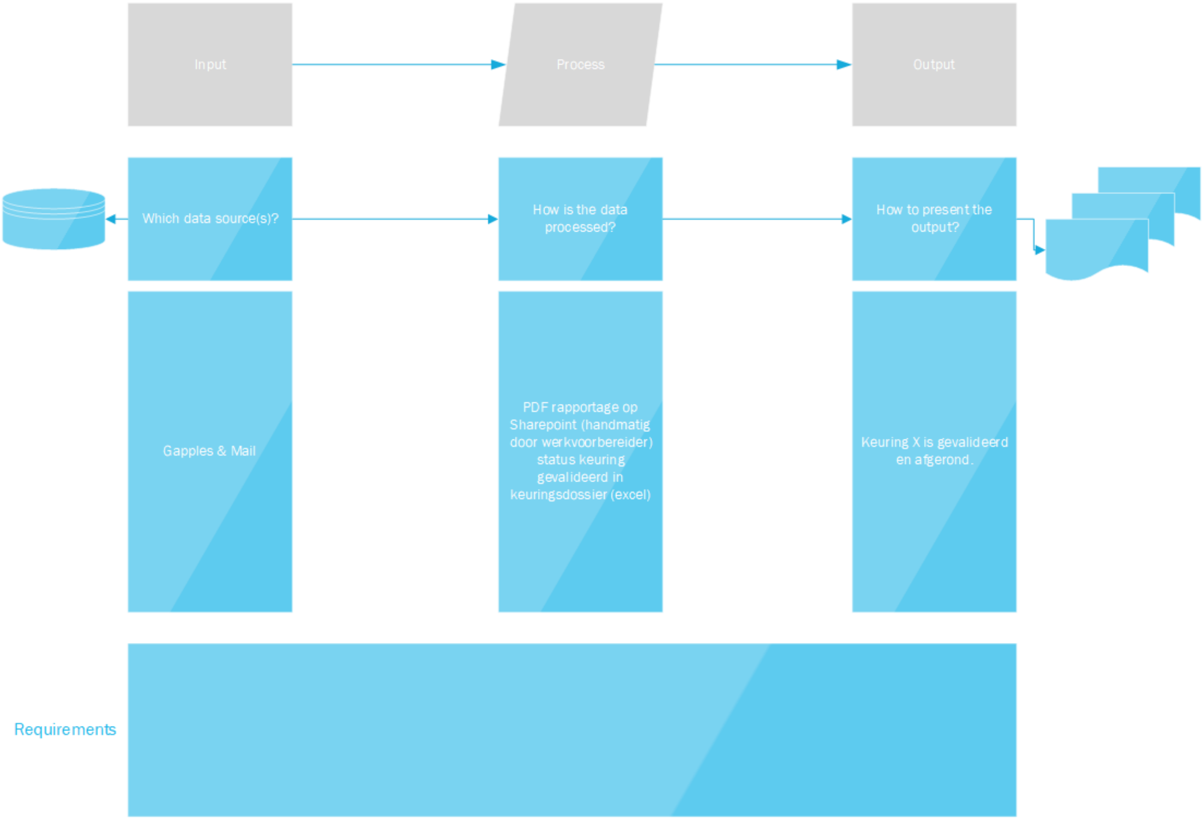
How is the data processed?

How to present the output?

Primavera planning

De geplande startdatum van activiteit X is?

Requirements



Input

Process

Output

Which data source(s)?

How is the data processed?

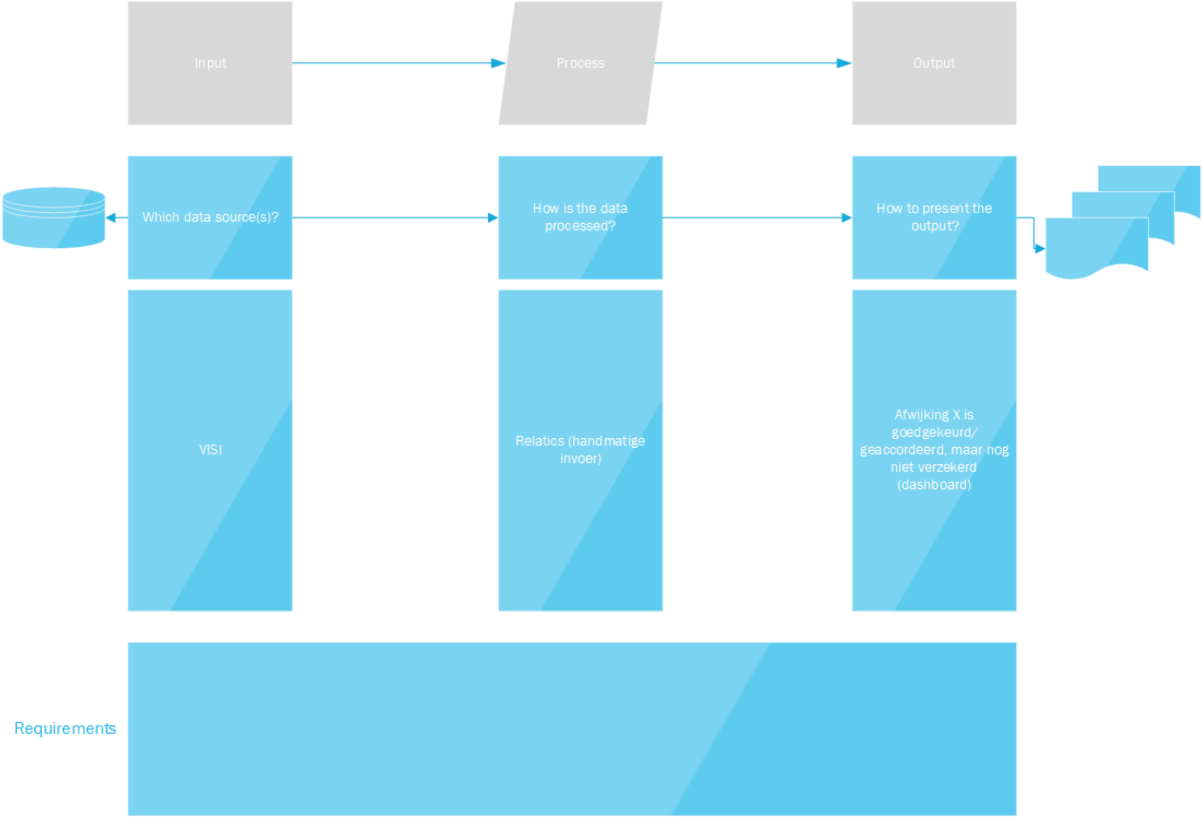
How to present the output?

Gapples & Mail

PDF rapportage op Sharepoint (handmatig door werkvoorbereider) status keuring gevalideerd in keuringsdossier (excel)

Keuring X is gevalideerd en afgerond.

Requirements



Input

Process

Output

Which data source(s)?

How is the data processed?

How to present the output?

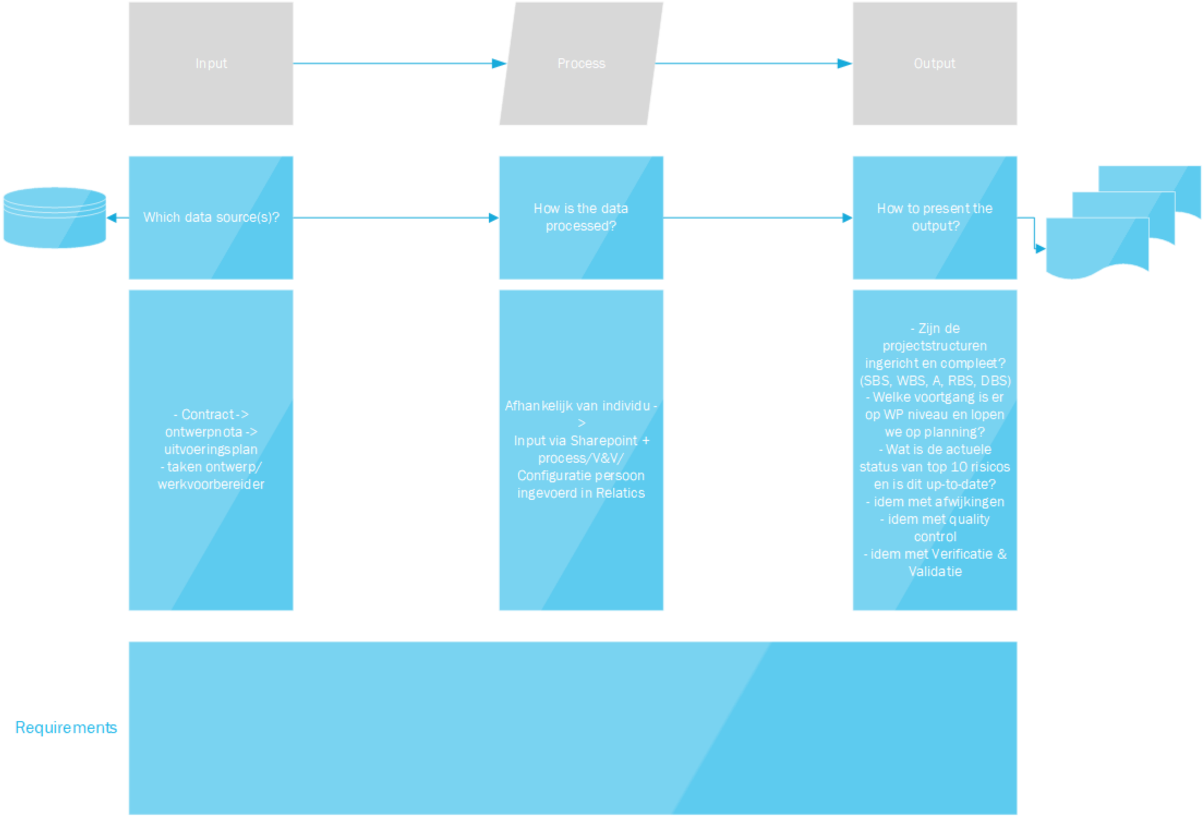
VISI

Relatics (handmatige invoer)

Afwijking X is goedgekeurd/ geaccordeerd, maar nog niet verzekerd (dashboard)

Requirements

Appendix 3: Poster interview session Results of VeenIX project



Input

Process

Output

Which data source(s)?

How is the data processed?

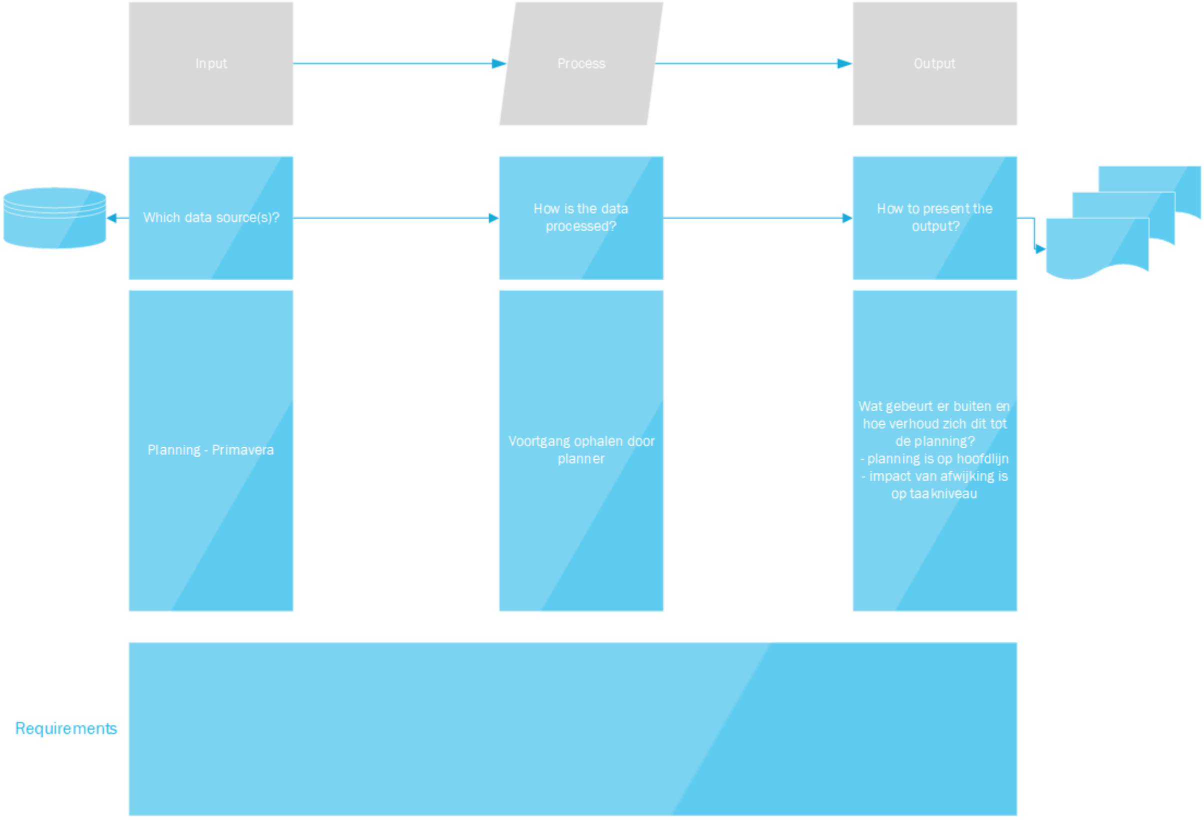
How to present the output?

- Contract -> ontwerpnota -> uitvoeringsplan
- taken ontwerp/ werkvoorbereider

Afhankelijk van individu -> Input via Sharepoint + process/V&V/ Configuratie persoon ingevoerd in Relatics

- Zijn de projectstructuren ingericht en compleet? (SBS, WBS, A, RBS, DBS)
- Welke voortgang is er op WP niveau en lopen we op planning?
- Wat is de actuele status van top 10 risicos en is dit up-to-date?
- idem met afwijkingen
- idem met quality control
- idem met Verificatie & Validatie

Requirements



Input

Process

Output

Which data source(s)?

How is the data processed?

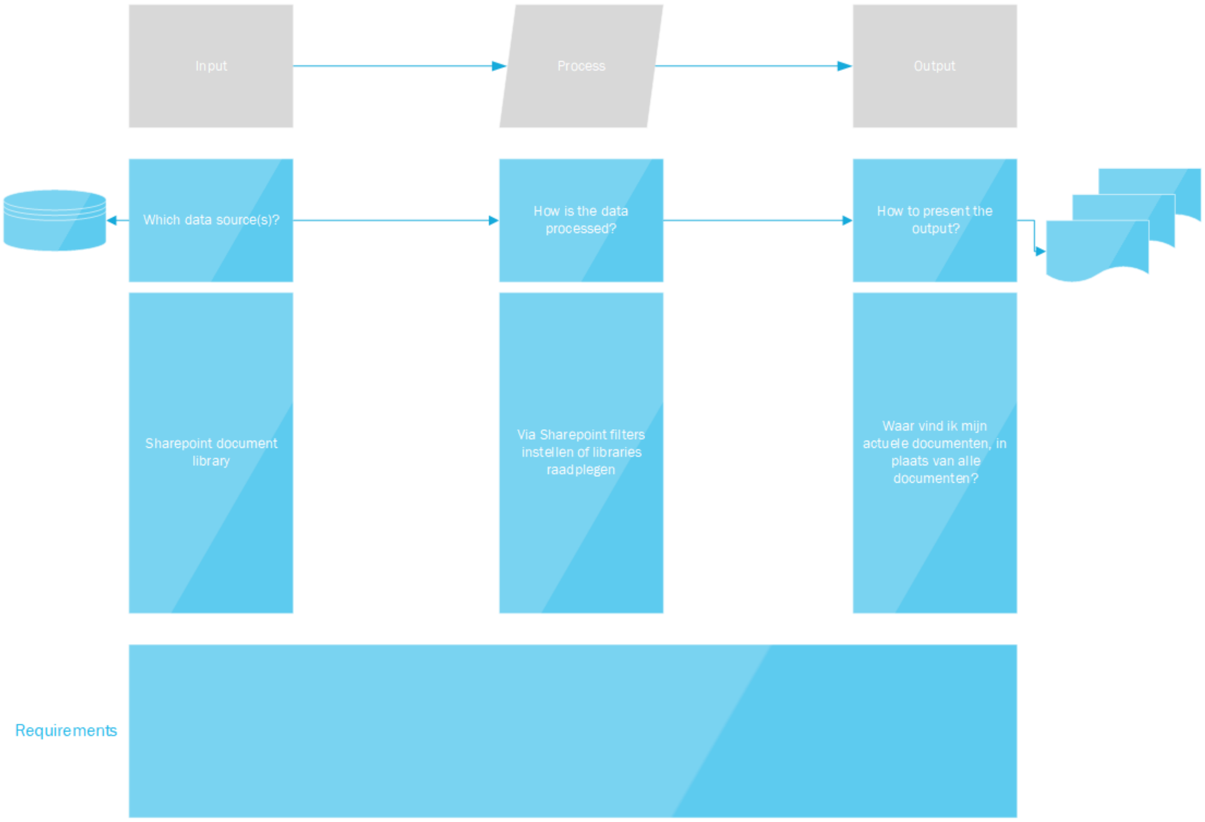
How to present the output?

Planning - Primavera

Voortgang ophalen door planner

Wat gebeurt er buiten en hoe verhoud zich dit tot de planning?
- planning is op hoofdlijn
- impact van afwijking is op taakniveau

Requirements



Input

Process

Output

Which data source(s)?

How is the data processed?

How to present the output?

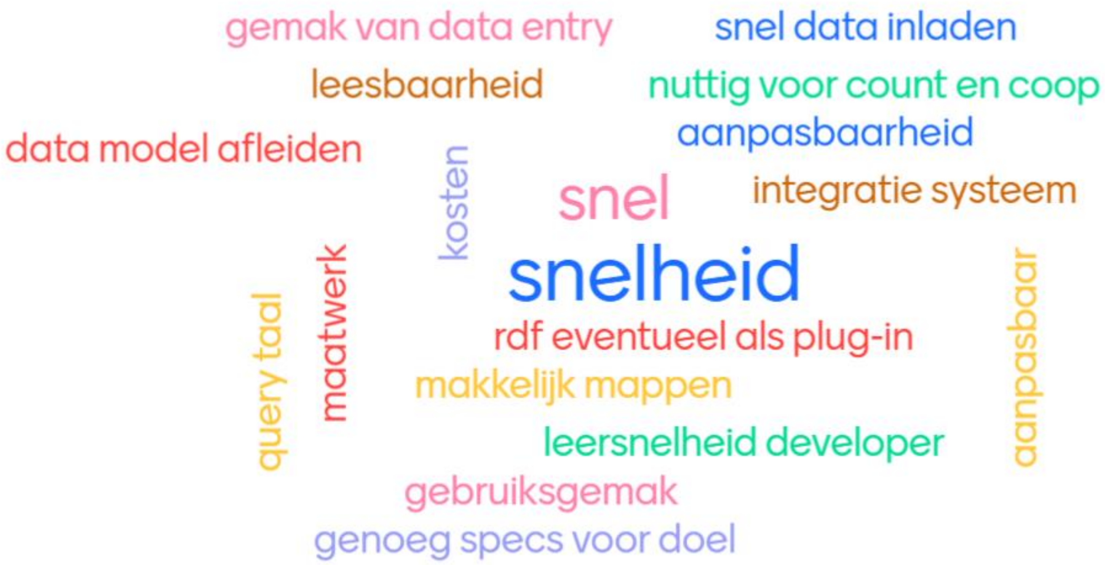
Sharepoint document library

Via Sharepoint filters instellen of libraries raadplegen

Waar vind ik mijn actuele documenten, in plaats van alle documenten?

Requirements

Appendix 4: Word cloud for business criteria


gemak van data entry
snel data inladen
leesbaarheid
nuttig voor count en coop
aanpasbaarheid
data model afleiden
kosten
snel
integratie systeem
snelheid
maatwerk
query taal
aanpasbaar
rdf eventueel als plug-in
makkelijk mappen
leersnelheid developer
gebruiksgemak
genoeg specs voor doel