Camilla Flataukan

Jowita Januszko

Data-driven construction

Data management practices towards improvement of the health and safety environment

AALBORG UNIVERSITY
STUDENT REPORT

The School of Engineering and Science
Study Board of Civil Engineering
Thomas Manns Vej 23
9220 Aalborg Ø
Phone 99 40 84 84
http://www.civil.aau.dk

Project title:

Data-driven construction

Subtitle:

Data management practices towards improvement of the health and safety environment

Project:

Master's thesis

Time period:

1st of February 2018 to
7th of June 2018

Participant(s):

Camilla Flataukan
Jowita Januszko

Supervisor(s):

Kjeld Svidt
Lene Faber Ussing

Editions: 02
Report pages: 90
Completed: 07-06-2018

Synopsis:

The number of occupational accidents and diseases in the Danish construction industry is relatively hight. There are several groups of aspects contributing to the occupational accidents, namely, individual, task, management and environmental. The construction companies might contribute to working environment improvement by carefully distinguishing and analysing these factors. In the era of advanced technological solutions, the new tools might be used in order to overcome health and safety issues on site.

The following report is using data management as the main tool for working environment improvement. Big amounts of data are generated through the whole construction project life-cycle. Data management provides great techniques for analysing and improving the working environment. All process of incorporating data into health and safety practices contain the number of various steps. The relevant data needs to be gathered, collected, analysed and applied in the proper way to provide a value for the company.

The report is providing an universal guidance which might be used by distinct construction companies. This framework is presenting all data management steps, by distinguishing the main technological solutions and real-world examples, when these techniques might be used and in which purpose.

Preface

Master's thesis is completed during the 4th semester of studies in Management in the Building Industry. The report is conducted by two students who adapted the knowledge, skills and experiences gained during the whole period of Master's studies at Aalborg University. The report is conducted within the topic: *Data-driven construction - Data management practices towards improvement of the health and safety environment.*

Acknowledgements

The authors of the report would like to express their very great gratitude for the supervisors who provided a guidance during the process of Master's thesis creation. The advice and constructive recommendations given by them were very helpful for the research work. Lene Faber Ussing supported the students with the great knowledge of the practices in the construction industry. She provided a great support while choosing the direction of the following report. Moreover, Kjeld Svidt supported the students within the technical part of the report. His suggestions and guidance helped to elaborate the topic in the most valuable way. Both supervisors largely supported the students during their last report in Master's programme, therefore, the students are addressing great thanks for their time and engagement.

Participants

Camilla Flataukan

Jowita Januszko

Abstract

The number of occupational accidents and diseases in Danish construction industry is relatively hight in comparison to the other industries. The high number of accidents it is not only an issue in Denmark, it is perceived as a global problem. Some universal patters or factors contributing to the occupational accidents might be distinguished. By knowing the essence why accidents are happening, proper actions might be taken to improve the health and safety practices during the construction projects. One of such tools is management of large amounts of data. The research is combining data management methods and techniques to improve the existing health and safety practices in Denmark. The report is elaborating four steps of data management process which are: data acquisition, access, analytics, and application. To better understand the essence of this process, the theoretical background of each step is presented with the examples form the global practices which might be adapted by the Danish construction companies. Moreover, the SMART model is demonstrated which distinguishes the path for data management process incorporation into the organisations. The data management framework presents the path of traditional health and safety environment transformation into data-driven one.

Keywords: health and safety, Danish working environment, occupational injuries, occupational diseases, data management, data gathering, data engineering, data analytics, data applying, strategy.

Table of contents

Part I

Preliminary considerations

The first part of the report introduces the focus of conducted research and describes an overall structure of the report. It creates a foundation and an explanation of the further research. The research is analysing, how big amounts of data might be collected, analysed, and applied by the construction companies to improve health and safety management procedures during the construction phase of the project. Chap. 1 is conducted in order to introduce the topic for the reader. Further, Chap. 2 provides an argumentation of the topic's importance and why it is relevant for the construction industry. Moreover, the chapter presents the specific problem formulation which is elaborated through the report. In order to conduct valuable research, the working process needs to be planned and organised in the proper way. Chap. 3 provides an extensive elaboration of research methodology which includes the selection of research approach, design and methods.

1 | Introduction

The Danish construction industry is receiving criticism due to insufficient health and safety management [European Commission, 2016; Dansk Byggeri, 2016]. The industry is the second most hazardous, when comparing the number of accidents and occupational diseases with the statistics from other industries in Denmark [Yearbook, 2017; Nielsen, 2007]. In 2017, 169,594 people were employed in the Danish construction industry, while 4,681 of them reported occupational accidents [Yearbook, 2017].

The main causes for occupational accidents in the construction industry, compared to more automated industries like manufacturing, are that the activities conducted during the construction phase of the project are complex and differing in a great extent, and that the construction site and labourers are only for a limited period of time [Liao and Perng, 2008; Kines, 2002; Larsson and Field, 2002; Sawacha et al., 1999]. To address this, the companies and safety managers have introduced risk-based approaches, which are based on proactive decision making. Even if the approaches have shown a positive effect, they are meeting some criticism, mainly because they are breaking down projects in smaller activities. By doing this, the efficient capturing of the transient and dynamic nature of construction work is prevented. Another aspect is that the data used within these approaches is often based on regulations, intuition, or judgement. This makes the data secondary and subjective [Tixier et al., 2017, 2016b; Nielsen, 2007]. These limitations needs to be addressed as occupational accidents cause serious problems for the companies. These issues contribute to loses in human resources as well as influence the budget of construction projects [Liao and Perng, 2008].

To address these limitation, and try to capture the dynamics of the construction site, several researchers has applied state-of-the-art data management techniques. Data management might be perceived as a new thing in the construction industry [Deutsh, 2015]. Moreover, the companies might have difficulties with distinguishing, which type of data is the most beneficial for health and safety improvement, how to capture required data, analyse it, and finally, use it in proper way. The process of data management consists of many techniques which in combination create one structured data environment. Despite diverse nature of data, as it might be captured in various forms, the available data engineering and analytics techniques provide a ways how to gain value form it [Deutsh, 2015]. It is important, that data itself is not as beneficial, only the final results of the whole data management process might improve the health and safety practices and enhance the decision making. It might be perceived as a complicated and a complex process, however, with the proper strategy, the desired results might be obtained.

The strategy of data incorporation into the company's operations is differing from the strategies of incorporation Building Information Modeling (BIM), as the latter basically relies on software and employees' experience with the new tools provided by BIM. In the case of data, the situation is slightly different. While incorporating the data, the software changes or additional technologies are needed, but also the change in general organisational culture regarding data awareness [Deutsh, 2015]. The employees need to understand and be convinced that the problems regarding health and safety practices can be solved within data management tools. Therefore, the report is presenting the possibilities within the data management process and guides the construction professionals on how this process should begin. Data in this report is perceived as a main driver of change. Now it is a time to make the step forward and improve the existing practices with the potential of existing data tools.

2 | Research focus

This chapter aims at introducing research focus, boundaries and problem statement of the following report. This is done to present the topic to the reader. First, the reason for choosing particular topic for the research is provided in Sec. 2.1. Further, the problem formulation is presented in Sec. 2.2, which defines the main issue insightfully analysed through the report. To narrow down the problem statement, research questions and set of objectives were made. These determine the boundaries of the research and help to control that the solution for the given problem is realistic and valuable.

2.1 Research topic

At the beginning of the research process, the initial research topic was chosen in order to set up the boundaries and the focus of further investigation. According to Robson and McCartan [2016], research focus supports the investigator to gather relevant information for the research from the very beginning of this process. In this report, data analytics was selected as an initial topic. The decision of choosing this subject was mostly driven by project groups' interests and concerns. Therefore, the motivation for obtaining the best final output is much higher, moreover, the research process is less discouraging in stressful situations which might be caused by time limitations [Robson and McCartan, 2016]. Additionally, some of the external constraints were taken into account while choosing the focus of the research, for instance, availability of the resources.

Based on knowledge obtained during the studies in Master's programme in Management in the Building Industry, also from experience and observations gained during the academic internships, the authors of the report observed that the construction industry is dealing with large amounts of data generated through the project life-cycle. Due to the extensive range of stakeholders and the complexity of operations, the process of effective data integration into day-to-day activities is becoming more difficult. The global construction industry is slowly entering into the new generation of data management. Many different applications of data are determined, for instance, for improving cost planning and scheduling, labour performance evaluation, safety enhancement on site [Motawa, 2017]. Nowadays, the technology is on a very advanced level and data collection is more extensive and common than ever before [Lavalle et al., 2011]. The question is if the organisations are getting and understanding possible advantages and benefits of using data. The benefits are not only the storage and easy access to data, the most important, data analytics provides extensive tools for analysis of information which contributes to the development of new solutions. By analysing big amounts of information, the real value might be captured and used in the proper way when needed [Koseleva and Ropaite, 2017].

To explore possible topics of data analytics in the construction industry, the literature review was made, see Sec. 3.4.1, and the results are presented in Fig. 2.1. While reviewing various literature, some patterns were distinguished and observed that the topics might be divided into four general areas which are: health and safety, project planning, productivity and environmental issues. By using data analytics in different areas of the construction project, many various issues might be resolved.

In order to select one topic for further investigation, first, the indicators affecting project success were determined. According to Chan and Chan [2004], the project success might be explained as the group of particular principles and specifications which enable to complete the project with the best final outcome. However, these principles and specifications might be differently understood by various parties involved in the project. Depending on the construction project phase the favourable outcome might differ in some extent, for instance, the objectives of owners, designers or consultants might be different than contractors or subcontractors. For some stakeholders, the aesthetics of construction is considered as the most important factor while for others it might be cost or construction time [Chan and Chan, 2004].

Data in the construction industry:	
1. Health and safety	• Safety analysis and work injuries • Social patterns in work groups • Identifying actions of workers and heavy machinery towards safety improvement in site • Analysis of human performance and behaviour
2. Project planning	• Identifying causes of construction delays • Legal decision support system • Cost and schedule integrated planning • Real-time access to cost, schedule, material utilisation and installation data • Risk imposed by schedule and workspace planning • Cost overruns and quality control
3. Productivity	• Construction data integration for enhanced productivity • Labour productivity
4. Environmental issues	• Energy sustainability and building performance • Waste utilisation • Site selection

Fig. 2.1: Overview of different topic choices related to data in the construction industry.

The most common and well-known concept of project success is called *iron triangle* which is represented by time, cost and quality [Williams et al., 2015; Atkinson, 1999; Hatush and Skitmore, 1997; Belassi and Tukel, 1996]. Due to the dynamic development of project management, it was suggested to supplement the basic concepts with some additional factors [Chan and Chan, 2004]. According to Pinto and Pinto [1991], despite time, cost and quality, the success of the project should be also measured through psychological lenses. Psychological factors indicate the satisfaction of stakeholders and quality of interpersonal relations between them. Satisfaction of project participants is a subjective measure and is also called a soft measure [Chan and Chan, 2004]. Further, Pocock et al. [1996], determined the lack of legal claims as an additional factor affecting project success. This factor is closely related to health and safety management on site since the injuries or accidents might cause legal claims, affect the financial side of the projects or even delay its completion. Many other researchers included health and safety as an important factor affecting project success [Lim and Mohamed, 1999; Kumaraswamy and Thorpe, 1996; Kometa et al., 1995]. However, in many situations, safety precautions are neglected by clients, project managers, and supervisors, additionally, the productivity enhancement is becoming emphasised more rather than safety [Leung et al., 2012; Hughes et al., 2004]. Due to the fact that health and safety management is the important factor affecting the success of the project, moreover, it is often ignored and not planed in the proper way, the authors of the following report decided to look more insightfully into this topic. Subsequently, the main focus of the following research was determined and developed more detailed by research problem formulation, see Sec. 2.2. The following section determines the specific problem area which is insightfully analysed through the report.

2.2 Problem formulation

The construction industry might be perceived as highly dangerous due to its complexity and rough work environment, moreover, the construction operations are often highly risky [Guo et al., 2017; Hallowell, 2012]. The high rate of injuries is recognised in the global construction industry, both in developing and developed countries, and Denmark is not an exception [Guo et al., 2017; Yearbook, 2017; Teran et al., 2015]. The government and the construction industry professionals are paying more attention in order to find strategies how to improve health and safety management and reduce the number of accidents. During the construction phase, the supervisors work towards improving working environment by using various forms of safety management, for instance, providing personal protective equipment, safety indications or training. However, these actions do not provide the satisfactory results towards safety improvement on site [Guo et al., 2017; Kim et al., 2017; Zhou et al., 2015]. Personal protective equipment only partly secures the workers, the safety indications often do not present the real world examples and safety regulations do not work in proper way, if these are incorporated by the companies in the wrong manner [Kim et al., 2017; Coglianese et al., 2003; Harvey et al., 2001]. Moreover, the nature of construction projects, harsh working environment, heavy machinery or noise does not help to minimise the negative outcome of these aspects [Kim et al., 2017].

Accidents during the construction phase of the project might be caused by various factors, for instance, inappropriate planning of site layout, incorrect usage of the materials, tools and machinery or trade workforces [Park and Kim, 2013]. Planning of safety management process is an important step in safety improvement, as it enables to effectively identify possible risk sources and encourage communication regarding safety issues between different parties [Wilson and Koehn, 2000; Sawacha et al., 1999]. The general tendency in the construction projects might be visible, that existing safety management practices do not reflect the real-time work environment [Park and Kim, 2013]. Usually, the data for safety planning on site is outdated and proper identification of risk sources is becoming more difficult [Golparvar-Fard et al., 2009]. Due to increasing trend of multiculturalism in construction projects, the safety management required additional visual-based training and tools for better understanding safety issues on site for foreign labour. Few of the available tools for safety management improvement are BIM, virtual reality or management of large amounts of data [Park and Kim, 2013; Kiviniemi et al., 2011; Hadikusumo and Rowlinson, 2002].

The global practices show that the data management process is positively influencing the construction companies to take proper actions, encourages effective decision making and increases overall competitive advantage [Marr, 2015]. Moreover, it minimises the cost and time of the construction process. The companies are generating large amounts of data however, they are not capturing the value from it. Data is available in different stages of the projects, however, not many data usage practices were developed particularly for management of health and safety practices. The industry professionals need to understand and be convinced that the problems regarding health and safety environment might be solved with the use of data management tools. Therefore, this repost is analysing the whole process of data management to see if the health and safety environment might be improved. The problem statement was made as follows:

What kind of data could be collected, and which tools within data analysis can be used to improve health and safety environment during the execution phase?

The report is conducted based on the data management framework presented by [Mello et al., 2014]. First, the overall situation of health and safety management in the industry is introduced. To understand the essentials of the topic, the explanation how data is perceived in this report is provided. Further, the methods of obtaining and analysing the big amounts of data are elaborated. The findings within data analytics are not valuable if the industry professionals are not convinced about the positive influence of the final outcome, therefore, some application methods of data are provided with specific examples. This topic is relatively new in Danish construction industry, so the initial steps how to start working with big amounts of data are presented. The problem statement is supported by research questions and objectives which are elaborated through the entire report, see Tab. 2.1.

Tab. 2.1: Representation of research questions and objectives.

Number	Research question	Research objectives
1	*What is the overall situation of health and safety management performance in the Danish construction industry?*	• Determine current practices used within the working environment • Identify the factors affecting occupational accidents occurrence
2	*What are the potentials of data which might be used in the construction industry?*	• Establish the explanation of data • Determine current state of data usage in the construction industry and within the working environment
3	*Which techniques and methods might be used for collecting and analysing data?*	• Distinguish what kind of data is useful for health safety management improvement • Identify the opportunities within each of the data management processes • Indicate the necessary resources in order to conduct the analysis • Present the real examples from the global construction industry
4	*What are the initial steps towards obtaining data-driven health and safety management practices?*	• Establish the strategy for industry professionals in order to start the process of data incorporation into safety management practices

The report is built based on the research questions, as following chapters are explaining each of the issue presented above. The objectives help to remain concentrated on the outcome, which contribute to solution determination for stated problem. The detailed explanation of the report structure is presented in Chap. 3.

3 | Methodology

This chapter aims at determining a path of research study process. In order to conduct the research, various preliminary considerations need to be taken into account. The chapter starts with the introduction of basic research approaches which indicate the framework of the research process. To enlighten the basic principles of the investigators, a philosophical worldview is presented in Sec. 3.2. Further, Sec. 3.3 elaborates the research design with the graphical representation of the overall structure. Moreover, methods used for literature collection are presented in Sec. 3.4. The chapter ends with the selection of approach for the following research, see Sec. 3.5. The framework for Chap. 3 is chosen based on Robson and McCartan [2016] and Creswell [2013].

3.1 Research approaches

In the academic literature, there is no clear definition what research is. Different parties understand it in various ways, however, some basic features of the research might be distinguished. **Research** might be explained as a systematic and methodical process of topic exploration and investigation. The main goal of conducting the research is to gain knowledge increase and provide a value for interested in the topic [Amaratunga et al., 2002]. Based on Buckley et al. [1976], the research should meet several conditions. First, the problem is investigated in organised way and elaborated by using relevant scientific methods, while the information gathered for the investigation, is relevant for analysed problem. The researcher should avoid the bias, which might affect the outcome of the investigation. Further, the final results are realistic and possible to validate in order to prove the value of the solution proposed for given problem. The authors of the following report were aiming for the creation of a valuable research, therefore, the main conditions, presented by Buckley et al. [1976], were taken into account.

According to Creswell [2013], **research approach** is a plan which contains a list of steps and methods used in order to move the investigation form the initial extensive assumptions to detailed and deliberative data selection, analysis and judgement. While selecting the approach, some decisions should be taken regarding philosophical assumptions, design, and methods. Three commonly known research approaches might be distinguished and these are qualitative research, quantitative research, and mixed research approaches [Bryman, 2016; Robson and McCartan, 2016; Creswell, 2013].

To gain a better understanding of particular research approaches and provide the reasoning for choosing approach for this report, each of them is briefly presented in the sections below.

Qualitative research approach

According to Creswell [2013], the qualitative approach is connecting questions and procedures in order to investigate the social problems. Qualitative research emphasises on observations and words which are indicating real world and natural situations of the individuals [Atkinson, 1999]. During the process of data collection and analysis, this approach highlights words rather than measurements [Bryman, 2016]. The research is handled by using observations of real-life situations. Data is perceived as qualitative if describes various issues of individuals, objects or situations [Atkinson, 1999]. The researchers which are working with this approach tend to use inductive research process, see Fig. 3.1, meaning, they concentrate on problem complexity and overall analysed situation [Creswell, 2013]. The inductive research process is created based on the relationship between theory and research, where theory is generated through the real world observations [Bryman and Bell, 2011].

Quantitative research approach

The quantitative approach has a strong tradition in academic environment, where the numbers are emphasised and research relies of hard data [Atkinson, 1999]. This approach is focusing on verification of objective theories, meaning, analysing the dependencies between different variables. The values of these variables are obtained by various measurements. The results are usually available in the numeric system, therefore, they might be analysed by statistical methods [Creswell, 2013]. By using quantitative approach, the researchers tend to test the theories in deductive way, see Fig. 3.1. Deductive research process is testing the theory by referencing to hypothesis which is made based on theoretical framework [Bryman and Bell, 2011]. Deductive research process enables researcher to avoid the bias or helps to reproduce the findings [Creswell, 2013]. In this research process, the existing theories are tested based on real-world situations [Dubois and Gadde, 2002].

Mixed methods research approach

Based on years of experience, scientific researchers concluded that the best outcomes might be gained if combined qualitative and quantitative research approaches together [Atkinson, 1999]. By using mixed approaches, the researcher might look at the same topic from different and new perspectives, therefore, it might be analysed in greater detail [Atkinson, 1999; Rossman and Wilson, 1985]. Mixed methods research approach connects various philosophical approaches and frameworks. The major advantage of choosing the mixed approach is that it provides better tools for problem understanding and analysing than if the researcher would use qualitative or quantitative approaches separately [Creswell, 2013]. Moreover, despite inductive and deductive research processes, the abductive process might be used in case of creative research [Spens and Kovács, 2006]. Abductive approach is helpful if research is aiming at designing new relationships between different variables. Development of new theories is emphasised more rather than validating existing ones [Dubois and Gadde, 2002].

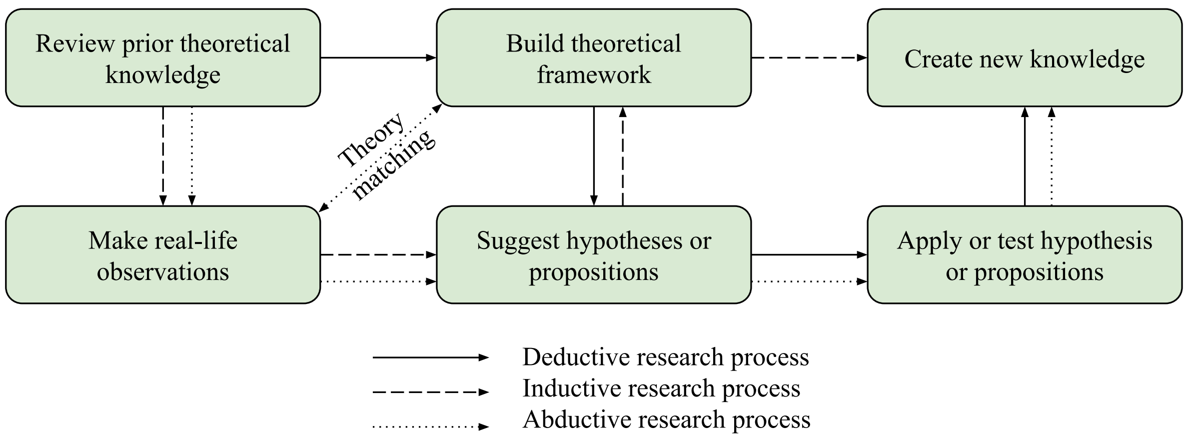


Fig. 3.1: Deductive, inductive and abductive research processes [Spens and Kovács, 2006, *ed.* Fig.1].

To design valuable research, where all parts are logically connected, various components of research approach should be distinguished and analysed. According to Creswell [2013], the research approach might be described in terms of philosophical worldviews, designs and methods, see Fig. 3.2. All the parts are connected together and these are interacting with each other.

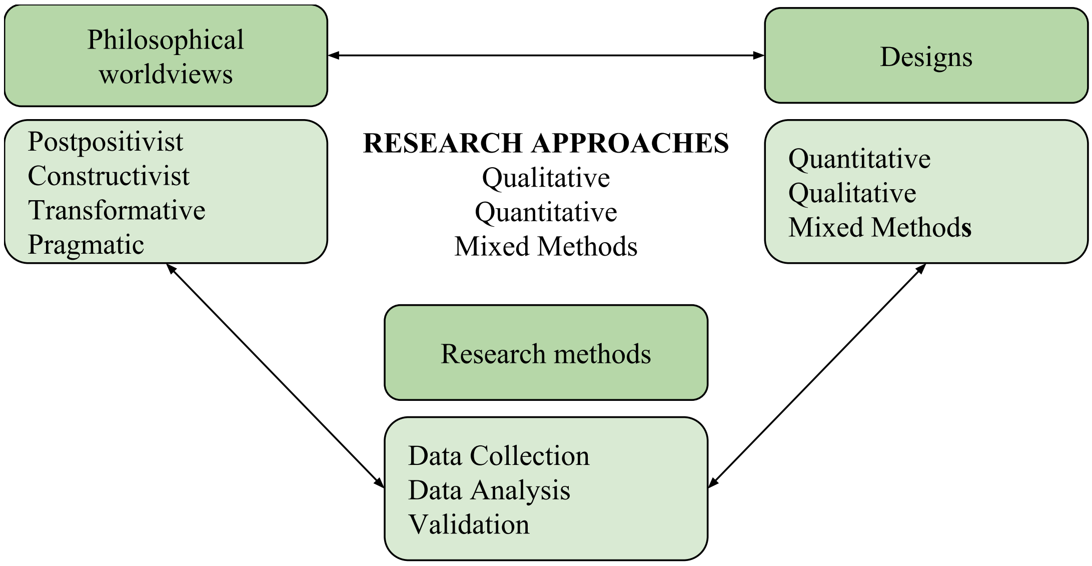


Fig. 3.2: The interconnections of worldviews, design and research methods [Creswell, 2013, *ed.* Fig. 1.1].

To select research approach, which is the most appropriate for conducting the following investigation, the components, which compose the approach are determined and analysed in the sections below. Based on following information and the adaptation of the existing research theories, approach and process for research creation, which are the most reflecting the purpose of the investigation, are chosen and presented in Sec. 3.5.

3.2 Philosophical worldviews

Philosophical perspectives are not directly exposed in the research, however, these might affect the choices of researcher regarding choosing approaches and methods. In this report, **worldview** is perceived as a group of fundamental beliefs that lead the further action of the researcher. In other scientific literature the worldview is also known as paradigm, epistemology or ontology [Creswell, 2013; Bryman and Bell, 2011]. The worldview might be influenced by the number of factors such as field of study, inclinations of the author's knowledge or past practices. Creswell [2013] distinguishes four well-known and widely discussed worldviews which are as follows:

- **Postpositivism** - it is the deterministic philosophy which emphasises that causes are directly connected with effects and final outcomes. It is also known as science research. The nature of research is more quantitative than qualitative.
- **Constructivism** - it is a worldview which is based on the understanding of the world in which individuals are living and working. The researchers tend to elaborate multiple theories and determine their own meanings. Constructivism is usually found in qualitative research.
- **Transformative approach** - an important part of this worldview is the linkage between politics and social actions. The approach in the great extent is power and justice orientated and tends to obtain solutions for change.
- **Pragmatism** - this approach enlightens actions and final consequences rather than causes. The understanding of the problem is the most important in the research. Pragmatism is usually applied in mixed research methods.

To better understand the way of working and conducting following research, the worldview of the authors is presented. The report is made by two students which have the educational background in civil engineering. Before the beginning the studies in current Master's programme of Management in the Building Industry, the worldview of project participants were considered as postpositivist. Due to engineering mentality, the authors emphasised the reasons for the final results as they are, for instance, in experimental works. The knowledge in terms of postpositivist worldview is obtained based on observations and measurements [Creswell, 2013]. However, the worldview of the authors has changed while the students started the Master's programme. The nature of project management discipline differs from civil engineering as well as the worldview of the researchers. During the last two years of study in this programme, the philosophical perspective of the authors switched from postpositivism to the combination of constructivism and pragmatism. The main features of these three approaches are visible in Fig. 3.3.

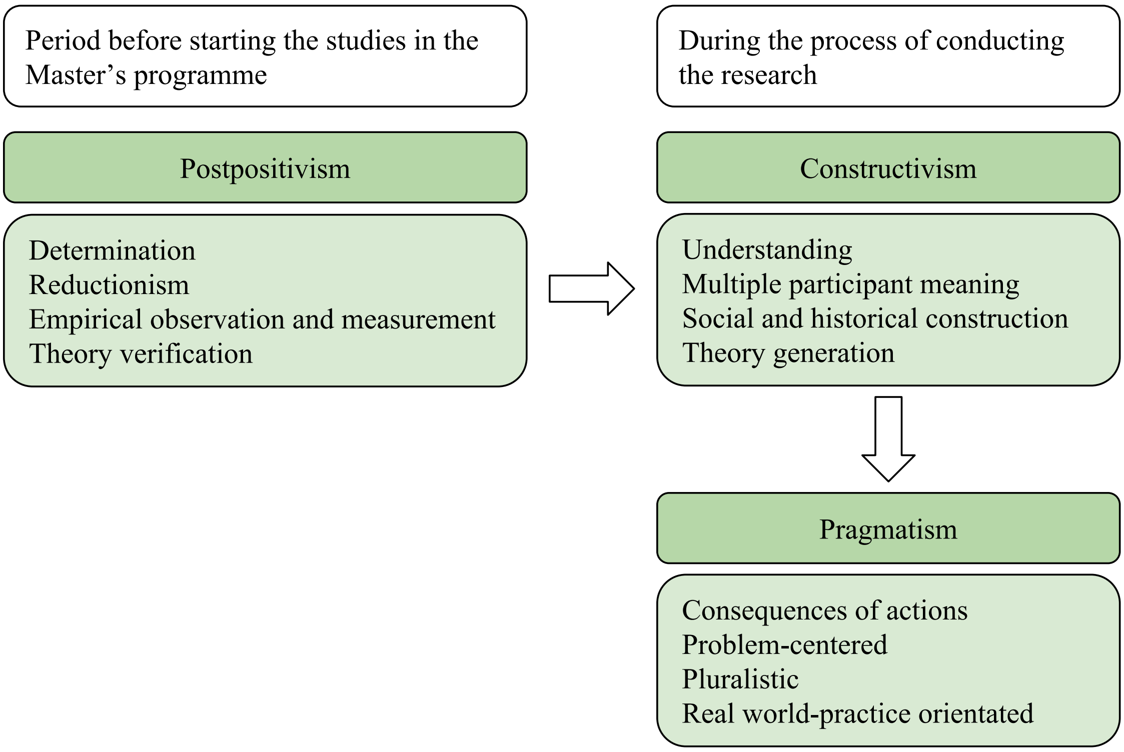


Fig. 3.3: Main features of postpositivism, constructivism and pragmatism [Creswell, 2013, *ed.* Tab. 1.1].

At the beginning of the following research process, the authors attempted to look into the research from constructivist perspective. The authors of the report did not have a specific problem formulation from the starting point. They had a specific experience and knowledge and tried to connect the different disciplines in one report. The initial idea for the project was the application of large amounts of data in the construction industry, known also as big data. The topic was elaborated form different views, and various meanings were analysed in order to find the best connection between data and processes in the building industry. The open-minded review of the topic helped to discover the complexity of views and ideas presented in the scientific literature. The authors looked into various explanations of data, types of data collection and analysis in all construction project phases.

After the initial research focus initiation, the project group realised that the research tends to be too broad and general. The topic of big data in the construction industry is extensive and addresses many different issues. It was obvious that with available resources it might be challenging to conduct valuable report with the clear solution in the end. The constructivist worldview was not effective enough and project group changed the pattern of conducting the research. The authors tried to look at the topic from pragmatic perspective. The major challenge at this point was narrowing the research focus and start to be more problem-specific. Therefore, the following report starts from the presentation of the research focus and the specific problem formulation. By using the list of research questions, presented in Tab. 2.1, the researchers are able to prepare the reliable solution at the end of the report. The authors chose the most appropriate methods, techniques, and procedures in order to meet the project purpose. Based on pragmatic worldview, the researchers are not using one strict approach to collecting and analysing data. The mixed research methods are used in order to obtain the best understanding of given problem [Creswell, 2013].

3.3 Research design

Besides the selection of research approach, the researcher needs to specify the type of study within qualitative, quantitative or mixed methods approach. **Research design** provides a particular direction how the research should be proceeded [Creswell, 2013]. It might be also perceived as a plan outlining actions needed to reach the final goal. This is a long path which includes many steps in between the beginning and last phase of the investigation [de Vaus, 2006]. Moreover, research design is a scheme that explains which questions are investigated, what is relevant for the research, which data to collect and how to analyse it to get the desired results [Philliber et al., 1980]. In the other scientific literature, the research design is also known as a strategy [Creswell, 2013; Denzin and Lincoln, 2011].

Tab. 3.1: Types of research designs [Creswell, 2013, Tab. 1.2]

Quantitative	**Qualitative**	**Mixed Methods**
Experimental designs	Narrative reasearch	**Convergent**
Nonexperimental designs, such as surveys	Phenomenology	Explanatory sequential
	Grounded theory	Exploratory sequential
	Ethnographies	Transformative, embedded or multiphase
	Case study	

According to Creswell [2013], research design might take one of three forms, which are qualitative, quantitative and mixed methods designs of conducting the research. Together with the development of new technologies, advanced data collection, and analysis, the research designs have been changing during through the years. New procedures might be distinguished in order to conduct effective research [Creswell, 2013]. Basic types of three research designs are presented in Tab. 3.1.

The following research is conducted based on mixed methods research design. Mixed design combines both qualitative and quantitative researches and types of data [Creswell, 2013]. By collecting data in qualitative and quantitative forms, the weaknesses and bias of particular data might be overcome as it complements each other. Due to increasing popularity of data triangulation, the new research design types were distinguished within mixed methods, see Tab. 3.1 [Creswell, 2013; Tashakkori and Teddlie, 2010]. Each design provides the different way of looking at data collection and analysis, for instance, in an explanatory sequential method, the quantitative data is selected and analysed first and afterwards followed up with qualitative data.

This report is conducted based on convergent parallel mixed design method, see Fig. 3.4. In this design method, the researcher combines qualitative and quantitative data in order to insightfully analyse the formulated problem. Moreover, both types of data are collected parallel through the research process. Both types of data are analysed separately and then studied in order to confirm if the results prove and verify each other [Creswell, 2013]. In this report, the desk information obtained through the literature review is supported by the different kinds of information obtained from the industry. Quantitative information, for instance, statistics or annual reports, are underpinned by knowledge of industry professionals and observations. Convergent parallel research design supports the main purpose of the research, and enables data validation which is used for valuable solution creation.

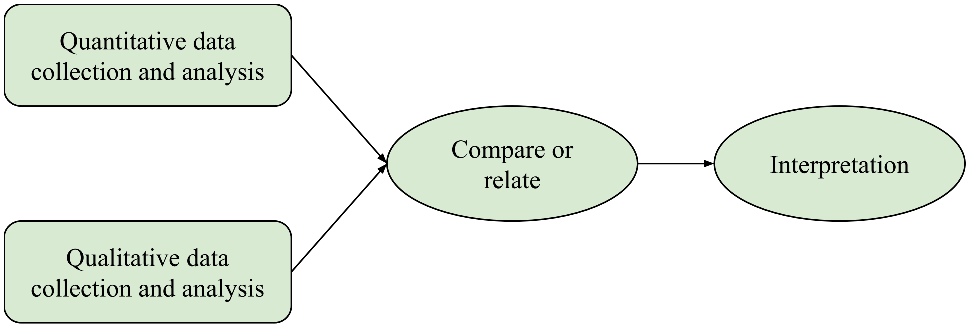


Fig. 3.4: Convergent parallel mixed methods design [Creswell, 2013, Fig. 10.1].

To better understand the research process flow and the structure of the report, the graphical representation of the following report is presented in Fig. 3.5. The report contains five main parts and these are as follows: preliminary considerations, topic review, solution, strategy, and ending. Part I introduces the reader to the topic analysed through the report. The path from initial research idea to the specific problem formulation is described here. Methodology demonstrates the framework, how research is conducted. The approach is insightfully elaborated with its main parts, which are worldview, design, and methods. Part I creates the foundation for the following report, as presents the way of working and the topic chosen for the investigation.

Part II of the report is analysing two main topics which are health and safety environment and data management. Chap. 4 describes the working environment in Danish construction industry and distinguishes the main factors contributing to occupational accidents. The analysis is conducted in order to discover the weak points and issues regarding safety during the construction projects. Further, Chap. 5 introduces data management, which is the foundation for solution creation. Moreover, the explanation, how data is perceived in the following report, is presented. Following part of the report responds to two first research questions.

After the insightful elaboration of the problem area, the solution is prepared in Part III. The following part of the report is conducted based on the steps of data management which are: data acquisition, access, analytics, and application. Chap. 6 is presenting the main techniques of data gathering for the further analysis. Further, Chap. 7 is elaborating the data engineering techniques which are: storage and processing. The next step in data management process is data analytics, and the topic is explained in Chap. 8. Finally, Chap. 9 is presenting the process of data application which is based on data visualisation, decision making, and taking action. All the techniques presented in the following part are underpinned by the real world examples from the global practices.

To help the construction companies to start their data management implementation process, the strategy based on the SMART model is presented in Part IV. Chap 10 is describing, the path of creating data-related strategy which could enhance health and safety environment improvement. The overall conclusions, discussion, and future perspectives of the following report are presented in Part. V.

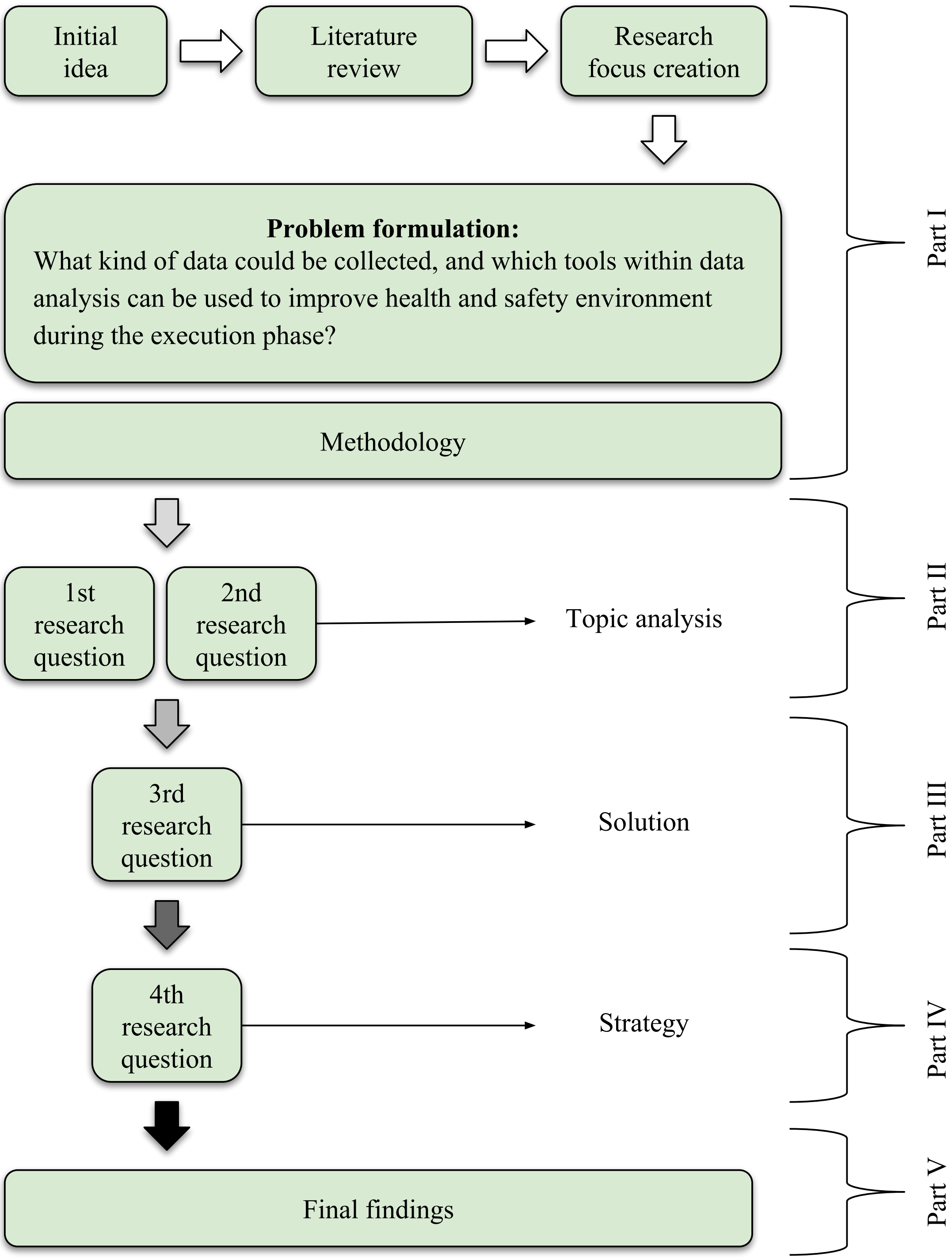


Fig. 3.5: Structure of the report and research process flow.

3.4 Research methods

Research methods represent the last element of research approach presented in Fig. 3.2. According to [Creswell, 2013], **methods** explain the way how researcher collected, analysed and explained data necessary for the project. Various research methods might be used for this purpose, as different research designs propose distinct methods.

3.4.1 Methods for data collection

According to Robson and McCartan [2016], appropriate literature is an important component of scientific research and it can be explained as the information already published. Researchers have to continuously determine, detect and analyse relevant information in different documents such as books, book reviews, research articles, reports, dissertations or other sources of information. Literature review might serve many purposes in the project. First, the literature review reveals the inconsistency of knowledge required for the project, moreover, indicates possible fields of study where uncertainty might occur. Further, by analysing multiple sources, the proper terminology might be determined [Robson and McCartan, 2016]. At the beginning of following research process *data* had various different meanings, however, by using appropriate literature review, the broad explanation of topic was reduced into one specific definition, see Chap. 5.

Literature review methods

In order to identify relevant literature, several literature review approaches were used at the beginning of the research process. At the early stage of planning the report focus, the electronic Aalborg University database Primo and Web of Science database were used. Both databases are covering the wide range of scientific fields. The advanced search facilities were used, in order to specify which materials are relevant for the report and also to avoid too extensive number of search results. The search was covering scientific journals and books, and it was conducted based on titles, and descriptions of the sources. This type of literature review helped to find the most relevant sources. The main keywords in initial literature review are presented in Tab. 3.2. The literature was searched based on the set of keywords. The keywords enable to describe the fundamental elements of the research [Robson and McCartan, 2016]. Afterwards, the search results were examined in more detail, to reject irrelevant literature.

Tab. 3.2: Results of initial literature review, the numbers in parenthesis are showing the amount or sources published in years 2014-2018.

Keyword	Electronic database	Number of results
title search "data" AND title search "construction" AND description search "big data"	Primo	84 (79)
title search "big" AND title search "data" AND title search "construction"	Web of Science	89 (85)

While conducting the first search, there were several articles not relevant for the construction industry. Irrelevant sources were screened in several ways, for instance, by checking the headline, abstract or by the brief review of the article's text. Only the most relevant articles which demonstrate the connection between management of big amounts of data and construction industry were insightfully checked and reviewed. Moreover, the articles were screened in order to check if there are the same positions from both Primo and Web of Science databases.

Authors of the report decided to use the latest sources in order to get an overview what is the current situation in case of data usage in the construction industry. It was assumed that 5 years timespan is the most effective in this case. From the total amount of found literature in Primo database, the 79 sources were published in years 2014-2018, and in Web of Science database - 85 sources. However, the literature published before 2014 were not neglected and also briefly analysed to get the essence of the topic. According to Oliver [2012], the year of publication and its value for the project depends on several circumstances. In some situations, it is even better to use older literature, as some of the subjects might be relatively old. In this report, data management is considered as the main point of analysis and this subject is present for many years in scientific literature. Therefore, for the general explanation of the topic, it is reasonable to use the sources published before 2014. However, in the building industry this topic is still recent, therefore, it was concluded to use the latest literature.

Based on initial literature review, possible topics for further investigation were distinguished, see Fig. 2.1. While familiarising with the subject, some dependencies were visible. Data is the broad topic which might be apparent in different stages of the project. It also covers the wide range of activities where the benefits of data usage might be visible. Due to several limitations, it was decided to narrow down the field of research. Authors of the project decided to look more insightfully into the interconnections between health and safety and data management in the construction industry. Therefore, a supplementary literature review was conducted in order to specify the topic. The electronic databases Primo and Web of Science were used as a primary literature sources. Due to the limited amount of literature found based on keywords "big" AND "data" AND "safety" AND "construction", the search was extended by looking into new technologies used for safety improvement. The keywords and search results are presented in Tab. 3.3.

Tab. 3.3: Results of supplementary literature review, the numbers in parenthesis are showing the amount or sources published in years 2014-2018.

Keyword	Electronic database	Number of results
topic search "big" AND topic search"data" AND topic search "safety" AND topic search "construction"	Web of Science	55 (28)
topic search "technology" AND topic search"health" AND topic search "safety" AND topic search"construction"	Web of Science	334 (156)
title search "technology" AND title search "safety" AND title search "construction"	Primo	600 (481)

The secondary literature review gave the wide range of information, however, not all the sources were relevant for investigation. The priority was given for latest publications, and these were analysed in the first place. In order to select relevant literature, the sources were screened in the same system as during the initial literature review. Despite the literature found during the searches in Primo and Web of Science databases, the articles founded in references are also used for topic argumentation. Additionally, the authors used hard copies of the literature from Aalborg University Library. The literature obtained during the initial and supplementary searches is creating the foundation for the following research.

Referencing system

According to Clauss et al. [2013], proper selection of referencing system is a necessary part of valuable research creation. Referencing of sources demonstrates the researcher's interpretation in accordance to the original data source. Therefore, the reader has a possibility to validate if the research has a scientific creditability. Moreover, in case if the reader is not familiar with study area, he has a possibility to investigate reference list more insightfully in order to gain greater knowledge and understanding of the report.

In this report, Harvard referencing system is used which is also known as parenthetical referencing [Clauss et al., 2013]. This style was chosen due to its readability. This report is mainly dedicated to the scientific audience and the industry professionals. Harvard style enables the continuous quality control of the research and easy access to to the source. The reader has a quick view of the author and the year of publication, which in some cases is important to validate given information, for instance, when describing newest IT solutions.

In the case when sources were used in particular order to carry out the whole chapter, the references are presented in the small introduction part written in italic in the beginning of the chapter. In other cases, the references are given in the text when these are validating the creditability and supporting the information provided in the project. In this report, active and passive references are used. In the passive referencing, the last name of the authors(s) and the year of publication are given in the parentheses as *[Author(s), Year]*. If the argumentation in the text refers to several sources, the publications are placed chronologically. In active references, only the year of publication is given in the parentheses and it looks as follows *Author(s)[Year]*. The reference list, which is given at the end of the report, is arranged in alphabetical order, based on authors(s) last name [Clauss et al., 2013].

3.5 Selection of research approach

The following report is conducted based on mixed methods research approach, as both qualitative and quantitative data was used in order to solve the problem statement which is supported by more detailed research questions, see Sec. 2.2. This approach was chosen due to its advantage of combining both types of research, therefore, overcoming the limitations of qualitative and quantitative researches. The authors of the following report gained more insightful understanding of the chosen problem, as different perspectives were analysed. The final outcome is more valuable as analysis of quantitative data is supported by qualitative information.

The process of the following research might be described as abductive. The process started by combing the theoretical knowledge and the observations of the research authors in order to suggest the problem for the following investigation. Further, the problem was insightfully analysed and tested for solution creation. Moreover, the research was conducted based on the pragmatic philosophical worldview. It was considered as important to emphasise the problem within the one area, and find the possible solution how these issues might be overcome. The whole report is underpinned by the real world examples and the best practices form the other companies.

Part II

Topic rewiev

Part II of the following report is analysing the current situation of health and safety practices in Danish construction industry, as well as the current state of data usage in the industry and basic principles how data is defined. This is done to answer the following research questions: *What is the overall situation of health and safety management performance in the Danish construction industry?* and *What are the potentials of data which might be used in the construction industry?* Chap. 4 is focusing on the identification of main factors contributing to occupational accidents, and is analysing Danish construction industry based on these factors. To gain an understanding what data is, the introduction to this topic is made in Chap. 5. The specific definition of data is provided to avoid misunderstandings and distinguish the opportunities of data regarding health and safety management.

4 | Health and safety

This chapter provides an extensive description of the Danish working environment and the factors contributing to the occupational accidents. This is done to gain a deeper knowledge within current practices, and to distinguish the aspects which are directly connected to health and safety issues on site. In Sec. 4.1, the Danish working environment and the main legislation is elaborated. Further, 4.2 describes the factors that are contributing to occupational accident.

4.1 Danish working environment

The working environment in Denmark is regulated by legislation and requirements which must be followed by the construction companies. The main objectives of the legislation are to ensure that the companies are contributing to health and safety rules and the constructions sites are organised in way that the environment is safe for the people working there. The following legislation includes Working Environment Act and the number of executive orders which complement the regulations from the Act [Sector Working Environment Committee for Building and Construction, 2016].

Working Environment Authority is an important player contributing to new legislation creation and the monitoring, that the companies are working regarding the Working Environment Act [Sector Working Environment Committee for Building and Construction, 2016]. The Authority is preparing the annual reports in order to recognise and analyse the occupational issues and risks, such as fatal or serious accidents, psychological risk factors or risk related to repetitive work [Christiansen and Pedersen, 2007].

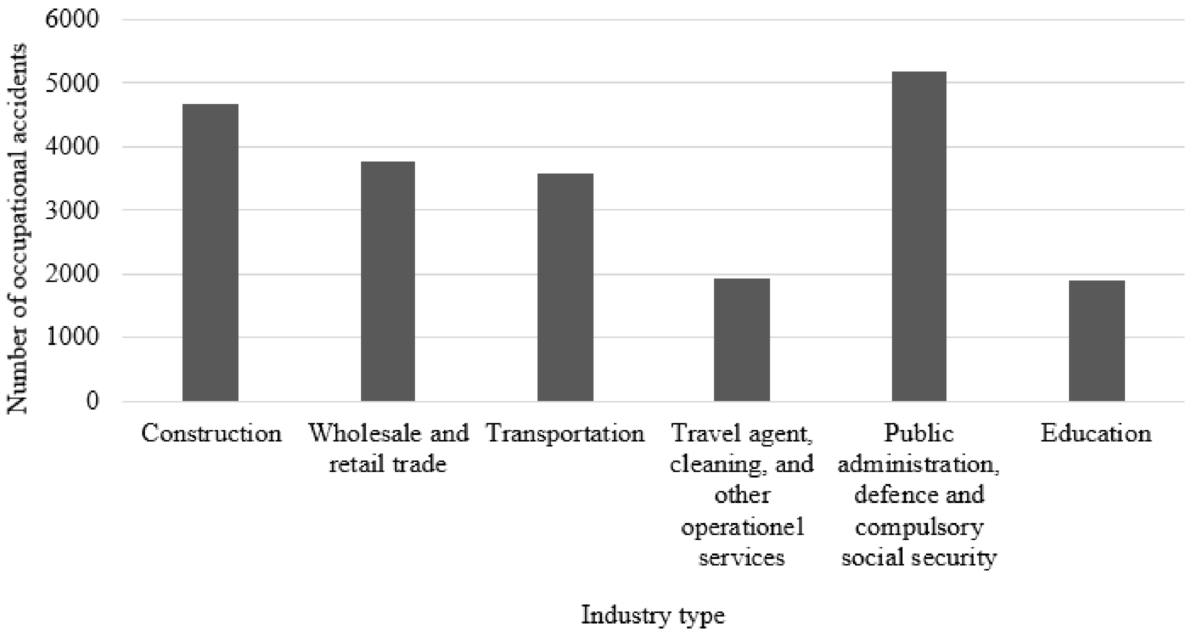


Fig. 4.1: Number of occupational accidents in 2015 by the industry type [Yearbook, 2017].

According to the Danish Working Environment Authority [2016], the total number of reported occupational diseases in all industries in 2016 decreased around 8 % from approximately 21,600 to 20,000 in comparison to 2015. However, if compared this number in the time period from 2011 to 2014, the number of reported cases increased by 24 %. Musculoskeletal disorders were the largest group among the occupational diseases. An occupational disease is explained as a disease which is wholly or partly caused by workplace effects both through short or long-term periods. Based on statistics from Danish Working Environment Authority [2016], the construction workers rank on second place in terms of occupational diseases.

In 2017, 169,594 people were employed in Danish construction industry, were 4,681 of them reported occupational accidents, see Fig. 4.1 [Yearbook, 2017]. According to Yearbook [2017], approximately ten construction workers die every year, wherein 2015, there were six fatal accidents reported in the industry. Based on statistics, the serious accidents are usually occurring in the small and medium-size construction companies. The most common accidents with serious outcomes are: hit by objects, fall from height, body cuts while working with saws or injuries caused by contact with high temperature and hazardous substances [Statistikken Arbejdsulykker, 2016]. In order to protect the workers from occupational injuries, safety measures should be determined and implemented. Moreover, some of the accident prevention methods could be planned already in the early stages of the project [Manase et al., 2004]. The high number of occupational accidents in Danish construction industry is an important issue, therefore, the Authority together with the construction enterprises are working toward safety improvement [Jensen and Koch, 2003].

Recently, there has been a debate in the media about foreigners in the Danish labour market. It has been argued in the debate that an increase in the number of foreigners in recent years in Danish construction industry may have contributed to the general increase seen in reported accidents at work in the construction sector. The number of serious occupational accidents involving death among foreigners has been highlighted, as many foreigners in Denmark are exposed to a higher risk of accident and generally having a poorer working environment than Danes [Danish Working Environment Authority, 2007].

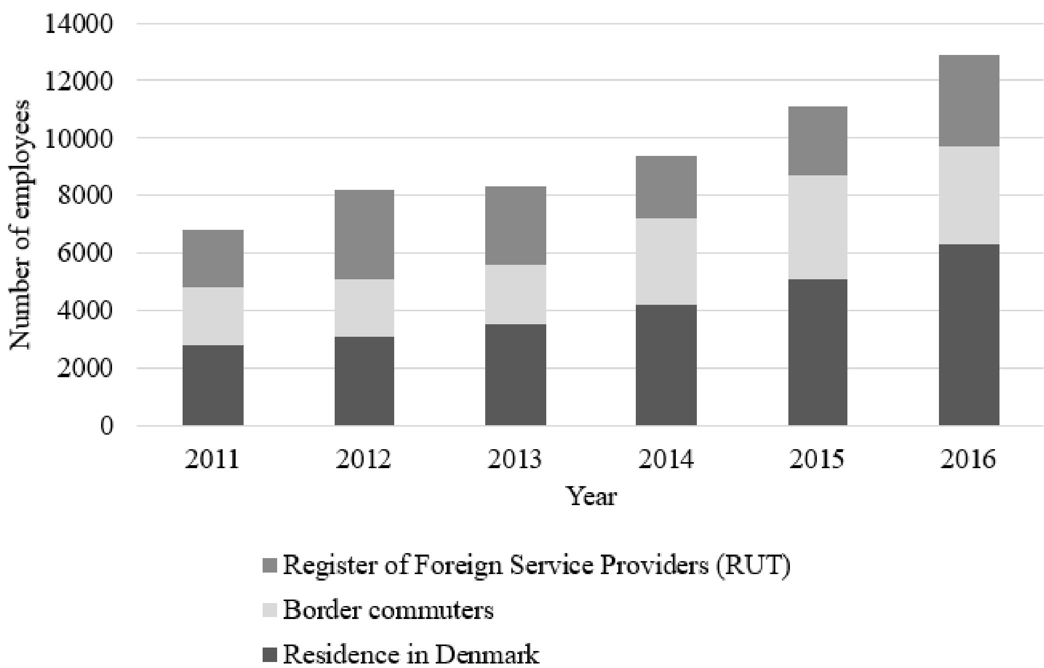


Fig. 4.2: Number of full-time foreign labour in Danish construction industry [Dansk Byggeri, 2017].

Based on statistics form 2016, 7.8 % of the total Danish construction industry labours, are the workers from foreign countries. This number includes all types of employees with Danish resident permit, and the workers that are conducting their work temporary in Denmark. The tendency is that the total number of foreign labour in Danish construction industry is steadily increasing, see Fig. 4.2 [Dansk Byggeri, 2017].

In order to develop the framework for the working environment improvement during the construction project, the most important factors contributing to occupational accidents are determined and analysed in sections below.

4.2 Factors contributing to occupational accidents

Danish construction industry might be perceived as multicultural environment since the number of foreign workers and foreign contractors is increasing every year [Dansk Byggeri, 2017]. To better understand the root causes of the occupational accidents, the different perspectives form distinct research papers are elaborated. Danish construction industry should be analysed from the broader perspective, as some of the factors are changing due to the increased number of labour outside Denmark. Moreover, some of the accident types are seen in the global construction industry, and Denmark is not an exception. The factors determined from the scientific literature are supported by the examples form Danish construction industry in order to ensure that the factors are relevant for the research.

Many scientific researches are elaborating the causes of injuries in the construction industry [Winge and Albrechtsen, 2018; Choudhry and Fang, 2008; Liao and Perng, 2008; Haslam et al., 2005]. According to Liao and Perng [2008], the variety of different factors should be taken into account when analysing the occupational injuries. These factors might be: type of the injury, experience of the workers, salary, age of the injured party, size of the enterprise, project budget or time of the day when the accident occurred [Liao and Perng, 2008; Chi et al., 2005; Hinze et al., 2005; Rabi et al., 1998]. Based on the list of sources, the assumption was made that all factors contributing to occupational accidents might be divided into four main groups, and these are: individual aspects, task aspects, management aspects and environmental aspects, see Fig. 4.3 [Alarcón et al., 2016; Bellamy et al., 2008; Camino López et al., 2008; Liao and Perng, 2008].

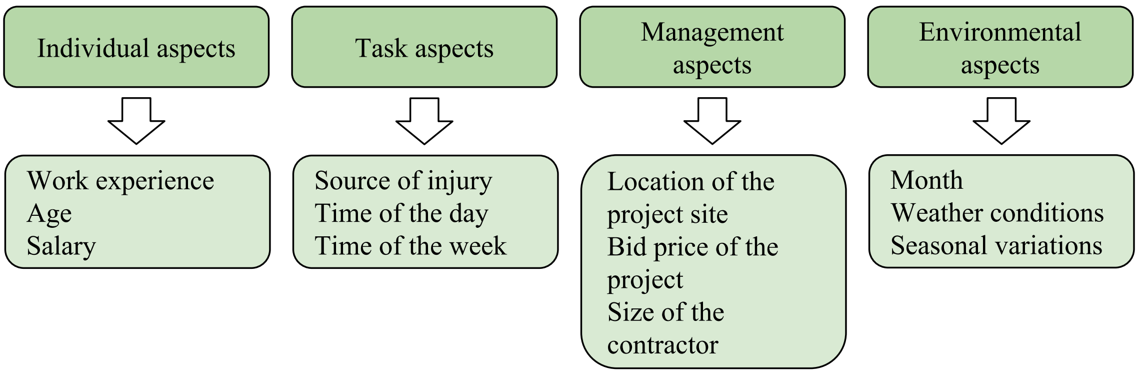


Fig. 4.3: Four groups of factors contributing to occupational accidents.

In order to understand the essence of following factors, each of them is insightfully elaborated in sections below. The analysis is supported by the description of practices from Danish construction industry.

4.2.1 Individual aspects

Individual aspects are determined based on qualifications, practices and overall background of the worker which had an occupational accident. These aspects might be explained as work experience, age of the worker or the work compensation size and method. The experience is measured by the timespan, which worker spent on the previous projects, and by the familiarity of particular work which contributed to the accident. Further, the salary is measured by the daily amount of money which employee is receiving and by the compensation type. An important aspect is age of the injured employee, as it might be connected to other health issues, or occupational diseases [Camino López et al., 2008; Liao and Perng, 2008].

The analysis of the individual aspects should be started from the macro perspective. To better understand the social and educational background of the individual workers, national educational programs should be taken into account [Spangenberg et al., 2003]. According to European Commission [2018a], Danish market is facing the skills shortage, especially it is noticeable in the construction industry. The shortage is partially caused by generation shift since elderly construction workers are retiring. Moreover, there is a deficit of younger workers with sufficient technical skills. Mučenski et al. [2015] found a dependency between the age of the workers and the number of occupational injuries. The results of study show, that the workers between 20-34 years and with the less experience than 4 years in the industry, are the most exposed group for occupational accidents. However, the study did not find the evidence that the elderly workers are facing higher risk of occupational injuries.

By analysing the Danish working environment the same results were found as in the study of Mučenski et al. [2015]. The statistical data from 2016 was examined, to show that the younger employees had a higher risk of being exposed to occupational injuries [Arbejdstilsynet, 2017]. The results of the analysed statistics, show that in the age group 20-34 years, the total number of reported occupational accidents reached 2046 cases. It is approximately 40 % of the total number of accidents. The least exposed group to the risk of injuries is in the age range 50-64 years. In this case, the number of accidents was the lowest. Based on statistics from Arbejdstilsynet [2017], it is clear that the younger workers suffer more from occupational injuries. It might be caused by insufficient training and lack of practical experience in the field. However, the results might be interpreted inexactly as the total number of worker based on age range was not taken into account.

Danish construction industry is especially struggling with the shortage of craftsman with particular skills, for instance, carpenters, bricklayers, plumbers, soil workers and concrete workers. Moreover, together with the increasing trend of sustainable construction, specialised workers in terms of energy-efficient constructions are needed [European Commission, 2018a]. European Commission [2018a] is estimating, that until 2020, the labour shortage will be visible among skilled and unskilled workers, also among high-skilled industry professionals. In order to emphasise and work toward shortage problem reduction, some initiatives and campaigns were arranged by the government. The purpose of these initiatives is to encourage young people to choose constitution-related professions and engage the construction companies to provide internships or training for the young people. One of such initiatives is "Thank you" campaign, and it is supporting the companies which are employing young people and providing them specific training [European Commission, 2018a].

According to Haslam et al. [2005], the majority of the accidents during the construction phase of the projects are caused by the workers or working groups by themselves. The most common mistakes in worker's behaviour are overlooking safety measures and protective equipment, prioritising work rather than safety, taking short-cuts at the construction sites or wrongly assessing possible risks. These unsafe practices might be explained by insufficient knowledge of the worker, lack of education or training [Haslam et al., 2005]. Therefore, it is especially important to emphasise safety training by the project managers and supervisors and provide the required information for safe task performance. Approximately 30 % of Danish adults from the construction industry are participating in the educational training while the average in European Union is 10 %. Moreover, the number of students in Denmark enrolled into the education related to the construction industry visibly grew in recent years [European Commission, 2018a].

The last factor which is forming the individual aspects, is the form of work compensation. The tasks during the construction projects are performed by employees which are usually working in small, organised gangs, and they are specialised in one working area, for instance, carpentering, concrete works or bricklaying [Ajslev et al., 2013, 2015]. The employees are compensated for their work based on one of several work compensation methods, which might be: performance based wage, time based compensation or fixed compensation. However, the most common compensation form in Danish construction industry is performance based wage. In this compensation method, the workers are paid based on work done or by produced unit [Ajslev et al., 2015].

The results of scientific studies show, that workers which are compensated based on performance system, are struggling more often with safety issues that workers with hourly wages [Ajslev et al., 2015; Ganster et al., 2011]. According to Yeh et al. [2009], personal burnout and work-relater tiredness are more likely in case of performance based wages rather than in the case of fixed salaries. These differences of health and safety conditions among employees might be caused by the fact, that the construction workers paid by piece rate wages are putting more effort into their tasks [Ajslev et al., 2015; Ackroyd and Thompson, 1999]. It is also related to the fact that these construction workers are more stressed about time pressure, therefore, are working harder and often skipping the protective equipment. Money factor is the main motivator for the employees which pushes their productivity [Ajslev et al., 2015, 2013].

According to Ajslev et al. [2015], the practices of increasing the work pace might cause serious health risks of the construction workers. The workers by pushing themselves for hard work, lifting heavy materials or by working in not ergonomic way, are exposed to musculoskeletal disorders more than workers from other industries [Ajslev et al., 2015; Andersen et al., 2007]. These disorders and the sick absences are more frequently noticeable among labour unskilled workers than among highly skilled ones [Andersen et al., 2011].

4.2.2 Task aspects

To better understand the causes of the construction accidents, the task specific aspects need to be analysed. These factors determine the specific features of the work performed by the injured employees. Task aspects might be described by direct cause of the injury, time of the day, or day of the week when activity was performed [Liao and Perng, 2008]. Causes of the injuries have various natures, as these might be objects or other substances which directly contributed to the injury [Liao and Perng, 2008; Oleske et al., 1989]. The objects in the construction projects are temporary constructions, scaffolding, bricks, heavy machinery, vehicles, soil or chemical substances. Moreover, time of the accident might have a crucial importance, since the activities could be performed in bad conditions, for instance, related to light accessibility [Liao and Perng, 2008].

Winge and Albrechtsen [2018] identified the list of common accident types in the construction industry based on reports from the Norwegian Labour Inspection Authority. The assumption was made, that the data presented by Winge and Albrechtsen [2018] might be interpreted also in terms of the Danish working environment. To prepare appropriate safety improvement plan, the essence of the risky tasks need to be understood. Three main groups of accidents, distinguished by Winge and Albrechtsen [2018], are elaborated in sections below.

Fall from height

Fall from height is the most frequent accident type in the construction industry [Winge and Albrechtsen, 2018]. Workers are usually falling from roofs, floors or platforms through unprotected openings [Winge and Albrechtsen, 2018; Hoła and Szóstak, 2017; Ale et al., 2008]. The majority of these accidents are occurring during new building erection or refurbishment since there are many temporary openings, for instance, unmounted staircases [Winge and Albrechtsen, 2018; Howell et al., 2002]. Due to strict schedules, workers are not paying attention to the proper protection of all open holes. These are only covered with thick places, moreover, these are not adequately attached. Based on reports from the accidents, in the most cases, the injured employees were not familiar with existing holes. To prevent this type of accidents, the hole's edges need to be sufficiently protected and marked in order to warn the people which are not aware of temporary openings [Winge and Albrechtsen, 2018].

Another type of accidents is falling from scaffolding. These accidents are happening due to insufficient conditions or improper attachment of the scaffolding when it was moved to another place [Winge and Albrechtsen, 2018]. While erecting, modifying, or moving the scaffold, the commissioning permit needs to be attached to it [Sector Working Environment Committee for Building and Construction, 2016]. In some of the reported cases, the workers slipped or forgot about missing parts of the scaffold. Some of the injuries are also caused by the irresponsibility of the workers as they are working outside the railing. Furthermore, the construction workers tend to use the ladder while working in height [Winge and Albrechtsen, 2018]. Based on Sector Working Environment Committee for Building and Construction [2016], ladders might be used only for the work in the short time, no more than 30 minutes. This activity is dangerous since in many cases the leaders are not attached in the proper way, both at the top and the bottom. Instead of ladders, the labourer might use movable scaffolding, platforms or telescopic lifts. Still, the majority of halls from the leaders are occurring among worker which are performing short-term jobs, for instance, labourer which are installing, controlling or dissembling parts of the construction [Winge and Albrechtsen, 2018].

The majority of fall accidents might be prevented by eliminating the risk at the source [Winge and Albrechtsen, 2018; Haddon, 1983]. The hazard might be detected and removed already in the early stages of the project planning. The job of designers is to ensure that the workers will not be exposed to sharp edges of the construction, and that they will have enough space to work. The work in height must be carefully protected since some of the accidents occur during the preparation works, for instance, when protecting the holes, mounting the scaffold or dissembling the platforms. Moreover, an appropriate allocation of labour might also help in this case, by choosing the trained and experienced workers to perform this type of tasks [Winge and Albrechtsen, 2018].

Contact with falling object

The accidents which occur through contact with falling objects are usually taking place during the demolition works or by moving large objects, as beams, poles, reinforcement or concrete elements, during the construction projects. In this case, the accidents might differ from small injuries to the fatal cases, depending on the features of the lifted object. It is hard to replace large objects in the construction projects, at the same time, it is hard to reduce risk at the source. Majority of the accidents are occurring while assembling of dissembling parts of the structure. The objects are falling due to insufficient attachments and bad weather conditions [Winge and Albrechtsen, 2018].

The big part of occupational accidents is caused by hits during the lifting of smaller objects by cranes, forklifts or other machinery. These types of accidents might be overcome by moving objects by vehicles on the ground instead of the lifting them. The accidents are commonly caused by loosing control of lifted object. In some cases, the workers were hit by the lifted objects in the height and it caused a fall from height [Winge and Albrechtsen, 2018]. According to Hale et al. [2007], these types of hazards are called domino accidents. Good prevention method, in this case, is using help from a flagger, which is making sure that workers are not in the danger zone [Winge and Albrechtsen, 2018].

Contact with moving parts of machine

Some of the workers in the construction industry are exposed to the accidents caused by contact with moving parts of the machines such as saws. There are not as many fatal accidents in this case, usually, workers are experiencing loss of fingers or deep cuts. To minimise the possibility of injury, the appropriate material choices should be made. It concerns both pre-cut materials and saws with modified blades. In some of the reported cases, workers were wearing inappropriate clothing or were not using pushing sticks which protect hands from the cuts. An important parts of individual protective equipment are glasses as these protect eyes from the injuries caused by pieces of cut materials thrown from the blade. Based on the accident reports, the younger and inexperienced workers are the most exposed group to this type of injuries [Winge and Albrechtsen, 2018].

Based on the study of Haslam et al. [2005], among the analysed occupational accidents, the majority of them is occurring not during the specific tack, rather it is during the work preparation or while moving around on site. The task itself is not as dangerous as the materials and the equipment the employees are working with. In many cases the lack of practical materials, insufficient equipment design or the bad condition of the tools are the main factors of the injuries [Haslam et al., 2005].

4.2.3 Management aspects

Management aspects are related to direct practices of the contractors, and their commitment to health and safety environment improvement. Moreover, the size of the company might have an impact on safety practices during the construction project. Size of the contractors is defined by the capital of the enterprise. Type of the project might also have a great impact on the accidents during the work. First, the location where the project is performed should be analysed, as some special risks might be determined. The price of the bid might be also an important factor as the resources for proper working environment planning might be limited [Liao and Perng, 2008].

To better understand the influence of management factors in terms of the Danish construction industry, some key industry drivers are presented. Based on European Commission [2018a], the total number of enterprises in the Danish construction industry was 70,120 in 2016. In comparison to the data from 2010, the number of the construction companies increased by 10 % in total. 77.8 % of total Danish work labour is employed in small and medium-size companies. The total number of bankrupted companies in the industry increased by 16.3 % in comparison with 2010. However, the number of new companies in the industry increased during these years. Hight rate of bankruptcies is partially caused by changes in the procedures for collecting old tax debt by the companies [European Commission, 2018a].

The dependencies between safety performance and safety management system within the construction companies were extensively analysed in the scientific papers [Forteza et al., 2017; Jørgensen, 2016; Niskanen et al., 2016; Wang et al., 2016; Bellamy et al., 2008]. When analysing safety performance, two main aspects might be distinguished: risk conditions, and safety participation. Risk conditions are connected to working environment conditions, safety protections, different safety procedures or guidelines, while safety participations might be explained as safety meetings, toolboxes or motivation of workers to adapt safety rules [Forteza et al., 2017; Wang et al., 2016]. The safety management is responsible for both safety performance aspects and it is a part of the broader management system of the organisation [Forteza et al., 2017].

According to Manu et al. [2010], construction project features might be associated with the occupational accidents in the great extent. Construction project features are described as aspects that determine and describe the construction projects, and these aspects might take one of several forms, for instance, organisational, operational or physical aspects. All of these aspects emerge from design or management decisions taken through the project life-cycle. Construction project features are establishing and determining the nature of the project, working environment, work extend, thereby, these influence the possible occupational accident causes [Manu et al., 2010].

The study of Forteza et al. [2017] confirms that the site complexity has the direct impact on possible safety risk. The complexity of the construction site is determined by project type, either it is renovation, repair, maintainable or new construction project, moreover, the main elements of the project such as work type, complexity of design, site planning[Forteza et al., 2017; Hon et al., 2010; Manu et al., 2010; Fang et al., 2004a,b]. The complexity of the project might cause the serious or even fatal accidents during the work, since, some project aspects might be planned in the bad meaner and cause, for instance, the possibility fall from high.

Forteza et al. [2017] in their research analysed what kind of influence on the safety performance has the structure of the organisation. There is no clear dependency on the size of the company and the number of accidents. Some of the research papers are arguing that that larger companies have more reported accidents, however, other scientific papers describe the opposite relationship between these two variables [Forteza et al., 2017; Holte et al., 2015; Camino López et al., 2011; Pérez-Alonso et al., 2011; Camino López et al., 2008]. However, it might be concluded that small size construction companies and non-governmental projects suffer more from occupational accidents. This is directly linked to smaller budges of private projects [Forteza et al., 2017; Cheng et al., 2010b].

The companies might minimise the risk of occupational accidents by proper relocation of the resources. The contractors regulate how the activities are performed, the number of foreman or workers on site and their cooperation. Moreover, the technical controls contribute to safe working environment creation [Forteza et al., 2017]. In the case when several different contractors and subcontractors are working on the same project, planning and management of the working environment become more complicated, therefore, the safety practices might be overlooked [Hinze et al., 2013; López-Alonso et al., 2013; Manu et al., 2013]. Manu et al. [2010] distinguished that as the project is more complex, and more stakeholders are involved, the risk that safety problems will occur, is much higher. The conclusion might be made that the structure and complexity of the project is negatively influencing the safety practices [Forteza et al., 2017].

Small and medium-size companies often struggle with safety planning, as usually, only one person is responsible for all safety issues within the company [Järvis and Tint, 2009]. In order to ensure the safe working environment for the workers, the health and safety planning need to be present in all day-to-day activities [Forteza et al., 2017]. Therefore, the safety managers, supervisors or even authorities have an important position in ensuring safety [Fang et al., 2004a,b]. Despite the supervision, also the materials, equipment, and machinery used during the construction project need to compile to required safety norms [Borys, 2012; Adam et al., 2009].

4.2.4 Environmental aspects

The last group of factors, affecting occupational accident's occurrence, are environmental aspects. This group is defined as basic external features of the working environment at the time, the accident occurred. The environmental aspects might be described through weather conditions or seasonal variations [Liao and Perng, 2008]. Seasonal changes are influencing the safety practices on construction site, as the working environment is differing in particular times of the year. The majority of injuries related to seasonal variations might be overcome by the effective planning of activities on site [Liao, 2012].

Some of the weather conditions particularly influence the possibility of accidents, for instance, rain, intensive heat or wind. The research of Liao and Perng [2008] shows that the possibility of the accident is much higher if it is raining. Moreover, the dependencies are visible that the workers tend to get injuries in the first days during the rainy periods. The workers want to finish their tasks, which they began in good weather conditions, therefore, they rushing more when it is starting to rain.

Heat stress is an important issue among the construction workers. Excessive exposure to heat might have serious outcomes and cause number for various heat diseases, where some of them might be even fatal [Rowlinson et al., 2014]. Fatal accidents are caused by reduced speed on respond, minimised reasoning ability in risky situations or reduced visual perception [Rowlinson et al., 2014; Chi et al., 2005]. Further, due to intensive workload in the high temperatures, other occupational accidents might occur as the capacity is reduced, and personal protective equipment is usually used in the improper way [Rowlinson et al., 2014]. Moreover, the number of newcomers in the construction industry, which are usually inexperienced, is higher during the summer period and it is contributing to the increased number of accidents [Liao, 2012]. Heat stress causes not only the safety issues but also influences the overall productivity and cost of the project [Rowlinson et al., 2014].

According to Rowlinson et al. [2014], personal protective equipment often is not adapted to the prevailing weather conditions. The work clothes and reflective vests, which are required on construction sites, are made from water-resistant materials which are blocking heat emission. Moreover, the protective helmets without ventilation impede the task performance, since the temperature within the helmet is much higher than the environmental temperature. Therefore, construction workers are not willing to wear protective equipment, which might prevent occupational injuries [Rowlinson et al., 2014]. The eye protection is also often neglected during the rain or hot days as the glasses are limiting visibility in harsh weather conditions [Choudhry and Fang, 2008].

Based on research of Huang and Hinze [2003], the risk of fall from height is much greater during the winter period. Due to low temperatures the movements of the workers are limited, the risk recognition is slower, and the surfaces are more slippery. During the winter period, unfavourable light conditions are influencing the work at construction site. While planning the project, the activities of the workers should be considered. Some of the accidents might be overcome, for instance, work in height might be planned for the summer period, when the weather conditions are more favourable for this type of job.

4.3 Occupational accidents prevention

By extensive analysis of all basic factors contributing to the occupational accident in the construction industry, the proper tools of working environment planning might be chosen. In this report, the framework for health and safety planning is created based on management of big amounts of data. Chap. 5 is conducted to better understand the opportunities and available tools of data and its relationship to practices in the construction industry.

5 | Data

This chapter aims at explaining the background and definition of data in the overall context as well as in the construction industry. This is done to define the word data and provide an introduction to the reader on what data is, and how it is used in the construction industry. The first section is defying the term data and its attributes, see Sec. 5.1. Further, Sec. 5.2 is examining data within the construction industry, while Sec. 5.3 is looking specific on the health and safety part.

5.1 Definition of data

Digitisation is one of the mega-trends in the global market since digital technologies are available in the majority of various organisations. The technologies, developed by digitisation processes, are enabling the automation and interconnection of many different processes form distinct areas. It is the next step in technological development which is closely related to the data usage in term of the process description. Millions of sensors are connected to the Internet which opens the new possibilities of information exchange. Different types of data are also generated by cameras, mobile devices, applications or services [Ylijoki and Porras, 2016]. According to Mayer-Schönberger and Cukier [2013], the process of intensive data generation is called datafication. Datafication describes the processes of value creation through data [Lycett, 2013]. By using the potentials of digitalisation and datafication, it is possible to capture occurring situations, activities or sequences of various events in distinct forms of data [Ylijoki and Porras, 2016].

The rapid growth in the computer technology has given access to enormous amounts of data [Storey and Song, 2017; Bilal et al., 2016; Deutsh, 2015]. The companies are capturing the information about their clients, suppliers or the operations which they are performing [McKinsey Global Institute, 2011]. Through various sources like the web, Internet of Things (IoT), sensor networks, and social media it is now possible to gather information rapidly and in bigger amounts than before [Storey and Song, 2017; Bilal et al., 2016]. The report of McKinsey Global Institute [2011] argues that data available today might be perceived as economy improvement driver, both in national and organisational levels. In perspective of the particular company, data enhances the productivity and increases its competitive advantage in the sector. Therefore, it has now come to a point where traditional management of data is no longer enough [Storey and Song, 2017]. Capturing and analysing these big amounts of data is not a new concept, many companies are handling large data sets every day, among them are companies like Amazon, Google, and Facebook [Maté et al., 2017; Marr, 2015]. These are gathering various types of data in the real time to enable task performance in an automatic way [Ylijoki and Porras, 2016].

A buzzword for this type of data is *Big Data*. The term is widely interpreted by researchers and there are many definitions of the word. In this report, the large amounts of data are referred to as **data**. For the purpose of the following report, the definition is: *Large amount of unstructured or structured data that people are able to capture and analyse* [Bilal et al., 2016]. It does not mean that the companies or individuals can understand everything within the data, but that they can see potential relationships and causes that can help the decision making [Deutsh, 2015].

To understand and describe big quantities of data, researchers often look at several attributes, in this case, they are often referred to as the *3V's*, *4V's* or *5V's* [Motawa, 2017; Storey and Song, 2017; Bilal et al., 2016; Sørensen et al., 2016]. Three fundamental attributes of data are *volume*, *velocity*, and *variety* which represent the 3V's definition [Laney, 2001]. Moreover, with the development of digital technologies, the companies have emphasised the *value* creation and *veracity* as an important part of data management

[Ylijoki and Porras, 2016]. This report uses the *5V's* since it is assumed to be the most covering. The components of the 5V's are described in sections below and presented in Fig. 5.1.

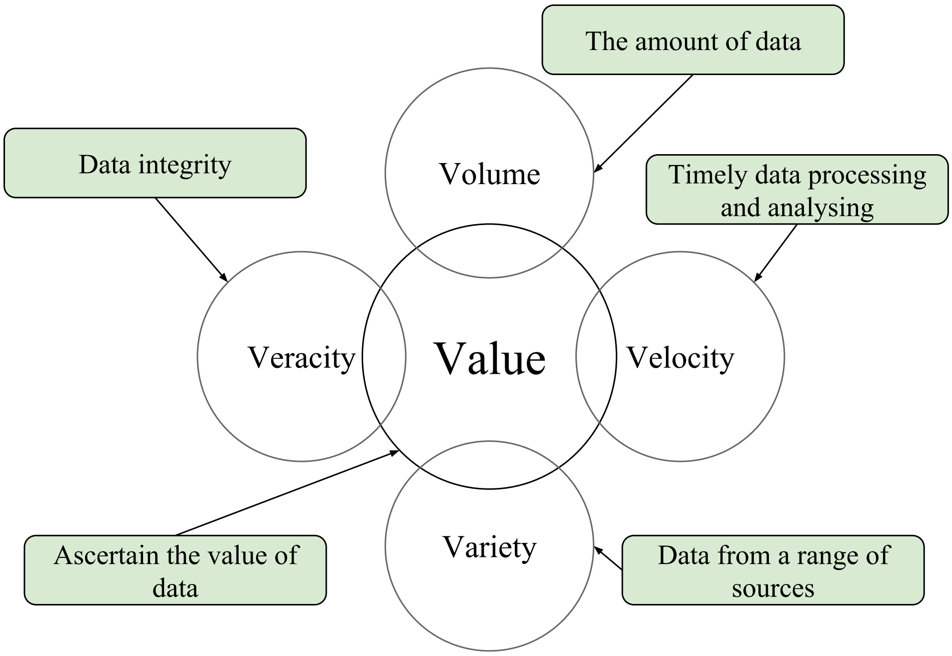


Fig. 5.1: Five attributes of data known as the 5V's [Song and Zhu, 2016, *ed.* Fig. 2].

Volume

Volume represents the size of data which can be measured from terabytes, petabytes and beyond [Bilal et al., 2016; Sørensen et al., 2016]. The increase of data volume is a very dynamic process, and the estimation is made that until 2020, the total volume of data will grow by 500 % in comparison to 2010 [Tabakow et al., 2014]. Extensive amounts of data might be perceived by the organisations as a tangible asset, therefore, it is important to properly store it. The companies need to find the ways how to store these increased volumes of data, for instance, by cloud storage. However, the organisations need to distinguish if all data is providing value for them, as with increasing volume of data, the value of particular data sets is decreasing [Laney, 2001].

Variety

Data is coming from a range of different sources, moreover, it is captured in various forms [Motawa, 2017; Bilal et al., 2016]. It might be saved in distinct models and expressed in many forms, for instance, text, audio, video or graphs [Bilal et al., 2016; Sørensen et al., 2016; Tabakow et al., 2014]. Such combinations of data give both structured and unstructured data sets [Bilal et al., 2016; Sørensen et al., 2016]. Some of the possible data sources are: internal databases of the organisations, external sources as the Internet, Deep Web or different kind of statistics [Tabakow et al., 2014]. According to Laney [2001], variety is the main issue in effective data management since information is available in incompatible formats and structures.

Velocity

Velocity is the speed of creating, capturing, extracting, processing, and storing data, like streaming [Motawa, 2017; Storey and Song, 2017; Sørensen et al., 2016]. The speed might vary from days, hours or remain as continuous process [Sørensen et al., 2016]. The most beneficial for the organisations is real-time data management. Data streaming requires sufficient computing power in order to enable fast decision making [Gierej, 2017].

Veracity

Veracity looks at the imprecision of the data, and its integrity [Motawa, 2017; Storey and Song, 2017]. It questions the meaning of the data, like quality, reliability, and uncertainties [Storey and Song, 2017]. Unnecessary and invaluable information should be avoided to not disturb the data analysis and to provide a reliable final results [Gierej, 2017].

Value

Value of data is difficult to ascertain. It provides a way to measure the focus areas as the return on investment (ROI), customer satisfaction, and other measures. This can be done in SQL-types queries, machine learning, data mining, etc. [Storey and Song, 2017]. Moreover, value expresses the unique informative function of large amounts of data. This function gives an opportunity to make the conclusions out of data which contribute to the improvement of activities on various levels of the organisation. By using data management, the problems occurring in the companies might be immediately accessed. Therefore, value is an important aspect of data as it is unique information contributing to the company's improvement [Tabakow et al., 2014].

5.2 Data in the construction industry

In the construction industry, there are large volumes of available data, and these amounts are expected to increase rapidly in the future [Storey and Song, 2017; Bilal et al., 2016; Marr, 2015]. Voluminous amounts of data are produced through the whole construction project life-cycle. The process of data creation and gathering is highly fragmented due to the extensive range of stakeholders. Some of the information is generated during the design stage, and some through the building maintenance, where data is produced by the end users [Motawa, 2017]. Today, the industry is handling some of this data by using Building Information Modelling (BIM) which captures CAD information to enhance the collaboration with stakeholders [Bilal et al., 2016]. However, when developing a continuous knowledge system that is designed to develop fast through adding information in the whole life-cycle of a building, knowledge capturing in digital format is problematic [Motawa, 2017]. It is believed that with the rapid development within the techniques and technologies to manage large quantities of data, knowledge, and information might be gathered in a more effective way [Motawa, 2017; Storey and Song, 2017].

When looking at the data accessible in the construction industry it is clearly evident that the industry meets all criteria of the 5V's principle. The volume and variety of data are especially visible in the industry since the construction projects are continuously generating the large amounts of structured and unstructured data [Motawa, 2017]. Data is voluminous due to the enormous availability of it in schedules, design or financial reports. It is heterogeneous, dynamic, and comes in various formats: DWG, DXF, PDF, ifcXLM [Bilal et al., 2016]. The dynamic nature of construction data is following the trends in society by streaming from sensors, and Building Automation System (BAS). To innovate and optimise the industry, it is seen as necessary to utilise the technology available today [Motawa, 2017; Storey and Song, 2017; Bilal et al., 2016].

In the construction industry, application of data might contribute to more efficient design, construction and maintenance practices. Accessible data collection tools enable extensive data gathering which describes the labour performance or effectiveness of operations. The comprehensive set of information might help the project stakeholders to better understand the overall project, and recognise the needs both of worker and end users. Current data gathering methods provide real-time information which contributes to the improved planning of project life-cycle [Motawa, 2017].

To start data-driven practices within the organisation, whole data management process needs to be understood. According to Mello et al. [2014], the procedure of data management might be described in four steps, see Fig. 5.2. First, the relevant data needs to be gathered from various sources, and this process is known as an acquisition. Further, captured data is stored in the proper way, and prepared for sharing with other parties. In order to catch the value in large data sets, these are handled and analysed through data analytics. The last step in this process is application of analysed data which is driving the decision making and actions of the company. All of the stages of this procedure are connected and influence each other since it is highly dynamic process [Motawa, 2017; Mello et al., 2014].

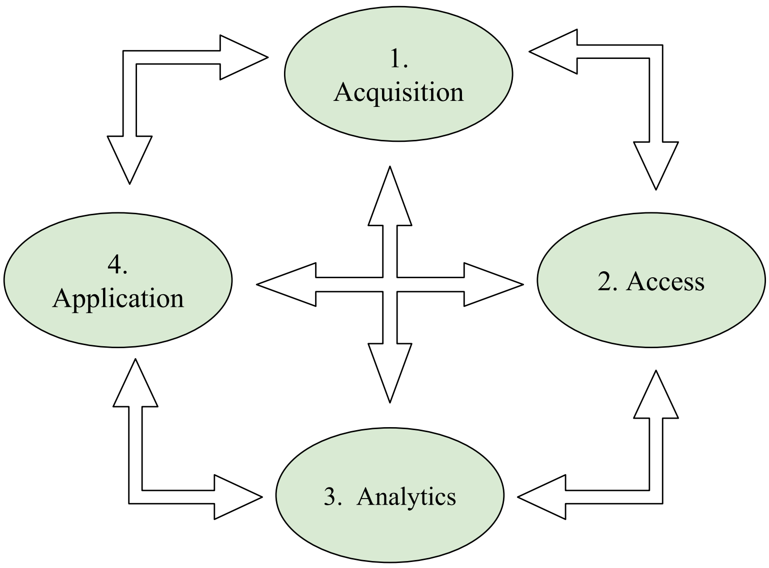


Fig. 5.2: Four steps of data management [Mello et al., 2014, *ed.*].

In the context of data management, the construction industry needs to adapt new tools for all stages of data management in order to convert data into valuable information [Motawa, 2017; Courtney, 2013; Tien, 2013; Waller and Fawcett, 2013]. Due to data variety, the construction companies need to find the way how to analyse both structured and unstructured data sets. The majority of analytical methods are handling only structured data sets. Therefore, there is a need to distinguish clear path who to analyse and get the value of the data from the industry [Motawa, 2017].

5.3 Health and safety improvement

The construction industry is slowly starting technology utilisation, although, the health and safety perspective is lacking behind. The industry is missing the systems which would help to predict risks in the working environment, thereby, would provide a safe workplace for the labourers. Not many health and safety environment managers are familiar with opportunities which data management provides [The International Institute of Risk and Safety Management, 2017]. However, while working with safety issues, the large amounts of information might are obtained and used for development of new solutions. Moreover, this kind of information meets the 5V's principle which is presented in Fig.5.3.

Volume	By looking at the endless inspections, both within behaviour-based safety and safety controls, large amounts of data can be collected. The data can take several forms as: reports from the workers and from the health and safety managers on near accidents, dangerous situations or various statistics. There is no doubt that there is a lot of data available today, however, it requires proper information reporting and storage.
Variety	A big step has been done due to inspections, from using paper reporting, to Internet of Things, sensors, and the Internet usage. It is now possible to do inspections on mobile devices or tablets, contributing to the extensive increase of variety data types.
Velocity	Today, a foreman can report a possible dangerous situation on the construction site by taking a picture or writing a comment on his or hers phone. It is almost a near-time manner before this information will be in a business intelligence tool, open for the stakeholders. It is not used in every company or by every foreman, but the opportunity is there which make the speed of data collection meet the velocity criteria.
Veracity	As with majority of data, the quality and accuracy can be both controllable and less controllable in terms of health and safety. The more data available, the more it can make up for the lack of quality or accuracy, but it should be addressed by a software solution.
Value	There are many ways to measure value within health and safety, and it can be done by controlling number of injuries, financial terms, productivity ect. The important thing is to address the value when using the data analytic tools and methods.

Fig. 5.3: Connection of health and safety environment data with the 5V's.

To understand and properly predict the future trends of health and safety practices during the construction projects, historical reports regarding the injuries, need to be analysed. The occupational accident reports are presenting factors which are connected to direct causes of the injuries, however, other aspects which might have an indirect influence, are not elaborated. These are reporting what happened, but not necessarily why it occurred. Usually the aspects such as weather, payroll or equipment features, are overlooked in occupational accident reports, however, these factors might be crucial in order to overcome the same accident in the future. To find the source of the problem, the construction companies need to identify the factors contributing to occupational injuries, afterwards, forecast the possibility of future risks [Deloitte, 2012].

To make clear, why particular accidents are happening, and if these types of injuries will continue in the future, the advanced data management methods need to be used by the construction companies [The International Institute of Risk and Safety Management, 2017]. Data management provides the range of tools which help to investigate the broader perspective of the working environment. By using data analytics, effective prevention plan might be made. However, the majority of construction companies have difficulties to start data-driven working environment practices since they are not familiar with current data tools and techniques [Deloitte, 2012]. The health and safety managers need to investigate their own structured and unstructured data, also find data sets from external sources which in some situations might bring the significant benefits [The International Institute of Risk and Safety Management, 2017].

Since the data management process is complex and extensive, the report is giving a guidance how to conduct this process form the first to the last step. The framework includes four stages of data management proposed by Mello et al. [2014] with the connection to the main factors contributing to the occupational accidents, see Fig. 5.4.

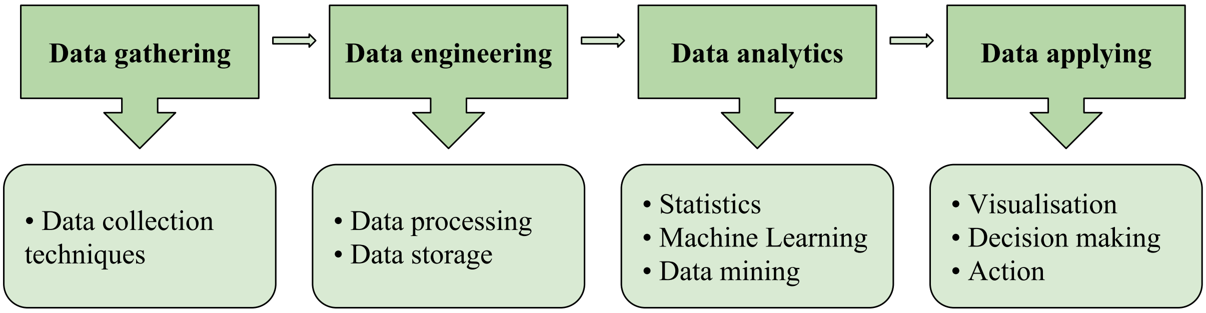


Fig. 5.4: The process flow of the data management framework creation.

Each stage is elaborated through the theoretical lances, moreover, underpinned by real industry examples, statistics, and cases. This combination of both theoretical and practical knowledge makes the report relevant from the wider group of industry professionals. The framework is presented in a way that the companies of distinct sizes and practices, might take what is the most relevant and beneficial for them.

Part III

Solution

Part III of the report is presenting the techniques within data acquisition, access, analytics and application. This is done to provide the answer for the following research question: *Which techniques and methods might be used for collecting and analysing data?* Chap. 6 is explaining the main data collection sources, and the techniques how to gather needed information. After the data is collected, Chap. 7 is explaining what are the possibilities of data storage and processing. Chap. 8 is elaborating the techniques of data analytics. The last chapter in this part presents the process of data application, and shows how data enhances decision making within the organisations, see Chap. 9.

6 | Data gathering

This chapter gives an introduction to the fist step of the data management process which is data gathering, see Fig. 6.1. This is done to provide examples of data collection techniques which might be used by the construction companies. Sec. 6.1 presents and explains different data collection tools, wile Sec. 6.2 is elaborating the benefits and challenges of this process.

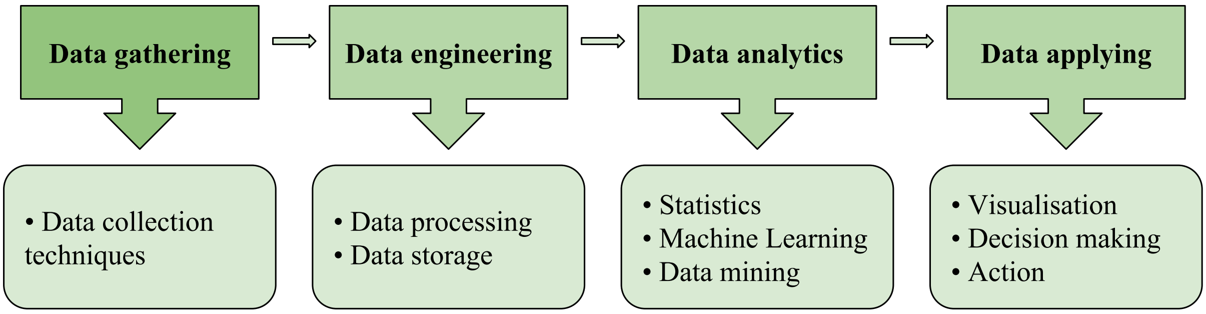


Fig. 6.1: First stage of the data management process.

6.1 Data collection

Data management is chosen as a tool of the construction industry improvement in terms of health and safety environment. The first step in data management process is data gathering, see Fig. 6.1. In the beginning, the sources of data need to be identified. There are two main types of data sources, namely, public and private. The difference between them is if data is gathered inside or outside of the organisation [Deutsh, 2015]. Moreover, the ways and methods of how data might be captured from these sources need to be distinguished, for instance, by sensors, drones or mobile devices. Many different formats of data might bring the benefits if the industry professionals know how to gather data, why it is relevant, and in which way it might be used.

For all aspects contributing to the accidents, the possible ways of data gathering are determined in the sections below. These show the wide range of private and public sources which might be used by the construction companies. It was decided by the authors of the report, to include the various data capturing means from the global practices. The research papers show that the construction companies worldwide are using data management solutions for improving construction processes with the satisfactory results [Zhang et al., 2018; Lee et al., 2017; Zhu et al., 2016; De Dominicis et al., 2013]. Therefore, the techniques of data capturing are presented in order to give an example for Danish construction companies. It is a way to increase the importance of digitisation health and safety environment.

Different data collection means were chosen in the way, that various companies might find the most appropriate ways to gather data with available resources. The larger companies, which have a possibility to invest in additional technologies, are having more possibilities to collect data regarding particular factors influencing the working environment. However, small and medium-size companies are also able to start their data collection process with less experience and resources. Moreover, Tab. 6.1 shows, by which data collection methods, the information for particular aspects might be collected. By choosing only some of the presented techniques, the majority of the factors might be covered. In some situations, the similar data might be gathered by using different means.

Tab. 6.1: Different types of data gathering techniques.

	Individual aspects	Task aspects	Management aspects	Environmental aspects
1. Data from client	+	+	+	+
2. Cameras	+	+	+	
3. Smart devices	+	+	+	+
4. Card-swipe readers	+		+	
5. RFID and wearable sensors	+	+	+	
6. The mobile sensing device	+	+	+	
7. Drones	+	+	+	
8. Public sources	+	+	+	+

Data collection techniques were chosen in the way that would cover the most important aspects, directly influencing the accident's occurrence. Tab. 6.1 is presenting only several data collection techniques. Danish construction industry is slowly starting the data management practices, therefore, it is important to be not overwhelmed by large amounts of information. However, as more factors are analysed, and more techniques are used, the better results are obtained in the end. The following sections are presenting chosen techniques supported by the examples from the industry.

6.1.1 Data from client

The least costly and the most simple way to gather data, is data collection from the client. The clients are responsible for ensuring that the contractors are able to conduct the work according to the health and safety rules [Sector Working Environment Committee for Building and Construction, 2016]. Despite, the Health and Safety Plan, and the notes form safety meetings, the contractors might be supported by some additional information about the project. The additional data from the client might contribute to the improved working environment planning. Moreover, the information concerning the client itself, such as organisation, the number of employees in different departments or the communication lines, might be obtained [Deutsh, 2015]. This type of information improve the communication between the parties, as in the case of issues, the contractors have a clear picture, who they should contact. Depending on the project and type of obtained data, the **individual**, **task**, **management**, and **environmental** aspects influencing the accidents, might be analysed.

6.1.2 Cameras

The cameras on-site help to collect data which is needed for warning system creation. The cameras enable to control the site from different angles. Captured video material is further transferred for the analysis. The material obtained by the cameras might be used for distinguishing the positions of particular workers, construction materials or machinery. Based set of videos, also the future positions of workers and machinery might be predicted. To capture the full view of the construction site it is important to ensure that the views from particular cameras are overlapping one on another [Zhang et al., 2018; Zhu et al., 2016].

First, the cameras are delivering the information about the overall situation on site. By having the clear picture, **management aspects** regarding project might be improved. **Individual aspects** of the workers, such as behaviour on site, taking short-cuts, or using safety precautions might be monitored. Finally,

Task related aspects might be improved by narrow analyses of the activities on site or the locations of particular work gangs.

6.1.3 Smart devices

The technologies of smart devices are broadly used by the individuals, and this fact contributes to the improvement of existing project management practices. Several features of smartphones and tablets are leading to their extensive use, and these are: compact size, lightness, ease to use, and wide availability in reasonable price. Due to easy access to the smart devices, not only the site managers are using this technology, but also regular workers on site. However, smart devices might be used not only for the communication process but also for data gathering about the workers and their positions on site. Nowadays, the smartphones are supported by native localisation techniques such as Global Positioning System (GPS), Assisted GPS (aGPS), Global System for Mobile communications (GSM) or Wi-Fi, therefore, data collection process is becoming more easier [De Dominicis et al., 2013].

Each of the localisation techniques has different level of precision and energy consumption. The GPS is using satellites in order to determine the localisation. The precision is relatively high, however, it is consuming large amounts of energy. Further, the aGPS depends not only on information from satellites, but it is also using the GSM antennas for triangulation of the location. Therefore, the aGPS is more precise even in surrounding of tall objects. Another localisation technique, the GSM, is using satellites supported by stationary the GPS stations. The main advantage of this localisation method is lower energy consumption, however, the accuracy in remote areas is low. Last smartphone localisation technique is Wi-Fi positioning system (WPS), which is supported by wireless access points. This type of localisation assistance provides Google [De Dominicis et al., 2013].

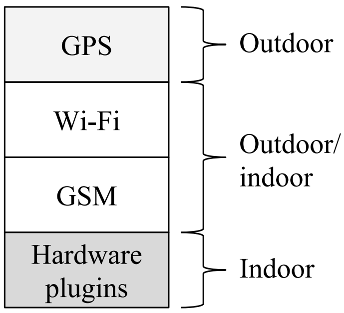


Fig. 6.2: Distinct localisation techniques [De Dominicis et al., 2013, *ed.* Fig. 5].

Due to extensive use of smart devices on sites, these might be used for everyday management improvement. The information about the labour location might be perfectly used for risk prevention planning. Hazardous positions of the workers are tracked in real time, hence, the accident prevention might be adapted more accurately. When analysing the native localisation techniques used in the smart devices, some limitations are distinguished. As it is shown in Fig. 6.2, GPS is assessing precise location only outdoors. The aGPS is more accurate, but still it needs unobstructed path to the satellite to assess location precisely. The GSM is working correctly in urban areas, where the distances to the base stations are not large, but in remote areas, it might be not precise. Wi-Fi positioning system basically relies on third party databases which might be not accurate in remote areas, the same as GSM localisation technique [De Dominicis et al., 2013].

The limitations of native location techniques in smart devices might be overcome by applying the combination of smart devices with Wireless Sensor Network (WSN) [De Dominicis et al., 2013]. WSN is a systematised wireless network which contains the number of sensor nodes. These nodes are collecting data about real-world objects [Wu et al., 2013]. The devices need to be connected to WSN to enable data merging and it can be done by supporting the smartphone with hardware plugins, such as Bluetooth or USB. In the study of De Dominicis et al. [2013], the localisation techniques were extended by using USB plugins. The benefit of using this combination is that data regarding labour location might be gathered both outdoor and indoor without any interruptions even with the 1 meter precision [De Dominicis et al., 2013]. This type of of data collection provides the direct information about **individual** and **task** aspects, further, this data might be used for planning and analysing **management** aspects. Despite the localisation possibilities, provided by smart devices, **environmental** aspects might be analysed in some extent by using appropriate applications.

6.1.4 Card-swipe readers

The card-swipe readers are one of the automatic identification technologies available in the construction industry. The card-swipe readers are working only by direct contact of the reader and the card, while some of the newer automatic identification technologies, such as voice recognition, do not require the direct contact of tagging and reading objects [Wu et al., 2013].

In some situations, card-swipe readers are providing information which is hard to collect by any other data collection tool [Deutsh, 2015]. In the construction industry, the card-swipe readers are often used at the entrance to the construction site. This information is especially important on large construction sites in the case of emergency when all the workers need to leave the site. It is useful to monitor if any workers are still on site. In some cases, data obtained through the card-swipe readers is not reliable, as some of the employees might swipe in while entering the construction site, but not necessary swipe out [Deutsh, 2015]. This data gathering technique helps to analyse **management aspects**, as the overall number of workers on site, and partially the **individual aspects** by distinguishing if the workers are paying attention to swipe in and out when entering or leaving the site. It might give a warning, that the additional training regarding safety rules on site is needed as the labourers do not understand the importance of information gained through the card-swipe readers.

6.1.5 Radio-frequency identification (RFID) and wearable sensors

With the increasing popularity of monitoring technologies in the construction industry, it is now possible to collect data about labour productivity and safety habits [Lee et al., 2017]. By utilizing this information, some working environment issues might be overcome. Monitoring technologies are used not only for observing the working environment, but also for personal monitoring. Advanced activity trackers are measuring the number of steps, distance or even the number of climbed stairs. Moreover, activity trackers, such as Fitbit, are monitoring heart-rate or elevation by integrated GPS navigation device [Lee et al., 2017]. This type of monitoring technologies are gaining more popularity in different industries due to their real-time labour health and safety monitoring and ability to provide accurate accident prevention methods [Lee et al., 2017].

Radio Frequency Identification (RFID) gained more popularity in the construction industry, especially for material tracking [Kelm et al., 2013]. The RFID is a technology which recognises the objects by using distinct frequency radio waves [Lu et al., 2011]. Despite the tracking of materials, the RFID is used to solve the important issue in the construction sites, which is wearing of the personal protective equipment by the employees. This factor relates to **individual aspects** of accidents occurrence. At the construction site, the workers are obligated to wear protective equipment which is delivered by the employer [Sector Working Environment Committee for Building and Construction, 2016]. The basic elements of protective equipment are: appropriate clothing, safety footwear, helmet, protective goggles, safety vests [Sector Working Environment Committee for Building and Construction, 2016; Kelm et al., 2013]. Based on the project, each working environment has different personal protective equipment rules, for instance, in the power-plant and residential construction projects, these restrictions might differ in some extent. Kelm et al. [2013] in their research identified the need of incorporating latest data gathering techniques into better working environment assessment, especially in terms of personal protective equipment.

Several studies suggested to use the RFID tags on personal equipment [Kelm et al., 2013; Costin et al., 2012; Lu et al., 2011]. First, the RFID might serve the function of identification cart. The RFID tag can be attached to almost almost all of the personal protective equipment pieces. When workers are passing the construction site gate, or the gates to the particular working areas, the tag is sending the information to the reader. The data about the necessary individual equipment is automatically transferred to the database. Moreover, this technology enables to provide the warnings, in the case if the worker is missing an important protective equipment. The warnings might take one of several forms, for instance, audio, vibration, automatic lock of the gate. The examples of RFID tags on the protective equipment are provided in Fig. 6.3.

Fig. 6.3: Personal protective equipment tagged with the RFID [Kelm et al., 2013, Fig. 10].

The health and safety improvement will not come if workers will only be using wearable sensors. First, it is more important to understand what are the root factors of the accidents or unsafe behaviour during the work. Monitoring technologies are collecting data on the individual worker level, afterwards, this informations might be used for reorganising the working environment in the way, that would contribute to labour safety and performance increase. Wearable sensors are collecting information which might be used regarding the **individual aspects**, **task aspects** and **management aspects** affecting accidents occurrence.

6.1.6 The mobile sensing device

The mobile sensing device, presented by Lee et al. [2009], is a part of broader safety monitoring system, and this device is developed for automatic data collection regarding hazards on site. The difference between the wearable sensors and the mobile sensing device is, that latter is not only collecting the information about the worker, it also warning and transmitting the voice messages by wireless connection to the site manager's office, see Fig. 6.4. This technology is especially used for prevention of fall accidents, for instance, when the workers are close to the danger zone, as unprotected building edge, the labourer receives the warring by a sound message. If the worker does not move from the danger zone, the information about his position is automatically sent to the office, where safety manager might take action to prevent the accident. In the case, if the safety manager is not at the office, he receives a text message about the risk on site [Lee et al., 2009].

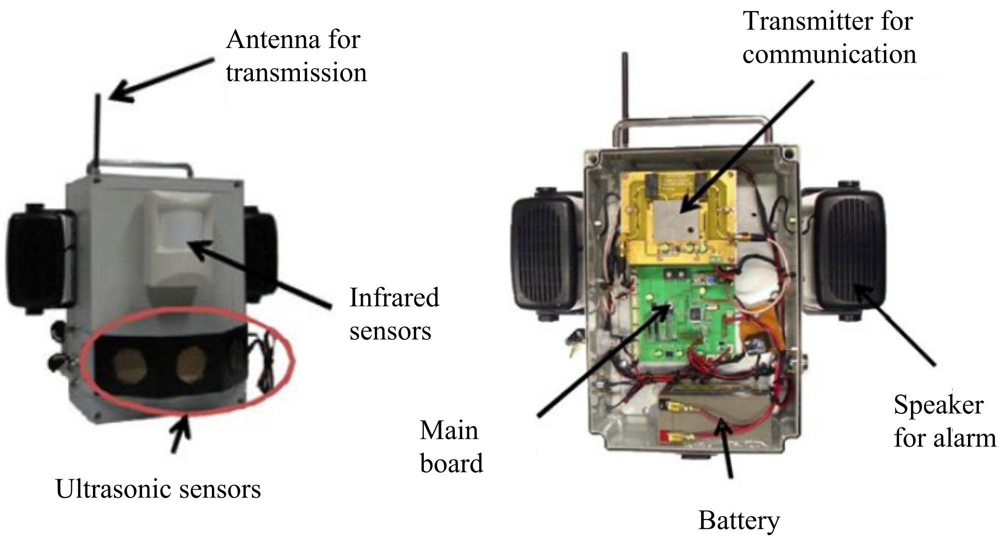


Fig. 6.4: Configuration of the mobile sensing device [Lee et al., 2009, Fig. 3].

In order to understand the working principle of the mobile sensing device, the main configurations of the mechanism are presented in Fig. 6.4. The device has two types of sensors which are responsible for different functions. On the construction site there are many temporary materials stored around, therefore, it is important to distinguish, either if it is stored material or worker in danger zone. Infra-red sensor is detecting body heat, thereby, distinguishing the labourer from the other objects. Further, ultrasonic sensor is measuring the distance from the worker to the risk zone, see Fig. 6.5. Based on the distance from individual to the hazard, the mobile sensing device is transmitting the signals as light, beep or voice message. By the main board, the information about the incident is sent to the site office [Lee et al., 2009].

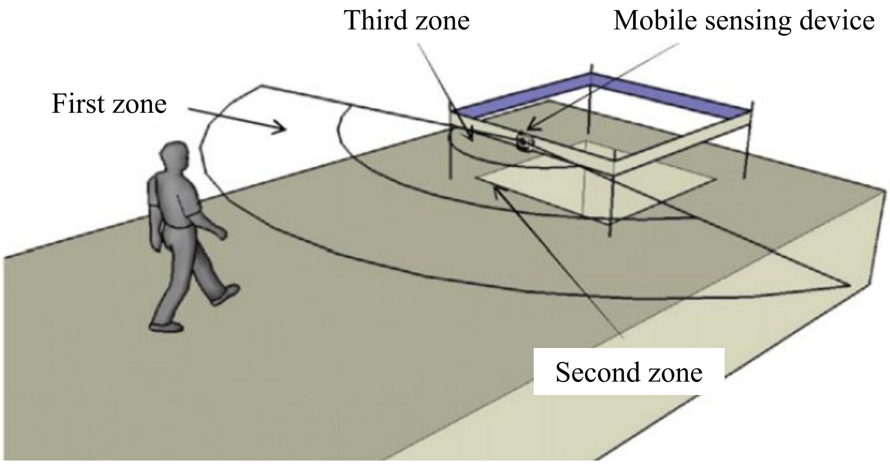


Fig. 6.5: Detection zones [Lee et al., 2009, Fig. 5].

The information about all dangerous situations are collected on the hardware and this data might be analysed afterwards. The combination of data gathering and real-time broadcasting provides a range of benefits. The company might analyse the positions of labourers during each hour of the day, thereby, properly plan the activities on site to protect the workers and enhance the productivity. However, the most important is that proper actions might be taken immediately. In some situations, if worker fall from a height, the real-time warning might prevent the worker from the accident [Lee et al., 2009]. By using the mobile sensing device, **individual aspects** of the workers might be analysed, as it shows the worker's behaviour. The study of Lee et al. [2009] shows, that some of the workers are entering the danger zone only due to curiosity. Further, **management aspects** might be improved by larger encouragement into safety planning. **Task aspects** might be improved by additional warning systems. While the worker has to perform the task near the risk zone, he will receive the warnings if will come too close to the hazard, which might be the unprotected hole, sharp element of construction or sticking reinforcement. The stored information might be analysed in the way to see how often, at what time of the day and how many workers need to pass or work close to the danger zone.

6.1.7 Drones

Drones are extensively known and used in different industries [Li and Liu, 2018]. The construction companies are starting the process of drones implementation into their practices [Li and Liu, 2018; Dupont et al., 2017]. Several kinds of drones exist, for instance, multi-rotor, single-rotor or fixed-wing, but the most advantageous is multi-rotor drone due to its robustness, mobility, flexibility, higher payload capacity and low investment cost, see Fig. 6.6. In the construction industry, it is a new technology, and full potential of drones is still uncovered. In terms of working environment, it is possible to monitor the activities on site and control if the labourers comply with the health and safety rules [Li and Liu, 2018]. This type of monitoring brings savings in construction cost, time and minimise the number of accidents, thereby, contributes to working environment improvement [Herrmann, 2016]. By increasing awareness of construction professionals about possible ways of using drones, they could benefit and change their practices in more effective way [Li and Liu, 2018]. The **individual**, **task**, and **management** aspects might be addressed by data gathered by drones.

Fig. 6.6: Multi-rotor drone [DJI].

There are various different applications of multi-rotor drones through the whole project life-cycle, however, only the most important functions which are contributing to working environment improvement are presented in this section. First, drones might be effectively used in construction logistics as the movement of materials and machinery might be monitored and analysed. The working locations of the employees and the storage areas of particular materials might be easily distinguished and reorganised to improve safety and productivity in site [Li and Liu, 2018]. In the more advanced applications, this technology enables to track the real-time locations of materials by using the combination of drones and RFID [Hubbard et al., 2015].

The construction management on site might be enhanced by application of drones. Based on video information provided by this technology, the 3D representation of the site is more accurate. The combination of drone technologies and augmented reality (AR) might contribute to improved visualisation techniques. By combining real and virtual representations of the site, the proper actions might be taken by the construction managers regarding labour and material planning or risk identification [Li and Liu, 2018].

Drone technologies are broadly used in construction land surveying. It is not only less costly and faster way to do this type of work but also improving the working environment for the surveyors [Li and Liu, 2018]. The surveyors are usually working in harsh weather conditions, moreover, they are exposed to additional risks related to sloped surfaces or surrounding of the heavy machinery [El Meouche et al., 2016]. It this case, drones are not only collecting the information but directly contributing to safety improvement of the surveyors. The other situation when drones are directly minimising the risk of the accident is by using this technology for structural inspections. The drones are minimising the risk of fall from hight, or hit by objects of the inspecting individual since this technology might monitor all the roof parts, high walls, towers or any kind of damages caused by fire or any other events [Mat Yasin et al., 2016].

Irizarry et al. [2012] in their study analysed the application of drones toward safety improvement on site. Based on the study conducted in the field, some recommendation for the safety managers were determined. The drones would improve the safety monitoring on site if these would meet three main features: autonomous navigation, voice navigation and extended battery life. Usually, the safety managers are controlling the drones manually, but it would be even more beneficial, if the site scanning could be performed in automatic way. The drone would follow earlier defined paths and locations on site without interaction of the people. Further, it would help to collect the data automatically without any disruptions. Despite the data collection feature, it could be beneficial, if the safety manager could have a possibility to talk with the workers through the drone device. In hazardous situation, it might be the fastest way to interact with the labourers and prevent the accident. The safety manager would have a possibility to give a guidelines based on the materials streamed from drone to his mobile device. Finally, third important requirement is longer battery life. To gather information from the whole construction site, the drone needs to have enough power to fly for a longer time [Irizarry et al., 2012]. The drone presented in Fig. 6.6, has 30 minutes flight time which might be enough to control all construction site [DJI].

Despite the long list of benefits, some limitations of using drones on the construction site need to be addressed. The most important issue is related to the legal regulations regarding usage of drones. The drone operations in Denmark are regulated by the multiple regulations [Dronebevis; Civil Aviation Administration - Denmark, 2004]. According to Dronebevis, there are distinguished two types of circumstances while operating with the drone: operations on the countryside and operations in urban areas. In case of using the drone in the urban areas, the individuals need to have a drone license which in Danish is called *Dronebevis*. Therefore, if the drone is used on the construction site, only professional drone operator can perform these tasks. Other issues, presented by Li and Liu [2018], are the distraction of the workers, the fly paths might be influenced by harsh weather conditions and affect the reliability of the information, or insufficient power capacity in case of using additional devices connected to the drones. Hence, the construction companies which are deciding to use drones to improve their health and safety practices need to address the additional resources.

6.1.8 Public and open data sources

Despite the private data sources, the data might be gathered also from public and open data accesses. Public data is supporting the construction professionals with large amount of information which might be used for assessing the trends, strategies, or just for effective project planning [Deutsh, 2015]. In terms of health and safety, Danish construction companies might gather public data in several ways. Some of the suggestions are presented below.

- https://www.dst.dk
- https://www.danskbyggeri.dk
- https://arbejdstilsynet.dk
- http://www.dmi.dk

Statistics Denmark provides the wide range of statistics about the society and industries in Denmark. The statistics provide a valuable information about the trends, and the number of employees with occupational injuries. This information might be used while planning the working environment, or by preparing health and safety management regulations within the organisations. The two other sources of statistics are available on *Dansk Byggeri* and *Danish Working Environment Authority* websites. These websites provide more detailed information regarding the Danish construction industry and the working environment. The statistics are open data sources and free of charge. The combination of public and internal data might provide a great input in appropriate health and safety planning.

The Danish Meteorological Institute (DMI) provides the meteorological service regarding the weather and environmental conditions [Hansen, 2007]. The data provided by DMI, might be used for analysing **environmental** aspects, such as weather forecasting or seasonal variations. DMI provides the background knowledge for the companies, and this information might be used as a help for the construction working environment planning.

6.2 Benefits and challenges of data collection

As it is shown in the sections above, several data gathering techniques are not only collecting distinct data regarding the working environment but also directly preventing the accidents. By combining data collection tools with the transmitters for communication, the workers might be warned in real-time, when they are exposed to the hazards. Some of the accidents might be prevented as the workers could receive the direct guidances regarding their activities. However, several limitations of presented techniques need to be taken into account concerning the Danish construction industry. Data collection techniques, presented in this chapter, are providing the examples from the worldwide practices. The strict law regarding personal data and monitoring of employees, might bring the challenges when adapting the presented techniques in Danish organisations. Moreover, the EU General Data Protection Regulation (GDPR), which came into the force in 2018, is regulating the personal data of the EU individuals [European Commission, 2018b]. The regulations as GDPR needs to be followed also by the construction company, therefore, some of the data collection techniques might be difficult to implement, even though these are providing value for the organisation.

7 | Data engineering

This chapter explains the methods and theories behind data engineering which is the second step of the data management process, see Fig. 7.1. This is done to introduce possible techniques for processing and storing data. First, Sec. 7.1 gives a brief introduction to the term. Further, Sec. 7.2 explains data processing and introduces methods which can be applied. Sec. 7.3 briefly goes through two different ways of storing data, while Sec. 7.4 elaborates a case study.

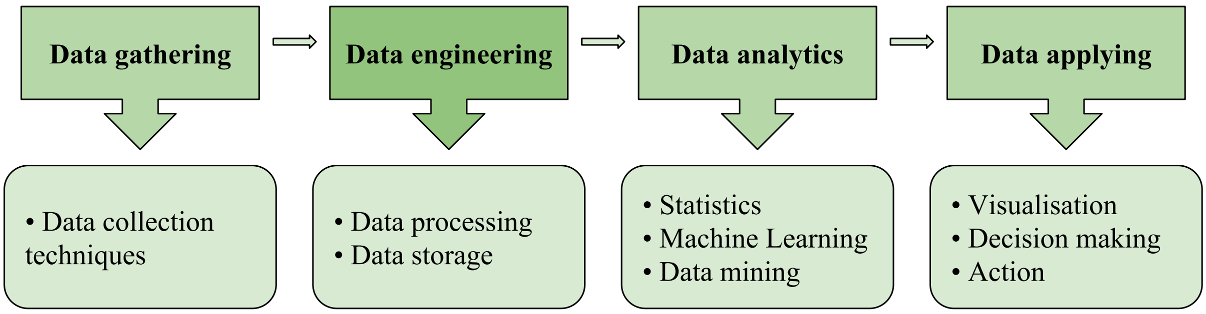


Fig. 7.1: Second stage of the data management process.

7.1 Possibilities of data engineering

The world today is changing rapidly when it concerns technology. Digital storage is now more cost effective than paper and media storing. The consumption and creation of data through devices and Internet of Things has exploded, and the growing need to predict and monitor competitors have addressed certain challenges in the society [Yu and Guo, 2016]. Many existing computer and software tools are not made for storing and processing voluminous amounts of data, which contains the variety of file formats, with a stipulated velocity, precise veracity, and to a reasonable value [Yu and Guo, 2016]. Data engineering is addressing the four V's, shown in Fig. 5.1, and concerns the receiving, transforming, storing and accessing of data. It is the foundation of data analytics that ensures that the data is properly handled before it goes out to others. For the data analytics methods, to be able to perform sufficient analysis, the need for scaling up hardware becomes necessary [Bilal et al., 2016; Singh and Reddy, 2015].

7.1.1 Scaling

To be able to process and store these large amounts of data, the system needs to have enough capacity. When a system meets increasing demands it needs to adopt these, and have the capability to process the data, this is called scaling [Singh and Reddy, 2015]. Different platforms have different ways of scaling, but from a broad perspective they can be categorised in two groups *horizontal scaling* and *vertical scaling* [Bilal et al., 2016; Singh and Reddy, 2015]. The difference between the two is shown in Fig. 7.2.

Horizontal scaling is also known as "scale out". It is a way of dividing workload by adding new servers or machines to an already existing server or machine to increase the processing capacity [Singh and Reddy, 2015]. This is the least costly way of increasing performance since it can be done step by step and just add the capacity needed. There are no limits on how much scaling can be done [Singh and Reddy, 2015]. The downside of horizontal scaling is that there are limited software frameworks that can take advantage of it.

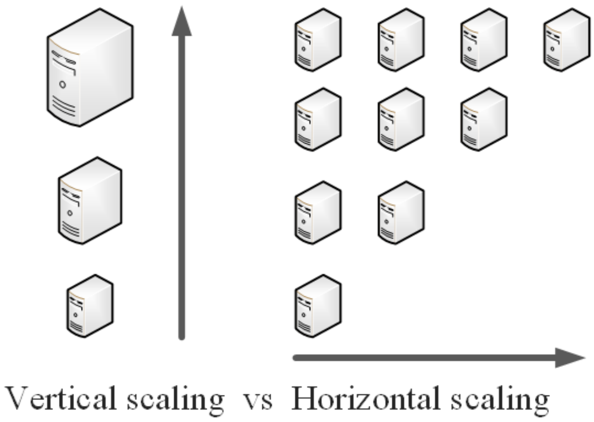


Fig. 7.2: The difference between vertical and horizontal scaling illustrated.

Vertical scaling, also called "scale up", is used to upgrade the hardware within one single server. This can be done by improving the processor, memory or disk [Bilal et al., 2016; Singh and Reddy, 2015]. The advantage of scaling up is that it is easy to manage and install, on the other hand, upgrading the hardware can be expensive, since the user has to pay for more power than needed. Taken into account that future workloads are more powerful, the need for always adding a stronger hardware that meets the current requirements, is necessary. It is also a question about capacity and that every single hardware or server have a limit on how much it can be increased [Singh and Reddy, 2015].

Both the vertical and horizontal scaling are dealing with different challenges which should be addressed when choosing a platform. Singh and Reddy [2015] say that three factors are important when choosing platforms: data size, speed or throughput optimisation, and model development.

- **Data size:** This is probably the most important factor. If the data can fit into one single storage, vertical scaling platforms will be the best fit since the entire data can be processed on one computer. If the system's memory is too small, horizontal scaling will be the most beneficial.
- **Speed or throughput optimisation:** It is the platform's ability to process data in real-time and how much data it can handle simultaneously. Here deciding if process data in real-time or simultaneously is the key when choosing. Vertical scaling is great for optimising small data sets for speed, while horizontal scaling is best for processing large amounts of data simultaneously on the behalf of time.
- **Developing a model:** When developing a model in data analytics this is often done offline. These processes often take a lot of time. Making the best model often requires a lot of data to train with, therefore, horizontal scaling is usually the best fit for these operations.

Based on the research of Singh and Reddy [2015], it is chosen to elaborate only on platforms that use horizontal scaling, like Hadoop and Spark. This is done because horizontal scaling is seen as the most appropriate for data within health and safety. Due to Fig. 5.3, large amounts of structured and unstructured data are needed to be processed, run simultaneously, and tested. These factors are seen as more important than the speed of the process.

7.2 Data processing

Within data engineering, data processing concerns the conversion of the raw data into meaningful information. As many different processes, inputs are given to a process to produce an output. It is a way of manipulating data to improve or solve a problem [Bilal et al., 2016]. Some of these are presented bellow.

7.2.1 MapReduce

MapReduce was introduced by a data platform called Hadoop [Steele et al., 2016]. They created an ecosystem or a framework that can scale thousand of nodes and are developed to make MapReduce programs executed successfully [Bilal et al., 2016; Steele et al., 2016]. MapReduce is a programming model that has two key functions *mapping* and *reducing* large amounts of data. The process can be divided into three steps, first, the mappers put the data into pairs that have the same key-value. Second, the shufflers take the pairs that have the same key-values and puts them in a single DataNode. At last, the reducers decrease the amount of data so it comes out as a list, statistics, graphs, tables or some kind of summarising [Steele et al., 2016]. The three steps of the MapReduce process are visualised in Fig. 7.3.

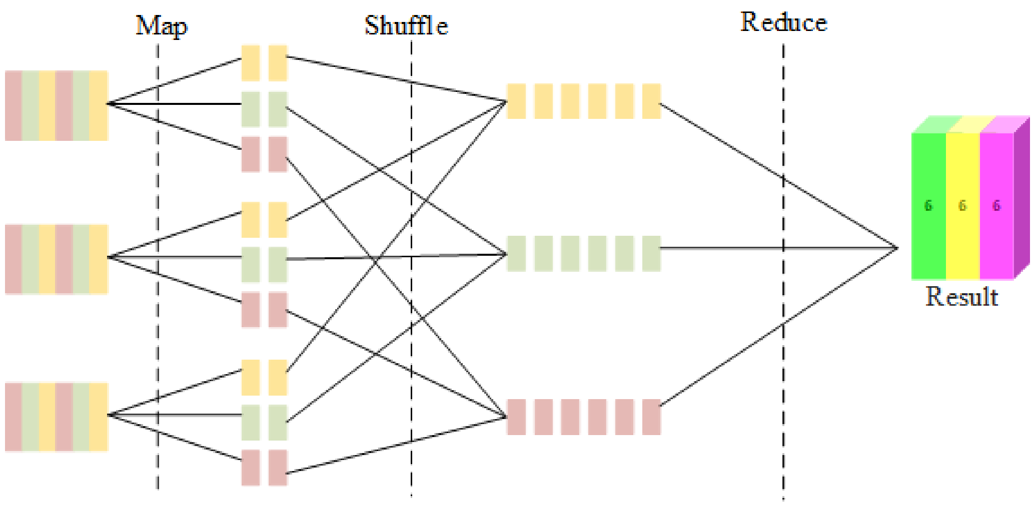


Fig. 7.3: MapReduce processing as a visualisation.

MapReduce programs are great for batch-processing tasks because the mappers and reducers often run several programs simultaneously [Bilal et al., 2016]. However, one of the biggest drawbacks of MapReduce is its struggle in running iterative algorithms and real-time graphs [Singh and Reddy, 2015]. Hadoop is trying to fix this by detaching some of the processing aspects of the MapReduce from the ecosystem and introduce Yet Another Resource Negotiator (YARN). YARN allocates resources and schedules the jobs across the cluster [Singh and Reddy, 2015]. It works as a resource allocation and management system, and gives Hadoop the opportunity to implement innovative applications.

7.2.2 Direct Acyclic Graphs (DAG)

Direct Acyclic Graphs (DAG) is another type of processing model, it takes data and represents its dependencies [Brito et al., 2016; Bilal et al., 2016]. The process takes bigger tasks, weight them, and make them into an arbitrary set of tasks showing the dependencies and communication between them [Khan and Ali, 2013]. Compared to Hadoop it can handle cyclic data and in-memory computing, this means that it can read files and save it as a memory [Storey and Song, 2017; Singh and Reddy, 2015]. Apache Spark is a computing platform that has advanced DAG. Many researchers argue that Spark is a more powerful tool that Hadoop, while others argue that it is a complement to it. However, Spark is reported to be 100 times faster than Hadoop on memory-resident tasks and 10 times faster on disk-resident tasks [Storey and Song, 2017; Bilal et al., 2016; Khan and Ali, 2013]. To create an enterprise software the support functions of Spark are vital, these are shown in Fig. 7.4.

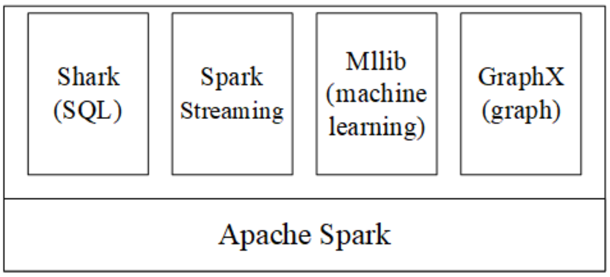


Fig. 7.4: Components of Apache Spark [Bilal et al., 2016, Fig. 3].

7.3 Data storage

Another aspect of data engineering is data storage. As the name indicates, this is where the data is stored and managed in a scalable way. The data storage should work in a way that satisfies the requirements of the operations that are going to use the data later. In this section, some of the data storage technologies are presented that address the 5V's mentioned in Chap.5. It is not covering relational database systems since these are too complex for data systems, and often get expensive and less efficient when handling the 5V's [Storey and Song, 2017; Marz and Warren, 2015]. This section is covering the current-state of data storage technologies that are capable of handling big amounts of data [Cavanillas et al., 2016].

7.3.1 Distributed file systems

A distributed file system is different from a local file systems normally found on computers because the data is stored in a cluster. It stores and manages large amounts of data across multiple nodes or computers. Even if the data is stored in different places, the data can be reached from each computer in the cluster, as if it was stored on only that computer. The pros about distributed file systems are that these provide data scalability, fault tolerance, and high concurrency through partitioning and replication of data on many nodes [Yu and Guo, 2016]. In this section two of the most common distributed file systems are discussed, HDFS and Tachyon.

Hadoop Distributed File System (HDFS)

Hadoop distributed file system (HDFS) is the second part of the Hadoop framework [Singh and Reddy, 2015]. It allows data to be stored across different clusters or machines, this makes it possible to process all the data in the different clusters and in parallel. This is the main reason for why it is so successful. The HDFS can typically store data of any size and format [Yu and Guo, 2016]. The HDFS architecture is based on one master node or NameNode, remaining nodes are called slave nodes or DataNodes. The DataNodes are responsible for storing the actual data, where the NameNode is responsible for managing and maintaining all the data and the DataNodes [Bilal et al., 2016; Yu and Guo, 2016]. Distribution and replication are the key in HDFS, and these make the system highly fault tolerant and prepared for hardware failure [Bilal et al., 2016]. However, it is not ideal in every area, as it is not suitable for storing large amounts of small data, because of the NameNode is managing the metadata. It is not optimal for doing many simultaneous modifications in random places in the data, and running high volumes of data with minimal delay [Bilal et al., 2016].

Tachyon

Tachyon is an in-memory distributed file system [Yu and Guo, 2016]. It enables to distribute reliable data through the cluster framework in memory speed [Bilal et al., 2016]. By using lineage information and memory, it increases the performance of the system. Tachyon is based on HDFS but can catch often read data and put it in memory, so that the files can be read in memory speed [Singh and Reddy, 2015]. One of the big advantages in Tachyon, beside the memory, is that it is compatible with Spark, HDFS, and MapReduce without changing any codes [Bilal et al., 2016; Singh and Reddy, 2015]. It can also remember specified parts or columns of big tables so that user can access these faster [Singh and Reddy, 2015].

7.3.2 NoSQL databases

NoSQL is in many ways a form of unstructured storage. The NoSQL was designed to meet the demands that are needed in data application by using different data technologies. It is an open source, distributed, and non-relational database [Storey and Song, 2017]. The horizontal scaling makes it designed for scalability and flexibility. Its aim is to avoid schema storage and focus on schema-less storage [Bilal et al., 2016]. It has four different data storage types which are:

- **Key-value store**: Key-value stores data sets in a schema-less way, in terms of keys and values [Storey and Song, 2017]. It is a simple way to store unstructured or structured data, where the keys are the way to access the data [Cavanillas et al., 2016]. The cons of key-value store are lack of consistency and that the underlying data is not described in a satisfactory way [Bilal et al., 2016].
- **Columnar Stores:** Stores data in columns and not rows. By creating columnar families and then use keys to compress sparse data in the columns. Columnar stores are not working optimal, when accessing all the columns at once. However, this hardly ever happens [Cavanillas et al., 2016].
- **Document Databases:** Document databases are, as the name implies, key-value stores where the value are documents [Storey and Song, 2017]. It is a way to store semi-structured data, this means that there are any requirements for a common schema [Cavanillas et al., 2016]
- **Graph Database:** Graph databases store data in a graphs structure that consists of sets of nodes, edges, and properties. It is perfect to store inter-connected relationship data such as social network graphs and communication patterns [Storey and Song, 2017; Cavanillas et al., 2016]. One of the downsides with this, is storing truly large data sets, for this the database is not yet suitable [Storey and Song, 2017; Cavanillas et al., 2016]

7.3.3 Cloud computing

Another way of storing data is by using a hybrid cloud approach. By storing data in a cloud, the data can rapidly be accessed with minimal management effort or service provider interactions. It is maintenance free, rapidly scalable, and bases the cost on the companies actual needs. With a hybrid cloud both sensitive and public shareable data can be stored. The cloud stores the sensitive data in a on-premise server and the public data in a cloud, see Fig. 7.5. This makes the process dynamic and highly changeable when it comes to workloads. The aim of the hybrid cloud is to provide businesses flexibility and give them access to more data management options. This is done by enable workloads to move between the cloud and the on-premise servers, as the companies computing requirements change. The limitations with cloud storing is that it requires good quality on the metadata, good data governance, and data integration processes [Storey and Song, 2017].

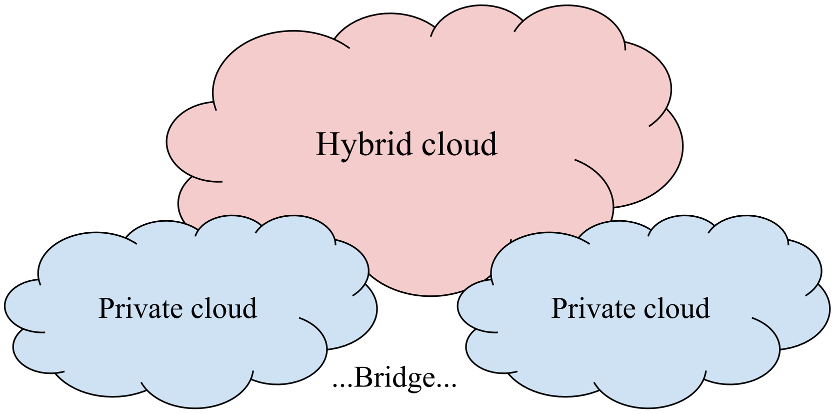


Fig. 7.5: Hybrid cloud storage [Storey and Song, 2017, Fig. 4]

7.4 Using of data engineering within health and safety

In this section, examples from case studies that have used data engineering, are presented. This is done to show opportunities within data engineering, and how it can be used to create value for the health and safety managers.

7.4.1 Hadoop Distributed File System (HDFS)

Guo et al. [2016] used HDFS to create a platform which classifies, collects, and stores data about workers unsafe behaviour. The case study was conducted on a metro project in Wuhan in China. The aim of the research was to create a platform that could identify patterns of unsafe behaviour. The study has taken the critics Behaviour-Based Safety (BBS) has received, and tried to make a risk knowledge base, which can help construction managers implement strategies and techniques to improve the safety. BBS, briefly explained, is an approach which is used to analyse workers behaviour and apply techniques that improve behaviour process [Li et al., 2015]. The analysis depend on observations done by project managers, and this is one of the main criticised areas. This is because the approach is time-consuming, expensive and can lead to decisions being made on subjective and bias opinions [Cameron and Duff, 2007; Li et al., 2015].

Research approach

The research was conducted on data from 15 stations, 8 tunnels and 1731 injury records from 2008 until 2014, which were examined to identify common recurring unsafe acts [Guo et al., 2016]. The platform creation consisted of three parts: *making a risk knowledge base*, *collect data*, and *store needed information in HDFS*. The risk knowledge base was made by using classification and coding.

The classifications were made by analysing standards and injury records, and by dividing them into unsafe behaviour types by using a work breakdown structure. Over 522 different unsafe behaviour types were identified. These were later categorised into 16 different themes, where each theme had several subdivisions with a number for the coding. The behavioural risk knowledge base consisted of these classifications and their coding. For explanation of available classification techniques see Chap. 8.

Further, the all data was collected by using intelligent video surveillance, which is a camera-based behaviour analysis technology. The technology usually contains some way of monitoring a motion or an object, which is able to identify a person or object, and track it. Camera-based behaviour can perform analysis on behaviour and activities [Bremond et al., 2006]. Other methods used were taking pictures of potential risks, on the daily safety check. Later, the images taken of unsafe behaviours, were collected and matched with unsafe acts stored in the knowledge base. The semantic structure used to determine the workers unsafe behaviour from the images, was providing the following information: action(s) executed on object(s)-worked-on using resource(s) at location(s) with nearby object(s) and nearby action(s) [Guo et al., 2016]. To test the validity of the semantic information extracted, project managers were set to label the images in the same way. Big amounts of data were needed to be stored, and for this HDFS meets the demands. It stored data from the intelligent video surveillance and the mobile applications, behaviour risk knowledge base, schedule management information and semantic information of identifying images.

Process for data storage

In the HDFS, the NameNode is the master as the DataNode is the slave node, see Sec. 7.3.1. The NameNode is storing the metadata and managing the DataNodes. The metadata includes data that attributes to the information, in this study it contained name of image files, the time the data was collected, location, who collected it, what kind of weather it was, and the files directory from the behavioural risk knowledge base. The DataNode is storing information in data blocks and offer service to the client. So when the client is uploading or retrieving unstructured or structured data, the NameNode will allocate this data to the free DataNode for these to create data blocks to store the data. In the research of Guo et al. [2016], the data stored was unsafe behaviour images with complete semantic information. Fig. 7.6 is presented to see the framework in HDFS for data sources and storage.

To find data stored in the HDFS the researchers could simply search in a filter where the data was sorted after different limitations such as weather, collector, injury, date, location, etc.

Limitations

Several limitations were found in the research, among them, the camera-based behaviour analysis technology. One single camera only fits to follow one specific kind of behaviour, so it is needed to change the specific rules for each camera in the different phases of construction. Another potential limitation is the language used, while collecting and analysing reports and different material. The language used to describe the pictures can differ, and this can cause confusion. The words can also have synonymous meanings which also can interfere with the result. The platform is a way of assisting with the storing and collecting of behavioural data among workers, it neglects human feelings like mood, emotions, and attitude. This is a limitation because these factors can affect peoples emotions, cognitive disposition, and action.

Conclusion

The collection of data, both video and mobile, was daily around 300GB, which is the same as 150.000 images. The image data was collected, by applying the semantic structure, completed with semantic information reflecting the workers unsafe behaviour. The number of the intelligent surveillance videos that were collected, are much higher than from the daily inspection photographs. The goal of the research was not to find a platform that worked all alone, but to make a helping platform for the managers. The study showed that by combining traditionally observations with data technologies, the limitations with BBS were overcome. The platform can be used as a reference for behaviour observation, and as a tool to proactive management of the health and safety on a construction site. It can help people to find patterns in unsafe behaviour and be a foundation for future, modifications in behaviour research.

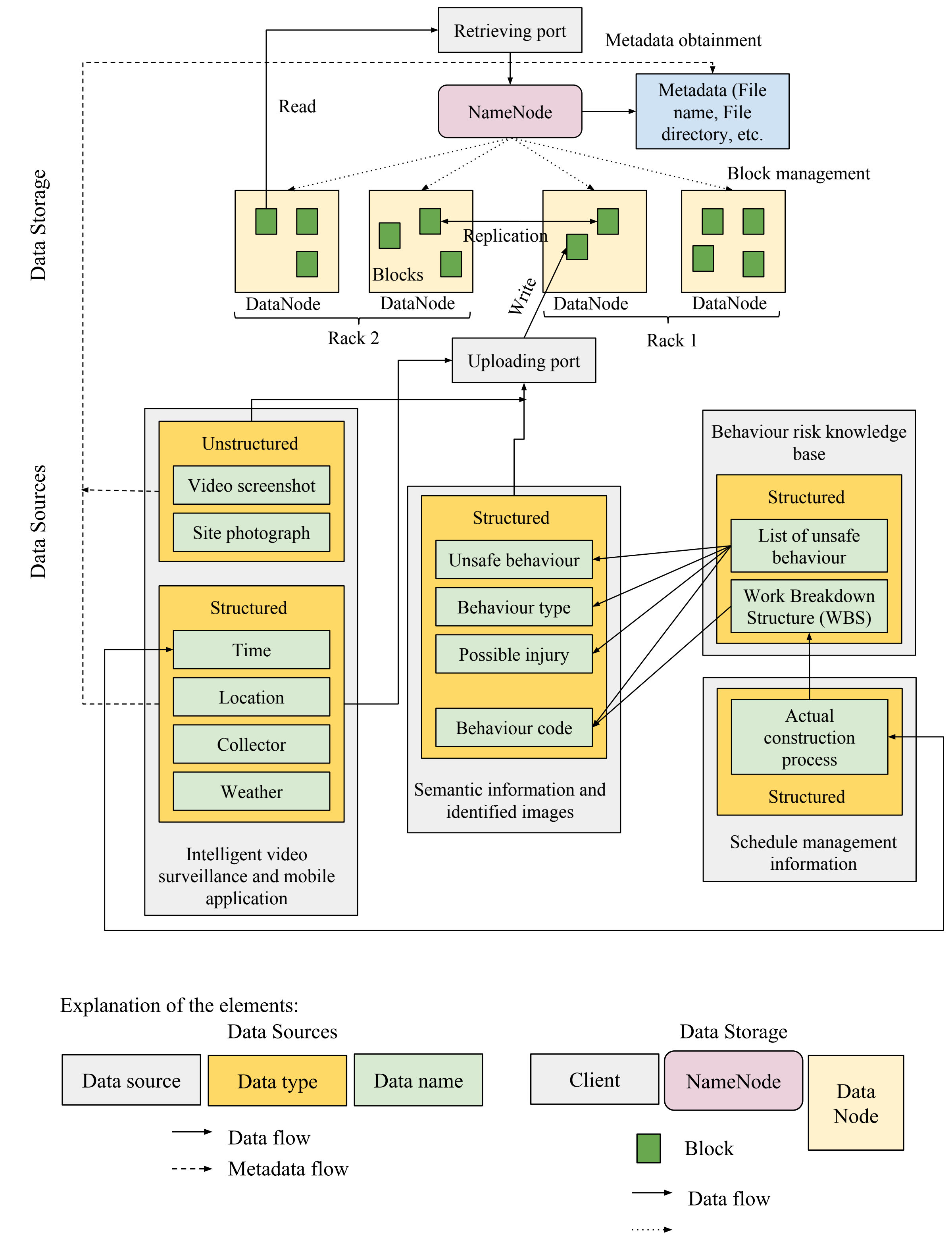


Fig. 7.6: The framework of data sources and storage in HDFS [Guo et al., 2016, *ed.* Fig. 5].

At last, HDFS showed itself as a high fault-tolerance data storage with high throughput. It improved the efficiency of the platform, and ensured fast response for uploading and retrieving of data. In the future, it can help managers to find the information they are looking for faster, and help them to make timely decisions.

8 | Data analytics

This chapter introduces the methods and theories behind data analytics which is the third step of the data management process, see Fig. 8.1. This is done to give a state of the art impression by incorporating many existing techniques for analysing large amounts of diverse data. First, Sec. 8.1 gives an introduction to the terms data analysis and analytics. Further, Sec. 8.2, 8.3, and 8.4 briefly explain data analytics techniques. Finally, Sec. 8.5 presents case studies where these tools were used.

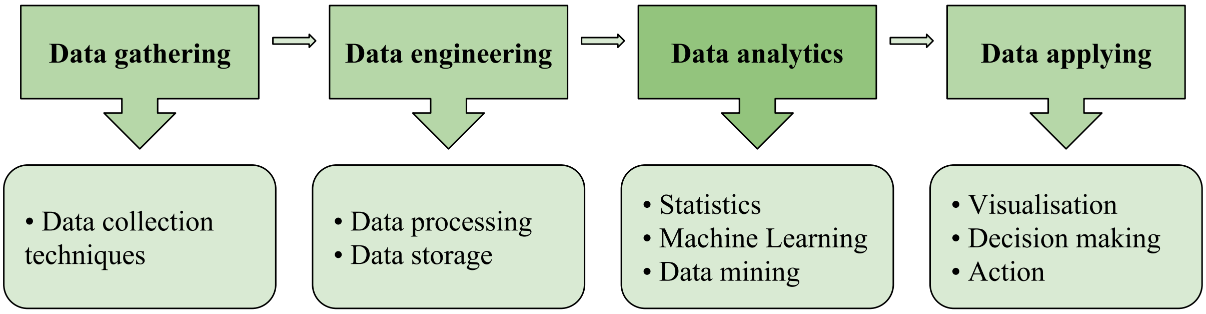


Fig. 8.1: Third stage of the data management process.

8.1 Data analysis and data analytics

Data analytics is a term that is often confused with data analysis. The term data analysis covers the entire process of acquisition, access, analytics and application of data, while data analytics is related to the tools and techniques used to achieve business objectives [Storey and Song, 2017; Yu and Guo, 2016]. Analysing data is not a new concept, companies have used the information in data to gain competitive advantage, reduce risk, and make more assured decisions for a long time [Deutsh, 2015]. As the accessibility and amount of data increases, see Chap. 6 and Chap. 7, current methods to utilise these are not designed to deal with the evolving streams. Therefore, improved data analytics tools and methods are needed [Storey and Song, 2017; Bilal et al., 2016; Yu and Guo, 2016]. There is no point in gathering and storing data if the companies do not use it or learn from it. When managing these amounts of data, data analytics can create value, and help companies understand their surroundings, and make decisions. Therefore, data analytics is addressing the last V, from Fig. 5.1, the Value [Marr, 2015].

8.1.1 Data analytics environment

Data analytics environment can be confusing and consists of many different disciplines [Bilal et al., 2016]. Fig. 8.2 presents a simplification of some of the disciplines and their connections. The figure is made, so it can be easier to understand the dependencies within the techniques that are elaborated bellow. These sections provide a brief explanation of the terms statistics, Machine Learning, and data mining.

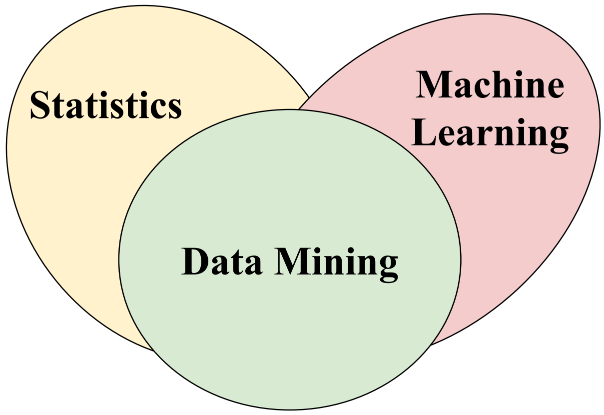


Fig. 8.2: The nature of data analytics [Dean, 2014, *ed.* Fig. 4.1]

8.2 Statistics

Various data analytics tools are borrowing statistical techniques to support their conclusion [Bilal et al., 2016]. It is a branch of mathematics, which concerns collecting, understanding, and accounting for the relevant uncertainties in data [Bilal et al., 2016]. Within data analytics, statistics is an important tool, to ensure that meaningful and accurate information is extracted from the data. It can be used to measure the uncertainties of data, assessing the quality of data, deal with missing data, and much more.

8.3 Machine Learning

Machine Learning (ML) is a branch within artificial intelligence. The techniques learn from identifying patterns in data. It helps people find hidden data and use it for opportunities and creating strategies within a specific area or for enhancing customer experience [Bilal et al., 2016; Cavanillas et al., 2016]. Some of the key ideas within Machine Learning techniques are learning from available data which implies taking that data and looking for a structure. Various techniques have been developed, all with different characteristics. Based on the way they learn about data these can be divided in three groups:

- Unsupervised learning
- Supervised learning
- Deeper learning

Unsupervised learning: Is about generating an algorithm that will look for patterns in data, without telling the algorithm which kind of data too look for, what kind of data it is, or label the data. The goal for the algorithm is to find patterns or clusters in the data, without human interruption [Yu and Guo, 2016].

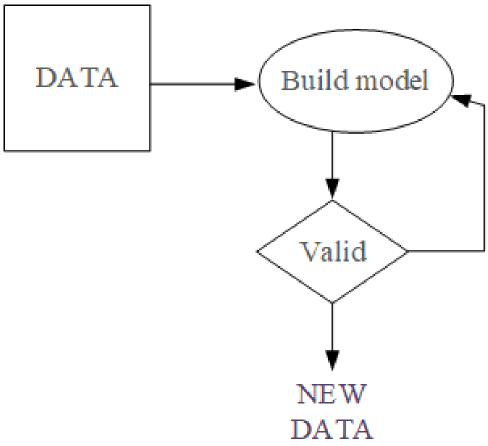


Fig. 8.3: Unsupervised learning schema [Deng, 2014].

Supervised learning: Is about classifying and regressing data by taking some known factors, label them, and let the algorithm learn from these labels for future data. It is different from unsupervised learning because it is assumed that there are already existing clusters, therefore, the labels are known. It consists of two steps: training and testing data. The training algorithm is constructed to recognise the key structure of the labels. After this is done, a test is made on new data to see how well the algorithm did [Yu and Guo, 2016].

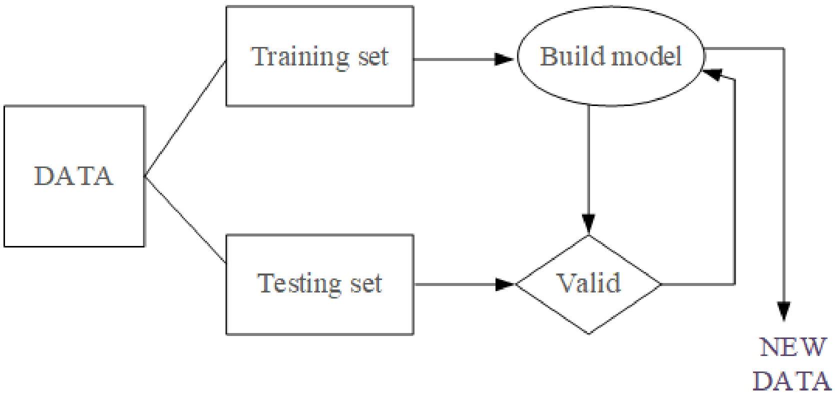


Fig. 8.4: Supervised learning schema [Deng, 2014].

Deep learning: Deep learning tries to learn on the same way as humans. It consists of several levels, and every level uses the information in all the previous levels to learn deeper. Deep learning is using neural networks as it is seen in Fig. 8.5. To understand deep learning, therefore, it is seen as important to give a brief explanation of what neural network is. A neural network works like any other network, its main function is to receive a set of inputs, perform complex calculations, and then use the output to solve a problem. Unlike many networks, a neural network is highly structured and comes in layers, an input layer, a hidden layer, and an output layer. In practice this means that a set of inputs are sent to the first hidden layer, and the activations from that layer are passed on to a new hidden layer and so on, until the output layer, where the results are determined [Deng, 2014].

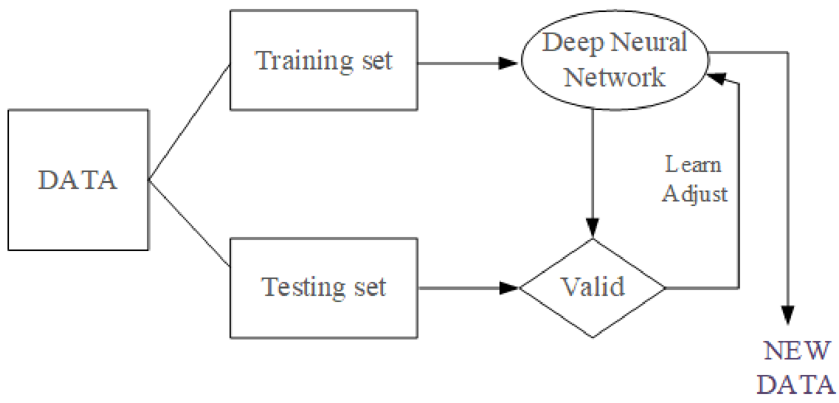


Fig. 8.5: Deep learning schema [Deng, 2014].

To gain a deeper knowledge on some of the Machine Learning techniques, these are briefly explained in the sections below.

8.3.1 Regression techniques

Regression is a supervised learning technique. It is a statistical tool used on a dependent target to find relationships with one or more independent variables. The regression techniques are good for indicating significant relationships between the target and variables, and indicating the strength of the variables on a target [Bilal et al., 2016]. It is used to determine if a hypothesis is true by seeing how one variable is affecting another, and measuring how close and predictable relationship the variables has [Marr, 2016].

8.3.2 Classification techniques

The goal of classification techniques is to predict a discrete value associated with some feature vector. In simpler words, it identifies new data and classifies it after something known. It is a supervised learning technique, which bases future decision on previous, correct decision. Since these techniques learn by examples, they are more fitted for single but more focused decisions [Bilal et al., 2016].

8.3.3 Information Retrieval

Information retrieval is a way to organise unstructured data, by finding material in the unstructured data that satisfy the information needed. It is retrieving relevant information from databases, usually textual [Bilal et al., 2016].

8.3.4 Natural Language Processing (NLP)

Natural language processing is combining computer science and linguistics. Its goal is to resemble the linguistic abilities of people, so that people can interact with computers without using an input device. NLP can be used for many purposes, for instance, summarise the most important information in a text, identify sentiment of a term or generate keyword topics from a document [Bilal et al., 2016; Olive et al., 2011]

8.3.5 Clustering techniques

Clustering is unsupervised technique which aims to partitioning data into sub-classes. Unlike the classification techniques where the labels are set for the algorithm to find similar, clustering will group object based on similarity. This means that the objects in one cluster are different from the objects in another clusters [Bilal et al., 2016].

8.4 Data Mining

Data mining uses techniques from statistics, ML, and pattern recognition to analyse the large amount of data. The aim is to discover relevant insights, patterns or relationships between variables to predict the future [Marr, 2016]. It is an automatic or semi-automatic analysis that extracts unknown patterns being used to improve results or performance [Bilal et al., 2016]. Some researches argue that it is the process of navigating through data, finding patterns before making all the relationships [Yu and Guo, 2016].

8.5 Using of data analytics within health and safety

In this section, examples from case studies that have used data analytics, are presented. This is done to show the opportunities within data analytics, and how it can be used to create value for health and safety managers.

Similar to the researchers mentioned in Chap. 4, the researchers presented in Tab. 8.1 believe, that specific factors exist that are contributing to occupational accidents. The different studies are emphasising different aspects of these factors which they believe have an impact on occupational injuries. To make an overview of the literature used in this section, and which aspects they are focusing on, Tab. 8.1 is presented.

Tab. 8.1: Literature overview.

Research	**Individual aspect**	**Task aspect**	**Management aspect**	**Environment aspect**
Cheng et al. [2010a]	x	x	x	x
Esmaeili and Hallowell [2011, 2012] and Esmaeili [2012]		x		x
Goh and Chua [2013]			x	
Desvignes [2014] and Prades Villanova [2014]		x		x
Tixier et al. [2016a]		x		x
Tixier et al. [2016b]		x		x
Tixier et al. [2017]		x		x
Poh et al. [2018]		x	x	x

8.5.1 Research background

As mentioned earlier in this report, the different data management techniques has not been widely used in the construction industry, and even less within the health and safety perspective. Researches are arguing that the health and safety performance has reached saturation before risk-based approaches started to emerge [Tixier et al., 2016b]. These approaches base their safety improvements on proactive decision making, which has shown positive results. However, these approaches are meeting some criticism because they are breaking down a project in smaller pieces for the sake of simplicity and to get a better overview [Tixier et al., 2017, 2016b]. The construction projects today are complex and dynamic, and by looking at the different parts in isolation, the interactions between factors that are affecting safety, might be lost. To perform a robust analysis that comprehensively captures safety, several numbers of units and interactions have to be taken into account [Poh et al., 2018; Tixier et al., 2017, 2016a,b]. Another limitation is the lack of empirical data in safety analyses, these are often based on subjective and expert opinions [Tixier et al., 2017, 2016a,b]. Therefore, several researchers have tried to address these limitations by using data analytics. Some of these studies are presented below.

8.5.2 Methods

The fact that health and safety managers look at the risks and tasks in isolation, and not at the dynamic nature of construction work, has made researchers wonder if there are better ways to look at it. Their hypothesis is that there might be patterns and trends in the data available, that can be used to prevent occupational accident [Poh et al., 2018; Tixier et al., 2017, 2016a,b]. Esmaeili and Hallowell [2011, 2012] and Esmaeili [2012] were the first who created a groundwork for a new conceptual framework. By using clustering techniques, they showed that injuries are the outcome of a series of joint events, also called fundamental attributes or injury processors. The approach allowed the user to extract structured information from unstructured injury reports. They were also the first who conducted an attribute level-risk analysis, where they examined 300 injury reports, and found 34 fundamental attributes [Tixier et al., 2016b]. Desvignes [2014] and Prades Villanova [2014] broadened Esmaeili and Hallowells list, by using the empirical data from 7.033 and 1.611 detailed injury reports. By quantifying the risk attributes, they found 80 fundamental attributes. The study classified the injury processors in three categories: upstream, transitional, and downstream. The upstream attributes can typically be prevented in the design phase, while transitional attributes base more on knowledge and can be prevented before the construction phase begins. Downstream attributes can be only detected in the construction phase, and are typically concerned with human behaviour. The studies were also able to extract various information from the report, like injury code, injury severity, body part affected, and energy source involved. With the ability to extract and code the information in the reports, data sets suited for data analytics can be made, which can lead to a deeper understanding of the underlying mechanisms in safety risks and injuries [Tixier et al., 2016a].

Cheng et al. [2010a] wanted to show the cause-and-effect relationship between occupational accidents by using data mining and statistical analysis. They meant that the excising statistical techniques were insufficient when it came to handling the complexity and large numbers of factors involved in accidents. By analysing 1.347 accidents in the Taiwan construction industry during the period 2000-2007, they found out that when certain combinations of factors are present, accidents tend to occur. The study found out that most of the accidents could be prevented because they steamed from a combination of: managers failing to protect workers from hazards in the working environment, mostly because they failed to have adequate safety measures. As well as the workers behaviour, and the unsafe acts they commit.

The limitations with the attribute-level safety analysis Esmaeili, Hallowell, Prades, and Desvignes contributed to, that it is extremely costly, time consuming, and requires allot of resources [Tixier et al., 2016a,b]. This is because they use manual content analysis. Another limitation is that with human coders it is impossible to ensure that they are entirely consistent in their choices [Tixier et al., 2016a]. The same limitations apply to Cheng et al. [2010a], where the analysis is not run fully unsupervised.

These researches have contributed much to health and safety understanding and management perspective. Several researchers have used their findings for future research to optimise limitations within the data analytics.

8.5.3 Standardisation of extracting value unsupervised

Based on the framework of Esmaeili and Hallowell [2011, 2012] and [Esmaeili, 2012], Tixier et al. [2016b] presented a natural language processing (NLP) system as an alternative to replace the manual content analysis of injury reports. The system showed promising results and a recall of 97% and 95% precision rate when scanning 101 unstructured injury reports. The system scanned for 80 validated attributes, 7 injury types, 5 body parts, and 9 energy sources, much similar with Desvignes [2014] and [Prades Villanova, 2014] findings, but with small changes.

To extend Tixier et al. [2016b] research, Tixier et al. [2017] applied state-of-the-art data mining techniques to identify safety incompatibilities among fundamental attributes. To extract attributes, the NPL system of Tixier et al. [2016b] was used on 5.298 raw accidents reports from 470 contractors, before applying clustering techniques and data mining. The study wanted to test how these techniques can be used to look at patterns, combinations, and relations in the extracted attributes, and if it can extract valuable new safety knowledge. Similar to Cheng et al. [2010a], the researchers believed that when some combinations of different fundamental attributes in the working environment are happening, the combination magnifies risk or new threats greater than one attribute in isolation. The researchers called these combinations *safety clashes*, which simply are situations where several attributions produce a greater risk than alone. Information like this is ideal for safety management on-site, but these binary input variables are also suited to be integrated in BIM. With the arising of new technologies within BIM, this information can be implemented and help to flag, address, and prevent safety-critical situations in design. The study was not aiming to address all safety clashes that would be needed to integrate in a BIM, but rather discover clashes that are not well known, and not easy for one person to identify alone based on experience. The findings in the report were organised in five themes, they are listed, and explained below:

- *Congested and confined workspace compound the risk of other attributes.* When the workspace were found congested and confined, a great variety of work situations turned hazardous. Because of the attributes great impact on the work around them, the study suggests that these always should be considered as main safety targets. The study also showed that by having a messy workspace the ergonomics of the workers got worse, since they had to adapt their body position to follow improper procedures. Conclusion were that the construction environment had to adopt the workers and not the other way and that safety and ergonomic shoul be to priority.
- *Flaggers are at greater risk for slips, trips, and fall due to their attention being caught by other stimuli.* This is one of the findings in the study that is previously undocumented as many of these accidents are considered as bad luck. However, when the researchers analysed the data these incidents accounts for hundreds of thousand lost work hours in a year. They are found to happen more frequently than assumed, and with many similarities. It is believed that these accidents are not random but are caused by the same underlying mechanism. This information can be used for further studies to find a way to correct these actions and improve health and safety.
- *Workers are unable to recognise immediate hazards due to poor visibility.* In this theme most of accidents happen because workers were unable to recognise hazard situations. The main finding was that the users of vehicle and equipment were unable to see workers on the ground, and not so much peoples skills in recognising hazardous situations due to lack of visibility. The study found another perspective that did not include equipment, but hidden dangers at the construction site. In this cases, a sign or a block of the area is highly recommended.
- *Existing equipment, vehicles, or work stations is safety critical.* Everyday activities may seem harmless due to more dangerous operations, but the study found out that also these attributes

make an important injury contribution. A shift in location can make workers situation awareness decrease. In other words, moving equipment from one location to another, lowers peoples awareness and increases the potential of injury.

- *Working with hazardous substance requires proper preparation and personal protection equipment (PPE), and following procedures.* The study found that working with hazardous substances requires a strict procedure and the workers full awareness and concentration. Even if the substance is seen as unharmful, other attributes like electrical tools, lack of PPE, etc., can effect the operation to be very harmful.

This study was focusing on hazardous situation but can also be used to visualise a better understanding of fundamental attributes in general. Limitations in this study are the number of reports analysed, and the methodology need to be tested further and on bigger data sets. However, the results in this study give a proof of the concept, and contribute to a standardised way of extracting safety knowledge by extracting raw data from injury reports. This will be a lift and help to address the limitations of the opinion-based safety analyses. Another finding in this study is, that the earlier beliefs that accidents are random by nature can be questioned. It is showing promising results for the theory that accidents are caused by disturbances in an underlying fundamental of attributes, and should be handled as empirical data. This statement requires further research, but the empirical evidences are promising.

8.5.4 Injury prediction

Tixier et al. [2016a] took the research presented above further, and tried to use the information extracted from the injury reports to create a model with high predictive skills. As mentioned earlier, the researchers' hypothesis assumes that accidents are non-random and can be studied and prevented. By using the attributes found by Esmaeili and Hallowell [2011, 2012]; Esmaeili [2012] and the framework by [Tixier et al., 2017], they tested different classification algorithms to find a way to predict safety outcomes.

The findings in Tixier et al. [2016a] shows that by applying these Machine Learning algorithms to the data set from the previous researches, the model could predict 3 out of 4 safety outcomes. The algorithm predicted the kind of injury type, which body part was injured, and energy type. The high predictive results enhances the hypothesis that accidents have underlying patterns and trends that can be extracted using statistical learning. It also strengths the theory that health and safety should be studied empirically and scientifically.

If these findings are corrected they can be a great complement to safety managers. When starting up an activity assessing the hazardous situation is important. By using this model, the algorithm can predict the likelihood of what type of body part, injury type etc. that is most likely to happen. This information can help the safety manager prevent this from happen by removing or replace attributes or plan the activity better. It can also be used in construction meetings where the algorithm can predict the likelihood of an incident. If it is higher risk for one type of accident the coming week, the safety manger can raise awareness and have a discussion with the workers on how this challenge can be prevented or what kind of procedures or equipment that need to be used. This model is also suitable for implementation with BIM. For example, it can assist in the design phase of a project so the risk in the best possible way can be avoided. The fact that health and safety in many ways is considered as a fragmented function that has little to do with time, cost, and planning, can hopefully be seen with new eyes, and give health and safety the ability to be a competing criteria to these measures.

8.5.5 Safety leading indicators

Similar to Cheng et al. [2010a], cause and effect research, Goh and Chua [2013] was one of the first to use neural networks on Occupational Safety and Health Management Systems (OSHMS). The study wanted to understand the relationship between OSHMS and safety performance. There research used audit and accident data from the construction industry in Singapore to predict accident occurrence and severity. The networked showed a predictive skill of 96 % for serious injuries, and 89 % for minor. It also identified three elements that were assumed to contribute to the accidents occurrence and severity, these were: *incidents investigation and analysing*, *emergency preparedness*, and *group meetings*. One limitation of this study was that some of the data were based on audit, which is not frequently available. The prediction model needs to be updated regularly to work perfectly. Another limitation was the lack of project- related attributes in the study, where these could address some of the accident occurrences and severities.

Poh et al. [2018] wanted to take Goh and Chua [2013] research further, and create a Machine Learning model that could predict an accidents occurrence and severity. The model was made to help construction companies to estimate safety risk with the help of validated safety leading indicators. According to Robson et al. [2017], leading indicators measure the activities, conditions and events on the construction site, like number of inspections, training effectiveness and more. These indicators are supported by lagging indicators which measures the construction site's health and safety outcomes, such as the injury rate or illness percentage. The leading indicators are seen as relevant since they can monitor and prevent the construction site's safety outcomes. Since lagging indicators are reactive and delayed, it is important that managers develop leading indicators that can help in preventing and assessing the risk on the site. Safety leading indicators should, therefore, be developed so the construction companies can prioritise there resources effectively.

The study conducted by Poh et al. [2018] points at several studies that indicates that the manger's safety perspective is important, if the managers lacks of safety knowledge or leadership, the safety culture and accident rate will be poor. The managers responsible for the portfolio in a construction company preventing risk by acting proactive is important, it is therefore, necessary that they understand the risk level in their projects. In many companies it is not the manager's lack of safety skills that is missing, but the limited resources available. Often the projects consist of a limited amount of assistant managers that do not have the time to look over the data. To solve this, the researchers suggest an optimisation of the use of human resources, where projects with higher risk, get more human resources. This can be done by establishing safety indicators as a tool for managers to use.

The study was conducted on an anonymous contractor in Singapore that provided data from 27 construction projects over 6 years. In addition, the research team got 785 records from safety inspections, and 418 accident cases and their attributes. By using Machine Learning approaches, the method selected 13 inputs variables out of 33 inputs variables. Out of the 13 selected, almost half were project related and 7 were safety-related attributes. The project related attributes were: project type, project ownership, contract sum, percent completed, magnitude of delay, and project manpower. The safety related attributes were: crane/lifting operations, scaffolding, mechanical-elevated working platform, falling hazards/openings, environmental management, good practices, and weight safety inspection score. These findings are interesting because they suggest that project related attributes have great influence on predicting the occurrence and severity of accidents.

9 | Applying data

The following chapter is elaborating the last step of the data management process which is formed by data visualisation, decision making and taking actions toward working environment improvement. This is done to apply analysed data to resolve particular health and safety issues or improve existing practices in the company. The whole process of data management contains of many steps and these are presented in Fig. 9.1. First, Sec. 9.1 gives a brief introduction to data applying, while Sec. 9.2 presents a case study.

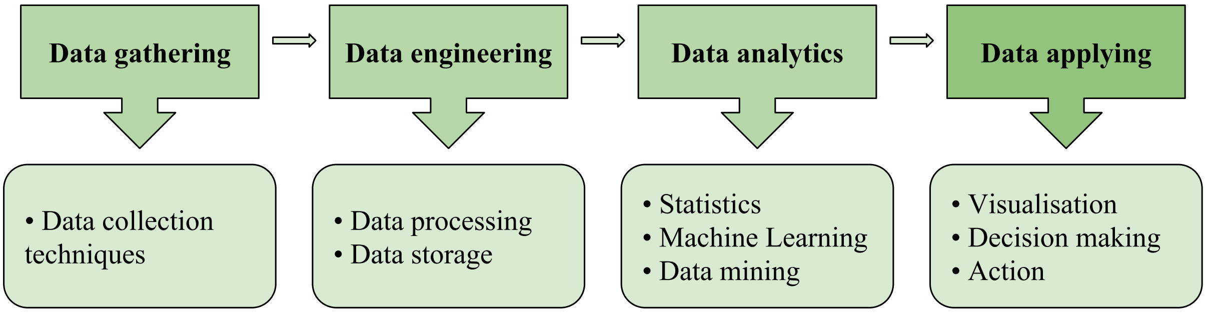


Fig. 9.1: Fourth step of the data management process.

9.1 Data in action

Due to fractured nature of the construction projects, the proper communication between different stakeholders is perceived as the project's success driver. In the classical understanding of the communication, the information during the construction projects is exchanged through paper documents, 2D drawings and verbal conversations [Cheng and Teizer, 2013]. The critical limitation of traditional communication is that the stakeholders are not always able to make correct and fast decisions due to insufficient real-time information [Cheng and Teizer, 2013; Hammad et al., 2009]. To improve the decision making at any management levels, the information regarding the project needs to be automatically updated and presented in the visual format [Cheng and Teizer, 2013].

Based on the current health and safety practices, the monitoring of the working environment, for instance, employees or materials, is conducted in a manual way [Cheng et al., 2011]. The safety on site is usually observed by the health and safety managers or supervisors which need sufficient knowledge within this field. Therefore, some risky situations might be perceived by them in a subjective way. Moreover, the information obtained from the manual observations is more difficult to share with the project team, especially in case of large projects. This fact is influencing proper decision making, as not all of the stakeholders are receiving the information regarding the working environment in the sufficient time [Cheng and Teizer, 2013]. Health and safety management during the construction projects might be positively influenced by the real-time resource tracking and visualisation. It is evident that the manual site monitoring would not be immediately replaced by automatic one, however, the implementation of these technologies might help the safety managers to improve their work and encourage proper decision making [Cheng and Teizer, 2013].

Deutsh [2015] argues that the first step in the process of applying data is to admit that the problem exists which needs to be solved. In terms of the working environment, Danish construction companies are facing the range of different tasks as risk assessment, health and safety planning, assessing specific production processes, preventing occupational accidents [Danish Construction Association, 2017]. Each of this task might bring some issues, or even do not work properly in the real world. Moreover, the commitment of top and middle-level management is an important aspect while starting the work with data. In some cases, if the company plans to completely transform their practices into data-driven ones, the additional resources, such as hardware, software or trained personnel, are needed. However, the companies might start this process also on their own, without the additional help of outside professionals. It just crucial in this case to understand, that the data is not something new in the industry, the information which might be used in everywhere around the company, in their activities, processes, and resources. The success of applying data is all about the merging the techniques provided by the data management with the capabilities of the organisation towards health and safety improvement [Deutsh, 2015].

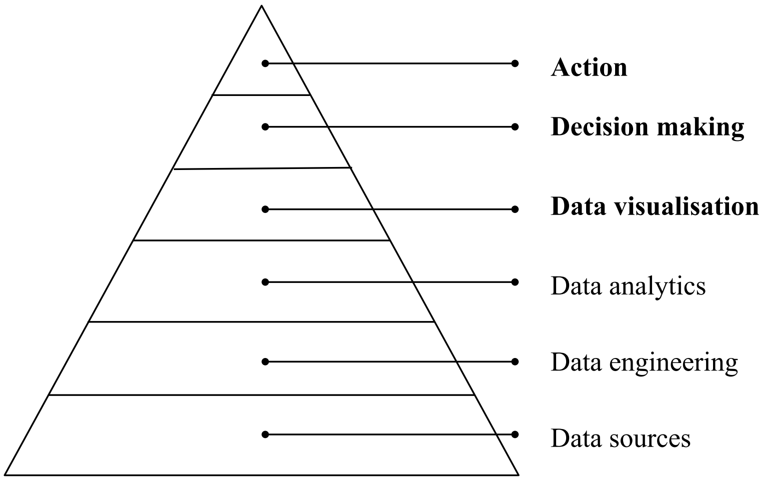


Fig. 9.2: Process of data application [Deutsh, 2015, *ed.* Fig. 6.1].

According to Deutsh [2015], the process of data application relies on the data collection and analysis in the great extent, see Fig. 9.2. The techniques presented in previous chapters creates the base for data application, and without them, the final actions would not bring as many benefits as the company would expect. The following sections are presenting three components of data application process which are contributing to the health and safety environment improvement. The following sections are addressing the last V, see Fig. 5.1, as by taking proper actions, the value might be captured.

9.2 Process of data application within health and safety

The health and safety environment planning is starting from the early stages of the project planning. In the construction industry, virtual reality (VR) is used to combine the design stage with the operational stage of the project. The main principle of VR is that virtual objects are combined with the real-world objects in the same environment [Cheng and Teizer, 2013]. VR has the broad range of applications in the construction industry. It might serve the educational purpose for the engineers and architects [Cheng and Teizer, 2013]. This tool enable to detect the construction issues in the very beginning of the project, moreover, helps contractors to manage the project is more effective and efficient way. The majority of VR applications are dealing with project design, cost, and schedule, however, there are not many research papers that are elaborating the connection of visualisation technologies with the real-time monitoring of resources on the task level [Cheng and Teizer, 2013]. The data collection and real-time visualisation still are not common in the construction industry [Cheng et al., 2011]. This is caused by fact that most of the VR models are using data which was obtained earlier or just simulated. Therefore, the changing character of the construction site could not be presented [Cheng and Teizer, 2013].

In the research of Cheng and Teizer [2013], real-time visualisation technologies were used to improve the monitoring of health and safety practices on site. The main objective of the research was to develop the technology which would enhance the project stakeholders to be more aware of the effective working environment planning. The technology presented by Cheng and Teizer [2013], is focusing on safety during both indoor and outdoor activities. The research is concentrating on four stages of real-time visualisation, which are: data collection, processing, visualisation, and decision making, see Fig. 9.3.

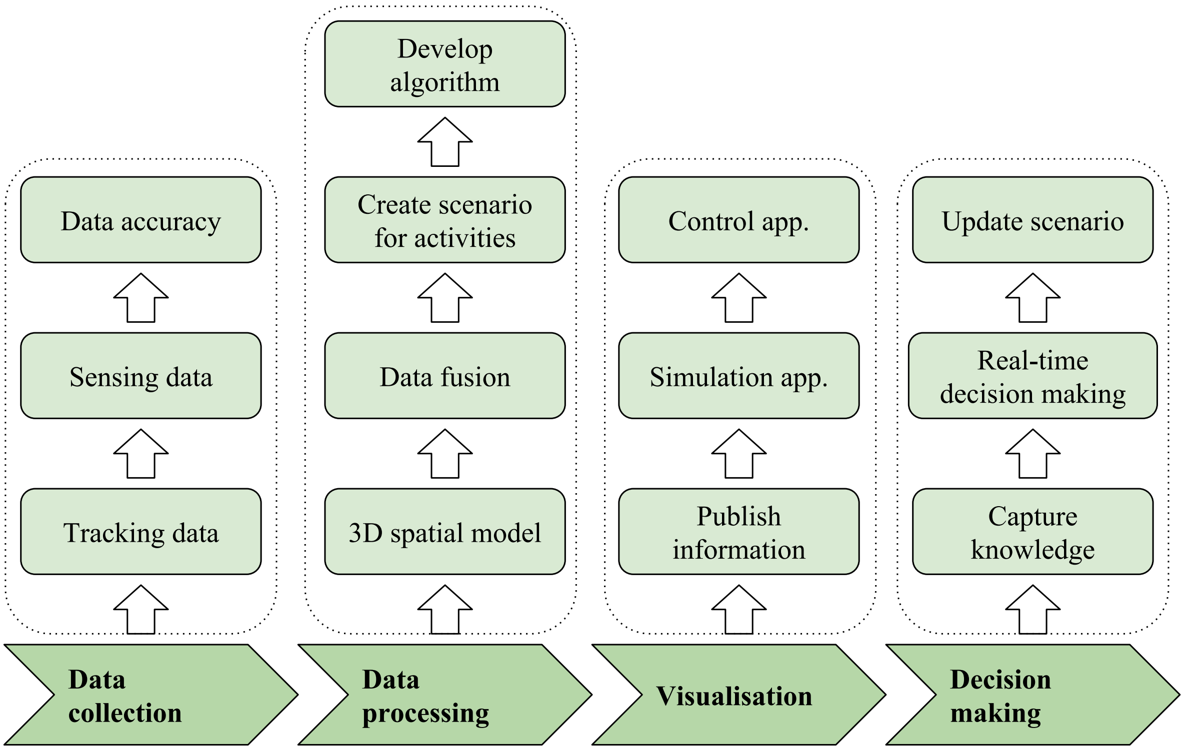


Fig. 9.3: Flowchart of real-time visualisation [Cheng and Teizer, 2013, *ed.* Fig. 1].

The research of Cheng and Teizer [2013] is limited in some extent, as it is analysing construction workers which are working on the ground. The work in height is not included in this case. The process of data visualisation is started from data collection in the field from dynamic resources, for instance, machinery or labour. Further, the information is filtered and processed to be able to stream it to the real-time VR. The assumption was made by Cheng and Teizer [2013] that the project stakeholders, which have the access to the real-time site visualisation, might help the decision making regarding safety on site, and to minimise the cost and time of this process.

The VR world representing the construction site was made with the help of laser scanning and available modeling tools. The VR was supplemented by the data obtained from different tracking and sensing techniques. The comprehensive explanation of data collection techniques is presented in Chap. 6. The information of tracing and sensing devices enable to determine the moving trajectory of particular resources. Further, the trajectory information needs to be processed, and in the last stage of the research, the algorithms are visualised in the VR environment.

In the case of health and safety improvement and accidents prevention, the real-time guidance and working environment assessment become a very important step. In the VR, the hazards might be distinguished, and if the dynamic resources are moving closer to the source of hazard, the alerts are visualised in the VR.

9.2.1 Data visualisation technology

The virtual representation of the reality in the research of Cheng and Teizer [2013] is called World Data Model (WDM). WDM incorporates the number of entities and properties which enable to present the working environment from the real perspective. To obtain real-world representation, the several of entities are used which are explained below.

- *Scene* is one of the commonly used tools for depicting the real-environment in virtual form, and it represents the interfaces and the modular components of this environment. The scene consists of surface, different types of objects, cameras or lights [Cheng and Teizer, 2013].
- *Surface and static objects*, in the research of Cheng and Teizer [2013], were represented by the laser scanning, which is one of the surveying technologies. The laser scanning is generating the range of points witch further are transformed into the triangular mesh. Each scan is creating its scan world, and it is described by the exclusive coordinate system. Therefore, several scan worlds might be combined and this system of scan worlds is known as project scan world.
- *Lightning* is a characteristic of the virtual environment, which in many cases it is represented by spot or conical lights. In some virtual environments, light might take directional form instead of static [Cheng and Teizer, 2013].
- *Objects* in the virtual environment are generated based on CAD geometry and level of details (LODs). The shapes might vary from the basic ones, as cubes or spheres, to the more complex ones, which has higher level of details. In the VR, the object might take static or dynamic form [Cheng and Teizer, 2013].
- *Viewing camera* defines the points from which the scene might be viewed. In the one scene, several virtual viewing cameras might be used to monitor the virtual environment. The viewing camera might be placed in the machinery, generated from the perspective of the construction worker, or from a bird's perspective [Cheng and Teizer, 2013].
- *Relations* are linking the particular objects within the scene. Relations are representing the interconnections which are distinguished based on real-world environment, for instance, distances between the objects [Cheng and Teizer, 2013].
- *Properties of dynamic objects* are obtained from real-time sensing devices. This data is describing the features of the objects. Further, the information regarding the location of particular resource might be captured by location tracking data [Cheng and Teizer, 2013].
- *Label* is the visualisation of the data processing results. In the research of Cheng and Teizer [2013], labels were used for depicting the distances between dynamic resources. During the process of decision making, the information in VR presented by different labels is more valuable than raw data.

The comprehensive explanation of the VRM with the main components is presented in Fig. 9.4. The World Data Model (WDM), presented by Cheng and Teizer [2013], combines the static and dynamic objects in the one virtual environment. This type of visualisation helps the stakeholders to understand the activities conducted during the construction project.

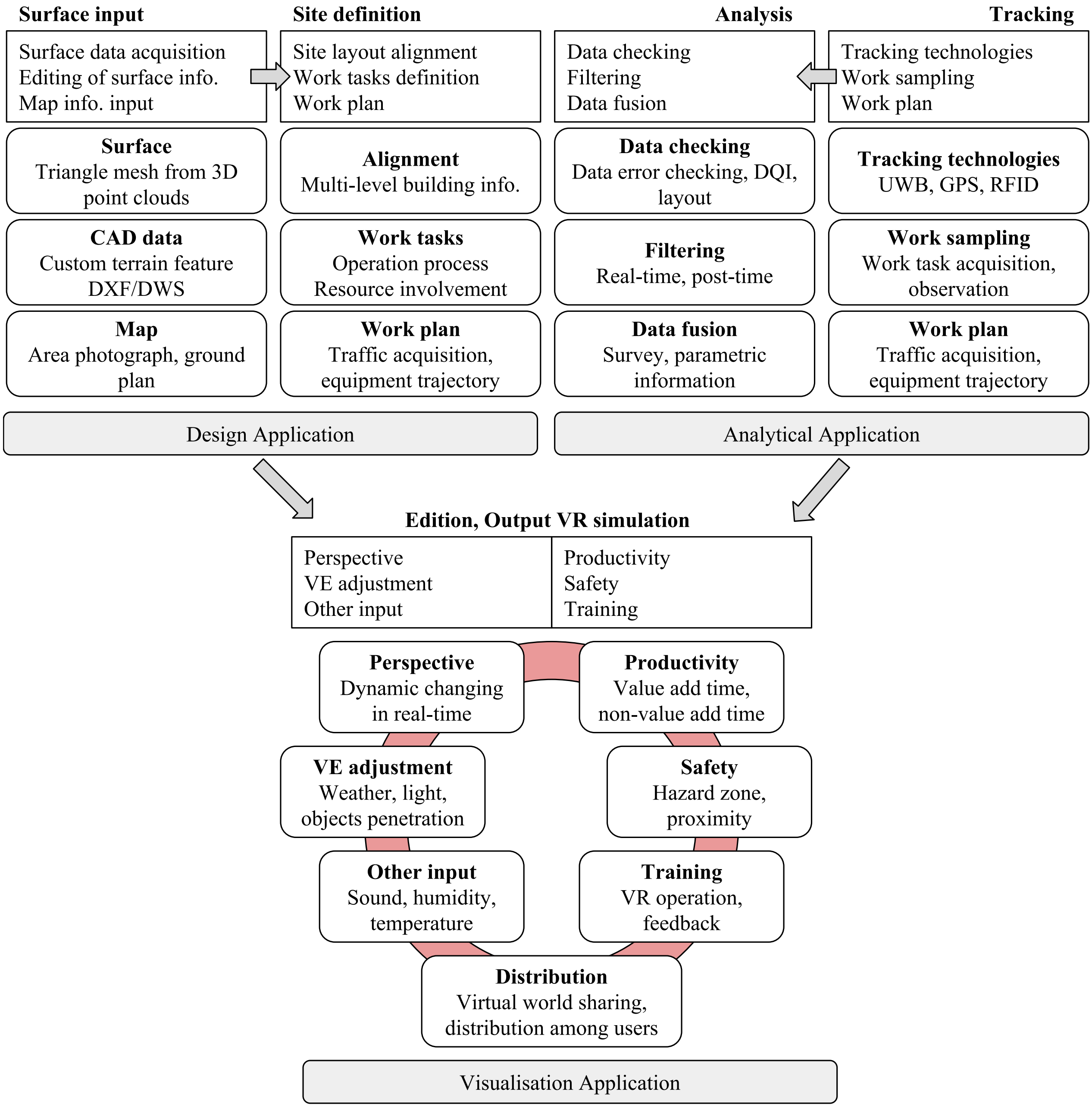


Fig. 9.4: Structure of real-time data tracking and data visualisation [Cheng and Teizer, 2013, *ed.* Fig. 3].

9.2.2 Decision making

The VR elements, presented above create the foundation for enhancing decision making. To enable that, the visualised information needs to be shared among the project team members by using a local server, but also by internet access. Therefore, the real-time information might be distributed also in remote areas [Cheng and Teizer, 2013].

The WDM enables the stakeholders to observe the real working environment, and increases their understanding regarding the activities conducted on site. The project team has a possibility to share the information faster and easier, thereby, the decision making is becoming more effective. Moreover, the WDM might be used for educational purpose, for training at the university level, but also for training the new employees. Such form of education increases the awarenesses of the site complexity, points the possible hazards and help to understand the construction operations [Cheng and Teizer, 2013].

9.2.3 Action

Cheng and Teizer [2013] conducted several cases to make sure that real-time data might be visualised and bring the benefits for the companies. The results of the cases show that the VR help to define risks regarding health and safety on site. The hazardous situations might be communicated through using the different colours indicating dangerous situation. Cheng and Teizer [2013] provided an example how colours are used in the VR environment. As it is presented in Fig. 9.5, the information regarding danger zones is visualised in the VR. The zones which are contributing to the risk are marked in green color, in this case, it is an area around the static loader. The distances between the particular objects and the danger zones are computed based on real-time data. Further, the distances are showed in the WDM labels. In the case, when the dynamic object is entering the potentially dangerous zone, it is visualised in the WDM by changed color of this zone from green to red. The information regarding possible risk is automatically visualised and proper actions might be taken by the project management group. This is an example, how data might be transformed in the real actions towards safety improvement. The entities regarding danger zones might be defined based on the health and safety regulations and rules. [Cheng and Teizer, 2013].

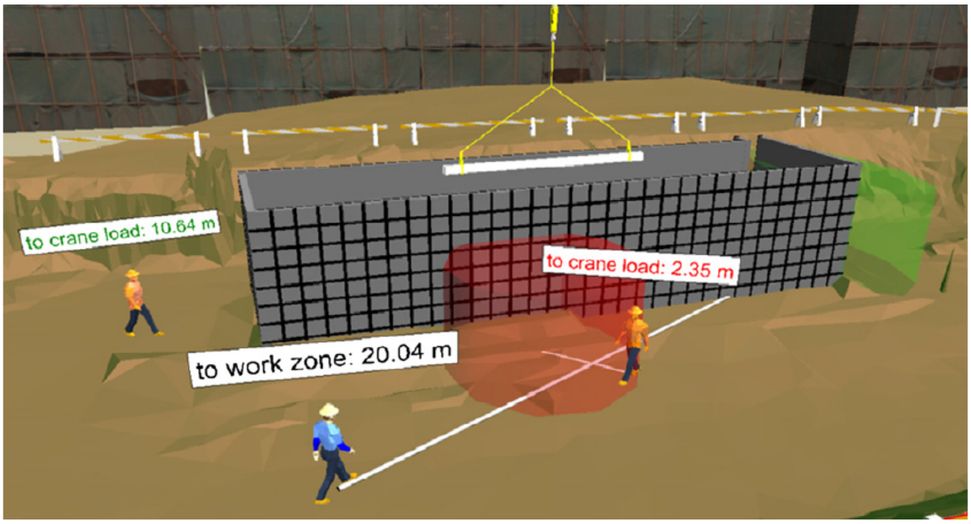


Fig. 9.5: Representation of the danger zones in WDM [Cheng and Teizer, 2013, Fig. 9].

Despite the control of the danger zones, the trajectory of the dynamic objects might be captured in WDM. To do that, sensors needs to be used, and should be tagged on dynamic and static objects to obtain tracking data [Cheng and Teizer, 2013]. More information regarding sensors and other location tracking techniques, is presented in Chap. 6. The visualised data of the changing locations of dynamic objects is supporting the decision making while planning the working environment. By analysing this data some patters might be distinguished regarding unsafe workers' behaviour and proper actions can be made towards working environment improvement.

Part IV

Strategy

Part IV of the report is presenting a guidance on how companies can leverage from the data management process. This is done to provide the answer to the following research question: *What are the initial steps towards obtaining data-driven health and safety management practices?* Chap. 10 provides a framework on how the companies can incorporate the data process described in the previous chapters.

10 | Data incorporation strategy

In this chapter, the aim is to show construction companies that data driven processes within their organisations are not complicated procedures, it just requires a framework that will guide them to find their exact needs. This is done to show small and big construction companies how they can use the information in this report, and utilise it in their company. Sec. 10.1 describes the steps of the framework.

10.1 The SMART model

A number of endless techniques, methods, and frameworks exist, for data acquisition, access, analytics and application. Some of these methods and techniques are presented in this report to give an introduction to the topic. For construction company it seems overwhelming to take in all of this new knowledge and it may lead to two reactions, either they start digging in too deep and get lost, or they run away and try to avoid it [Marr, 2015]. The problem with these two reactions is that by digging in to deep, the company may get to much information, and struggle with finding out what is relevant for them and use a lot of resources on things that are not necessary. On the other hand, by avoiding the data revolution, in the long term they might loose competitive advantage and miss opportunities.

To solve these challenges and provide a step back from the hype and noise around data, Marr [2015] has created a **SMART** framework. In many cases the companies are struggling with understanding that despite how much data is available, the value might be captured if the information is collected in the **SMART** way [Marr, 2015]. The framework provides a structured way on how construction companies by small steps might start process of data implementation. The SMART stands for: **S**trategy, **M**easure, **A**nalyse, **R**eport, and **T**ransform. A visualisation of the **SMART** model is presented in Fig.10.1.

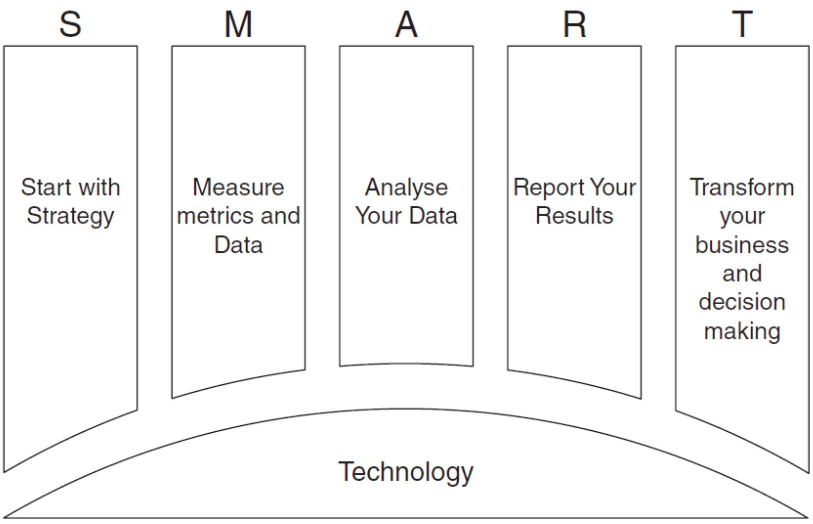


Fig. 10.1: The SMART model [Marr, 2015].

According to Deutsh [2015], to create a good strategy for data management, the organisations need to understand the essentials of the data, possible techniques and methods, and finally be convinced that data might provide a solutions for their problems. By starting with creating a strategy, the focus will immediately be leaded to the bigger picture and what the real goal with the data implementation is. It provides a guidance for the next steps, so it is easier for the company to see what is required and not to get overwhelmed by the possibilities. When the company knows what they are looking for, they can start looking on how to access that information and extract value from it.

10.1.1 Start with strategy

If the construction company wants to change their health and safety practices, a good way is to start by evaluating or changing the strategy. It does not matter if the company's size is small or large, in both cases the objectives, what are desired to achieve, need to be distinguished. The idea of managing all data within the organisation is challenging, and should not be the goal when starting implementing data techniques. The goal should be the opposite, to create a straightforward and clear database that contains only the data needed to achieve the objective. If the construction company wants to prevent falling accidents, it is not necessary to collect and analyse data about injuries where workers has been harmed do to cuts.

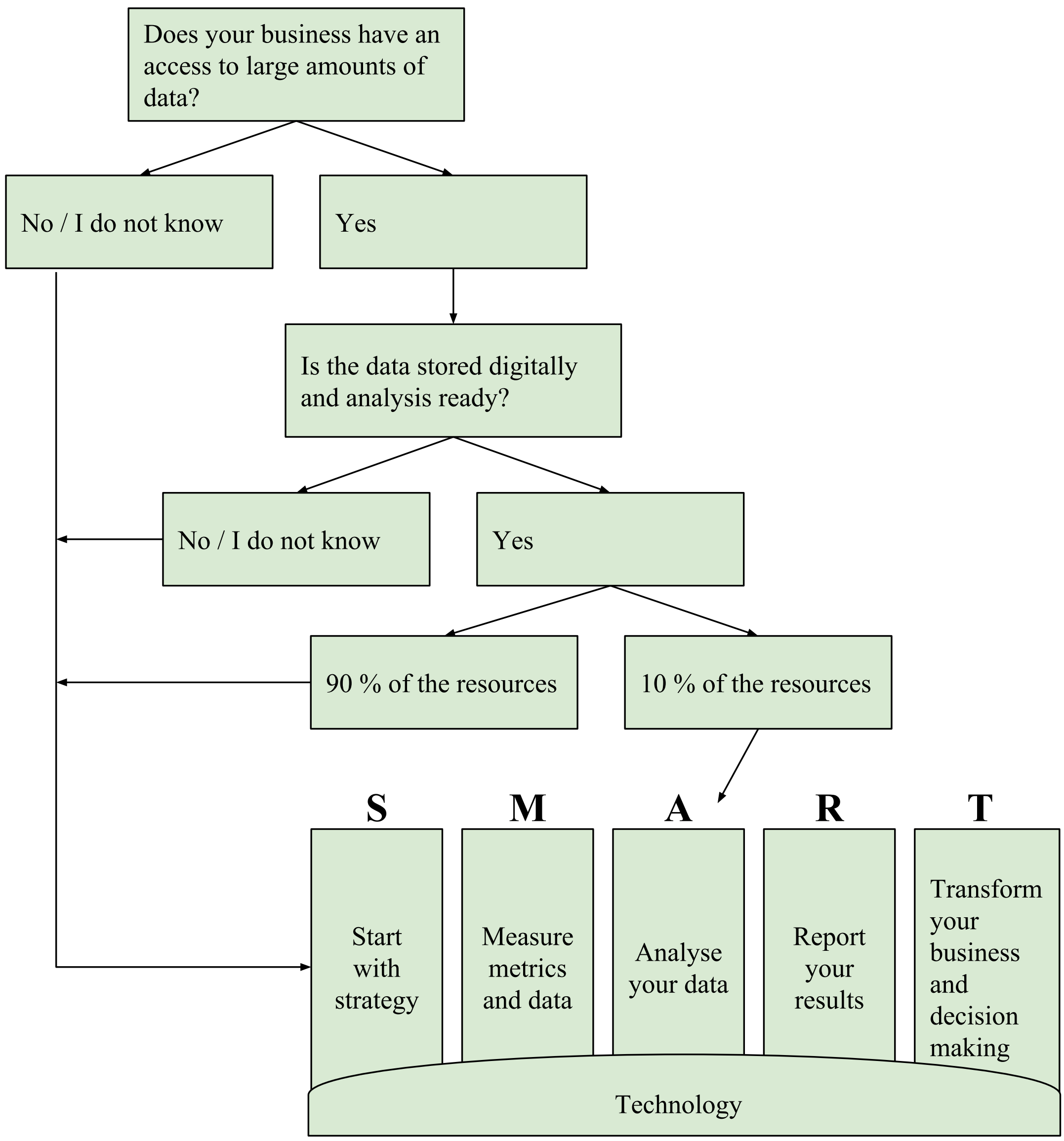


Fig. 10.2: The SMART framework of data management implementation [Marr, 2015, *ed.* Fig. 2.1].

Fig. 10.2 shows an evaluation form for companies that wants to start using data technologies. As it is seen in the figure, it does not matter if the company has an access to data or not. Strategy should in most of the cases be the first step. The exception is if the company is on more advanced level with data management, and has digitally stored larger amounts of data. Then the company might spend some resources on analysis. However, still a proper strategy is required. Marr [2015] suggest to spend 10 % of the total resources for the data discovery through data analytics, and the rest 90 % of resources for creating the strategy. The company should keep in mind that data has a shelf life [Marr, 2015]. It is maybe not the best idea to use huge amount of resources to extract value from injury reports that were made 15 years ago, since worker's behaviour and the construction site's health and safety policies have changed since.

Deutsh [2015] points several aspects which are important in the beginning of data incorporation process, and these are: capability of the organisation, the proper mindset, organisational culture, the appropriate people, support from the top management and the acceptance of change of the all employees in the organisation. While crating the strategy, it is beneficial to take a look on the practices from other companies, also from distinct countries witch have longer experience with data management. It is a good way to get an inspiration for strategy creation.

The **SMART** strategy board, inspired from Marr [2015], shown in Fig. 10.3, provides a structured way for identifying the companies' analytics strategy. The strategy board contains of five panels, each of them is designed to trigger four or five **SMART** questions. The board helps to realise and understand what is the wanted goal, and that the data itself does not provide the value. The value is within the actions which are taken by the company with the help of data. The five panels are elaborated in the sections bellow.

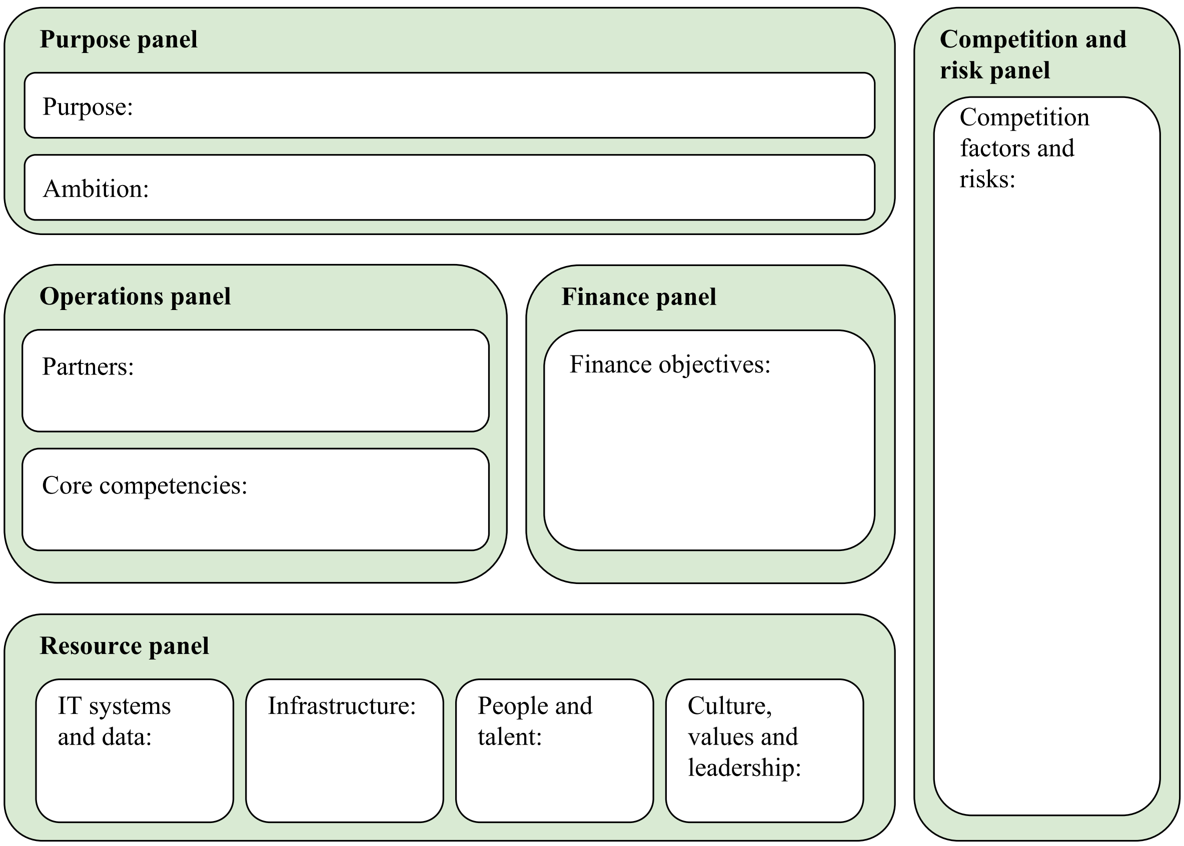


Fig. 10.3: Template for the SMART strategy board [Marr, 2015, *ed.*].

Purpose panel

The aim of the purpose panel is to create an overall context of the strategy or what the company wants to achieve. This panel is, unlike the other four, made for an overall context instead of developing SMART questions. This is done to show which direction the rest of the board is going. The best way to fill out this panel is by detailing the mission an vision for the construction company [Marr, 2015]. The purpose refers to the mission statement, while the ambition to the vision. The purpose in this case might be: *to move towards the working environment improvement*, wile the ambition: *to minimise the number of occupational injuries and diseases*.

Operations panel

The operations panel determines what is needed in terms of the internal strategy delivery. The following panel distinguishes the main partners of the organisation, and the core competences. The main partners in this case are subcontractors, the clients, and the hardware and software suppliers. Moreover, the company needs to consider their competences, and if there are some gaps which need to be filled [Marr, 2015]. The competences describes the internal processes which will contribute to strategy implementation. SMART questions here might be: *How can the company optimise the health and safety information to the subcontractors?* and *Which products are the safeties to use?*

Resource panel

The resource panel in the template, represents the main resources needed for creating a new practices with the usage of data. It contains of four components: *IT systems and data*, *infrastructure*, *people and talent*, and *culture*. IT systems are an important resources, as these enable effective data storage, processing or analysis. Moreover, the infrastructure describes what is needed of machinery, equipment, and property. People and talent concerns having the right people for the job, and if they are not within the company where they can be found or what kind of training can be provided. For the culture section is important to evaluate what kind of values are needed, and what kind of culture and leadership is needed for achieving the strategy [Marr, 2015]. SMART questions might be: *What is the perception of the management style? Where does the company find the most suitable health and safety manager? What is the most optimal usage of the construction site?*

Finance panel

The finance panel indicates the financial objectives of the strategy that enable to implement the stated purpose in the real life. It considers the fact that the strategy will bring the additional money, as well as how much the company needs to invest into it. The objectives might be described by revenue, profit, cost of additional software or hardware. In construction companies it can be looked at from a cost-benefit perspective. How much does it cost, but also how much does it help preventing occupational accidents [Marr, 2015]. SMART questions here can be: *What is the optimal investment strategy in terms of cost benefit? What can be changed today without effecting cost and safety? What is the optimal procedure to maximise profit and minimise risks?*

Competition and risk panel

Risk panel helps the companies to predict the possible risk sources and the competition from the industry to predict the success of the project. Some circumstances exist, which the companies do not have influence on, such as new law, however, the organisation might predict risk sources and adopt their practices to overcome there factors [Marr, 2015]. The risks might be analysed in both internal and external levels. Pssible SMART questions might be: *Which legislation can prevent the company from gathering data? What are the key financial risks?*

By answering the SMART questions the companies get an insight on what exactly they need and gain an understanding of the relevance of data. Few companies has the money, technology, capability or talent to analyse enormous amounts of data and just hope to find answers to questions they did not need to ask or care about [Marr, 2015]. Keep it simple and only identify small numbers of SMART questions for each panel. To create a better understanding engage key people from each of the panels. To avoid confusion, make the SMART questions clear and concise. Further, it is important when moving forward in the framework to always use the SMART questions as a guide, by doing this the company will get relevant and meaningful information from the analysis.

10.1.2 Measure metrics and data

After making the strategy, the company needs to measure the metrics and data. As elaborated in Chap. 5, many types and forms of data exist. These quantities of data are often described with the 5V's. This section explores four of these V's: *volume*, *velocity*, *variety*, and *veracity*. Every day enormous amounts of data get available through social media, sensors, ITO, etc. [Storey and Song, 2017]. For companies that want to leverage of this data it is important to be able to sort out and measure the data they need, or else they can be overwhelmed by the amount [Marr, 2015]. In this section a way to measure metrics and data is described, to easily get an overview of what kind of data is needed for what.

This framework for measuring is inspired by Marr [2015], and is made for companies to simplify how they look at data. It is easy to get overwhelmed by the enormous amounts and get lost when trying to use it. By using the knowledge gathered in the strategy section, this framework will provide a simple but effective way of choosing which data sets the company should pursue. The framework consist of five steps, where the SMART questions from the earlier section make the core. The questions will be a guidance for the companies so they can easily manoeuvre through the data, and find exact, and specific data they can use. By evaluating each SMART question individually, a better overview might be got. Going through each of the panels they can access different data sets that can help answering the questions. When seeing the questions separately it is easier to know exactly what to look for. A framework inspired by Marr [2015] is presented in Fig. 10.4

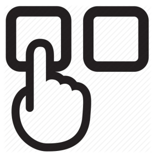
STEP 1  Identify needs	When the SMART questions are identified, the company needs to figure out what kind of metrics and data that is needed to answer the questions in order to accomplish the strategy.
STEP 2  Understand data	First step is understanding the various type of data and metrics that are available. When this is done the company can start consider which data could ideally help the company answer the SMART questions. It is also important to understand and look at how the data and metrics can be collected. How to access them is not important in this step, the focus is on what ideally could be accomplished.
STEP 3  Smart strategy	When the data and metrics are identified, run each of the key areas through the SMART strategy board. This process shows what type of data and metrics that are required to answer the company's resources, competition and risk questions. It is important to complete the data sheet for each of the areas identified.
STEP 4  Access the data	Based on the SMART board, the company now needs to look at how easy and cost effective the data and metrics can be accessed. It is recommended to start looking at the internal and structured data since these are usually cheaper and easier to access and analyse.
STEP 5  Choose the best option	When the data sheets is completed, it is it time to evaluate the data options. This is based on how quick, and cost effective the data can be collected, and how it will create value for the company. When this is assessed the company can pursue with the best option.

Fig. 10.4: Five steps on how to measure metrics and data.

10.1.3 Analyse data

This section provides the fifth V, the *value*. When the strategy is set and the SMART questions are answered, it is time to use the data gathered and turn it into meaningful information. This information can help the companies execute the strategy and improve their performance. This section gives some advice on what companies need to think about when applying data analytics.

Much like the measuring of metrics and data, the first step in analysis is to understand the best possibilities available for answering the SMART questions. In Chap. 8, some of the data analytics tools, that may be used for health and safety, are explained. These analytics tools can help a company to see patterns in data that can help them get an understanding of a specific topic or in decision making. It is important to look at the SMART questions and use them as a guidance when applying analytics. This is because it is easy to be seduced by fancy tools, rather than seeing what is needed to accomplish the strategy.

Another thing, is to be aware of when using analytics is ethical. There are endless ways of creating value from data, but it may not always be a good idea. As much as methods and techniques seem endless, it is also endless concern. Peoples private life is one of these. For example, tracking workers or similar can give access to a lot of sensitive data, this kind of data may be illegal to use. But on the other hand, some of the information extracted from personal data may have great value for a company. For example, some of the researches in this report implies that workers age or salary might contribute to occupational accidents. If factors like these can add value, for instance, in terms of preventing accidents to happen, this data may be a factor that can help the industry evolve. The company must, therefore, emphasise to be as transparent as possible about what data that is collected and how the data is used. This should not only be a matter of private policy but in general with data.

When using data analytics the outcome should always add value to the company and to the workers. Marr [2015] sees this as one of the most important points. The data should always benefit the parties involved, and it is important to be transparent on how the data is being used to add value. If the parties understand the value of methods and techniques used, it is more likely they will have a positive perspective of the process [Marr, 2015].

Privacy and transparency is not the only challenge when applying data analytics. Some researchers believes that the data techniques are leading to more causalities in favour of correlations [Marr, 2015]. For example, in lack of data, a company runs an algorithm based on the assumption that all of the workers behaviour are identical in order to predict hazardous situations. The algorithm may conclude that the work operations are too dangerous for a worker to conduct. However, in reality this worker has long experience within this field, and is trained especially for this operation. The problem here is that the company may take assumptions and decisions based on probability and not reality, which can lead to complications.

All of these dilemmas have to be taken into account when the company wants to extract value from data analytics. Therefore, it is important that the company answer the SMART questions and only apply analytics to those data sets. By doing this, they add value for themselves and the parties involved. If the parties see the value of the data collected, it is more likely they will share the information. Having a transparent process will help companies in the long run, and make it easier for the parties to look at it as an fair and worthwhile exchange [Marr, 2015]. It should also be though about if predictive analytics should be applied, and if it is used, it should be used appropriate. Do not make assumptions on automated untested data, to tell people what to do or not. Only use it to cluster people for further analysis and aim for a win win situation.

10.1.4 Report the results

After applying analytics to create value, the next step is to report or present the results in a way people will understand. As it is presented in the previous section, data analytics can contribute to a lot of findings, when it comes to understanding, monitoring and predicting situations. However, if the results are presented in a bad way, to the wrong person, the findings will loose its importance and in worst case, be lost. Therefore, this section gives some advice for companies when reporting results.

There are many ways to present the results from the data analytics and some of these are presented in Chap. 9. Even if many ways of presenting findings exist, it does not mean that all of them are suitable or that company should use all of them. It is important to think about how the results can be effectively communicated so it can create value and support decision making. Marr [2015] has presented a few guidelines on how successfully present the results.

Identify your target audience

The first thing that is needed to be taken into consideration, is who is going to receive the information, is it workers, managers or the client. Take into consideration their knowledge of the issue, what does the receiver need to know and how the party is going to use it Marr [2015]. For example, if the information is going to be presented on a health and safety meeting with the workers, or if the health and safety managers are using it in their statistics.

Customise the data visualisation

When the receiver is identified, the presentation should be customised to meet the specific requirements. Keep it simple to avoid confusion Marr [2015]. Often companies share the report with to many parties, just in case it is useful, or divide the presentation in many parts and send it to different people. By doing this, information can be lost or not reach the target group. Therefore, the presentation should be put into context and include only the necessary information for the intended audience. At a safety meeting, the workers do not need to know about which tools or methods that are used to calculate the result, but rather which hazardous challenges that are needed to be taken into account the coming week. On the other hand, the safety managers will need to see which attributes can cause the hazardous situation and how they are calculated by the algorithm.

Give the data visualisation a clear label or title

It is important to present the results as plainly as possible, not in a cryptic or clever way Marr [2015]. The main point is for the receiver to understand the visualisation, so it is important to explain graphs or numbers, and what they do. This helps to put the presentation into context.

Link the data visualisation to your strategy

If the results are answering one of the SMART questions, include the question in the presentation. By linking the results to the strategy, the receiver can shortly see the relevance and value of the data findings. That way, they are more likely to engage and use the information wisely Marr [2015]. For example, if the company has a zero accident policy, explain the result and how it can effect this goal.

Choose the graphics wisely

The graphics should be chosen so the presentations is as simple and short as possible. This means that using visuals just because it looks good is no go. Visuals that provides relevance for the receiver should be the only information included. This also implies that if there is empty space on a page, it should not be filled up, this creates confusion. Be careful with the use of colours, different colours have different meaning. It is also important to remember to not present many different types of graphs, charts or graphic Marr [2015]. For example, if a worker or a safety manager get presented with different attributes that creates a safe work environment, presenting these in the colour red is unfortunate. This is because red implies danger or warning.

Use headings to make the important points stand out

Time is valuable in the construction industry and by using headings to mark the important points, project managers can scan the document when they are in a hurry. That way they can get a quick overview of the context, before heading out, and later come back to read it.

Add a short narrative where appropriate

Graphs and charts often just give a snapshot, and it, therefore, may be needed a narrative to explain the findings in depth. It also helps the receiver understand the key points of the result Marr [2015].

10.1.5 Transform the business and decision making

As much of the research in this report suggests, the accessibility of data and the development of data techniques are transforming the world Marr [2015]. However, it is the companies' choice how much they want to use these techniques. This section provides some advice when transforming the company.

First off all, when the company has decided to use some data techniques, it is important to see which opportunities these techniques can give. Both, for what the objective for the implementation is, and within other areas Marr [2015]. When the companies have used the SMART framework they will gain insight on how they can improve their health and safety operations. But it is most likely to reap also huge benefits in other areas of the company.

When the health and safety strategy is implemented, the company has to see this as an ongoing commitment and not a one time process Marr [2015]. The access to data never stops, and can provide information that will make it easier and more effective to reach the health and safety objectives. Further, it is important to use the findings gathered. By doing this the company is a roll model for they employees and encourage them to change their way of thinking when making decisions.

Additionally, the company should always look for new opportunities that may emerge from the data Marr [2015]. This can be while collecting data for the health and safety objectives. Suddenly, the company can have a lot of valuable data that can improve the decision making or understanding within another area in the company.

Part V
Ending

The last part of the report is introducing the findings of the research. This is done to answer the problem statement and provide a direction for future research. First, Sec. 11.1 presents and discusses the findings and their limitations. Further, Sec. 11.2 demonstrates the conclusions and provides the answer for the problem statement. Moreover, the future perspectives were distinguished in Sec. 11.3 which presents the other possible directions of the research continuation.

11 | Final findings

The chapter aims at introducing the findings and the results of the conducted research. This is done to validate of the research outcome brought the answer for the research question. First, Sec. 11.1 presents the discussion of the results. Further, Sec. 11.2 demonstrates the overall conclusions of the research by answering the problem statement. Last section introduces the possible directions of future research since during the report work, some possible paths of further investigation were found, see Sec. 11.3.

11.1 Discussion

When looking at the findings in this report, there are a lot of opportunities within data acquisition, access, analytics and application. It is evident that the health and safety environment within the construction industry has a big potential when it comes to the usage of data techniques to gain better understanding of the processes and while decision making. The data available today creates a great foundation for extracting data knowledge, especially from injury reports and pictures. However, there are many techniques which are not so common, and that have the potential of adding value for the companies and workers. Many of the state-of-the-art techniques, elaborated in this report, consider tracking of workers on the construction site. Today, smart technologies like Internet of Things, sensors, and robots make it possible to access much more information than earlier. Most likely, this information can revolutionise the industry in terms of health and safety improvement.

The working environment in the Danish construction industry was analysed to gain deeper knowledge abut the overall situation but also to distinguish the main factors affecting occupational accidents. Due to limited number of English literature concerning the Danish working environment and the issues related to this topic, some assumptions were made by the authors of the report. It was decided to introduce the Danish construction industry, however, underpin this information by the research articles from other countries. The assumption was made that some universal aspects contributing to the occupational accidents exist, and these are also relevant for Denmark. The issues within health and safety in the construction industry are apparent in global perspective as different developed countries are facing similar problems. Moreover, due to the fact, that Danish construction industry is becoming more international and multicultural, it is not possible to distance themselves from the heath and safety practices from the other countries or cultures. Despite the extensive range of distinct literature regarding this topic, the analysis was conducted in a way to avoid the facts which are irrelevant for Denmark. By elaborating the topic from different perspectives, it was tried to obtain the insightful analysis of the root causes of the accidents.

The results of of data acquisition analysis show that gathered data might be used in many different ways, and that similar information might be collected by using diverse data gathering techniques. This report presents several techniques which are present in the global construction industry as an example for the Danish companies. However, there are some pitfalls and concerns that the companies need to be aware of. Especially this concerns the private policies of the workers and legislation. By utilising these techniques in the company, it is important to remember to stick within the law, and use the data only for the intended purpose within the organisation.

During the past decade, in the field of data engineering, many concerns of processing and storing data has emerged. Especially, when it comes to the size and cost of storing data. The study in this report presented Hadoop as an alternative for processing and storing data. However, there are many other options, for instance Spark which is gaining larger popularity, due to its real-time and parallel processing capabilities. The evolution of the in-memory distributed file system, Tachyon can also be an alternative, especially since the platform is compatible with Spark, HDFS, and MapReduce without changing any codes. The open source, distributed, and non-relational database NoSQL is suited to use in data analytics, the only drawback is the lack of standard query language and discipline for modelling. Cloud storage is a easy and good alternative for companies with little resources and knowledge within data techniques. However, the issues regarding the data protection need to be addressed in the future. These are only few platforms available today, and it is the users needs that decides which platform to use. Little research form the construction industry exists where these techniques are used in the real world. This might be because the construction industry has been slow in adopting data techniques, therefore, the need for storing big amounts of data has not been present. However, it is important to keep in mind the safety of data, so it do not leek and spread sensitive information.

Data analytics is an area with a rapid development. The construction industry has utilised little of these techniques in their work. Even if they have started using some of these, the health and safety environment has been a bit neglected. This may be due to the believe that accidents happen randomly, and that experience and expert opinions are the key drivers to prevent them. However, some researchers have made several remarkable findings by applying data analytics. Their hypothesis is that occupational accidents may not be random but an event of underlying mechanisms. They want to prove that accidents are caused by a combination of specific attributes and can be prevented by identifying patterns and trends in the data. Tixier et al. [2017] tried to create a standardised way of extracting value form injury reports unsupervised. Their research made promising results, and indicates that accident is not random and should be handled as empirical data. Tixier et al. [2016a] took this research further, and used the information extracted from the injury reports to predict accidents. Their findings were remarkable, the algorithm they created was able to predict 3 out of 4 safety outcomes. Poh et al. [2018] wanted to create Machine Learning approach so they could develop leading indicators that could classify construction sites according to risk. Unlike Tixier et al. [2016a], these researches included not only the attributes but also project management factors in their prediction. From 13 input variables, almost half of them turned out to be project related, which indicates that project management factors have an influence on accidents. By including these, the researchers found out that most likely severity of accidents can be predicted, this contradicts Tixier et al. [2016a] which believed that this was random.

The studies shows a promising future for predicting accidents and extracting value from injury reports. This can be a great supplementary to safety managers within decision making and when understanding the work environment. However, these studies are conducted on limited amount of injury reports and need to be tested further before it can become a standardised practise. Another perspective is that these researchers want to make the methods integrated in BIM, as designers and consultants can plan the building and get notifications when they are planning. In this way, many risks can be prevented before the execution phase, it will also make it easier to manage risk and plan what to do in the next week. In the future, a drone can fly over the construction site and scan the area, the data will update the BIM model, and the project managers will always know where they are compared to planning. Concerning the safety managers, they will get a notification whenever the model spots and predicts new hazardous situations, so these situations can be prevented.

The results of data visualisation show that the real-time status of the construction site might be obtained through World Data Model. This model is combining the VR with the real-time data, which might be used for health and safety monitoring by the construction project team. The analysis shows that this type site representation is encouraging the decision making regarding additional safety precautions and overall working environment planning. However, the study of Cheng and Teizer [2013] does not provide the long term results of this model effectiveness. Moreover, it might be considered to combine real-time visualisation with the active real-time warning techniques, as it might bring larger benefits and some accidents could be avoided by real-time actions.

Much of the literature and studies in this report were not conducted in Denmark. This can be problematic due to many factors, like culture, economy etc.. However, the authors of this report do not see it as very problematic. A lot of studies, used in this report, are not about a specific country, but rather about creating data techniques that can extract and predict value adding knowledge from the input data. This input data will be different from country to country, but in the end it is the user of the techniques that decides which data that will be provided. It is seen as important to look outside Denmark to be inspired, learn, and be updated on the technology that is available. All of the researchers have the same goal, and that is to minimise the number of accidents as much as possible and make the construction industry a safer place for people to go to work. There is always room for improvement.

11.2 Conclusions

The report provided a guidance through the data management process which was conducted to answer the following research question:

What kind of data could be collected, and which tools within data analysis can be used to improve health and safety environment during the execution phase?

As it is presented in Chap. 6, many different data types and data collection techniques exist in the construction industry. First, data might be roughly divided in structured and unstructured data sets, which might take a form of picture, drawing, report, graph, video, voice record etc. The same situation is with the data collection tools, many different techniques are available. Data might be gathered either inside or outside of the organisation, and it is usually apparent that the company's private data sets would be much larger than the public ones. In this report, many different data gathering techniques are presented, however, there is no clear way on how this process needs to be conducted, as the factors within the particular companies and projects, needs to be addressed. The companies need to distinguish by their own, on what kind of data that is the most beneficial and in their case would bring the largest value. It was determined that data regarding the individual and task aspects of the workers are influencing the accidents occurrence directly, therefore, it is more beneficial for the small or medium size companies to focus more on these aspects. Moreover, some researchers suggest that the management aspects are influencing the working environment in the great extent, it is ,therefore, beneficial to obtain as much information regarding the nature of project as the resources allow. Even if it is possible to collect almost all kinds of data, some data can not be collected by using the techniques presented in this report. A good example are emotions of the workers, which may influence their behaviour and reactions.

Further, many different tools within data analysis exist. Some of these tools are presented in this report, and cover the data storing, analytics, and applying, see Chap. 7, Chap. 8 and Chap. 9. Through the report it became clear that many tools that can be applied exist, and that some of these are more fitting than others. It was assumed that maybe it would be possible to find a standardised way on how to extract value from the data collected within the health and safety area. This assumption was wrong. What turned out to be the case was that the amount of tools and techniques can be overwhelming for companies that want to utilise the advantages of data management. Therefore, a framework is suggested that can make it easier for companies, no matter the size, to figure out what exactly is needed for them to reach their objectives. The aim of the data management process is not to rapidly change all existing health and safety practices, rather is it showing the possible ways on how the common practices might be improved and provide value for the company as well as support the safety managers in decision making. Data should not be perceived as an additional thing that the company need to deal with but it should be merged into the daily-operations.

11.3 Future perspectives

The research introduced the potential applications of the data management for the health an safety environment improvement, which gives an foundation for further research in this area. In the early stage of research process, the project group decided to conduct the report based on available desk information. With the passing of time, the benefits of having cooperation with industry professionals was identified. It was perceived that particularly the implementation part of the report would gain more value by using the real world example. Therefore, one of the possible continuation paths of this topic is implementation of developed data management techniques on the real case. The research would distinguish what are the potential struggles of the Danish companies in data incorporation process. Moreover, the report analyses the majority of practices from the global perspective, however, it is not sure that these techniques might be implemented in Danish companies with such positive results. The further research might take this issue further.

Due to the one of constraints which is insufficient knowledge of the authors in software development, the next research could focus on developing new tools that matches companies' needs. The whole process from data gathering until analysis are described in the following report, however, the companies need to invest their resources in additional software which is not particularly created for health and safety environment. The other obstacle is that safety managers are not ready for the radical chance of their practices, and the data incorporation process is closely related to the personal aspects. The possible data implementation strategies and techniques might be elaborated particularity for the people which are responsible for the safety on site during the construction projects.

Another area that has more potential is the utilising of data from the individual aspects. The analysis in Chap. 4 showed that the individual aspects have a large impact on occupational accidents. Therefore, it should be included when assessing attributes that contributes to risks. The aspect is maybe the hardest to get access to, due to legislation and private policies, and this may be the reason for the lack of studies which include it. As seen in this report, few of the studies are including these aspects when they are trying to extract information and predict occupational accidents. Therefore, a further study would be to include these factors, and see what kind of impact they have.

Bibliography

Ackroyd and Thompson, **1999**. Stephen Ackroyd and Paul Thompson. *Organizational Misbehaviour*. Sage Publications, 1999. ISBN 0-8039-8736-6.

Adam et al., **2009**. Jose M. Adam, Francisco J. Pallarés and Pedro A. Calderón. *Falls from height during the floor slab formwork of buildings: Current situation in Spain*. Journal of Safety Research, 40(4), 293–299, 2009. ISSN 00224375. doi: 10.1016/j.jsr.2009.07.003.

Ajslev et al., **2015**. Jeppe Zielinski Nguyen Ajslev, Roger Persson and Lars Louis Andersen. *Associations between Wage System and Risk Factors for Musculoskeletal Disorders among Construction Workers*. Pain Research and Treatment, 2015. ISSN 20901550. doi: 10.1155/2015/513903.

Ajslev et al., **2013**. Jeppe Z.N. Ajslev, Henrik L. Lund, Jeppe L. Møller, Roger Persson and Lars L. Andersen. *Habituating pain: Questioning pain and physical strain as inextricable conditions in the construction industry*. Nordic Journal of Working Life Studies, 3(3), 195–218, 2013. ISSN 2245-0157. doi: 10.19154/njwls.v3i3.3018.

Alarcón et al., **2016**. Luis Fernando Alarcón, Diego Acuña, Sven Diethelm and Eugenio Pellicer. *Strategies for improving safety performance in construction firms*. Accident Analysis and Prevention, 94, 107–118, 2016. ISSN 00014575. doi: 10.1016/j.aap.2016.05.021.

Ale et al., **2008**. B. J M Ale, L. J. Bellamy, H. Baksteen, M. Damen, L. H J Goossens, A. R. Hale, M. Mud, J. Oh, I. A. Papazoglou and J. Y. Whiston. *Accidents in the construction industry in the Netherlands: An analysis of accident reports using Storybuilder*. Reliability Engineering and System Safety, 93(10), 1523–1533, 2008. ISSN 09518320. doi: 10.1016/j.ress.2007.09.004.

Amaratunga et al., **2002**. Dilanthi Amaratunga, David Baldry, Marjan Sarshar and Rita Newton. *Quantitative and qualitative research in the built environment: application of "mixed" research approach*. Work Study, 51(1), 17–31, 2002. ISSN 0043-8022. doi: 10.1108/00438020210415488.

Andersen et al., **2007**. Johan H. Andersen, Jens P. Haahr and Poul Frost. *Risk factors for more severe regional musculoskeletal symptoms: A two-year prospective study of a general working population*. Arthritis and Rheumatism, 56(4), 1355–1364, 2007. ISSN 00043591. doi: 10.1002/art.22513.

Andersen et al., **2011**. Lars Louis Andersen, Ole Steen Mortensen, Jørgen Vinsløv Hansen and Hermann Burr. *A prospective cohort study on severe pain as a risk factor for long-term sickness absence in blue- and white-collar workers*. Occupational and Environmental Medicine, 68(8), 590–592, 2011. ISSN 13510711. doi: 10.1136/oem.2010.056259.

Arbejdstilsynet, **2017**. Arbejdstilsynet. *Interaktive tabeller over arbejdsulykker*, 2017. URL `https://arbejdstilsynet.dk/da/statistik/arbejdsskader/interaktive-arbejdsskadetal/arbejdsulykker`.

Atkinson, **1999**. Roger Atkinson. *Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept the other success criteria*. International Journal of Project Management, 17(6), 337–342, 1999. ISSN 0263-7863.

Belassi and Tukel, **1996**. Walid Belassi and Oya Icmeli Tukel. *A new framework for determining critical success/failure factors in projects*. International Journal of Project Management, 14(3), 141–151, 1996. ISSN 02637863. doi: 10.1016/0263-7863(95)00064-X.

Bellamy et al., 2008. Linda J. Bellamy, Tim A.W. Geyer and John Wilkinson. *Development of a functional model which integrates human factors, safety management systems and wider organisational issues*. Safety Science, 46(3), 461–492, 2008. ISSN 09257535. doi: 10.1016/j.ssci.2006.08.019.

Bilal et al., 2016. Muhammad Bilal, Lukumon O. Oyedele, Junaid Qadir, Kamran Munir, Saheed O. Ajayi, Olugbenga O. Akinade, Hakeem A. Owolabi, Hafiz A. Alaka and Maruf Pasha. *Big Data in the construction industry: A review of present status, opportunities, and future trends*. Advanced Engineering Informatics, 30(3), 500–521, 2016. ISSN 14740346. doi: 10.1016/j.aei.2016.07.001.

Borys, 2012. David Borys. *The role of safe work method statements in the Australian construction industry*. Safety Science, 50(2), 210–220, 2012. ISSN 09257535. doi: 10.1016/j.ssci.2011.08.010.

Bremond et al., 2006. Francios Bremond, Monique Thonnat and Marcos Zuinga. *Video Understanding Framework For Automatic Behavior Recognition*. Behavior Research Methods, Psychonomic Society, 3(38), 416 –426, 2006.

Brito et al., 2016. Jaqueline Joice Brito, Thiago Mosqueiro, Ricardo Rodrigues Ciferri and Cristina Dutra De Aguiar Ciferri. *Faster cloud Star Joins with reduced disk spill and network communication*. Procedia Computer Science, 80, 74–85, 2016. ISSN 18770509. doi: 10.1016/j.procs.2016.05.299.

Bryman, 2016. Alan Bryman. *Social research methods*. Oxford University Press, 5th edition, 2016. ISBN 978-0-19-968945-3.

Bryman and Bell, 2011. Alan Bryman and Emma Bell. *Business research methods*. 3rd edition, 2011. ISBN 9788578110796. doi: 10.1017/CBO9781107415324.004.

Buckley et al., 1976. J. W. Buckley, M. H. Buckley and H. F. Ching. *Research Methodology and Business Decisions*. National Association of Accountants and the Society of Industrial Accountants of Canada, 1976.

Cameron and Duff, 2007. Iain Cameron and Roy Duff. *A critical review of safety initiatives using goal setting and feedback*. Construction Management and Economics, 25(5), 495–508, 2007. ISSN 0144-6193. doi: 10.1080/01446190701275173.

Camino López et al., 2008. Miguel A. Camino López, Dale O. Ritzel, Ignacio Fontaneda and Oscar J. González Alcantara. *Construction industry accidents in Spain*. Journal of Safety Research, 39(5), 497–507, 2008. ISSN 00224375. doi: 10.1016/j.jsr.2008.07.006.

Camino López et al., 2011. Miguel A. Camino López, Dale O. Ritzel, Ignacio Fontaneda González and Oscar J. González Alcántara. *Occupational accidents with ladders in Spain: Risk factors*. Journal of Safety Research, 42(5), 391–398, 2011. ISSN 00224375. doi: 10.1016/j.jsr.2011.08.003.

Cavanillas et al., 2016. José María Cavanillas, Edward Curry and Wolfgang Wahlster. *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*. New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe, pages 1–303, 2016. doi: 10.1007/978-3-319-21569-3.

Chan and Chan, 2004. Albert P. C. Chan and Ada P. L. Chan. *Key performance indicators for measuring construction success*. Benchmarking: An International Journal, 11(2), 203–221, 2004. ISSN 1463-5771. doi: 10.1108/14635770410532624.

Cheng et al., 2010a. Ching W. Cheng, Chen Chung Lin and Sou S. Leu. *Use of association rules to explore cause-effect relationships in occupational accidents in the Taiwan construction industry*. Safety Science, 48(4), 436–444, 2010. ISSN 09257535. doi: 10.1016/j.ssci.2009.12.005.

Cheng et al., 2010b. Ching Wu Cheng, Sou Sen Leu, Chen Chung Lin and Chihhao Fan. *Characteristic analysis of occupational accidents at small construction enterprises*. Safety Science, 48(6), 698–707, 2010. ISSN 09257535. doi: 10.1016/j.ssci.2010.02.001.

Cheng et al., 2011. T. Cheng, M. Venugopal, J. Teizer and P. A. Vela. *Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments*. Automation in Construction, 20, 1173–1184, 2011. ISSN 09265805. doi: 10.1016/j.autcon.2011.05.001.

Cheng and Teizer, 2013. Tao Cheng and Jochen Teizer. *Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications*. Automation in Construction, 34, 3–15, 2013. ISSN 09265805. doi: 10.1016/j.autcon.2012.10.017.

Chi et al., 2005. Chia Fen Chi, Tin Chang Chang and Hsin I. Ting. *Accident patterns and prevention measures for fatal occupational falls in the construction industry*. Applied Ergonomics, 36(4 SPEC. ISS.), 391–400, 2005. ISSN 00036870. doi: 10.1016/j.apergo.2004.09.011.

Choudhry and Fang, 2008. Rafiq M. Choudhry and Dongping Fang. *Why operatives engage in unsafe work behavior: Investigating factors on construction sites*. Safety Science, 46(4), 566–584, 2008. ISSN 09257535. doi: 10.1016/j.ssci.2007.06.027.

Christiansen and Pedersen, 2007. Rune Holm Christiansen and Henrik Stener Pedersen. *Trends for risk factors in the Danish work environment*, European Foundation for the Improvement of Living and Working Conditions, 2007.

Civil Aviation Administration - Denmark, 2004. Civil Aviation Administration - Denmark. *Regulations for Civil Aviation - BL 9-4 Regulations on unmanned aircraft not weighing more than 25 kg*, 2004.

Clauss et al., 2013. Marcus Clauss, Dennis W. H. Müller and Daryl Codron. *Source References and the Scientist's Mind-Map: Harvard vs. Vancouver Style*. Journal of Scholarly Publishing, 44(3), 274–282, 2013. ISSN 1198-9742. doi: 10.3138/jsp.44.3.005.

Coglianese et al., 2003. Cary Coglianese, Jennifer Nash and Todd Olmstead. *Performance-based regulation: prospects and limitations in health, safety, and environmental protection*. Administrative Law Review, 55(4), 705–729, 2003. ISSN 0001-8368.

Costin et al., 2012. Aaron Costin, Nipesh Pradhananga and Jochen Teizer. *Leveraging passive RFID technology for construction resource field mobility and status monitoring in a high-rise renovation project*. Automation in Construction, 24, 1–15, 2012. ISSN 09265805. doi: 10.1016/j.autcon.2012.02.015.

Courtney, 2013. Martin Courtney. *Puzzling out big data*. Engineering & Technology, 7(12), 56–60, 2013. doi: 10.1049/et.2012.1215.

Creswell, 2013. John W. Creswell. *Research design : qualitative, quantitative, and mixed methods approaches*. SAGE Publications Inc., 4th edition, 2013. ISBN 978-1-4522-2610-1.

Danish Construction Association, 2017. Danish Construction Association. *Working environment policy of the Danish Construction Association*, 2017.

Danish Working Environment Authority, **2007**. Danish Working Environment Authority. *Analyse af stigning i anmeldte arbejdsulykker 2003-2006*, 2007.

Danish Working Environment Authority, **2016**. Danish Working Environment Authority. *Arbejdstilsynets årsopgørelse 2016*, 2016.

Dansk Byggeri, **2016**. Dansk Byggeri. *Forecast for the Construction Sector*, 2016.

Dansk Byggeri, **2017**. Dansk Byggeri. *Notat 04. juli 2017*, 2017.

C. M. De Dominicis, A. Depari, A. Flammini, S. Rinaldi and E. Sisinni, *2013*. C. M. De Dominicis, A. Depari, A. Flammini, S. Rinaldi and E. Sisinni. Smartphone based localization solution for construction site management. In *2013 IEEE Sensors Applications Symposium Proceedings*, pages 43–48, 2013. ISBN 9781467346351. doi: 10.1109/SAS.2013.6493554.

de Vaus, **2006**. David de Vaus. *Research Design: Sage benchmarks in social research methods.* SAGE Publications Inc., London, 2006. ISBN 0761944877.

Dean, **2014**. Jared Dean. *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*. John Wiley & Sons, Inc., 2014. ISBN 9781118618042. doi: 10.1002/9781118691786.

Deloitte, **2012**. Deloitte. *Workplace safety analytics: Save lives and the bottom line*, 2012.

Deng, **2014**. Li Deng. *Deep Learning: Methods and Applications*. Foundations and Trends® in Signal Processing, 7(3-4), 197–387, 2014. ISSN 1932-8346. doi: 10.1561/2000000039.

Denzin and Lincoln, **2011**. Norman K. Denzin and Yvonna S. Lincoln. *The Sage handbook of qualitative research*. Thousand Oaks : Sage, 2011. ISBN 9781412974172.

Desvignes, **2014**. Matthieu Desvignes. *Requisite Empirical Risk Data for Integration of Safety with Advanced Technologies and Intelligent Systems*, 2014.

Deutsh, **2015**. Randy Deutsh. *Data-Driven Design and Construction*. John Wiley & Sons. Inc, 1. edition, 2015. ISBN 978-1-118-89870-3.

DJI. DJI. *Phantom 4 Pro Reality Capture Package*. URL `https://store.dji.com/product/phantom-4-professional-reality-capture-package`.

Dronebevis. Dronebevis. *Rules and regulations of drone flying in Denmark*. URL `http://dronebevis.dk/english/`.

Dubois and Gadde, **2002**. Anna Dubois and Lars Erik Gadde. *Systematic combining: an abductive approach to case research*. Journal of Business Research, 55(7), 553–560, 2002. ISSN 01482963. doi: 10.1016/S0148-2963(00)00195-8.

Dupont et al., **2017**. Quentin F.M. Dupont, David K.H. Chua, Ahmad Tashrif and Ernest L.S. Abbott. *Potential Applications of UAV along the Construction's Value Chain*. Procedia Engineering, 182, 165–173, 2017. ISSN 18777058. doi: 10.1016/j.proeng.2017.03.155.

R. El Meouche, I. Hijazi, P. A. Poncet, M. Abunemeh and M. Rezoug, *2016*. R. El Meouche, I. Hijazi, P. A. Poncet, M. Abunemeh and M. Rezoug. UAV photogrammetry implementation to enhance land surveying, comparisons and possibilities. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XLII-2, pages 107–114, 2016. ISBN 16821750 (ISSN). doi: 10.5194/isprs-archives-XLII-2-W2-107-2016.

Esmaeili, **2012**. Behzad Esmaeili. *Identifying and quantifying construction safety risks at the attribute level*. ProQuest Dissertations and Theses, page 208, 2012.

Esmaeili and Hallowell, **2011**. Behzad Esmaeili and Matthew R. Hallowell. *Using Network Analysis to Model Fall Hazards on Construction Projects*, 2011.

Behzad Esmaeili and Matthew R. Hallowell, *2012*. Behzad Esmaeili and Matthew R. Hallowell. Attribute-Based Risk Model for Measuring Safety Risk of Struck-By Accidents. In *Construction Research Congress, Construction Challenges in a Flat World*, West Lafayette, Indiana, 2012. American Society of Civil Engineers.

European Commission, **2016**. European Commission. *European Construction Sector Observatory. Country profile Denmark*, 2016.

European Commission, **2018a**. European Commission. *European Construction Sector Observatory. Country profile Denmark*, 2018a.

European Commission, **2018b**. European Commission. *What does the General Data Protection Regulation (GDPR) govern?*, 2018. URL `https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern{_}en{#}references`.

Fang et al., **2004a**. D. P. Fang, F. Xie, X. Y. Huang and H. Li. *Factor analysis-based studies on construction workplace safety management in China*. International Journal of Project Management, 22(1), 43–49, 2004. ISSN 02637863. doi: 10.1016/S0263-7863(02)00115-1.

Fang et al., **2004b**. D.P.a b Fang, X.Y.c d Huang and J.d Hinze. *Benchmarking studies on construction safety management in China*. Journal of Construction Engineering and Management, 130(3), 424–432, 2004. ISSN 07339364. doi: 10.1061/(ASCE)0733-9364(2004)130:3(424).

Forteza et al., **2017**. Francisco J. Forteza, Jose M. Carretero-Gómez and Albert Sesé. *Effects of organizational complexity and resources on construction site risk*. Journal of Safety Research, 62, 185–198, 2017. ISSN 00224375. doi: 10.1016/j.jsr.2017.06.015.

Ganster et al., **2011**. Daniel C. Ganster, Christa E. Kiersch, Rachel E. Marsh and Angela Bowen. *Performance-Based Rewards and Work Stress*. Journal of Organizational Behavior Management, 31 (4), 221–235, 2011. ISSN 01608061. doi: 10.1080/01608061.2011.619388.

Gierej, **2017**. Sylwia Gierej. *The main barriers to effective use of Big Data/ Główne bariery w efektywnym wykorzystaniu Big Data*. Zarządzanie. Teoria i Praktyka, 20(2), 39–45, 2017. ISSN 2449-9730.

Goh and Chua, **2013**. Yang Miang Goh and David Chua. *Neural network analysis of construction safety management systems: a case study in Singapore*. Construction Management and Economics, 31(5), 460–470, 2013. ISSN 01446193. doi: 10.1080/01446193.2013.797095.

Golparvar-Fard et al., **2009**. Mani Golparvar-Fard, Feniosky Peña-Mora, Carlos A. Arboleda and SangHyun Lee. *Visualization of Construction Progress Monitoring with 4D Simulation Model Overlaid on Time-Lapsed Photographs*. Journal of Computing in Civil Engineering, 23(6), 391–404, 2009. ISSN 0887-3801. doi: 10.1061/(ASCE)0887-3801(2009)23:6(391).

Guo et al., 2017. Hongling Guo, Yantao Yu and Martin Skitmore. *Visualization technology-based construction safety management: A review*. Automation in Construction, 73, 135–144, 2017. ISSN 09265805. doi: 10.1016/j.autcon.2016.10.004.

Guo et al., 2016. S. Y. Guo, L. Y. Ding, H. B. Luo and X. Y. Jiang. *A Big-Data-based platform of workers' behavior: Observations from the field*. Accident Analysis and Prevention, 93, 299–309, 2016. ISSN 00014575. doi: 10.1016/j.aap.2015.09.024.

William Haddon, *1983*. William Haddon. Approaches to Prevention of Injuries. In *American Medical Association Conference on Prevention of Disabling Injuries*, Miami, 1983.

Hadikusumo and Rowlinson, 2002. B. H.W. Hadikusumo and Steve Rowlinson. *Integration of virtually real construction model and design-for-safety-process database*. Automation in Construction, 11(5), 501–509, 2002. ISSN 09265805. doi: 10.1016/S0926-5805(01)00061-9.

Hale et al., 2007. A. R. Hale, B. J.M. Ale, L. H.J. Goossens, T. Heijer, L. J. Bellamy, M. L. Mud, A. Roelen, H. Baksteen, J. Post, I. A. Papazoglou, A. Bloemhoff and J. I.H. Oh. *Modeling accidents for prioritizing prevention*. Reliability Engineering and System Safety, 92(12), 1701–1715, 2007. ISSN 09518320. doi: 10.1016/j.ress.2006.09.025.

Hallowell, 2012. Matthew R. Hallowell. *Safety-Knowledge Management in American Construction Organizations*. Journal of Management in Engineering, 28(2), 203–211, 2012. ISSN 0742-597X. doi: 10.1061/(ASCE)ME.1943-5479.0000067.

Hammad et al., 2009. Amin Hammad, M Asce, Hui Wang and Sudhir P. Mudur. *Distributed Augmented Reality for Visualizing Collaborative Construction Tasks.Pdf*. Journal of Computing in Civil Engineering, 23(6), 418–428, 2009. ISSN 0887-3801. doi: 10.1061/(ASCE)0887-3801(2009)23:6(418).

Hansen, 2007. Niels Hansen. *Introduction to DMI*, 2007. URL `http://www.dmi.dk/en/about/profile/introduction-to-dmi/`.

Harvey et al., 2001. Joan Harvey, Helen Bolam, David Gregory and George Erdos. *The effectiveness of training to change safety culture and attitudes within a highly regulated environment*. Personnel Review, 30(6), 615–636, 2001. ISSN 0048-3486. doi: 10.1108/EUM0000000005976.

Haslam et al., 2005. R. A. Haslam, S. A. Hide, A. G.F. Gibb, D. E. Gyi, T. Pavitt, S. Atkinson and A. R. Duff. *Contributing factors in construction accidents*. Applied Ergonomics, 36, 401–415, 2005. ISSN 00036870. doi: 10.1016/j.apergo.2004.12.002.

Hatush and Skitmore, 1997. Zedan Hatush and Martin Skitmore. *Evaluating contractor prequalification data: Selection criteria and project success factors*. Construction Management and Economics, 15(2), 129–147, 1997. ISSN 01446193. doi: 10.1080/014461997373024.

Michele Herrmann, *2016*. Michele Herrmann. Unmanned aerial vehicles in construction: an overview of current and proposed rules. In *Construction Research Congress 2016*, San Juan, Puerto Rico, 2016.

Hinze et al., 2005. Jimmie Hinze, Xinyu Huang and Lani Terry. *The Nature of Struck-by Accidents*. Journal of Construction Engineering and Management, 131(2), 262–268, 2005. ISSN 0733-9364. doi: 10.1061/(ASCE)0733-9364(2005)131:2(262).

Hinze et al., 2013. Jimmie Hinze, Samuel Thurman and Andrew Wehle. *Leading indicators of construction safety performance*. Safety Science, 51, 23–28, 2013. ISSN 09257535. doi: 10.1016/j.ssci.2012.05.016.

Hoła and Szóstak, 2017. Bozena Hoła and Mariusz Szóstak. *Methodology of Analysing the Accident Rate in the Construction Industry*. Procedia Engineering, 172, 355–362, 2017. ISSN 18777058. doi: 10.1016/j.proeng.2017.02.040.

Holte et al., 2015. Kari Anne Holte, Kari Kjestveit and Hester J. Lipscomb. *Company size and differences in injury prevalence among apprentices in building and construction in Norway*. Safety Science, 71, 205–212, 2015. ISSN 18791042. doi: 10.1016/j.ssci.2014.01.007.

Hon et al., 2010. Carol K H Hon, Albert P C Chan and Francis K W Wong. *An analysis for the causes of accidents of repair, maintenance, alteration and addition works in Hong Kong*. Safety Science, 48 (7), 894–901, 2010. ISSN 09257535. doi: 10.1016/j.ssci.2010.03.013.

Gregory A. Howell, Glenn Ballard, Tariq S. Abdelhamid and Panagiotis Mitropoulos, 2002. Gregory A. Howell, Glenn Ballard, Tariq S. Abdelhamid and Panagiotis Mitropoulos. Working near the edge: a new approach to construction safety. In *10th Annual Conference of the International Group for Lean Construction*, 2002.

Huang and Hinze, 2003. Xinyu Huang and Jimmie Hinze. *Analysis of Construction Worker Fall Accidents*. Journal of Construction Engineering and Management, 129(3), 262–271, 2003. ISSN 0733-9364. doi: 10.1061/(ASCE)0733-9364(2003)129:3(262).

Bryan Hubbard, Heng Wang, Michael Leasure, Tim Ropp, Tamara Lofton and Sarah Hubbard, 2015. Bryan Hubbard, Heng Wang, Michael Leasure, Tim Ropp, Tamara Lofton and Sarah Hubbard. Feasibility Study of UAV use for RFID Material Tracking on Construction Sites. In *51st ASC Annual international Conference Proceedings*, 2015.

Hughes et al., 2004. Shawn W. Hughes, Donald D. Tippett and Warren K. Thomas. *Measuring project success in the construction industry*. Engineering Management Journal, 16(3), 31–37, 2004. ISSN 10429247. doi: 10.1080/10429247.2004.11415255.

Irizarry et al., 2012. Javier Irizarry, Masoud Gheisari and Bruce N. Walker. *Usability assessment of drone technology as safety inspection tools*. Journal of Information Technology in Construction (ITcon), 17, 194–212, 2012. ISSN 1874-4753.

Järvis and Tint, 2009. Marina Järvis and Piia Tint. *The formation of a good safety culture at enterprise*. Journal of Business Economics and Management, 10(2), 169–180, 2009. ISSN 1611-1699. doi: 10.3846/1611-1699.2009.10.169-180.

Ester Jensen and Christian Koch, 2003. Ester Jensen and Christian Koch. Safety efforts on large construction projects - the Copenhagen Metro. In *Third Nordic Conference on Construction Economics and Organization*, Lund, 2003.

Jørgensen, 2016. Kirsten Jørgensen. *Prevention of "simple accidents at work" with major consequences*. Safety Science, 81, 46–58, 2016. ISSN 18791042. doi: 10.1016/j.ssci.2015.01.017.

Kelm et al., 2013. Agnes Kelm, Lars Laußat, Anica Meins-Becker, Daniel Platz, Mohammad J. Khazaee, Aaron M. Costin, Manfred Helmus and Jochen Teizer. *Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment*

(PPE) on construction sites. Automation in Construction, 36, 38–52, 2013. ISSN 09265805. doi: 10.1016/j.autcon.2013.08.009.

Khan and Ali, **2013**. Rafiqul Zaman Khan and Javed Ali. *Use of DAG in Distributed Parallel Computing*. 2(11), 81–85, 2013.

Kim et al., **2017**. Kinam Kim, Hongjo Kim and Hyoungkwan Kim. *Image-based construction hazard avoidance system using augmented reality in wearable device*. Automation in Construction, 83, 390–403, 2017. ISSN 09265805. doi: 10.1016/j.autcon.2017.06.014.

Kines, **2002**. Pete Kines. *Construction workers' falls through roofs: Fatal versus serious injuries*. Journal of Safety Research, 33(2), 195–208, 2002. ISSN 00224375. doi: 10.1016/S0022-4375(02)00019-1.

Kiviniemi et al., **2011**. Markku Kiviniemi, Kristiina Sulankivi, Kalle Kähkönen, Tarja Mäkelä and Maija Leena Merivirta. *BIM-based safety management and communication for building construction*, 2011.

Kometa et al., **1995**. Simon T. Kometa, Paul O. Olomolaiye and Frank C. Harris. *An evaluation of clients' needs and responsibilities in the construction process*. Engineering, Construction and Architectural Management, 2(1), 57–76, 1995.

Koseleva and Ropaite, **2017**. Natalija Koseleva and Guoda Ropaite. *Big data in building energy efficiency: understanding of big data and main challenges*. Procedia Engineering, 172, 544–549, 2017. ISSN 18777058. doi: 10.1016/j.proeng.2017.02.064.

Kumaraswamy and Thorpe, **1996**. Mohan M. Kumaraswamy and Antony Thorpe. *Systematizing Construction Project Evaluations*. Journal of Management in Engineering, 12(1), 34–39, 1996. ISSN 0742-597X. doi: 10.1061/(ASCE)0742-597X(1996)12:1(34).

Laney, **2001**. Doug Laney. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group Research Note, 949(February), 2001. ISSN 09505849. doi: 10.1016/j.infsof.2008.09.005.

Larsson and Field, **2002**. Tore J. Larsson and Brian Field. *The distribution of occupational injury risks in the Victorian construction industry*. Safety Science, 40(5), 439–456, 2002. ISSN 09257535. doi: 10.1016/S0925-7535(01)00015-7.

Lavalle et al., **2011**. Steve Lavalle, Eric Lesser, Rebecca Shockley, Michael S Hopkins and Nina Kruschwitz. *Big Data, Analytics and the Path From Insights to Value*. MIT Sloan Management Review, 52(2), 21–32, 2011. ISSN 15329194. doi: 10.0000/PMID57750728.

Lee et al., **2009**. Ung Kyun Lee, Joo Heon Kim, Hunhee Cho and Kyung In Kang. *Development of a mobile safety monitoring system for construction sites*. Automation in Construction, 18, 258–264, 2009. ISSN 09265805. doi: 10.1016/j.autcon.2008.08.002.

Lee et al., **2017**. Wonil Lee, Ken Yu Lin, Edmund Seto and Giovanni C. Migliaccio. *Wearable sensors for monitoring on-duty and off-duty worker physiological status and activities in construction*. Automation in Construction, 83, 341–353, 2017. ISSN 09265805. doi: 10.1016/j.autcon.2017.06.012.

Leung et al., 2012. Mei Yung Leung, Isabelle Yee Shan Chan and Jingyu Yu. *Preventing construction worker injury incidents through the management of personal stress and organizational stressors*. Accident Analysis and Prevention, 48, 156–166, 2012. ISSN 00014575. doi: 10.1016/j.aap.2011.03.017.

Li et al., 2015. Heng Li, Miaojia Lu, Shu Chien Hsu, Matthew Gray and Ting Huang. *Proactive behavior-based safety management for construction safety improvement*. Safety Science, 75, 107–117, 2015. ISSN 18791042. doi: 10.1016/j.ssci.2015.01.013.

Li and Liu, 2018. Yan Li and Chunlu Liu. *Applications of multirotor drone technologies in construction management*. International Journal of Construction Management, 2018. doi: 10.1080/15623599.2018.1452101.

Liao and Perng, 2008. Chia W. Liao and Yeng Horng Perng. *Data mining for occupational injuries in the Taiwan construction industry*. Safety Science, 46(7), 1091–1102, 2008. ISSN 09257535. doi: 10.1016/j.ssci.2007.04.007.

Liao, 2012. Chia Wen Liao. *Pattern analysis of seasonal variation in occupational accidents in the construction industry*. Procedia Engineering, 29, 3240–3244, 2012. ISSN 18777058. doi: 10.1016/j.proeng.2012.01.473.

Lim and Mohamed, 1999. C. S. Lim and M. Zain Mohamed. *Criteria of project success: an exploratory re-examination*. International Journal of Project Management, 17(4), 243–248, 1999. ISSN 02637863. doi: 10.1016/S0263-7863(98)00040-4.

López-Alonso et al., 2013. Monica López-Alonso, Maria Pilar Ibarrondo-Dávila, María Carmen Rubio-Gámez and Teresa Garcia Munoz. *The impact of health and safety investment on construction company costs*. Safety Science, 60, 151–159, 2013. ISSN 09257535. doi: 10.1016/j.ssci.2013.06.013.

Lu et al., 2011. Weisheng Lu, George Q. Huang and Heng Li. *Scenarios for applying RFID technology in construction project management*. Automation in Construction, 20, 101–106, 2011. ISSN 09265805. doi: 10.1016/j.autcon.2010.09.007.

Lycett, 2013. Mark Lycett. *'Datafication': Making sense of (big) data in a complex world*. European Journal of Information Systems, 22(4), 381–386, 2013. ISSN 0960085X. doi: 10.1057/ejis.2013.10.

D. Manase, L. Mahdjoubi and V. Ahmed, 2004. D. Manase, L. Mahdjoubi and V. Ahmed. Accident Prevention On Construction Sites: Towards A New Approach. In *4th International Postgraduate Research Conference in the Built and Human Environment*, 2004.

Manu et al., 2010. Patrick Manu, Nii Ankrah, David Proverbs and Subashini Suresh. *An approach for determining the extent of contribution of construction project features to accident causation*. Safety Science, 48(6), 687–692, 2010. ISSN 09257535. doi: 10.1016/j.ssci.2010.03.001.

Manu et al., 2013. Patrick Manu, Nii Ankrah, David Proverbs and Subashini Suresh. *Mitigating the health and safety influence of subcontracting in construction: The approach of main contractors*. International Journal of Project Management, 31(7), 1017–1026, 2013. ISSN 02637863. doi: 10.1016/j.ijproman.2012.11.011.

Marr, 2015. Bernard Marr. *Big data : using smart big data, analytics and metrics tomake better decisions and improve performance*. John Wiley & Sons. Ltd., West Sussex, 2015. ISBN 9781446287484.

Marr, **2016**. Bernard Marr. *Key Business Analytics The 60 business analysis tools every manager needs to know*. 2016. ISBN 978-1-292-01743-3.

Marz and Warren, **2015**. Nathan Marz and James Warren. *Big Data: Principles and best practices of scalable real-time data systems*. Manning Publications Co., Shelter Island, 2015. ISBN 9781617290343.

Mat Yasin et al., **2016**. Mohd Fadzil Mat Yasin, Mohd Azian Zaidi and Mohd Nasrun Mohd Nawi. *A review of Small Unmanned Aircraft System (UAS) advantages as a tool in condition survey works*. MATEC Web of Conferences, 66, 2016. ISSN 2261-236X. doi: 10.1051/matecconf/20166600038.

Maté et al., **2017**. Alejandro Maté, Juan Trujillo and John Mylopoulos. *Specification and derivation of key performance indicators for business analytics: A semantic approach*. Data and Knowledge Engineering, 108(January), 30–49, 2017. ISSN 0169023X. doi: 10.1016/j.datak.2016.12.004.

Mayer-Schönberger and Cukier, **2013**. Viktor Mayer-Schönberger and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and T*. New York: Houghton Mifflin Harcourt, 2013. ISBN 978-1848547926. doi: 10.2501/IJA-33-1-181-183.

McKinsey Global Institute, **2011**. McKinsey Global Institute. *Big data: The next frontier for innovation, competition, and productivity*, 2011.

Raquel Mello, Luciana Rosa Leite and Roberto Antonio Martins, *2014*. Raquel Mello, Luciana Rosa Leite and Roberto Antonio Martins. Is Big Data the Next Big Thing in Performance Measurement Systems? In *Proceedings of the 2014 Industrial and Systems Engineering Research Conference*, 2014. ISBN 9780983762430.

Motawa, **2017**. Ibrahim Motawa. *Spoken dialogue BIM systems – an application of big data in construction*. Facilities, 35(13/14), 787–800, 2017. ISSN 0263-2772. doi: 10.1108/F-01-2016-0001.

Mučenski et al., **2015**. Vladimir Mučenski, Igor Peško, Jasmina Dražić, Goran Ćirović, Milan Trivunić and Dragana Bibić. *Construction workers injury risk assessment in relation to their experience and age*. Procedia Engineering, 117, 530–538, 2015. ISSN 18777058. doi: 10.1016/j.proeng.2015.08.205.

Nielsen, **2007**. Jesper Nielsen. *Struggles for health and safety in the Danish construction industry*. International Journal of Occupational and Environmental Health, 13(1), 21–26, 2007. ISSN 10773525. doi: 10.1179/107735207800245018.

Niskanen et al., **2016**. Toivo Niskanen, Kyösti Louhelainen and Maria L. Hirvonen. *A systems thinking approach of occupational safety and health applied in the micro-, meso- and macro-levels: A Finnish survey*. Safety Science, 82, 212–227, 2016. ISSN 18791042. doi: 10.1016/j.ssci.2015.09.012.

Oleske et al., **1989**. Denise M. Oleske, Robert D. Brewer, Pamela Doan and Jerome Hahn. *An epidemiologic evaluation of the injury experience of a cohort of automotive parts workers: A model for surveillance in small industries*. Journal of Occupational Accidents, 10(4), 239–253, 1989. ISSN 03766349. doi: 10.1016/0376-6349(89)90019-9.

Olive et al., **2011**. Joseph Olive, Caitlin Christianson and John MacCary. *Handbook of Natrual Language Processing and Machine Translation*. Springer, 2011. ISBN 9781441977120.

Oliver, **2012**. Paul Oliver. *Succeeding With Your Literature Review: A Handbook For Students*. Open University Press, 2012. ISBN 9780335243686.

Park and Kim, 2013. Chan Sik Park and Hyeon Jin Kim. *A framework for construction safety management and visualization system*. Automation in Construction, 33, 95–103, 2013. ISSN 09265805. doi: 10.1016/j.autcon.2012.09.012.

Pérez-Alonso et al., 2011. José Pérez-Alonso, Ángel Carreño-Ortega, Ángel J. Callejón-Ferre and F. Javier Vázquez-Cabrera. *Preventive activity in the greenhouse-construction industry of south-eastern Spain*. Safety Science, 49(2), 345–354, 2011. ISSN 09257535. doi: 10.1016/j.ssci.2010.09.013.

Philliber et al., 1980. G. S. Philliber, M. R. Schwab and G. Samsloss. *Social research: Guides to a decision-making process*. 1980.

Pinto and Pinto, 1991. M. B. Pinto and J. K. Pinto. *Determinants of cross-functional cooperation in the project implementation process*. Project Management Journal, 22(2), 13–20, 1991.

Pocock et al., 1996. James B. Pocock, Chang T. Hyun, Liang Y. Liu and Michael K. Kim. *Relationship between project interaction and performance indicators*. Journal of Construction Engineering and Management, 122(June), 165–176, 1996. ISSN 0733-9364/96/0002-0165-0176.

Poh et al., 2018. Clive Q.X. Poh, Chalani Udhyami Ubeynarayana and Yang Miang Goh. *Safety leading indicators for construction sites: A machine learning approach*. Automation in Construction, (October 2017), 0–1, 2018. ISSN 09265805. doi: 10.1016/j.autcon.2018.03.022. URL `http://dx.doi.org/10.1016/j.autcon.2018.03.022`.

Prades Villanova, 2014. Marc Prades Villanova. *Attribute-based Risk Model for Assessing Risk to Industrial Construction Tasks*, University of Colorado, 2014.

Rabi et al., 1998. Atallah Z. Rabi, Linda W. Jamous, Burhan A. AbuDhaise and Rafi H. Alwash. *Fatal occupational injuries in Jordan during the period 1980 through 1993*. Safety Science, 28(3), 177–187, 1998. ISSN 09257535. doi: 10.1016/S0925-7535(98)80007-6.

Robson and McCartan, 2016. Colin Robson and Kieran McCartan. *Real world research : a resource for users of social research methods in applied settings*. Chichester : John Wiley & Sons, Inc, 4th edition, 2016. ISBN 978-1-118-74523-6.

Robson et al., 2017. Lynda S. Robson, Selahadin Ibrahim, Sheilah Hogg-Johnson, Ivan A. Steenstra, Dwayne Van Eerd and Benjamin C. Amick. *Developing leading indicators from OHS management audit data: Determining the measurement properties of audit data from the field*. Journal of Safety Research, 61, 93–103, 2017. ISSN 00224375. doi: 10.1016/j.jsr.2017.02.008.

Rossman and Wilson, 1985. Gretchen B. Rossman and Bruce L. Wilson. *Numbers and Words: Combining Quantitative and Qualitative Methods in a Single Large-Scale Evaluation Study*. Evaluation Review, 9(5), 627–643, 1985.

Rowlinson et al., 2014. Steve Rowlinson, Andrea Yunyanjia, Baizhan Li and Carrie Chuanjingju. *Management of climatic heat stress risk in construction: A review of practices, methodologies, and future research*. Accident Analysis and Prevention, 66, 187–198, 2014. ISSN 00014575. doi: 10.1016/j.aap.2013.08.011.

Sawacha et al., 1999. Edwin Sawacha, Shamil Naoum and Daniel Fong. *Factors affecting safety performance on construction sites*. International Journal of Project Management, 17(5), 309–315, 1999. ISSN 0263-7863. doi: 10.1016/S0263-7863(98)00042-8.

Sector Working Environment Committee for Building and Construction, **2016**. Sector Working Environment Committee for Building and Construction. *Manual – building and construction working environment*. 1st. 2016. ISBN 978-87-7952-248-0.

Singh and Reddy, **2015**. Dilpreet Singh and Chandan K. Reddy. *A survey on platforms for big data analytics*. Journal of Big Data, 2(1), 1–20, 2015. ISSN 21961115. doi: 10.1186/s40537-014-0008-6.

Song and Zhu, **2016**. Il-Yeol Song and Yongjun Zhu. *Big data and data science: what should we teach?* Expert Systems, 33(4), 364–373, 2016. ISSN 02664720. doi: 10.1111/exsy.12130.

Sørensen et al., **2016**. Anette Østbø Sørensen, Nils Olsson and Andreas D. Landmark. *Big Data in Construction Management Research*. 405-416, pages 1–12, 2016.

Spangenberg et al., **2003**. Søren Spangenberg, Charlotte Baarts, Johnny Dyreborg, Lars Jensen, Pete Kines and Kim L. Mikkelsen. *Factors contributing to the differences in work related injury rates between Danish and Swedish construction workers*. Safety Science, 41(6), 517–530, 2003. ISSN 09257535. doi: 10.1016/S0925-7535(02)00007-3.

Spens and Kovács, **2006**. Karen M. Spens and Gyöngyi Kovács. *A content analysis of research approaches in logistics research*. International Journal of Physical Distribution & Logistics Management, 36(5), 374–390, 2006. doi: 10.1108/09600030610676259.

Statistikken Arbejdsulykker, **2016**. Statistikken Arbejdsulykker. *ARBEJDSULYKKER 2015*, 2016.

Steele et al., **2016**. Brian Steele, John Chandler and Swarna Reddy. *Algorithms for Data Science*. Springer, 2016. ISBN 978-3-319-45795-6. doi: 10.1007/978-3-319-45797-0. URL `http://link.springer.com/10.1007/978-3-319-45797-0`.

Storey and Song, **2017**. Veda C. Storey and Il Yeol Song. *Big data technologies and Management: What conceptual modeling can do*. Data and Knowledge Engineering, 108(February), 50–67, 2017. ISSN 0169023X. doi: 10.1016/j.datak.2017.01.001.

Tabakow et al., **2014**. Marta Tabakow, Jerzy Korczak and Bogdan Franczyk. *Big Data - Definitions, challenges and information technologies*. Business Informatics, 1(31), 2014. ISSN 1507-3858. doi: 10.15611/ie.2014.1.12.

Tashakkori and Teddlie, **2010**. Abbas Tashakkori and Charles Teddlie. *SAGE handbook of mixed methods in social & behavioral research*. Thousand Oaks, Calif. : SAGE Publications, 2010. ISBN 9781412972666.

Teran et al., **2015**. Suzanne Teran, Hillary Blecker, Kelsie Scruggs, Javier García Hernández and Barbara Rahke. *Promoting adoption of fall prevention measures among Latino workers and residential contractors: Formative research findings*. American Journal of Industrial Medicine, 58 (8), 870–879, 2015. ISSN 10970274. doi: 10.1002/ajim.22480.

The International Institute of Risk and Safety Management, **2017**. The International Institute of Risk and Safety Management. *Big data is for safety professionals*, 2017. URL `https://www.iirsm.org/big-data-safety-professionals`.

Tien, **2013**. James M. Tien. *Big Data: Unleashing information*. Journal of Systems Science and Systems Engineering, 22(2), 127–151, 2013. ISSN 10043756. doi: 10.1007/s11518-013-5219-4.

Tixier et al., 2016a. Antoine J.P. Tixier, Matthew R. Hallowell, Balaji Rajagopalan and Dean Bowman. *Application of machine learning to construction injury prediction*. Automation in Construction, 69, 102–114, 2016. ISSN 09265805. doi: 10.1016/j.autcon.2016.05.016.

Tixier et al., 2016b. Antoine J.P. Tixier, Matthew R. Hallowell, Balaji Rajagopalan and Dean Bowman. *Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports*. Automation in Construction, 62, 45–56, 2016. ISSN 09265805. doi: 10.1016/j.autcon.2015.11.001.

Tixier et al., 2017. Antoine J.P. Tixier, Matthew R. Hallowell, Balaji Rajagopalan and Dean Bowman. *Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes using Data Mining*. Automation in Construction, 74, 39–54, 2017. ISSN 09265805. doi: 10.1016/j.autcon.2016.11.001. URL `http://dx.doi.org/10.1016/j.autcon.2016.11.001`.

Waller and Fawcett, 2013. Matthew A. Waller and Stanley E. Fawcett. *Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management*. Journal of Business Logistics, 34(2), 77–84, 2013. ISSN 21581592. doi: 10.1111/jbl.12010.

Wang et al., 2016. Fan Wang, Lieyun Ding, Peter E.D. Love and David J. Edwards. *Modeling tunnel construction risk dynamics: Addressing the production versus protection problem*. Safety Science, 87, 101–115, 2016. ISSN 18791042. doi: 10.1016/j.ssci.2016.01.014.

Williams et al., 2015. Paul Williams, Nicholas J. Ashill, Earl Naumann and Eric Jackson. *Relationship quality and satisfaction: Customer-perceived success factors for on-time projects*. International Journal of Project Management, 33(8), 1836–1850, 2015. ISSN 02637863. doi: 10.1016/j.ijproman.2015.07.009.

Wilson and Koehn, 2000. Joe M Wilson and Enno "Ed" Koehn. *Safety Management: Problems Encountered and Recommended Solutions*. Journal of Construction Engineering and Management, 126(1), 77–79, 2000. ISSN 1943-7862. doi: 10.1061/(ASCE)0733-9364(2000)126:1(77).

Winge and Albrechtsen, 2018. Stig Winge and Eirik Albrechtsen. *Accident types and barrier failures in the construction industry*. Safety Science, 105, 158–166, 2018. ISSN 18791042. doi: 10.1016/j.ssci.2018.02.006.

Wu et al., 2013. Weiwei Wu, Huanjia Yang, Qiming Li and David Chew. *An integrated information management model for proactive prevention of struck-by-falling-object accidents on construction sites*. Automation in Construction, 34, 67–74, 2013. ISSN 09265805. doi: 10.1016/j.autcon.2012.10.010.

Yearbook, 2017. Statistical Yearbook. *Statistics Denmark - Statistical Yearbook 2017. 7 June 2017*. 2017. ISBN 978-87-501-2274-6.

Yeh et al., 2009. Wan Yu Yeh, Yawen Cheng and Chiou Jung Chen. *Social patterns of pay systems and their associations with psychosocial job characteristics and burnout among paid employees in Taiwan*. Social Science and Medicine, 68(8), 1407–1415, 2009. ISSN 02779536. doi: 10.1016/j.socscimed.2009.01.031.

Ylijoki and Porras, 2016. Ossi Ylijoki and Jari Porras. *Perspectives to Definition of Big Data: A Mapping Study and Discussion*. Journal of Innovation Management Ylijoki, Porras JIM, 4(1), 69–91, 2016. ISSN 2183-0606.

Yu and Guo, **2016**. Shui Yu and Song Guo. *Big Data Concepts, Theories, and Applications*. Springer International Publishing, Cham, 2016. ISBN 978-3-319-27761-5. doi: 10.1007/978-3-319-27763-9.

Zhang et al., **2018**. Bingfei Zhang, Zhenhua Zhu, Amin Hammad and Walid Aly. *Automatic matching of construction onsite resources under camera views*. Automation in Construction, 91, 206–215, 2018. ISSN 09265805. doi: 10.1016/j.autcon.2018.03.011.

Zhou et al., **2015**. Zhipeng Zhou, Yang Miang Goh and Qiming Li. *Overview and analysis of safety management studies in the construction industry*. Safety Science, 72, 337–350, 2015. ISSN 18791042. doi: 10.1016/j.ssci.2014.10.006.

Zhu et al., **2016**. Zhenhua Zhu, Man Woo Park, Christian Koch, Mohamad Soltani, Amin Hammad and Khashayar Davari. *Predicting movements of onsite workers and mobile equipment for enhancing construction site safety*. Automation in Construction, 68, 95–101, 2016. ISSN 09265805. doi: 10.1016/j.autcon.2016.04.009.