Introducing process mining for AECFM: three experimental case studies

S. van Schaijk, stijnvanschaijk@gmail.com
Stam en De Koning, Eindhoven, The Netherlands

L.A.H.M. van Berlo, leon.vanberlo@tno.nl
Netherlands organization for applied scientific research TNO, Delft, The Netherlands

ABSTRACT: The research field of process mining is relatively new and not been applied often in the Architecture, Engineering, Construction and Facility Management industry (AECFM). Process mining uses databases of existing IT systems to gain major insights in processes. Currently the AECFM industry increasingly adapts IT systems within all phases of the process. This creates the possibility to use process mining techniques to gain insight in the processes of construction projects. This paper introduces process mining by presenting three experimental case studies which are conducted in order to study the applicability of process mining in the AECFM. Studies are done in the design-, build- and operational phase. The study has proven to provide useful insight and potential applications. The research method that was used does not allow generalisation of the conclusions for the whole industry. Additional research is needed to study the potential of integrating different data sources from several phases.

Keywords: Data Mining, Process Mining, BIM, Systems Engineering, Facility Management.

1 INTRODUCTION

During the ‘data explosion’ from last decades the capabilities of information systems expanded rapidly. As a result the digital universe and the physical universe are becoming more and more aligned. The growth of the technological possibilities with RFID (Radio Frequency Identification), GPS (Global Positioning System), Intelligent Imaging Camera systems, and sensor networks like the Internet of Things will stimulate further alignment of the digital and physical universe. Even in the construction industry, which generally is perceived as ‘old fashioned’, these new technologies are implemented. This data expension development makes it possible to record physical events as ‘event logs’ and analyse them digitally.

Process mining is all about exploiting event data in a meaningful way in order to provide insights, identify bottlenecks, anticipate problems, record policy violations, recommend countermeasures, and streamline processes. (van der Aalst, 2011)

Process mining uses databases of existing IT systems to gain major insights in processes. Currently the AECFM adapts IT systems within all phases more and more. Which gives the possibility to use process mining techniques to gain insight in the processes of construction projects.

In general these IT systems are not used to extract event logs, but are used to run daily activities. Thereby companies do not realize that they are creating valuable data.

Currently a lot of construction companies trying to optimize their processes with help of Building Information Modelling (BIM) and LEAN approaches. Companies claim to be LEAN. Where they try to optimize processes in order to increase customer satisfaction. One of the fundamentals of LEAN is knowing what you are doing and have insight into your own processes (Sayer & Williams, 2013). Business Process Management (BPM) and Process Intelligence (PI) approaches are used to monitor and control LEAN process flows.

Interviews with practitioners conducted by Quirijnen & van Schaijk, (2013) revealed that these monitoring systems mainly depend on human written notes. It was noticed that despite the importance, the current practices of monitoring systems are still-time-consuming, costly and prone to errors. Thereby monitoring is rarely be done and the data is nearly used.

Given (a) the interest in monitoring, observation of construction projects and the interest in construction process models, (b) the limited quality of current monitoring and observation tools and the unrealistic hand-made models, and (c) the possibilities to

autonomous create, store, and extract event data from IT systems, it seems legit to study the value of applying process mining in the construction. Therefore this research will explore the possibilities of improving construction projects with help of process mining.

This paper starts with elaborating the research goals and methods. There after process mining and event logs are explained more detailed. Followed by a description of the case studies which are conducted. The papers finishes with a conclusion, discussion and recommendations for further work.

2 RESEARCH

2.1 *Research goals*

The main goal of this research is to improve construction projects by providing insight in the construction process. This study explores if process mining techniques are suitable for providing insights which lead to more efficient processes in the AECFM.

This study targets to introduce process mining in the AECFM. Thereby it explores especially the applications and usability of process mining. In addition it gives insight in opinions about process mining of people working in the industry. But it does not give details about process mining statistics or algorithms. When one is interested in more details it is recommended to read the report BIM based process mining by van Schaijk (2016).

2.2 *Method*

Three experimental case studies are conducted in order to research out the applicability of process mining in the AECFM. The cases are executed in the design-, construction- and operational phase. Thereby a large part of the construction process is covered and a good overview can be created of the applicability of process mining.

Firstly a study is done within the design phase of a construction project. A systems engineering database of a civil project of a large contractor is analysed with process mining techniques.

Secondly a study is done within the construction phase of a project. As-planned- and as-built Building Information Models are used to create a process-oriented data warehouse which is analysed with process mining techniques.

Lastly a study is done within the operational phase of a building. This experimental study explored if facility management data is suitable for analysing processes around building elements with process mining techniques.

3 BACKGROUND

3.1 *Process mining*

Process mining provides an approach to gain insight and improve processes in a lot application domains. The goal of process mining is to gain event data, extract process-related information and discover a process model. Most organizations detect process problems based on fiction rather than facts. Van der Aalst (2011) describes process mining as an "*emerging discipline providing comprehensive sets of tools to provide fact-based insights and to support process improvements*" (p. 7). In comparison with other data mining techniques, like for example Business Process Management (BPM) and Business Intelligence (BI), process mining provides a full understanding of as-is to end-to-end processes. BI focuses on dashboards and reporting rather than clear process insights. BPM heavily relies on the experts which are modelling the to-be process and not help organizations to understand the actual as-is processes. (van der Aalst, 2011)

Little research is done in studying the potential of process mining in construction projects. A small study is executed by Terlouw and Mulder (2014) who explored the potential of process mining with the ISO standard for communication VISI. They used the events extracted from VISI archives in order to gain insight into communication processes in civil projects. They gained insights within three projects from a social interaction perspective and concluded that the potential of this data analyzing approach in the construction industry is high.

When looking at other industries several process mining studies have been done. Applications of process mining can be found in various economic sectors and industries like healthcare, governments, banking and insurance, educational instances, retail, transportation, cloud computing, capital goods industry. (van der Aalst, 2011)

The idea of process mining is not new. The roots of the research field can be found a half century ago. (Nerode, 1958) already presented an approach to synthesize finite-state machines from traces, (Petri, 1962) introduced the first modelling language capturing concurrency and (Gold, 1967) first explored different notions of learnability. While data mining gained more attention during the nineties little attention was given to process related mining. Since the first survey on process mining in 2003 (van der Aalst et al., 2003) a lot of progress in the research field has been made. Several techniques have been developed and various tools have come into existence. A comprehensive overview of the state-of-the-art in process mining is given in the book of Wil van der Aalst "Process Mining Discovery, Conformance and Enhancement of Business Processes".

3.2 *Event logs*

The starting point of process mining is typically a raw data source. A raw data source may be every file, for example an Excel spreadsheet, transaction log, or a database. Often the necessary data is scattered over different files or data sources. For cross-organizational mining those sources may be distributed over multiple organisations. The raw data source has to be converted into event logs in order to be suitable for process mining analytics. The availability of high quality event logs is essential in order to enable process mining (van der Aalst, 2011).

3.2.1 *Event log structure*

Using figure 1 as a reference, the following is assumed regarding event logs (van der Aalst, 2011):

- A process consists of cases.
- A case consists of events such that each event related to precisely one case.
- Events within a case are ordered.
- Events can have attributes. Examples of attributes are activity, time, costs, and resource.

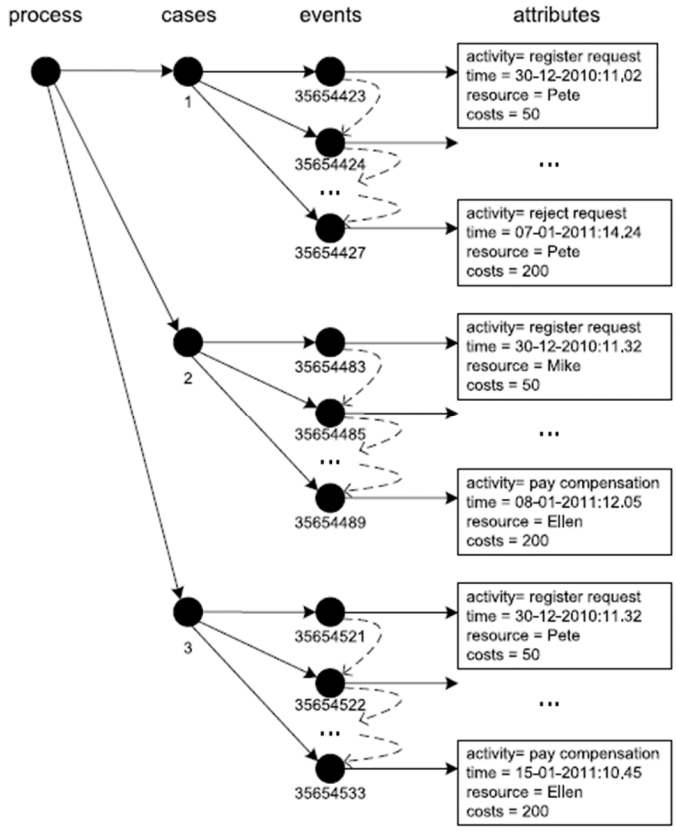

Figure 1. Event log structure (van der Aalst, 2011)

The bare minimum to enable process mining requires a Case ID and an Activity log:

1. Case ID: A case identifier, also called process instance ID, is necessary to distinguish different executions of the same process. What precisely the case ID is depends on the domain of the process. For example in a hospital this would be the patient ID. In the case of construction process mining with BIM the case ID would be the Global Unique Identifier (GUID) of the building element when one is interested in this specific process. When someone is interested in the generic process of a project he could use the unique building ID as a case ID.
2. Activity: There should be names for different process steps or status changes that were performed in the process. If you have only one entry (one row) for each process instance, then your data is not detailed enough. The data needs to be on the transactional level (you should have access to the history of each case) and should not be aggregated to the case level. In the case of construction process mining of a building element there should be a history of this element.

To analyze performance related properties additional attributes are useful in the event log like:

3. Timestamp: At least one timestamp is needed to bring events in the right order. This time stamp is also needed to identify delays between activities and bottleneck identification.
4. Resources: For example the person or the company who executed the activity.
5. … More attributes can be added…

However the structure and information within the event logs may be the same, event logs can be described with different data formats. Detailed information about those formats can be read in (van der Aalst, 2011, Chapter 4), but the most occurring ones are CSV and XES (eXtensible Event Stream).

3.3 *Process mining analytics*

Mature process mining tools are available and offer advanced analytics. In this study process mining tools Disco (https://fluxicon.com/disco/) and MyInvenio (https://www.my-invenio.com/) are mainly used for analytics. However scientific researchers may prefer using ProM (http://www.promtools.org/) which gives an extensible framework where one can contribute by developing plug-ins.

Event logs enable three types of process mining (van der Aalst, 2011). The first one is *discovery* where an event log is taken as input and a process model (i.e. a Petri net or BPMN model) is generated as output. In addition it is also possible to discover resource-related models, such as social networks, when the log contains information about resources.

The second type of mining is *conformance*. This method compares reality with an existing process

model using event logs of the process. In addition conformance checking can be used to detect, locate and explain deviations, and to measure the value of these deviations.

The last type of process mining is *enhancement*. This type focuses on extending or improving existing processes using event logs. Where conformance checking measures alignment, enhancement focusses on repairing, changing or extending the model. A model can be extended by adding performance data. Doing this enables one to show for example bottlenecks, service levels, throughput times, and frequencies.

4 CASE STUDIES

4.1 *Construction design mining*

Techniques that improved performance in other industries are more often adopted in the construction industry such as Systems Engineering (SE). Interviews with SE experts conducted within this study have indicated that SE IT systems supports with managing their project and helps to prove that the clients specifications are realized. However it is not known if those systems support an efficient process. Practioners indicated that a lot of people are involved in such projects who all work in parts of the IT systems but nobody has a clear overview of the total process. This case study explores the possibilities of discovering parts of the design process with process mining techniques. By use a large civil project the potential of process mining within the design phase of construction projects is discovered.

A database of the project is exported from SE tool Relatics (http://www.relatics.com/). It took some time to transform the data into event logs. But thereafter valuable analytics could be done with process mining tools Disco and MyInvenio

An as-desired process model of the project was made by the process engineers of the contractor. It is compared to the process model discovered with process mining. It seemed that both models did not match at all. In addition several process steps occurred repeatedly after each other. Some process-loops occurred more than 500 times, which was much more than expected.

Real bottlenecks are found in document flows, process variants are discovered, and social networks are exposed which proved that some people did not use the SE tool much. With this information improvements in the design process can be managed by the process engineers.

Due to this experimental case study the process engineers realized that the organization does not control the field of information- and data modelling enough. It is concluded that it is valuable to use process mining to give continuously feedback to the project managers at contractors. Process mining gives unique visualisations which enables refreshing insights. In addition process mining analytics did realize that some IT systems are used to store information, but those systems do not automatically facilitate an efficient process.

4.2 *Construction site process mining*

This case study targeted to get insight in the deviation of the as-planned and as-built process on construction sites. Since hardly any as-built data was available the researchers conducted an on-site research and collected the data.

It is seen that BIM based planning tools can be used to make as-planned models including planned tasks linked to building elements. Several monitoring technologies are available to capture the progress and planning deviation on construction sites. In this case drones flew repeatedly over the site. The pictures created with the drones are used to generate as-built point clouds models, which are compared to as-planned BIM models. The deviations of the as-planned model and as-built process where measured. By use of BIMserver (Beetz & Berlo, 2010) the as-planned model is translated into event logs. The as-built data is merged with those logs resulting in as-built event logs.

By use of the Plan-Capture-Analyse-Reuse workflow and several BIM and process mining tools, continuously learning loops for construction companies where realized. As part of this workflow the Planning consult software tool is developed. By use of this tool it is proven that event logs from previous projects can be reused in order to advice construction planners and identify risks in early phases of new to build construction projects. Thereby it serves as a basis for continuously learning and enables construction project time decrease.

It is noticed that with process mining tools bottlenecks from the event logs can be discovered. This case study proved that it is possible to gain insight in planning deviation, social networks (see figure 2) and bottlenecks (see figure 3) during the construction phase. In addition the technologies and methods proposed in this research can be implemented in current processes of construction companies. They will be useful and enable to reuse all captured data. Therefore it will form the basis for knowledge reassurance and fact based improving.

Based on the case study it can be said that the insights gained with the proposed methods have positive influence on managing construction projects.

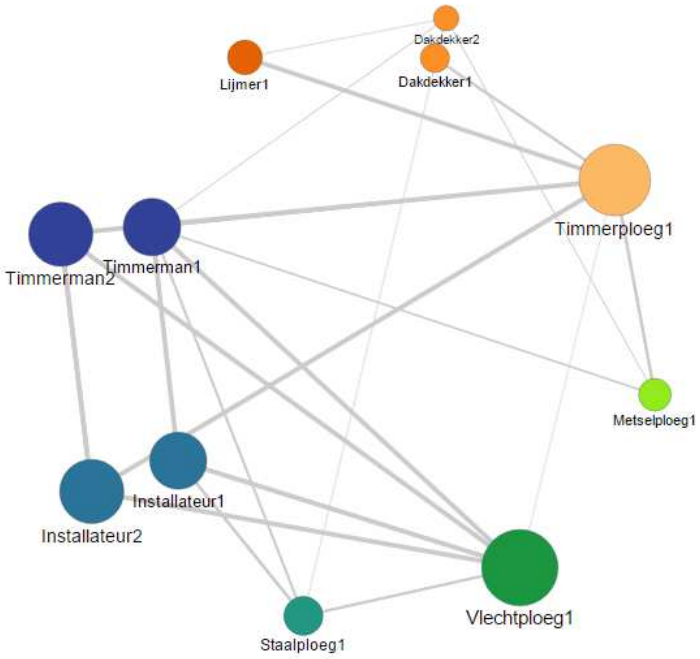

Figure 2. Social network visualizing the links between different people of the project. More lines means a bigger role in the project.

4.3 *Facility management mining*

This experimental study explored if facility management data is suitable for analyzing processes around building elements with process mining techniques. By use of a case study at a hospital the potential of combining process mining with maintenance data is discovered. It can be concluded that with some data transformation maintenance data is suitable for process mining. Moreover the facility managers were surprised by the visualization techniques and they gained clear insight in the error handling process. As a result they discovered problematic building elements and odd processes. In addition as a result of this analysis the facility managers were surprised about the amount of money which was spend on to short (and unrealistic) jobs and they are going to monitor those errors for next months to figure out how this maybe can save them money. A notifiable quote was mentioned by one of the facility managers, she said *"We can probably save more money with investing in data analytics, than with fir-*

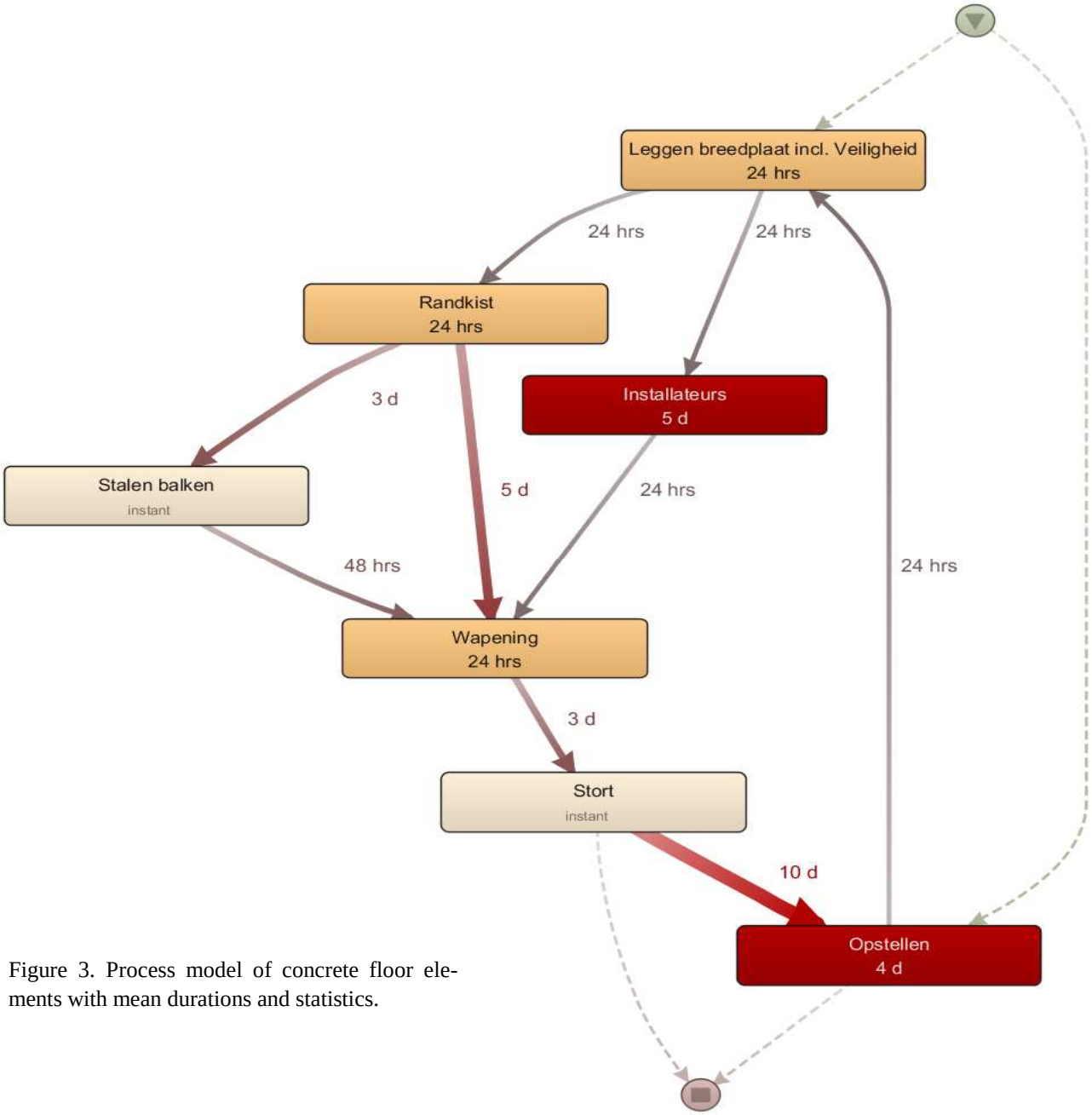

Figure 3. Process model of concrete floor elements with mean durations and statistics.

ing our own people".

The facility managers where definitely interested in using those kind of analytics more in the future. However this case study just explored the topic of facility management based process mining and has proven some useful applications. More (case-) studies should be elaborated to indicate the potential of this topic. In addition it would be useful to study the potential of integrating different data sources of other phases, for example the design -or construction-phase, in building elements lifecycle in order to gain insight in the process on a longer time span.

5 CONCLUSION

Companies in the AECFM do not always have clear sight on their processes (van Berlo et al, 2012). A lot of those processes are managed with IT systems. Especially in the plan- and maintain phase of buildings several tools are already applied. As a result of the experimental case studies within this research it can be concluded that some of these tools do not have functionality to give clear insight in overall processes. However the data which is in those tools can be extracted and analysed with process mining approaches.

Applying process mining can have different reasons. Organizational analytics, process variants, bottlenecks, planning deviation or conformance checking. Those applications are seen as valuable and have potential to help the construction industry understand and actually improve his processes.

It is seen that data preparation can be challenging. The used IT systems are not able to export clear event logs. This can cause a threshold to start with process mining experiments. However since event logs have a straightforward structure the researchers managed to transfer the data from the IT systems into event logs. In addition when one wants to do analytics on a more regular basis it is possible to automate the process of event log generation.

6 DISCUSSION AND FURTHER WORK

The research method that was used (experimental use cases) does not allow the conclusions to be applied to the whole industry. Additional research is needed to study the potential of integrating different data sources from several phases

More specific it is indicated that it can be useful to study the possibility of the integration of event log databases from different phases of construction projects. Meaning that event logs from a design phase, construction phase and operational phase could be merged in order to give unique insight in the lifecycle of a construction object. However, in contradiction to the plan and operate phase, event logs from the construction phase are difficult to find or generate (van Berlo & Natrop, 2015).

Interviews with domain experts revealed that several companies are interested in process mining services. When complex projects occur where the overall picture is difficult to manage, then process mining applications will be most promising. Since data is often available and existing algorithms are suitable such services are directly applicable (van Berlo & Bomhof, 2014), however companies have to be convinced of the power of those analytics. Therefore it is recommendable for further work to quantify the value of such analytics.

Event logs is a data structure which enables useful analytics, but storage of this data may be difficult. A possible solution as indicated in (van der Aalst, 2011) is one where companies can create a *process-oriented data warehouse* containing information about relevant events happened in the company. Possibilities for companies in the AECFM industry such as contractors can be studied as well.

Researchers which are interested in this subject can start with exploring the report and dataset (containing BIMs, Event logs, Point clouds and more) which is published along by van Schaijk (2016) and is freely downloadable.

REFERENCES

Beetz, Jakob, Berlo, LAHM van. "BIMserver. org–An open source IFC model server." Proceedings of the CIB W78 conference. 2010.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*. http://doi.org/10.1016/S0019-9958(67)91165-5

Nerode, A. (1958). Linear automaton transformations. *Proceedings of the American Mathematical Society*, *9*(4), 541–541. http://doi.org/10.1090/S0002-9939-1958-0135681-9

Petri, C. A. (1962). Kommunikation mit Automaten. *Fakultät Für Mathematik Und Physik*.

Quirijnen, R., & van Schaijk, S. (2013). *Meten = Weten*. Avans Hogeschool, Tilburg.

Sayer, N. J., & Williams, B. (2013). *Lean voor Dummies* (second edi). Amsterdam: Pearson Benelux bv.

Terlouw, L., & Mulder, H. (2014). Schatgraven in bouwprojecten. *Informatie*, *1408*, 32 – 29.

van Berlo, L. A. H. M., Dijkmans, T. J. A., Hendriks, H., Spekkink, D., & Pel, W. (2012). BIM quickscan: benchmark of BIM performance in the Netherlands.

Van Berlo, L. A. H. M., & Bomhof, F. (2014). Creating the Dutch national BIM levels of development. American Society of Civil Engineers (ASCE).

van Berlo, L. A. H. M., and Mathijs Natrop. BIM on the construction site: providing hidden information on task specific drawings. Department of Computer Science, 2015.

van der Aalst, W. M. P. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. *Media* (Vol. 136). http://doi.org/10.1007/978-3-642-

19345-3
van der Aalst, W. M. P., Van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., & Weijters, a. J. M. M. (2003). Workflow mining: A survey of issues and approaches. *Data and Knowledge Engineering*, *47*(2), 237–267. http://doi.org/10.1016/S0169-023X(03)00066-1
van Schaijk, S. (2016). *BIM based process mining*. Eindhoven university of technology. Retrieved from http://www.slideshare.net/StijnvanSchaijk/bim-based-process-mining-master-thesis-presentation