



<input checked="" type="checkbox"/>	Master's thesis
<input type="checkbox"/>	Licentiate's thesis
<input type="checkbox"/>	Doctor's thesis

Subject	International Master in Management of IT	Date	5/27/2017
Author(s)	Wei Zhong	Student number	736493
		Number of pages	80
Title	The Integration of Process Mining and Lean Six Sigma		
Supervisor(s)	Dr. E.A.M. Caron; Prof. A. Rutkowski; Prof. Hannu Salmela		

**Abstract**

Process mining is a technique that specifically aims at analyzing process data, which can substantially assist in business process management. On one hand, process mining is a relatively young disciplinary that not many enterprises practice it in real projects. Lean Six Sigma, on the other hand, is a methodology that is popular in improving process performance in various enterprises. However, the complication within the tools and techniques used in Lean Six Sigma greatly interferes with the efficiency and effectiveness of its practicality. Process mining, as an independent technique, is by nature a potential fit as the tools to support Lean Six Sigma. By integrating process mining and Lean Six Sigma, it provides a way to simplify the implementation of Lean Six Sigma project. As a result, enterprises would easily practice Lean Six Sigma in their process management project. Simplification is not the only advantage that process mining brings to the implementation. Process mining offers an automatic way to reorganize the large volume of process data, rendering it a clear structure for analysis. Defected processes can be more easily tracked than other process analysis techniques. This feature considerably enhances the efficiency.

This thesis aims at prove the concept of integrating process mining in the Lean Six Sigma and ideally shows the superiority of the integrated methodology over the regular one. A step-by-step explanation upon how the integration is proceeded is provided and then it is followed by a business case that examines the validity of this integration.

Key Words	Process Mining, Lean Six Sigma, process modeling, business process management
Further info.	



# **THE INTEGRATION OF PROCESS MINING AND LEAN SIX SIGMA**

**With a case study on the improvement of an accounting reporting process**

Master's Thesis  
in International Master in Information  
Technology Management (IMMIT)

Author (s):  
Wei Zhong

Supervisors:  
Dr. E.A.M. Caron  
Prof. A. Rutkowski  
Prof. Hannu Salmela

15.5.2017  
Tilburg

The originality of this thesis has been checked in accordance with the Tilburg University quality assurance system.

## Table of Contents

1	INTRODUCTION .....	9
1.1	Background and Motivation.....	9
1.2	The Study And Research Question .....	12
1.3	Research Relevance.....	13
1.3.1	Business Relevance.....	13
1.3.2	Scientific Relevance.....	13
1.4	Process Mining Project in Company ABC.....	14
1.5	Research Structure and Methodology .....	14
2	LITERATURE REVIEW .....	15
2.1	Lean thinking in finance.....	15
2.1.1	The major drivers of lean finance project .....	15
2.1.2	Lean Six Sigma in Finance .....	19
2.2	Process Mining.....	20
2.2.1	Event logs.....	21
2.2.2	Bridge between process mining and event logs .....	23
2.2.3	Discovery, Conformance and Enhancement.....	24
2.2.4	Process Mining Algorithm.....	24
2.2.5	Process Mining Software tools .....	25
2.2.6	Process Map Representation .....	25
2.2.7	Process Mining Implementation .....	26
3	INTEGRATE LEAN SIX SIGMA WITH PROCESS MINING .....	29
3.1	Define Phase.....	29
3.2	Measure Phase.....	32
3.3	Analysis Phase.....	35
3.4	Improve Phase .....	37
3.5	Control Phase .....	38
3.6	Validating the integrated model .....	39
4	CASE STUDY ON IMPLEMENTING PROCESS MINING IN LEAN SIX SIGMA PROJECT .....	42
4.1	Background .....	42
4.2	Define phase.....	44
4.3	Measure Phase.....	47
4.4	Analysis Phase.....	51
4.5	Improve Phase .....	53

4.6	Control Phase .....	54
4.7	Interview .....	54
5	LIMITATIONS OF THE INTEGRATED METHODOLOGY .....	55
5.1	Limitations on the project .....	55
5.2	Limitations on the event log.....	56
5.3	Limitations on the software tools .....	57
6	CONCLUSION AND DISCUSSION .....	58
7	RECOMMENDATIONS.....	60
7.1	Recommendations for business practice .....	60
7.2	Recommendations for further research .....	61
	REFERENCE.....	62
	APPENDIX.....	65
	Appendix 1 .....	65
	Appendix 2 .....	66
	Discovery .....	66
	Conformance .....	66
	Enhancement .....	66
	Appendix 3 .....	68
	Appendix 4 .....	70
	Appendix 5 .....	72
	Appendix 6 .....	74
	Appendix 7 .....	75
	Appendix 8 .....	77
	Appendix 9 .....	79

## List of figures

Figure 1	The normal BPM life-cycle analysis.....	11
Figure 2	Process Mining Analysis .....	11
Figure 3	Major obstacles in implementing lean finance.....	16
Figure 4	DMAIC Structure .....	20
Figure 5	Example of an event log .....	21
Figure 6	Hierarchical structure of an event log .....	22
Figure 7	Process Mining structure overview .....	23
Figure 8	CRISP-DM overview .....	27
Figure 9	Sample SIPOC table .....	30
Figure 10	Relation between SIPOC table and event log.....	31
Figure 11	Overview of three-level process map.....	32
Figure 12	Example of a cause and effect diagram.....	36
Figure 13	Example of a cause and effect matrix.....	36
Figure 14	Example of a control plan.....	38
Figure 15	Visual link between DMAIC and CRISP-DM.....	40
Figure 16	High level structure of MJV process .....	42
Figure 17	Project charter.....	45
Figure 18	SIPOC table for MJV posting.....	45
Figure 19	A fraction of event log.....	46
Figure 20	Sample event log .....	46
Figure 21	Level 1 process map .....	47
Figure 22	A fraction of level 2 process map.....	48

Figure 23	Statistical summary .....	49
Figure 24	Level 3 process map .....	50
Figure 25	Baseline process Analysis .....	51
Figure 26	Inefficient process pattern .....	51
Figure 27	Source of inefficient processes .....	52
Figure 28	Cause and effect diagram .....	52
Figure 29	Inefficient process mapping .....	53
Figure 30	Resource-centered process mapping .....	60
Figure 31	List of organizations .....	65
Figure 32	Petri net example .....	73
Figure 33	Six Sigma .....	77
Figure 34	Sigma to DPMO conversion.....	78



# 1 INTRODUCTION

## 1.1 Background and Motivation

*Lean thinking* is a business methodology that aims to provide a new way to think about how to organize human activities to deliver more benefits to society and value to individuals while eliminating waste.<sup>1</sup> The term *lean thinking* was coined by James P. Womack and Daniel T. Jones to capture the essence of their in-depth study of Toyota's famed *Toyota Production System*. Lean thinking is a new way of thinking any activities and seeing the waste inadvertently generated by the way the process is organized by focusing on the concepts of value, value streams, flow, pull, and perfection.<sup>2</sup>

Lean thinking is now widely implemented in different industries and in various functions.<sup>3</sup> In Appendix 1, readers will find a list of most representative organizations that successfully implemented Lean Finance project in recent 10 years. Its application in finance transformation is especially noticeable. PwC acknowledges that organizations are squeezing their finance and accounting functions like never before (Finance transformation: A Lean Approach to Increase Value, 2013). The driving force behind this trend lies with the accelerating speed of business transactions crossing time and geography.<sup>4</sup> However, the widespread of lean finance application does not yield the desired outcome quite often. In many organizations, the initiatives of moving to lean finance transformation starts on the production level and then extend its application concept to operational layers such as "procure to pay" and "order to cash." But lean thinking is not necessarily equally effective everywhere. Due to the different natures between finance transactions and production-level transactions, finance does not recognize the Lean thinking in the same way as manufacturing process does.<sup>5</sup> The consequence is predictable; application of lean finance is not as good as everyone expected, and often ends up with failure.

Obstacles that blocks Lean finance's way to success come from various sources. Most prominent one is from the change management. Either the employees are resistant to changes or the management behavior does not support changes. Problems with the

---

<sup>1</sup> Source: Womack, James P., Daniel, T. Jones (1996) *Lean Thinking*

<sup>2</sup> [https://en.wikipedia.org/wiki/Lean\\_thinking](https://en.wikipedia.org/wiki/Lean_thinking)

<sup>3</sup> Furterer, S. L. (2009). *Lean Six Sigma in Service : Applications and Case Studies*. Boca Raton: CRC Press.

<sup>4</sup> Rowland Hayler, Michael D. Nichols (2007), *Six Sigma for Financial Services*

<sup>5</sup> K Dillon - Hill Wiley (2015), *Lean in Finance, Journal of Corporate Accounting and Finance* 26(3):15 1044-8136

change management are mostly due to the inconsistent project goals as well as miscommunications in project team.<sup>6</sup> Various methodologies are then introduced to deal with those obstacles in lean project and the most accepted one is named Lean Six Sigma.

Lean Six Sigma is the combination of Six sigma and Lean Enterprise, aiming at improving quality, reducing variation, and eliminating waste in an organization.<sup>7</sup> A large portfolio of tools and techniques is included in this methodology, such as Input-Process-Output(IPO) diagram, Flow Diagram, CTQ(Critical to Quality) Tree and Project Charter. There is a drawback that is shared by those tools, however. Since those tools and techniques mostly originate from control charts and statistical process control <sup>8</sup>, they are based on data-centered analysis. Due to which, the process model has to be disassembled first and then information and analysis take place upon the respective disassembled components of the process. After browsing through dozens of case studies in this field, I found that this sort of delineated methodology creates a great number of trouble in terms of collecting and analyzing data. Process mining technique, on the other hand, is process oriented and prioritize the process stream over process components at analytical level. <sup>9</sup> Statistical analysis in process mining is always practiced upon the circulating process stream instead of process components. For instance, when applying process mining on a transaction logging system, the information regarding to each step of the logging process is collected but the analysis is not taken out yet. Instead, those information will be directly assembled to formulate a process stream for the whole logging transaction and then analysis is practiced upon the whole process stream instead of directly upon the data regarding to those separated steps. This feature perfectly fit in the concept of Lean thinking which constantly emphasize the continuous process streams. In data oriented analysis, data regarding to each component of process is first collected and analyzed respectively. Figure 1 illustrates what such a data oriented analysis looks like in a typical Business Process Management (BPM) life-cycle. Data and model is positioned in separated analysis phase. Process models play a dominant role in the (re)design and configuration/implementation phases, whereas data plays a dominant role in enactment/monitoring and diagnosis/requirements phases.

---

<sup>6</sup> Mckinsey Quarterly Transformation Executive Survey, 2008. Next Generation PMO KIP Team

<sup>7</sup> Taylor, Gerald. *Lean Six Sigma Service Excellence H/C*, edited by Gerald Taylor, J. Ross Publishing Inc., 2008.

<sup>8</sup> "Xerox cuts popular lean six sigma program". *Democrat and chronicle*. Retrieved March 10, 2015.

<sup>9</sup> Wil M. P. Van Der Aalst (2011), *Process Mining: Discovery, Conformance and Enhancement of Business Processes*.

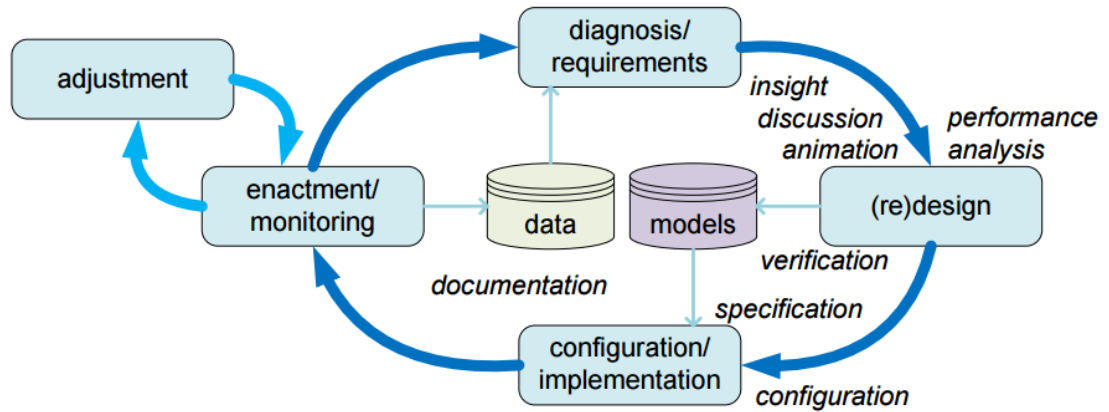


Figure 1 The normal BPM life-cycle analysis.<sup>10</sup>

As a comparison, in process oriented analysis, information related to the process components is collected and then directly placed and combined in the process stream via process mining algorithm. After which, analysis and evaluation take place on the whole process stream. As Figure 2 indicates, unlike BPM-life cycle analysis in which modeling derives from man-made summaries of the data, modeling in process mining is directly driven by data. This sort of modeling process ensures the consistent alignment between data and models. Therefore, model design, configuration, monitoring and diagnosis can all be based on the same foundation, models generated directly from data. Such attribute can potentially accelerate and calibrate the process management in lean project and make process mining a suitable and powerful tool for implementing Lean Six Sigma project in lean finance.

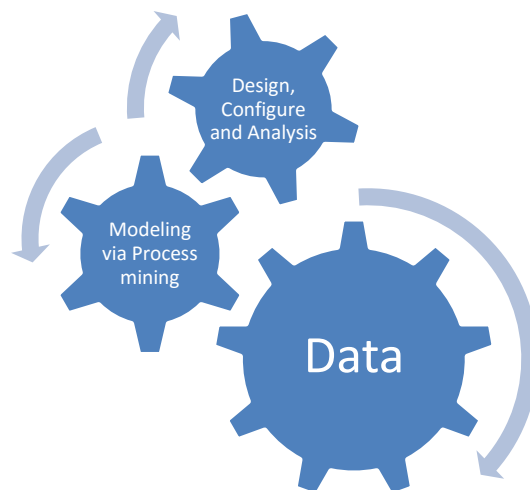


Figure 2 Process Mining Analysis<sup>11</sup>

<sup>10</sup> Source: [processmining.org/\\_media/processminingbook/process\\_mining\\_chapter\\_01\\_introduction.pdf](http://processmining.org/_media/processminingbook/process_mining_chapter_01_introduction.pdf)

<sup>11</sup> The principle idea of this figure is that data is the driven force of process modeling.

## 1.2 The Study And Research Question

This thesis seeks for an effective way to integrate process mining techniques in the conduction of business process improvement project. Considering the regular practice of continuous business process improvement in nowadays business, various methodologies are created to guide through those improvements. Lean Six Sigma is one of most common methodologies that are being used by quite a few enterprises. Despite of the widespread of Lean Six Sigma methodology, enterprises still encounter confusion and difficulties in implementation due to its dazzling amount of concepts and tools within this methodology.<sup>12</sup> This thesis aims to indicate the natural synergy between Lean Six Sigma and process mining and how process mining can replace the majority of tools and techniques within Lean Six Sigma and simplify its implementation. Hopefully at the end of this thesis, the practicality of integrating process mining in Lean Six Sigma can be validated and proven to be superior over the regular implementation of Lean Six Sigma.

The main research question is:

*“How to integrate process mining technique and the Lean Six Sigma methodology?”*

To answer this question, there are several sub-questions that need to be answered:

- What is process mining?
- What is Lean Six Sigma?
- How to integrate process mining to Lean Six Sigma?
- How to validate the quality of this new integrated model? Does it do better compared to Lean Six Sigma project without process mining?
- How to apply the integrated model on a Lean Finance project?
- What are the limitations on employing process mining in the project?

In this study, a business case is provided to validate the feasible integration of process mining in Lean Six Sigma. This business case is initiated during the writing of this thesis. Since a normal lean project on average takes more than two years to invoke observable effects on performance improvement, it is impossible to verify this integration simply by evaluating the outcome of this business case during the writing of this thesis. Instead, in this thesis the integration is validated by interviewing the practitioners<sup>13</sup> of this business case, asking them about the pros and cons of the integration. Until now this is the only case that process mining technique is integrated in Lean Six Sigma in practice. Statistically speaking, even if this case proves to be positive regarding to the integration, it can-

---

<sup>12</sup> Pepper, M. P. J., & Spedding, T. A. (2010). The evolution of lean Six Sigma. *International Journal of Quality & Reliability Management*, 27(2), 138-155.

<sup>13</sup> In this whole thesis, practitioners would be referred to the people who conduct and execute the Lean Six Sigma project.

not conclude the success of integration. But the purpose of this thesis is to create a framework for business to implement Lean Six Sigma via process mining techniques. To validate this framework, more follow-up studies and business cases are needed.

## **1.3 Research Relevance**

### ***1.3.1 Business Relevance***

Process mining as a young interdisciplinary between data mining and machine learning is not widely employed in business practice, because it is not a well-known subject in all disciplines as well as its lack of business case support. However, this does not distort the fact that process mining is an advanced and mature technique that can potentially smoothen the Business Process Management and assist to continuously improve the business processes. This thesis utilize Lean Six Sigma, which is a mainstream methodology to improve business processes, as an example to illustrate how the integration of process mining technique can potentially save a considerable amount of efforts and time in data collection and data analysis, since data collection and data analysis takes up the majority time of practicing process improvement.

This thesis aims to prove the effectiveness of process mining via a real business case and provide a guideline for business to employ process mining technique.

### ***1.3.2 Scientific Relevance***

Since the term ‘process mining’ is identified as a discipline by Van Der Aalst(2003)<sup>14</sup>, it is in most cases studied as an independent methodology that assists to manage process improvement. However, process mining in fact is a great tool that can be utilized in other more well-known and well-accepted methodologies instead of an independent methodology. By relating the process mining technique to other well-known methodology such as Lean Six Sigma, it will make business practice with process mining much easier and comprehensive. This thesis specifically add a business context to the utilization of process mining and produce a link between process mining and lean six sigma methodology. The illustration in this thesis hopefully can further enhance the effectiveness and usability of process mining theory.

---

<sup>14</sup> W.M.P. van der Aalst, J. Desel, and A.H.M. ter Hofstede, B. Kiepuszewski, and A.P. Barros. Workflow Patterns. *Distributed and Parallel Databases*, 14(1):5 – 51, 2003.

## **1.4 Process Mining Project in Company ABC**

The research was performed as part of my internship at Company ABC. The business case derives from a proof-of-concept research within Company ABC examining the feasibility of utilizing process mining technique in lean finance project. Process mining is only a tool to assist the general conduct of lean project. Therefore, process mining can only strengthen the performance of lean project but not directing the implementation of lean project. That is why Lean Six Sigma is selected as the methodology to conduct the project and then process mining is studied as a potential tool to assist Lean Six Sigma implementation. Lean Six Sigma is not the only methodology for Company ABC in lean finance project. But regardless of the methodologies, conceptual foundations for lean project are common. In other words, if it is proven feasible to integrate process mining in Lean Six Sigma, Company ABC can dynamically expand the practice of process mining to other methodologies as well.

## **1.5 Research Structure and Methodology**

Thesis will start by explaining all the concepts and terminologies that are related to process mining technique and Lean Six Sigma. However, not all the details are included. Both process mining and Lean Six Sigma have abstract interface that only provide users the logical functions, hiding the implementation details behind. Users only need to understand those logical structures to practice both techniques. Therefore, only those concepts relevant to those logical structures would be explained in detail. As for the implementation details, if readers are interested, they are also briefly explained in appendixes.

Following the concept explanation, synergies between process mining and Lean Six Sigma are studied through scientific induction. The indicated synergies would be validated by a business case conducted in Company ABC.

This is going to be a qualitative research, and information will be collected from both interview and existing database. Qualitative dissertations will include descriptive material that extracts from interviews, conversations, and documents. There are also some quantitative data that will be collected from database to support process mining analysis. Hence, some quantitative analysis will take place upon those data. The result of those quantitative analysis, however, will contribute to the overall qualitative analysis against the effectiveness of process mining practice in lean six sigma methodology.

## **2 LITERATURE REVIEW**

### **2.1 Lean thinking in finance**

#### ***2.1.1 The major drivers of lean finance project***

To fully understand the challenge in implementing Lean in finance, it is essential to first comprehend the role of finance service and the driving force of lean finance. Financial services can be considered in two terms, internal financial service and external financial service. They are very similar in terms of the nature of services. The only difference is the competitive environment they are facing. For external financial services, it is an open service provided to the other organizations, in which sense the service's end user is the service provider's customers. As a comparison, the internal financial service is bonded specifically to only one organization and it is either provided by this organization itself or by a long-term contractor. In other words, the end user of internal financial services is only one organization.

Regardless of who the end user is, both internal and external financial services are pursuing perfection and timeliness. But a considerable amount of problems appear to hinder their pursuit. Figure 3 summarized major problems that cost the efficiency and effectiveness of finance functionality. In general, finance processes all contain three major activities, Management & Reporting, Cost Optimization and Operational Finance. Each activity has its own unique cause for inefficiency and ineffectiveness, tearing down the overall performance of finance processes.

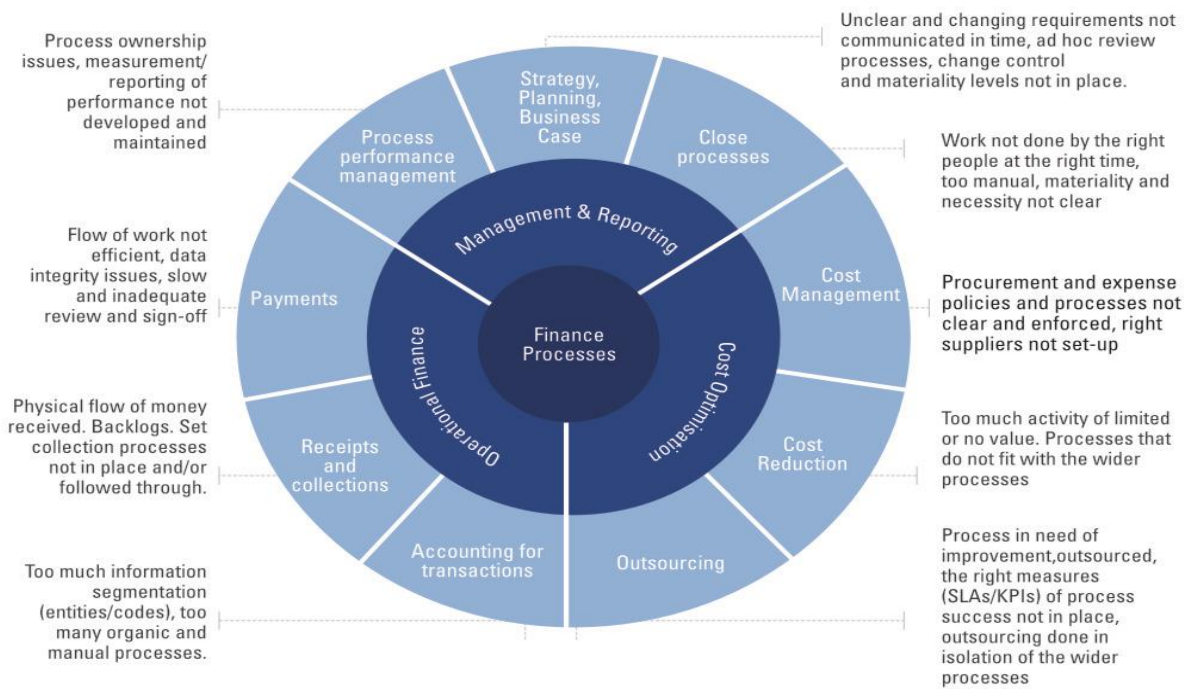


Figure 3 Major obstacles in implementing lean finance<sup>15</sup>

Along the development of both internal and external financial services both services are facing practically the same problems, such as low efficiencies, high defection rate and repeated iterations. As a popular encounter to those problems, lean thinking was introduced in financial service in the late 90s. Now it has become an emerging trend to apply lean concept in financial service industry.

There are mainly 6 drivers that popularize lean thinking as a solution. Since these 6 drivers are either the requirements from the stakeholders or requirements from the executions, it is more understandable to illustrate them in two categories, stakeholder-related and execution-related.

### Stakeholder-related drivers

In a typical financial service, there are three key stakeholders, end users, shareholders, and regulators. This section will start with the end users, which leads to the first key driver, “The Need to Continually Satisfy End Users”.<sup>16</sup> On one hand, in this industry end users’ requirements change rapidly. On the other hands, finance processes are mostly in

<sup>15</sup> Source: <https://home.kpmg.com/in/en/home/services/advisory/management-consulting/financial-management/lean-finance-enhancement.html>

<sup>16</sup> Rowland Hayler, Michael D. Nichols (2007), *Six Sigma for Financial Services*



stagnant patterns.<sup>17</sup> In addition, due to the widespread of IT implementation in finance, those patterns are integrated in the finance system, further staple the way financial service is provided. For the same reason, financial services implicitly moved to IT-centered instead of customer-centered.<sup>18</sup> Eventually most financial services that utilize IT systems lose sight of the changing client demands and as a result, end users start to suffer unpleasant experience, which pushes the organization to resort to lean finance for improvement. The second driver is “The Need to Create Sustainable Shareholder Value”. From shareholder’s perspective, financial service ought to contribute to the final revenue but unfortunately it is often not the case. Kevin Dilton <sup>19</sup> described finance as NNVA (Necessary Non-value Added) in one of his articles. His description is a clear reflection of the nature of finance. It is an indispensable function to sustain business but neither does finance really add any consumable value to company’s end products or services. To those NNVA, it is commonly realized that the resource they consume should be minimized.

The central idea of lean finance is to identify and eliminate waste and redistribute or save the resource that is spent on those waste. However, identifying defects and waste in a service and transactional environment is often very difficult and poses a significant challenge in quantifying and targeting improvement efforts. Here are the key reasons why it is the case<sup>20</sup>:

1. *The output is less tangible.* The output of financial processes are often information, money, or an experience.
2. *The process is less visible.* Since financial processes are working with information which sometimes does not take any physical forms (only in electronic formats). When finance people handles data, they typically have very little idea of where the transaction subsequently goes and what happens to it next.
3. *The distance between the root cause of the defect and its place of discovery is often very great.* For instance, some erroneous accounts might have already been created for several months before an error message is presented by the system. This kind of delayed error message imposes a great difficulty in spotting the real cause of the mistake.
4. *Performance data are not so readily available.* Service processes are commonly not well measured, and available data are often tied up in IT systems, from where it is generally challenging to distill meaningful operation information.

---

<sup>17</sup> K Dilton - Hill Wiley (2015), *Lean in Finance*, *Journal of Corporate Accounting and Finance* 26(3):15 1044-8136

<sup>18</sup> This is purely my own observations based on countless case studies I read, as well as my own experience

<sup>19</sup> Dilton- Hill, Kevin. "Lean in Finance." *Journal of Corporate Accounting & Finance* 26.3 (2015): 15-25.

Lean finance provides a few tools and techniques that can either solve those obstacles or work around them.

The third driver comes from regulators, more specifically, the need to comply with regulatory requirements. Finance is strictly regulated ever since its creation. After IT started to play vital roles in finance, intensive regulatory requirements such as Sarbanes Oxley (SOX), Gramm-Leach-Bliley and Basel II imposed very detailed control mechanism in terms of process management. As a result, organizations must document information about their internal controls in order for internal and external auditors to be able to validate management's statements of compliance effectiveness. This drives organizations to practice a business management approach that enables them to understand the transaction trail as a process and as the basis for establishing appropriate controls at appropriate point.

### **Execution-related drivers**

Besides the driving force from the stakeholders, the substantially enlarged scale of finance also drives organization to look for improvement in the finance processes<sup>21</sup>, leading to the forth driver "The need to address scale and complexity". As digital technology brings more variables and unpredictable elements to the competitive business environments, organizations, especially multinational enterprises, have to process a tremendous amount of transactions in a fast-pace manner, which puts a great pressure on the finance department. The amount and variety of those transactions create such a complexity that finance is always questioned whether they are doing their jobs in an appropriate way.

The fifth driver is the Need to Operate across an Extended Enterprise. This driver emerges from the popular trend of Business Process Outsourcing (BPO) in recent years. Offshoring the whole or part of finance department has become increasingly attractive to organizations with the cost advantages. Even though outsourcing finance releases organizations from the workload, it does not necessarily mean there are improvements in finance functionality. Without proper synchronization between the organization and outsourcing party, neither cost reduction nor effectiveness will be realized. Under such condition, organization needs an approach to continuously coordinate and align their finance demands with the finance service they obtain from the third party.

The sixth driver is the Need to Leverage Enabling and Accelerating Technology. Most people consider the digital technology as the motivation behind the current change in finance. However, it is not completely true. Technology is simply a tool that empower and accelerate finance processes. A flexible finance processes that can best combine and integrate technology in a continuous manner will grant organizations enormous advantage over their rivalries.

---

<sup>21</sup> K Dilton - Hill Wiley (2015), *Lean in Finance, Journal of Corporate Accounting and Finance* 26(3):15  
1044-8136

Understanding those drivers are extremely important to the finance project. After all, practitioners need to keep in mind why the lean finance project is needed and conducted. The next section would temporarily leave the topic of lean finance and focus on the concepts of process mining methodology. It is going to be the tools that accelerate the achievement of objectives set by those introduced drivers.

### **2.1.2 *Lean Six Sigma in Finance***

Lean Six Sigma appears to be a popular approach to conduct process-improving project since the late 90s. Since then, countless applications in different industries consistently proved its effectiveness in improving quality, reducing variation, and eliminating waste in the organization.<sup>22</sup> Lean Six Sigma is actually the combination of two improvement programs: Six Sigma and Lean Enterprise.<sup>23</sup> The former is a quality management philosophy and methodology that focuses on reducing variation; measuring defects (per million output/opportunities); and improving the quality of products, processes, and services. The term “Lean Enterprise” is a broaden version of lean thinking that expands scope of lean concept from merely manufacturing to entire enterprise.

The adoption of Six Sigma and Lean Enterprise do not necessarily take place in conjunction. Some companies prefer to implementing these two programs separately, which is perfectly fine. However, considering the large overlapping and complementary functionalities of these two, increasing amount of companies decide to combine both concepts together, as Lean Six Sigma program.

The early implementer of Lean Six Sigma is an aircraft-engine-controls firm called BAE System Controls. After they started the project, within 5 years they improved their productivities by 97% and customer lead time by 90%. The success of early adoption inspired numerous followers. Since then the methodology to adopt Lean Six Sigma gradually became mature.

The major methodology in Lean Six Sigma is called DMAIC (Define, Measure, Analysis, Improve and Control phase). Each of DMAIC phases are well defined and standardized, but steps carried out in each phase can vary based on the reference used. Since this thesis is dealing with lean finance, the setup of DMAIC phases will be specific to lean finance field. The table below shows the general structure of this methodology.

---

<sup>22</sup> Furterer, S. L. (2009). *Lean Six Sigma in Service : Applications and Case Studies*. Boca Raton: CRC Press.

<sup>23</sup> George, M. L., & George, M. (2003). *Lean six sigma for service* (p. 273). New York, NY: McGraw-Hill.

<b>Define</b>	<b>Measure</b>	<b>Analysis</b>	<b>Improve</b>	<b>Control</b>
1. Develop project charter	6. Define the current process	11. Develop cause and effect relationships	14. Identify breakthrough & select solutions	20. Measure results & manage change
2. Identify stakeholders	7. Define detailed VOC	12. Determine and validate root causes	15. Perform cost/benefit analysis	21. Report scorecard data & create process control plan
3. Perform initial VOC and identify CTS	8. Define the VOP and current performance	13. Develop process capability	16. Design future state	22. Apply P-D-C-A process.
4. Select team and launch the project	9. Validate Measurement System		17. Establish performance targets, project scorecard	23. Identify replication opportunities
5. Create project plan	10. Define COPQ and Cost/Benefit		18. Gain approval to implement and implement 19. Train and execute	24. Develop future plans

Figure 4 DMAIC Structure<sup>24</sup>

In Figure 4, there are some acronym, like VOC (Voice of Customer). CTS (Critical to Satisfactory) and COPQ (Cost of Poor Quality). These concepts would be explained in detail in the next chapter since they are very closely related to the mechanism that assists to integrate process mining and Lean Six Sigma.

## 2.2 Process Mining

Process mining has been developed for only a decade and it is considered an interdisciplinary combining machine learning, data mining, and process modeling analysis. Pro-

---

<sup>24</sup> Furterer, S. L. (2009). *Lean Six Sigma in Service : Applications and Case Studies*. Boca Raton: CRC Press.

cess mining aims at discovering, monitoring and improving real processes (i.e. not assumed processes) by extracting knowledge from event logs readily available in today's IT systems.

### 2.2.1 Event logs

Before introducing the details about analysis in process mining it is vital to understand the data that is analyzed in process mining, namely event logs. Unlike normal data mining, process mining analysis is always based on a specific form of data which is basically a tracking log that records the transactions in reality. As the digital universe becomes increasingly aligned with the physical world. Nowadays most information systems implemented in business are able to log enormous amounts of events. Classical WFM systems (e.g. Staffware and COSA), BM systems (e.g. Filenet, Global 360 and Teamwork by Lombardi Software), ERP systems (e.g. SAO Business Suite, Oracle E-Business Suite, and Microsoft DynamicsNAV), CRM systems (e.g. Salesforce), middleware (e.g. IBM's WebSphere), and hospital information systems (e.g. Sieman Soarian) all have the ability to closely track all the activities that are taken out by operations in the systems. Those tracking information is stored in a dataset which is called event log. The Figure 5 below is an example of a typical event log from a financial system.

RequestKey	Status	Action	ActionStartTime	ActionEndTime
REQ201612005066	Acknowledged and Closed	Request Submitted for Automated Posting	12/8/2016 8:45	12/8/2016 8:45
REQ201612005066	Acknowledged and Closed	Request submitted and approved by Requester	12/8/2016 8:45	12/8/2016 8:46
REQ201612005066	Acknowledged and Closed	Validation Failed, because a error verb was found in the log re	12/8/2016 8:46	12/8/2016 8:46
REQ201612005066	Acknowledged and Closed	Request Automated Posting Failed	12/8/2016 8:46	12/8/2016 8:47
REQ201612005066	Acknowledged and Closed	Request Submitted for Automated Posting	12/8/2016 8:47	12/8/2016 8:47
REQ201612005066	Acknowledged and Closed	Request submitted and approved by Requester	12/8/2016 8:47	12/8/2016 8:47
REQ201612005066	Acknowledged and Closed	Posting-request succeeded with SAP DocumentNr: 100026470	12/8/2016 8:47	12/8/2016 8:47
REQ201612005066	Acknowledged and Closed	Request Automated Posting Succeeded	12/8/2016 8:47	12/8/2016 8:47
REQ201612005066	Acknowledged and Closed	Request Closed	12/8/2016 8:47	#NULL!

Figure 5 Example of an event log

Event logs are constituted by several elements and those elements form a hierarchical structure, as the Figure 6 indicates.

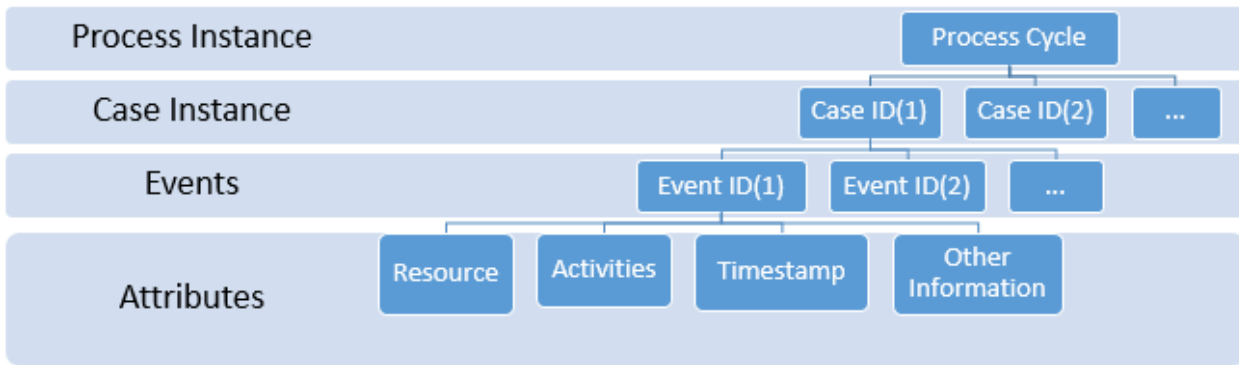


Figure 6 Hierarchical structure of an event log

Not all event logs necessarily have the same format as the above example shows. However, the hierarchical structure as well as the elements contained within this structure must be defined in the event log. A process instance is a-priori information that is set by the functionality of the system. An example is a transaction approval system, for which the process instance is the procedures that are taken to approve or disapprove various transactions. A process instance is simply an abstract statement of the system's objectives. A case instance represents the basic unit of a process cycle. In the previous example of the transaction approval system, those various transactions are considered case instance. The event log must have a column that contains a unique identifier for each transaction. In Figure 5, "RequestKey" can be used as Case ID since every single transaction in that event log all has its very own "RequestKey". Events are equivalent to all the activities that are taken out within a single transaction. Still using that transaction approval system as an example, events are the steps that are taken to either approve or disapprove the transactions. Each event does not necessarily have event ID in the event log, but each event contains a vector of attributes such as activity (i.e. the actions that are implemented upon the transactions), resource (i.e. the subjects who take out the activity), timestamp (i.e. the timeline that the activity is taking place). Hence, event is easily identifiable without an event ID. In most cases, the event log should at least contain unique case ID for each case instance, resource, activities and timestamp to derive valuable analysis via process mining. Event log might include some other information on the "Attribute" level of the hierarchy structure basing on the system functions. Those additional information might be useful depending on the purpose of process mining. But the idea is that all the data in the event log must be clarified and placed in that hierarchy structure. A well-defined event log is indispensable for the follow-up analysis.

### 2.2.2 Bridge between process mining and event logs

Process mining creates a statistical way to connect the real world information captured in information system's event logs to data analysis for performance management. Unlike normal process analysis steps which are mostly based on the presumed process model, process mining directly extracts and summarize the process model from event logs that keep tracking what happen in reality. Figure 7 below comprehensively represent such structure. All mature financial systems in business can directly provide such event logs. However, considering the fact that financial transactions are very fragmented in terms of process streams, e.g., event data is scattered over many tables or needs to be tapped off from subsystems exchanging messages. In such cases, even though event data does contain all the information, effort is needed to assemble and summarize the data. Data extraction is an integral part of any process mining effort. As this Figure shows, once the transactions in real world are recorded in the software system's event log, three types of process mining can be practiced.

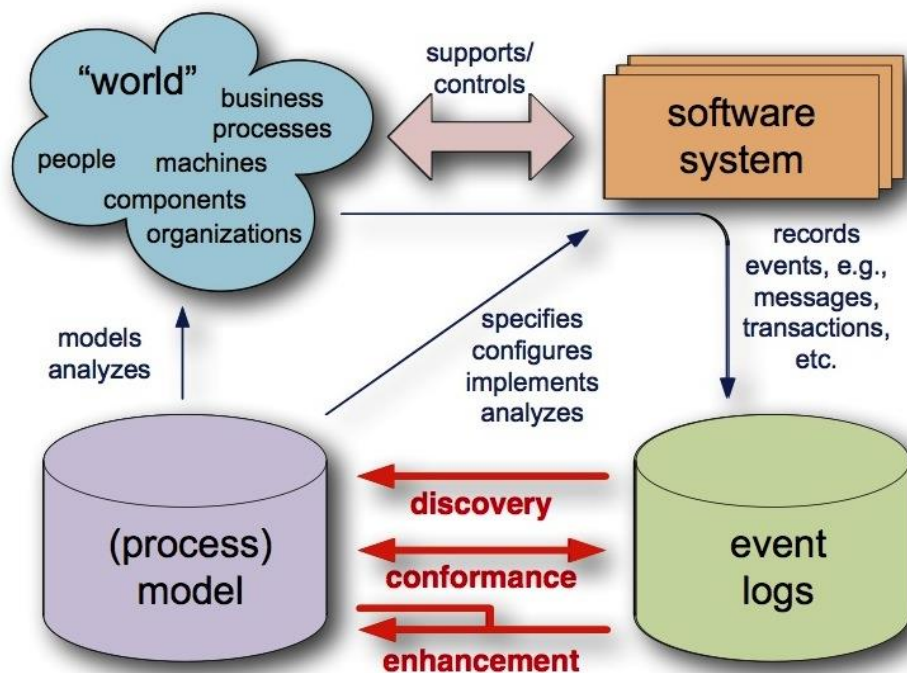


Figure 7 Process Mining structure overview<sup>25</sup>

<sup>25</sup> Wil M. P. Van Der Aalst (2011), *Process Mining: Discovery, Conformance and Enhancement of Business Processes*.

### 2.2.3 *Discovery, Conformance and Enhancement*

There are three types of process mining that sums up all the objectives of process mining analysis. The detail definition of these three types is provided in Appendix 2. Briefly speaking, discovery type of process mining aims at exploring the patterns of process streams among all the events that are captured in the event log. It takes the existing event logs as input and produce a complete process map via various algorithms. Conformance is applied when an organization needs to check the alignment between the pre-defined process model and the reality recorded in the event log. In this case, both the predefined process model and event log are taken as input for process mining and an analysis report is generated with regards to whether real operations comply to the design of the process. Enhancement is implemented to extend or improve an existing process model using information about the actual process recorded in some event log. This might sound very similar to conformance but there is a subtle difference. In conformance, the existing event log is analyzed to check the alignment between reality and process model but in enhancement, existing event logs are used to explore the spots that can potentially be improved or repaired in the predefined process model.

### 2.2.4 *Process Mining Algorithm*

The Apriori Algorithm<sup>26</sup> is one of the best known algorithms in computer science, that also forms the fundamental elements of its implementation in the process mining context. The apriori algorithm has the capability to accelerate the specification of association rules by exploiting the following two observations<sup>27</sup>:

1. If an item-set is frequent (i.e., an item-set with a support above the threshold), then all of its non-empty item-sets  $X, Y$ : if  $Y \subseteq X$  and  $N_X/N \geq \text{minsup}$ <sup>28</sup>, then  $N_Y/N \geq \text{minsup}$ .

---

<sup>26</sup> DaWaK (Conference) Vienna, Austria). (2012). *Data warehousing and knowledge discovery : 14th international conference, daWaK 2012, vienna, austria, september 3-6, 2012. proceedings* (Lecture notes in computer science, 7448). Berlin: Springer.

<sup>27</sup> Perego, R., Orlando, S., & Palmerini, P. (2001, September). Enhancing the apriori algorithm for frequent set counting. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 71-82). Springer Berlin Heidelberg.

<sup>28</sup> *Minisup* is a data mining concept, standing for minimal support. This is a parameter which indicates the minimum frequency ratio between a specific item set and all possible item sets that can define this specific item set as an association rule.



2. If, for any  $k$ ,  $I_k$  is the set of all frequent item-sets with cardinality  $k$  and  $I_l = \emptyset$  for some  $l$ , then  $I_k = \emptyset$  for all  $k \geq l$ .

The detail explanation of implementing this algorithm in process mining is given in Appendix 3. Since this is not the focus of this thesis, it is not necessary to spend time and words in explaining the detail of algorithm here. But understanding the basics of those algorithms can substantially help the practitioners to correctly implement process mining techniques. In general, it is crucial to comprehend one concept,  $\alpha$ -algorithm which can be seen as the modified version of Apriori Algorithm explained previously. (In fact,  $\alpha$ -algorithm is derived from Apriori Algorithm) There are 8 steps in  $\alpha$ -algorithm that can take a stream of activities captured by event logs and discover the pattern of those activities, through which a process map can be produced. This process map is able to show the concurrency and order of the activities that are taken to achieve the final objective. The detail of those 8 steps are also included in Appendix 3.

### **2.2.5 Process Mining Software tools**

The above introduction of algorithms might be a bit confusing to people with no data science background. However, this does not prevent people to utilize the process mining technique to solve problems. Nowadays there is a variety of software that have built-in algorithm to automatically generate process mining analysis results. Those software, even though might still run with some minor glitches, can fulfil most of the job. The two most famous ones are Disco and ProM.

Basically the practitioners of process mining merely need to cleanse and preprocessed the existing event log into the correct format and then import the event log in the software. With a few clicks, the software would calculate and generate the process map via the built-in algorithm. These two tools both have their own pros and cons. The detail comparison of these two tools are made in Appendix 4.

For the majority of this thesis, Disco is the tool to be used, due to the fact that the functions provided by Disco are sufficient to accomplish most of the objectives.

### **2.2.6 Process Map Representation**

Once the process model is generated via process mining algorithms, a proper notation of the process map is needed to express the final result. Notation is extremely important in terms of modeling and communication. An appropriate notation should enable practitioners to structure discussions and document procedures. Most importantly, it must include the complete results of the process mining analysis. Currently there are various process

modeling notations being practiced. Sometimes the plethora of notations is referred to as the new “tower of Babel”. Despite of the variety of notations, most of them can be easily converted to each other. For example, ProM is able to exploit  $\alpha$ -algorithm to generate Petri Net models and Petri Net can also be easily converted to other notations such as BPMN model, BPEL model and UML Activity Diagram. Hence, it does not really matter which kind of notations the practitioners implement in process mining as long as one uniform notation is consistently used and clarified to all the stakeholders throughout the whole process mining project.

In the practice of this thesis, Petri Net would be the only choice for modeling notation, since Petri Net is very well known for its simplicity and representativeness. The detail explanation of Petri Net is given in Appendix 5.

In real practice, practitioners should choose whichever is more understandable to them. In the end, the purpose of notation is always communication and documentation.

### **2.2.7 Process Mining Implementation**

The implementation of process mining in project follows the same procedures for data mining analysis, which is the well-known CRISP-DM (Cross Industry Standard Process for Data Mining)<sup>29</sup>. However, implementing process mining with CRISP-DM does not change the fact that CRISP-DM is still a data centered framework. It concentrates on the short term performance revealed by data instead of long term improvement via continuous process optimization. CRISP-DM is constantly based on the statistical attributes of the data rather than the process stream that generates those data. Therefore, implementing process mining via CRISP-DM cannot take advantage of all process mining’s features. Hence, Lean Six Sigma becomes the choice as the methodology to implement process mining in this thesis. But it is still worthwhile to take a look at CRISP-DM. As the Lean Six Sigma is going to be introduced in next section, readers will find significant synergies between CRISP-DM and Lean Six Sigma.

---

<sup>29</sup> Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.

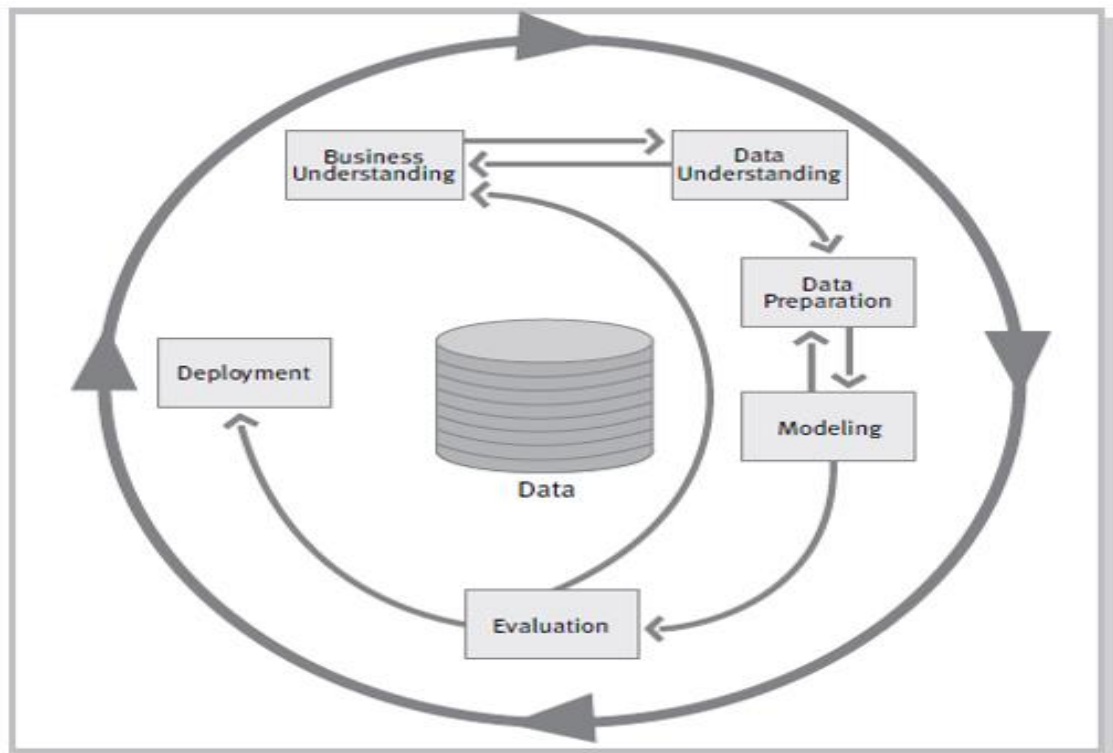


Figure 8 CRISP-DM overview<sup>30</sup>

The initial data collection stage involves business understanding, data understanding and data preparation.

1. Business understanding: Basically it is setting the objective for process mining, determining what eventually this process mining project wants to achieve as well as the evaluation standard for success of the project. This phase would directly instruct the next one,
2. Understanding data: it intends to determine what kinds of data should be included in analysis and how to interpret the data from an objective-oriented point of view. Data understanding can also contribute some valuable insights for business understanding. In many practices, initial raw data understanding can help to further extract the scope of the project.
3. Data preparation: it is primarily about data collection and data construction. The relevancy of raw data is further refined in this stage and some pre-processing of data takes place to make it ready for modeling in next stage. One noticeable remark here is that in real practice on average appropriately more than half of the time is spent on the first three phases. In other words, unlike most of people think

<sup>30</sup> Source: Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29-39).

modeling, evaluation and deploy actually take up less project timeslots than the initial three phases.

4. Modeling: this is where process mining algorithms step in. Model is generated and assessed in this phase.
5. Evaluation: this phase is conducted on the model as well as its analysis result. Some success criteria is built to validate and approve the model. It is possible that the model and analysis turn out to be not aligned with the initial business understanding. If that happens, all those processes have to be iterated to locate the cause of this un-alignment. If the model survives through the evaluation stage, a deployment plan is then constructed.
6. Deployment: Basing on the types of process mining, Discovery, Conformance, and Extend, different factors will be attributed to the deployment plans. Monitoring and maintenance plans are also essential parts of deployment stage.

The above introduction is just a very brief browse through CRISP-DM. By no means it covers all the details of this framework. However, like previously mentioned at the beginning of this section, Lean Six Sigma is the framework that is going to be used to implement process mining in this thesis. A brief introduction of CRISP-DM will be benevolent to reveal the synergy between these two frameworks.

### 3 INTEGRATE LEAN SIX SIGMA WITH PROCESS MINING

Lean Six Sigma is a project management methodology which provides context to the implementation of tools such as process mining. In general, Lean Six Sigma is composed by 6 phases, Define Phase, Measure Phase, Analysis Phase, Improve Phase, and Control Phase. Due to the process-aware nature of process mining technique, each phase of Lean Six Sigma all have the potential to integrate process mining tools to enhance the efficiency and effectiveness of project management in Lean, since all six phases of Lean Six Sigma is process-centered. The following sections would explain each phase as well as the integration of process mining in order.

#### 3.1 Define Phase

The Define phase is where the scope of the project charter is developed. The purpose is to delineate the business problem and scope of the project and the process to be improved. This phase corresponds to the “Understanding Business” phase in CRISP-DM. The objectives for process mining analysis are set and an approximate standard for success criteria is also defined.

A project charter will be produced in this phase. Like any other project preparation, a clear *project overview* including *project background* and *basic assumptions* related to the project should be present. Most importantly, a clearly defined *problem statement* is put in place, describing how this finance process is related to business and what the challenge or the problem faced by the business is. The *problem statement* need to specify the process that is affected and define the *measurable* impact of the problem accordingly. Then the clients who are the *end users* of the finance process should be identified as well as all relevant *stakeholders*. For each kind of stakeholders *CTS* (Critical to Satisfactory) can be identified by what is critical to quality (defects), delivery (time), and cost. Based on defined CTS, a *project goal* can be set. This goal does not have to be quantifiable yet. It should serve as a guideline to direct where the lean finance project is going.

Lastly, a *scope statement* should strictly address the boundaries of the project. There are two major kinds of scope, temporal scope which defines the timing of the process-data collection activities, and delivery scope, which specifies what should be delivered from the project, such as improvement recommendations and the implementation plan. Due to the limitation of the scope of this thesis, the delivery scope will be always set as improvement recommendations. In other words, the lean finance project illustrated in this thesis aims at providing a structural guideline to improve the existing process.

There are other aspects that are usually in the Define phase such as benefits, potential risks and project resources. However, the scope of this thesis is limited to the implementation of process mining and those other aspects are not essential to this topic.

After the project charter is completed, SIPOC table can be used to incorporate the information from the project charter and show the interrelationships between the clients (end users) and suppliers, and how they interact with the process.

Suppliers	Inputs	Process	Output	Customers
City employees	Time reports	Payroll	Checks, pension reports, taxes paid	City employees, taxing authorities, state, county
Vendors, city employees	Invoices, requests	Accounts payable	POs, checks	Vendors
State, county	Checks, direct deposits	Accounts receivable	Funds available or invested	City departments
City departments	Financial transactions, receipts, checks, invoices, bank statements	Monthly reconciliation	Balanced accounts, adjustments, financial reports	Finance director, council
City departments	Budgeting needs	Budgeting	Budget, appropriations	Council and citizens

Figure 9 Sample SIPOC table<sup>31</sup>

The above table is an example of SIPOC table which shows a financial service provided by a government department. In most cases, finance processes have a high level process map set up initially since financial service is highly regulated and structured. This is a very important information to be collected. Those high level process maps represent the pre-defined model. They serve as the basis for analysis and improvement. Compared to product-based processes, the supplier and customer in a finance process are very different. The substance that is flowing along the process is not physical products. Instead it is information and transaction that are exchanged within the process. Suppliers are the finance information providers or transaction executers. And the customer in finance process is the client (end user) who situates at the end of the process and they only serve as a receiver of information, not sending anything back in return. Correctly defining the end users of finance process is very significant. End users might validate the information but should never take the roles of respondents who contribute to the process in return. If the end user is found giving back adjustment or contributing to the process, then they are not really the end users or at least not purely end users. Instead they should be included in the

---

<sup>31</sup> Source: Furterer, S. L. (2009). *Lean Six Sigma in Service : Applications and Case Studies*. Boca Raton: CRC Press.

process. In addition, the end user is not necessarily human. They can be database or system as well. Another term that is worth mentioning in this SIPOC table is the process. The process in SIPOC table is defined as an individual object that is most representative in the corresponding process. This might sound a little bit confusing. But once readers start to link this SIPOC table to the event log that is introduced in the previous section, it is very comprehensive as it is shown in Figure 11. Suppliers are the owners of each process and when it is translated in the hierarchical structure of the event log, they are the resource of the process. Inputs and outputs are components of Activities, but inputs and outputs are not necessarily described directly in event logs. Processes in SIPOC table are in fact the case ID in event log, since case ID is the identifier of the subject of each process. Each subject of each process would take the input and generate the output for the next process subject. Customers are sometimes included in “Other information” in the event log (not always the case). Customers in the define phase can help to set up the CTS (critical to satisfactory). CTS and project goal should be included in the output column of the SIPOC table.

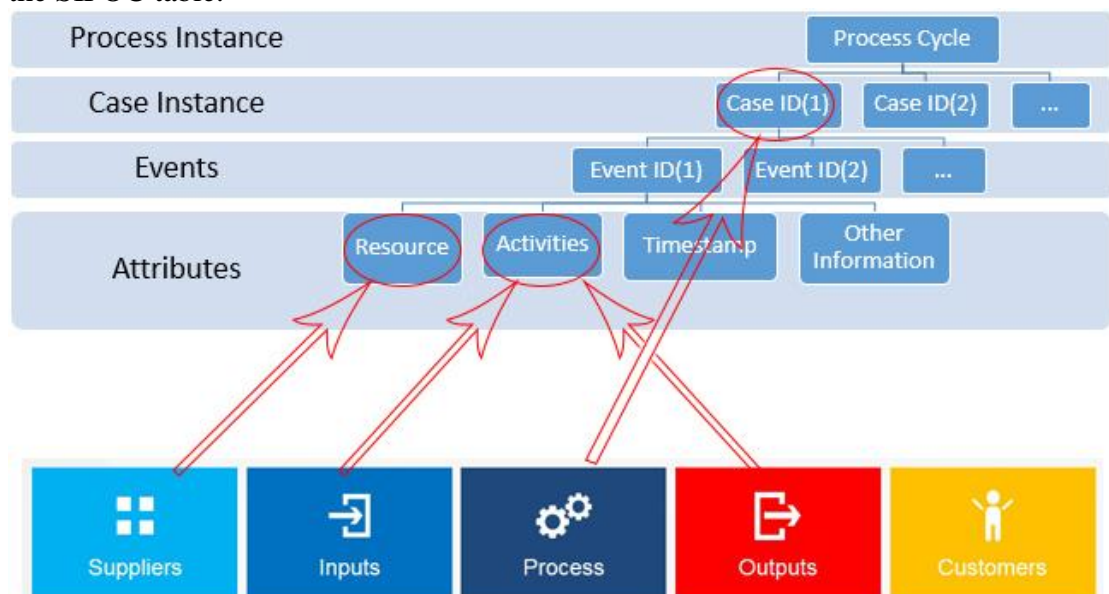


Figure 10 Relation between SIPOC table and event log

In Figure 10, a typical financial process normally use the account names to represent the process, as accounts are most common subject of a finance process. A clearly defined SIPOC table can give a great insight in exploring the event log of the system and it also guarantees the alignment between the reality captured in the event log and pre-defined high level process model. After the SIPOC table is developed, practitioners can start to draw event logs from the system and define the hierarchical structure of the event log through SIPOC table.

### 3.2 Measure Phase

The goal of the Measure phase is to understand and baseline the current process. This is where process mining technique makes a difference in comparison with the standard Lean Six Sigma methodology. In usual case, the result data in this phase is collected and then measured via statistical methods.<sup>32</sup> With process mining technique, instead of data centered, it is process centered. All the measurement will be made on the process level, aligning the pre-defined processes with the real process maps. Three levels of process maps are used in a process improvement effort:

Level	Type/Name	Purpose
Level 1	Macro or high level	Scope the improvement project Provide project and process boundaries Provide a high-level view of the process
Level 2	Process map	Identify process improvement areas Identify process inefficiencies Identify waste
Level 3	Process map or process flow chart	Identify improvement area Identify value vs. nonvalue-added activities Provide detailed how-to (almost procedural level)

Figure 11 Overview of three-level process map<sup>33</sup>

The Level 1 process map in most cases is predefined in financial process, since finance has its own operational protocols. Those protocols set the objectives for the financial processes. Because of which, no matter how IT transforms the finance process, the core of finance operation is unchanged in nature. It only acts as accelerator for finance processes. Therefore, the level 1 process map works as a directory of categorizing the objectives of those finance functions. Practitioners can roughly locate the deficit processes based on the objectives defined in level 1 process map. Level 2 and level 3 process mapping are where the process mining algorithms joins the practice.

The Level 1 process map scopes the approximate the area for improvement and then this information can be used to determine which system's event logs should be developed for further analysis. The level 2 process map is established upon extracted event logs. Practitioners can implement the software like Disco and ProM to generate the level

<sup>32</sup> Ron Basu, Chapter 5 - Tools for measurement, In *Implementing Six Sigma and Lean*, Butterworth-Heinemann, Oxford, 2008, Pages 64-88

<sup>33</sup> Furterer, S. L. (2009). *Lean Six Sigma in Service : Applications and Case Studies*. Boca Raton: CRC Press.



2 process map automatically with the corresponding event logs. An example of such process map is shown in Appendix 6. The information contained on this level starts to become concrete and practitioners will be able to more precisely locate the inefficiencies and waste.<sup>34</sup>

Readers should notice that it is a common mistake that people confused the concept of inefficiencies with waste. In a process, inefficiency refers to the procedures or transactions that do add value to the output but consume too much resource (such as time and money) along the way. Waste, on the other hand, refers to the non-value-added procedures that should be eliminated or avoided in the process. For instance, in an accounting report process, the company board member only requires the account with values over 100 euro to be recorded in the general ledger. However, systems would still automatically produce the reports that included accounts that are below 100 euro and those reports are also reviewed and approved as the other regular reports. All the processes that are taken to produce, review and approve those reports are considered as wastes. In the CTS (Critical to Satisfaction) point of view, the result of those processes do not contribute anything to the objectives. As a comparison, there are other reports that contain more than 100 euro value. But those reports are produced in the wrong format and constantly missing some information. Therefore, they are rejected in the approval processes and returned to the initial producing processes. This sort of iterating processes are considered inefficiencies. Even though inefficiencies are non-value-added processes, they are procedural attempts towards the fulfilment of the objective.

The detail definition of inefficiencies and wastes are determined by the reference point, since financial functions take up different roles in different organization's business strategies. To be more specific, those reference points derives from two attributes, VOC (Voice of Clients) and CTS (Critical to Satisfaction). The detail methods to define VOC and CTS for a finance process are explained in Appendix 7. Generally speaking, VOC is an expression that can capture the end users' requirements for the finance service. VOC emphasizes values and priority of the values. For instance, in finance process, some end users might attach great importance to responsiveness and some others might put more expectations on precision. Both precision and responsiveness are values that should be delivered and if possible, there should be a setting for priority of these two values, answering which value is more important. CTS can be regarded as a numerical measurement for VOC in the process. CTS is set to capture the measurable factors in the process that have direct impact on delivering those values expressed in VOC. For example, if the VOC prioritize the precision of the financial report, CTS can be the First-Time-Right rate. Ideally, CTS should be able to be assessed in a statistical way that practitioners can set some

---

<sup>34</sup> Necessity and inefficiency in the generation of waste : A thermodynamic analysis. (2003). *Journal of Industrial Ecology*, 7(2), 113-123. doi:10.1162/108819803322564389

concrete numerical ranges that can define the inefficiencies and wastes mentioned earlier. In previous example, First-Time-Right rate is mentioned as a measurement for the precision of financial report. Practitioners can set a threshold range for First-Time-Right rate. All the finance reporting processes that give a lower-than-threshold rate should be considered inefficient. Those processes that do not contribute any efforts to First-Time-Right rate are regarded as wastes. Wastes should basically be eliminated from the processes without second thought, while inefficiencies are the major focuses of the process improvement.

Once the inefficiencies and wastes are specifically defined accordingly, practitioners should summarize all the abnormalities in the process maps, relating those inefficiencies and wastes to the events shown in the event log. With the help of the software, all those inefficiencies and wastes can be represented by some specific patterns in the process maps. All the processes that follow such patterns can be filtered out via software tools. These filtered out processes will compose the level 3 process map.

Level 3 process map is a further filtered version of level 2 process map. In level 2 process map, the event logs of the system are imported to generate process maps, but as readers can observe from the Appendix 6, the process map at this level can only locate the system-defined errors which is very natural since event logs are generated by system and system can only show the errors that are programmed in the codes. Those system-defined errors are simply the outcome of either wasteful processes or inefficient activities behind the scene. Therefore, level 2 process map can only help to locate the source of errors but not reveal the cause and effects of those errors. This is precisely why level 3 process map is needed. Process mining software is capable of filtering the paths that all the erroneous processes take and regenerating a new process map that is only composed by all those abnormal processes. This capability considerably reduce practitioners' workload in seeking for causes and effects behind. Now practitioners need to develop a checking table that lists all the possible causes behind those erroneous processes and interview the users of each processes, checking the sanity of each cause.

Erroneous processes are not the only attributes for filtering. Practitioners can go back to level 2 process map and seek for all those processes that considerably outperform the others according to the defined CTS. For instance, in the same example of that finance reporting process, all those processes that obtain prominent First-Time-Right rates can also be filtered out. All those outperforming processes can be used to generate another process map via software tools. Those exponential processes will form the baseline processes, which can also be investigated with the same checking table that is developed in the previous step. However, this time practitioners, instead of looking for mistakes, can analyze what factors that fabricate the performance of those processes. This analysis can be a very good baseline process for continuous improvement of the other underperformed processes.

### 3.3 Analysis Phase

In the Analysis phase, the data collected in the Measure phase is analyzed to identify the root causes of problems. Basically there are three steps in Analysis Phase <sup>35</sup>:

- Develop cause and effect relationship
- Determine and validate the root cause
- Develop process capability

In the previous measurement phase, information regarding to the root causes are collected and verified. Those causes need to be linked to the effects in order to visualize those process factors that influence the performance. Before linking the causes with the effects, it is very helpful to categorize the causes into small groups. Those categories can give insight to what effects are corresponding to which cause. A cause and effect diagram is developed for this purpose. Figure 12 shows an example of a cause and effect diagram. Due to the fact that the diagram's shape resembles a fishbone structure, it is also named fishbone diagram. Causes and effects are not necessarily paired up in a one-to-one manner. In most cases, many effects are linked to the same root cause. In addition, effects should find their ways in the event logs. Practitioners must understand how those effects are represented in the event logs and how they are captured by the event logs. The inter-connection between the effects and event logs would serve as the foundation for measuring the effectiveness of the improvement made in the new Improve phase. The cause and effect matrix (George, Rowlands, Price, and Maxey 2003) is the most used tools to correlate the same root cause to multiple effects. Figure 13 shows a template of such a cause and effect matrix. The principle of this matrix is that the relation between causes and effects is measured in a scale of 9 (with "loosely related" as 1 to "closely related" as 9 ). The executers of those processes need to evaluate those relations according to the CTS that is defined in previous phase and the end users or the clients of the finance processes should measure the relative weighting of each cause, which shows the importance of each cause. The relative weighing of those causes would set the priority for the process improvement.

---

<sup>35</sup> Furterer, S. L. (2009). *Lean Six Sigma in Service : Applications and Case Studies*. Boca Raton: CRC Press.

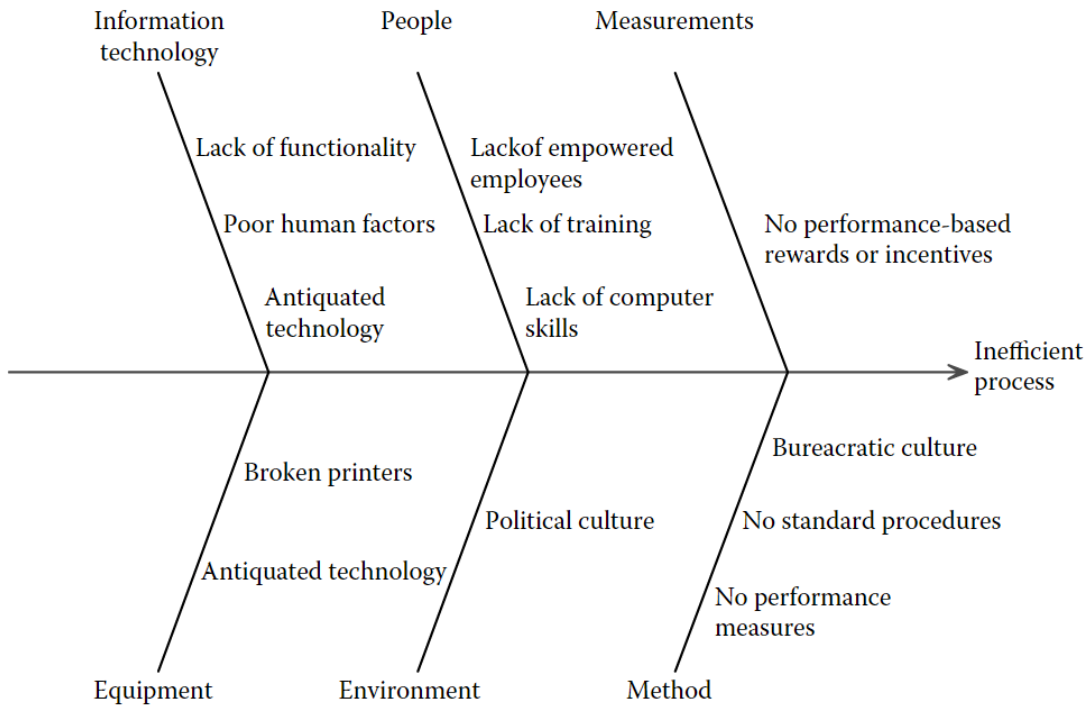


Figure 12 Example of a cause and effect diagram<sup>36</sup>

	Effects					
	Y1	Y2	Y3	Y4	Total	Relative Weighting
<b>Causes/Importance</b>						
<b>X1</b>						
<b>X2</b>						
<b>X3</b>						
<b>X4</b>						
<b>X5</b>						

Figure 13 Example of a cause and effect matrix<sup>37</sup>

Once causes and effects are linked, these links should be routed back to the Level 2 process map in order to confirm those links are valid and aligned with reality. Effects are not always recorded in the event logs, hence, additional information outside the event logs need to be collected to validate the authenticity of the links between cause and effect. Furthermore, the effect must find its way to represent itself in the developed process map.

<sup>36</sup> <http://asq.org/learn-about-quality/cause-analysis-tools/overview/fishbone.html>

<sup>37</sup> Furterer, S. L. (2009). *Lean Six Sigma in Service : Applications and Case Studies*. Boca Raton: CRC Press.

If those effects and causes are validated and confirmed by the reality, an overview of process capability can be then developed. This overview would set a very clear goal for improvement. In principle, it will indicate the upper boundary of the potential improvement of the process. In Lean Six Sigma, this goal is represented by DPMO (Defects Per Million Opportunity). DPMO basically is a statistical calculation of defect rate of a product or a service. The name “Six Sigma” actually originates from this calculation. In a Lean Six sigma project, the decrease of defects is the ultimate indicator of process improvement. Sigma “ $\sigma$ ” indicates the tightness of the distribution. The detail illustration of this concept is given in Appendix 8. In the development of process capability overview, practitioners need to understand two things, the current state of the process in terms of defect rate and the potential upgrade of the defect rate at the end of this project. The advantage of using  $\sigma$  here is to clarify the current state of the process performance and the potential improvement of the process with a concrete statistical value.

Process mining in this phase actually forms the foundation for process streaming and tracking the inefficiencies and wastes. In other words, process mining provides the raw material for analysis in Lean Six Sigma project. It might seem that process mining does not paly significant roles in Analysis phase. However, compared to regular ways of Lean Six Sigma, the raw material is much more refined and prepared for continuous analysis.

### **3.4 Improve Phase**

In the Improve phase, the improvement recommendations are developed and implemented. In the previous phase, the cause and effect analysis shed light on the detail reasons for the inefficiencies and wastes. Solutions need to be developed accordingly. Ideally practitioners can create a sandbox to text those solutions in simulated context and generate new event logs. Those event logs can be imported to the process mining software again to produce a new process map. This new process map can be used to make comparison with the previous process maps in order to check those located inefficiencies and wastes are indeed solved. In addition, the baseline processes (indicated by the process maps that are generated with those exceptionally outperformed processes in the measurement phase) can also be utilized as a comparing sample. Solutions should lead the new process map to approach to those baseline processes. There are many tools that can be exploited in this phase, such as DOE (Design of Experiment), scoreboards and COPQ

(Cost of Poor Quality) assessment. Those tools are very regularly used in almost all lean projects and if readers are interested, please refer to other Lean project books. <sup>38</sup>

### 3.5 Control Phase

The goal of the Control phase is to ensure that improvements had a positive impact and that they will be sustained and controlled. The key success factor of this phase is to develop a thorough control plan which document all the changes that are implemented in the process and locate those changes in the process stream. A sample control plan is shown in Figure 14.

Process step	Control mechanism	Measure/ metric	Criticality (High, Medium, Low)	Action to be taken if problems occur	Responsibility

Figure 14 Example of a control plan

This control plan should be constantly referred back to the process map generated in previous phase. Process step indicates the location in the process map where inefficiency and waste are located. Since the improvement targets at those areas of process maps, the changes made in the improve phase will take major effects on those areas. They should be closely monitored once the improvement is deployed. Control mechanism is in fact a group of indicators that can visualize the change. They work as the representatives of the changes that are made on the process map. Measure/ metric is a standard threshold that can evaluate the extent of changes that are taking place in the process step. Criticality is the predefined priority that is set up in the Analysis phase. Finally, ‘Actions to be taken if problems occur’ and ‘Responsibility’ are related to the personnel that are corresponding to each change that is made in the improvement. The information regarding to those personnel should be added in the event logs and represented in process map, which can make the tracking of responsibility much clear.

Control phase is not only responsible for sustaining the benefits of the change but also documenting the changes and related attributes for the purpose of potentially expanding the solutions to other similar problems in the organization. Another purpose of this

---

<sup>38</sup> Readers might notice that the illustration of improve phase is relatively simplified compared to previous phases. It is due to the irrelevancy of process mining techniques regarding to this phase. Process mining can spot and help to analyze the problems but solving them is highly context dependent.

documentation is to identify what lessons can be learned from this project and the next opportunity for improvement.

### 3.6 Validating the integrated model

Previous sections introduced the practical details about how to utilize process mining in DMAIC phases. This thesis intends to prove the integration of process mining can bring substantial ease to DMAIC phases. There are two approaches to achieve this intention. One is through business case. By practicing this methodology in reality, the observation upon the project execution and the interview with the practitioners can all provide significant insight regarding to the effectiveness of integration. Next chapter will completely devote to this approach. Another approach is through theoretical induction. CRISP-DM is the original methodology to implement process mining.<sup>39</sup> For each CRISP-DM step, there are very specific objectives and scopes. In theory, if those objectives and focuses can also coincide the objectives and focuses of each DMAIC phase, it is legitimate to conclude that process mining does have the natural fitness to be integrated in DMAIC phases.

In fact there is indeed a significant synergy between DMAIC phases and CRISP-DM. The define phase actually includes all the procedures that compose Business Understanding and Data Understanding in CRISP-DM since the define phase focuses on scoping the project and browsing through the stakeholders' demands and expectations upon the projects which are exactly the objectives of Business Understanding and Data Understanding. But the define phase incorporates some other additional content. For instance, the define phase already starts to define the process stream that connect the suppliers of the process to the end clients. This kind of process stream is in fact a sort of relationship among various datasets. But the Understand Business and Understanding data concentrate on the selection of right data, omitting the analysis of data relationship. In other words, the first two steps of CRISP-DM are embedded in the Define phase. Figure 9 represents this relation by circling these two steps and connecting them to the define phase.

Data preparation and Data Modeling are included in the Measure phase in the same logic. This also means that in Measure phase a complete process model is already produced via process mining algorithm and this serves as the foundation that guarantees all the analysis are deployed upon the whole process instead of a single procedure.

---

<sup>39</sup> Aalst, W. (2011). *Process mining : Discovery, conformance and enhancement of business processes*. Heidelberg etc.: Springer.

The Evaluation step of CRISP-DM becomes part of Analysis phase. But if reader goes back to the explanation of Evaluation step in previous section, it might seem inconsistent to the objective of Analysis phase. Evaluation step intends to validate the generated model via some established criteria. Analysis phase, however, aims to locate the root cause of the inefficiency and waste in the process. There is indeed a significant difference between Analysis phase and Evaluation step in CRISP-DM. Instead of validating the model, Analysis phase will use the established criteria to assess the generated model from previous phase and identify the root cause through this assessment. Simply speaking, the procedures that take place in Evaluation step are the same as the Analysis phase in DMAIC method, but they serve different purposes. One is to validate the process, and the other one is to assess the process and locate the root cause accordingly.

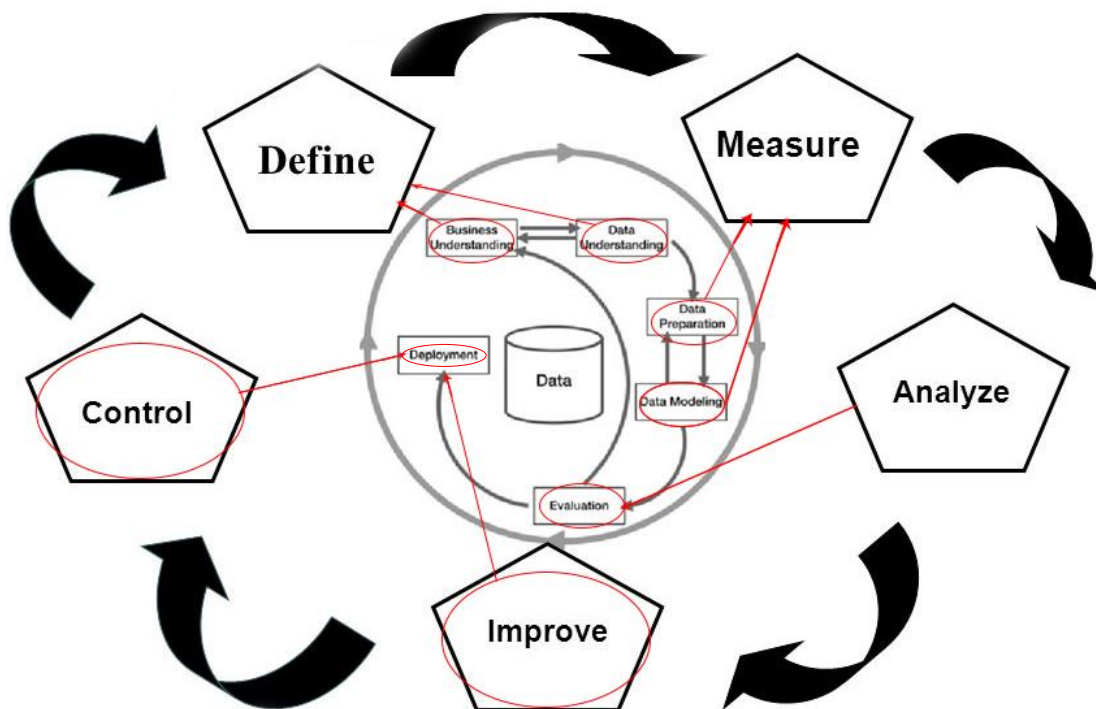


Figure 15 Visual link between DMAIC and CRISP-DM

The Improve and Control phases combined are actually equivalent to the deployment step in CRISP-DM. Improve phase starts to implement the result of analysis on the process and assess the potential improvement via statistical calculation and control phase keeps tracking and monitoring the implementation of process improvement. Deployment step also starts to test the improved process model and monitoring the implementation. They both provide maintenance to sustain process improvement, sharing exactly the same objectives. In addition, both methodology forms the circulating processes. In CRISP-DM, after deployment phase, the improvement is under scrutiny again and re-enter the business understanding phase to identify new needs. As a comparison, in DMAIC phases, control



phase also monitors and tracks the improvement, motivating the re-initiation of Define phase to recognize the changing demands of the stakeholders as well as the deficiencies of the previous improvement.

The synergies do not state that these two methodologies are interchangeable. CRISP-DM in theory is still a waterfall method. It focuses on studying data relations and utilize the analysis to indicate the correlations within the dataset. Practicing process mining through CRISP-DM can effectively seek for process defects that affect the short-term performance. By fixing those defects, CRISP-DM will iterate the steps to look for other defects. DMAIC phases, on the other hands, provide a more profound vision on process. It recognize the data relations in the beginning by identifying the process stream instead of separate dataset. Rather than improving the processes step by step as CRISP-DM does, it directly set the long-term performance as the goal and analyzing the process stream (which is in fact already a data structure composed by multiple related datasets).

## 4 CASE STUDY ON IMPLEMENTING PROCESS MINING IN LEAN SIX SIGMA PROJECT

This section would start by introducing some background information regarding to the project and then illustrate the practice of integrated method based on the order of each DMAIC phase.

### 4.1 Background

The implementation of process mining project is upon the accounting reporting processes in Company ABC. Company ABC at present concurrently implement several systems to handle different accounting transactions. For instance, the pay-roll accounts are managed by a specific system and the manufacturing accounts are processed by a different system. All those systems will automatically generate accounting data and feed those data to the SAP database, creating General Ledgers (GL) for accounting purposes. Despite of the extensive use of information systems, there is still a small portion of accounts that have to be created manually. When it comes to MEC (Monthly End closing) or YEC (Yearly End Closing), all accounts, before entering SAP database, need to be checked for their correctness as well as being adjusted for some change. Figure 16 shows an overview of the process.

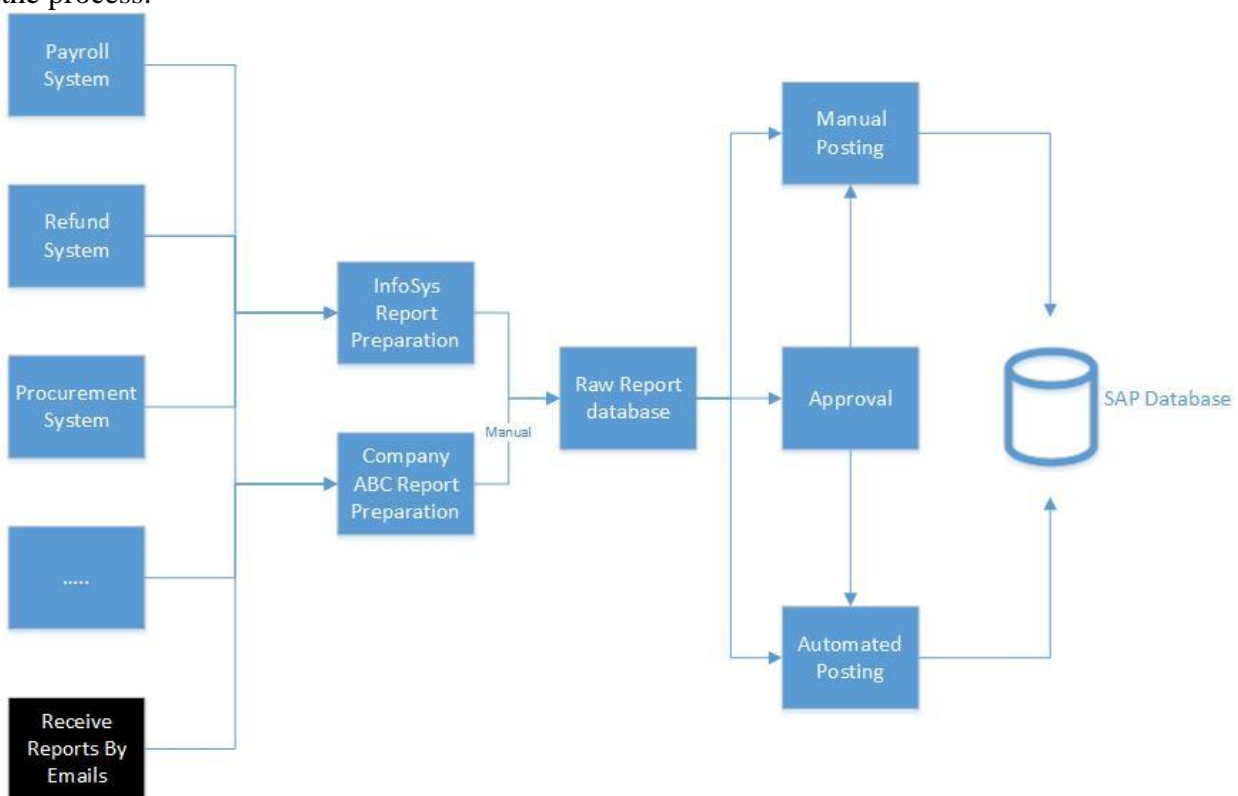


Figure 16 High level structure of MJV process

All those different systems need different MJVs to adjust or correct their own unique accounts. There are two approaches to generate MJVs. One is through third party support, namely InfoSys. Company ABC outsourced the majority of MJV posting processes to third party. When there are adjustments needed, Company ABC would inform InfoSys and InfoSys would prepare those MJV accordingly. The other approach is through Company ABC's Record-to-Report team. After the MJV is prepared, it will not directly enter SAP database. They have to go through approving steps. If the MJV is disapproved, the MJV would be returned to posting party (either InfoSys or Company ABC Record-to-Report team). They have to review the MJV and repost the corrected MJV. If posted MJV successfully get approved, they will be posted in the SAP database either automatically or manually (which depends on the nature of MJV since some MJVs' content contain some unusual information and those information needed to be sorted manually). For example, some "account receivable" are created by the Payroll System on a regular working day. During the MEC, this sort of account needs to be cleared so a checking procedure is performed and a manual journal entry is created to register the adjustment of this account in the database (General Ledger). First, procurement system would inform InfoSys. Then InfoSys would create the MJV and post it in the raw report database. Company ABC would review those MJVs and if everything is fine, it will approve it and have it automatically posted in SAP database.

The above example is a typical MJV process that incur no errors. The majority of the MJV posting processes is done automatically by the system. But there are still around 10% of cases that can only be handled manually. Even though manual work only accounts for 10% of the total workload in this process, it is still very costly in terms of time and resources. In addition, the manual work is error-prone. The company currently only has a high level process map which does not offer capability to document all the detailed activities in performing this work. It creates annoying disturbances in business process management in terms of locating the inefficiencies and wastes of the process. By implementing Lean Six Sigma combined with process mining techniques, the goal is to summarize the relevant data to build a thorough view on the entire processes and once those activities are modeled, the whole process can be documented and diagnosed from different perspectives. For instance, process bottleneck diagnosis can be performed, identifying what kind of resources are accountable for the errors or the slow pace of the process. Process models can also be used to detect what kinds of behaviors are undesired or unnecessary in the process.

## 4.2 Define phase

The major goal of this project is to increase the first-time-right rate of MJV process, accelerate the MJV posting processes and reduce the cost of errors that appear in the MJV. As introduced in the last section, Manual Voucher Journey (MJV) is a financial reporting process that handles the accounts that are not automated by finance IT system. The only purpose of MJV is to record the finance transactions that are not incorporated in the implementation of finance IT system. Ideally the whole concept of MJV should disappear in the finance process as all the financial accounts are supposed to be automated via IT. However, that ideal situation can hardly be realized since many unforeseen abnormal transactions would require human interference. In addition, currently the logics encoded in finance system cannot include all the situations that may happen in the posting process yet. For instance, different regions have their own unique assets that are unforeseen to the system designers. When transactions relevant to those unique assets happen, the finance reporting regarding to those transactions have to be dealt with manually and then MJV appears. Another problem regarding to MJV is that the supporting files related to those recorded transactions have to be edited manually as well. As all accounting report needs to be intensely audited, not only the outcome of the transactions is recorded as financial report, but also the tracking information concerning those transactions have to be put in the database, as the supporting files. For those transactions handled by finance system, their supporting files are instantaneously generated by system as well. But the transactions recorded by MJV do not have the same convenience. All the supporting files need to be added by hands. Reducing the efforts in making supporting files for MJV becomes another objectives of this process improvement.

The project charter can be developed as following:

<b>Project name:</b>	MJV process improvement
<b>Problem Statement:</b>	The finance department of Company ABC identified the need to streamline the MJV posting processes and improve the efficiency. The finance clerks complained about the difficulties of registering MJV in the database and inconvenience of modifying the records. The supporting files are hard to upload in the database due to the incorrect formatting and missing information from the local operators. The current MJV process time ranges from 1 second to 3 days and 5 hours and on average it takes 2.3 hours. In some cases, once the MJV posting failed the approving procedure, the reposting took as long as two days, which considerably oppress the general efficiency.

<b>Stakeholders:</b>	Financial department, local operating unit, outsourced company (InfoSys)
<b>CTS (Critical to Satisfactory):</b>	First-Time-Right rate, Amount of MJV
<b>Goal of project:</b>	To streamline MJV processes, reduce cycle time, increase first-time-right rate, decrease the amount of MJV.
<b>Scope statement:</b>	The MJV posting processes, including the supporting files posting.

Figure 17 Project charter

An SIPOC table can then be put in place:

Suppliers	Input	Process	Output	Clients
Finance System	Generating the raw finance data	General MJV Posting processes	Raw finance data	Report Consolidation team
Accounting Specialists	Determining the content of MJV		Refined items of finance data	
Account preparers	Recording and manually produce and post MJV		Standard formatted MJV report	
Account Reviewers	Revising and categorizing the MJV		Categorized MJV	
Account Approvers	Approving the MJV		Approved MJV that is ready for posting or Disapproved MJV that needs to be revised	

Figure 18 SIPOC table for MJV posting

This SIPOC table would be utilized to define the hierarchical structure of the event log data that is going to be drawn from the system. First, Company ABC fortunately has already set up a financial system that contains event logs that collectively capture all relevant system procedures that are taken in MJV posting processes. Therefore, with very little effort, a ready-for-use event logs can be directly drawn from the finance system. A fraction of that event log is shown in Figure 19. Due to information security reason, it is not possible to present the complete data. In Figure 19, there are two columns of data. One is named RequestKey which represents each posting request made by account specialists. This RequestKey is unique to every request, in other word, unique to every posted MJV. Considering all the subjects that receive input from suppliers are MJV, the case ID

is supposed to be unique in order to identify the complete set of procedures that happen to every MJV. In this case, RequestKey is the perfect case ID for this project.

RequestKey	Action
REQ201612005066	Request Submitted for Automated Posting
REQ201612005066	Request submitted and approved by Requester
REQ201612005066	Validation Failed, because a error verb was found in the log re
REQ201612005066	Request Automated Posting Failed
REQ201612005066	Request Submitted for Automated Posting
REQ201612005066	Request submitted and approved by Requester
REQ201612005066	Posting-request succeeded with SAP DocumentNr: 100026470
REQ201612005066	Request Automated Posting Succeeded
REQ201612005066	Request Closed
REQ201701003469	Request Submitted for Automated Validation
REQ201701003469	The validation-request was validated succesfull, but warnings
REQ201701003469	Request Automated Validation Succeeded
REQ201701003469	Request submitted for first approval
REQ201701003469	Request Approved By First Approver
REQ201701003469	Request Submitted for Automated Posting
REQ201701003469	Posting-request succeeded with SAP DocumentNr: 100028210
REQ201701003469	Request Automated Posting Succeeded
REQ201701003469	Request Closed
REQ201701004237	Request Submitted for Automated Posting

Figure 19 A fraction of event log

The other column in Figure 19 is called Action which records every procedure that happens to the corresponding MJV. This information is the major analytic subject for discovering patterns in processes. Each RequesKey is corresponding to a series of Action. In other words, each MJV goes through a series of procedures. According to the hierarchical structure of event log, Action is the activities that are on the Attribute level, under the case ID.

Action	ActionDoneBy	ActionStartTime	ActionEndTime
Request Submitted for Automated Posting	hannie.r	12/8/2016 8:45	12/8/2016 8:45
Request submitted and approved by Requester	hannie.r	12/8/2016 8:45	12/8/2016 8:46
Validation Failed, because a error verb was found in the log re	System.I	12/8/2016 8:46	12/8/2016 8:46
Request Automated Posting Failed	System.I	12/8/2016 8:46	12/8/2016 8:47
Request Submitted for Automated Posting	hannie.r	12/8/2016 8:47	12/8/2016 8:47
Request submitted and approved by Requester	hannie.r	12/8/2016 8:47	12/8/2016 8:47
Posting-request succeeded with SAP DocumentNr: 100026470	System.I	12/8/2016 8:47	12/8/2016 8:47
Request Automated Posting Succeeded	System.I	12/8/2016 8:47	12/8/2016 8:47
Request Closed	System.I	12/8/2016 8:47	#NULL!
Request Submitted for Automated Validation	arti.lend	1/4/2017 11:01	1/4/2017 11:01
The validation-request was validated succesfull, but warnings	System.I	1/4/2017 11:01	1/4/2017 11:01
Request Automated Validation Succeeded	System.I	1/4/2017 11:01	1/4/2017 11:01
Request submitted for first approval	System.I	1/4/2017 11:01	1/4/2017 11:02
Request Approved By First Approver	mahesh.	1/4/2017 11:02	1/4/2017 11:02
Request Submitted for Automated Posting	mahesh.	1/4/2017 11:02	1/4/2017 11:03

Figure 20 Sample event log

Figure 20 shows another part of the same event log. There are three other columns, 'ActionDoneBy' which indicates the suppliers of the process, 'ActionStartTime' and 'ActionEndTime' which are the timestamp of the process. All these above columns would form the basic hierarchical structure of the event log. In the following phases, event log would be imported to the software according to this structure.

### 4.3 Measure Phase

The current MJV posting processes need to be baselined in this phase via the three levels process map. The first level, which is the most abstract one is already provided in the background section. Now we take a closer look at this map.

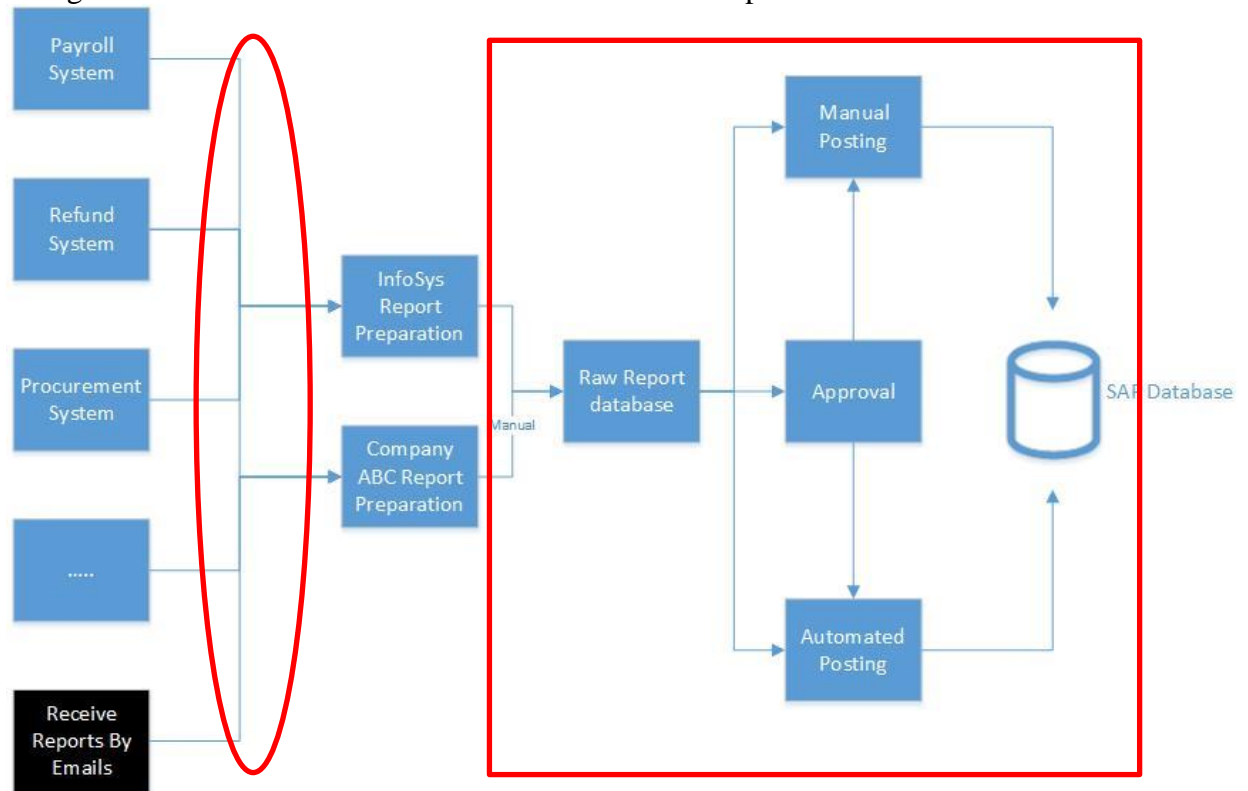


Figure 21 Level 1 process map

The scope of this project is bounded to MJV posting processes. Genuinely speaking, all the processes included in Figure 21 are supposed to be included in the posting process. However, the system that produces the event logs only captures the paths contained in the red rectangle. For the paths within the red circle in Figure 21, they are either manually done or not recorded in the system event log. In this case, manual data collection is needed to prepare the information regarding to those processes. However, due to the time limit of this thesis, it is impossible to take out data collection plans to prepare data for these processes. Therefore, this business case would only take advantage

of the existing event logs to study the paths contained in the red rectangle. But readers should get the idea that level 1 process map would indicate the completeness of data collection. Since event log originally is not for the purpose of process mining, it is very important to validate the information in the event log is sufficient to analyze the targeted processes.

Before proceeding to level 2 process map, it is now important to define inefficiency and wastes in MJV posting processes. A clear definition can help to locate the defected processes. The CTS is the First-Time-Right rate and the amount of MJV. Stakeholders is expecting the MJV is posted in very small amount and in the meantime can be posted first time right without any further modification. The inefficiency and waste can be defined as the following:

- **Inefficiency:** Error appears in the posting processes and require review and further modification. Those errors would either get rejected in the MJV preparing stage or get disapproved in the approving stage.
- **Waste:** The content of MJV is not necessary for the consolidation team, for which they should not be prepared in the first place and never enter the posting processes. In addition, there are some MJVs that have the potential to be replaced by automation, hence, do not need manual interference.

These two definitions should have their own patterns within the process data recorded in the event log. In level 2 process map, practitioners would find those patterns and locate inefficiencies and wastes accordingly.

With help of software, Disco, the level 2 process map can be generated by simply importing the event log in the software. What a level 2 process map looks like is already shown in Appendix 6. It is very difficult to present the complete process map in this thesis because of its massive size of the graph. Figure 22 is a fraction of the level 2 process map.

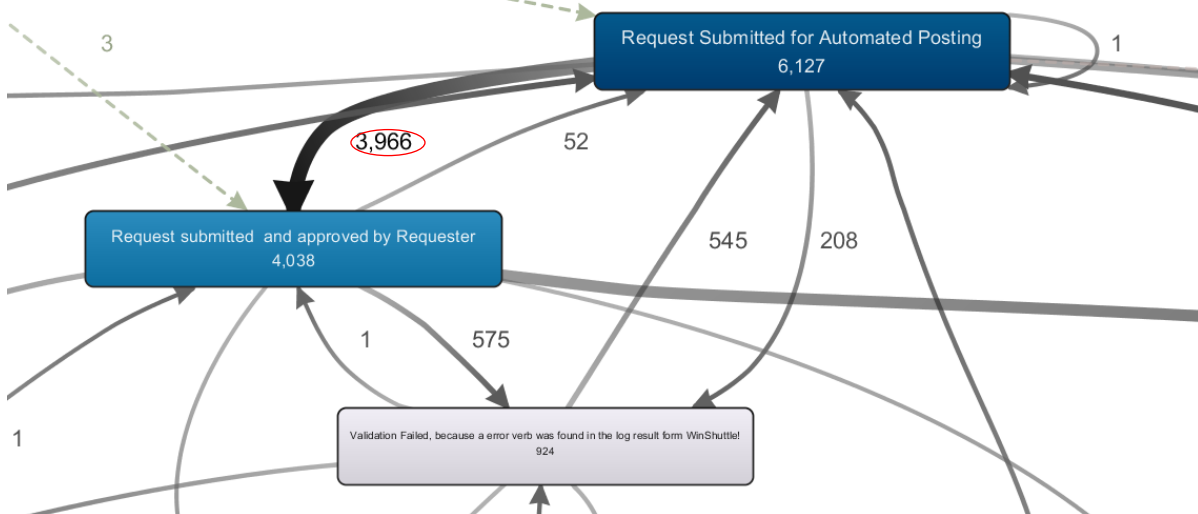


Figure 22 A fraction of level 2 process map



In figure 22, basically the words within the rectangle of the graph describe the procedures that can happen to each MJV and this information is extracted from 'Action' column of the event log. The paths in the graph represents the order of the procedure happens to each MJV. There is also a number situated beside the path, which indicates the amount of MJVs that pass through those procedures in the same order. For instance, the number circled in Figure 22 shows there are 3,966 MJVs that go through "Request submitted for automated posting" first and then take the action "Request submitted and approved by Requester". In this level 2 process map, the pattern of procedures that happen to every MJV is revealed via this sort of graphical notation. Disco can not only graphically represent those patterns but also summarize those patterns in terms of statistical analysis. This statistical analysis provides a shortcut to link the defined inefficiency and waste to the procedure pattern.

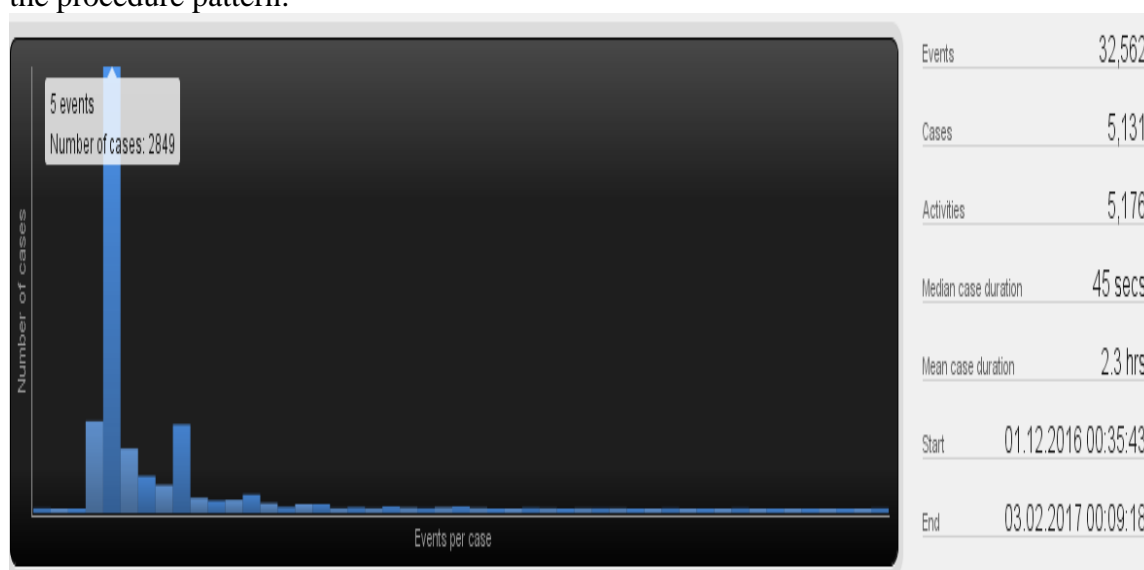


Figure 23 Statistical summary

Figure 23 shows the average time for MJV posting is 2.3 hours. However, the median duration is only 45 seconds, which means half of the MJV posting takes less than 45 seconds, in the meantime, the average is 2.3 hours. There must be some MJV posting that takes significantly long time, driving the average time up to 2.3 hours. In figure 23, the graph indicates the relation between the amount of MJV and the amount of procedures that happen to the MJV. This graph claims that the amount of MJV that took 5 procedures for posting is 2849, which represents the most frequent scenario. But as the blue area shows, there still exists some MJVs that go through a lot more than 5 procedures. For those MJVs, there must be some iteration of processes, such as getting rejected by reviewers, being disapproved and being invalidated. This sort of MJVs is a typical indicator of inefficiency. Therefore, in level 3 process map, MJVs contain such inefficiency features should be sorted out and specifically targeted upon. As for waste, according to the

previous definition, it is determined by the content of MJV. This information unfortunately is not captured in the event log and due to which process map cannot show the waste processes. Therefore, additional information need to be collected regarding to the content of each MJV and then develop another measurement to identify those waste MJV and observe the patterns of those MJV in terms of procedure order. Then those patterns can be referred back to the level 2 process map and be sorted out to compose the level 3 process map. In this business case, the waste is mainly represented by the MJV that recorded the values lower than 100 euro. The client of MJV posting processes, which is the consolidation team, is not interested to incorporate the transactions that contain lower-than-100-euro values. Therefore, all the reviews and approving regarding to those MJVs are considered wastes. This standard can be used to sort out waste processes in level 2 process map.

All those inefficient and waste processes can be easily filtered out via process mining software, such as Disco. After the pattern of inefficiency and waste is identified, the level 3 process map is:

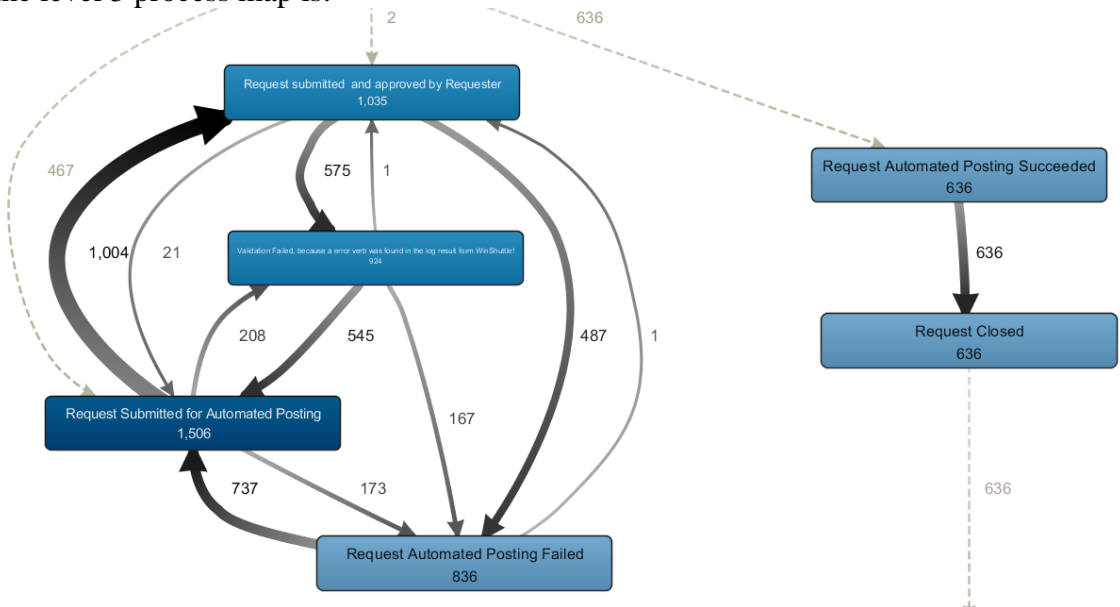


Figure 24 Level 3 process map

Figure 24 is the complete version of the level 3 process map, not just a fraction. Compared to Level 2 process map, it already becomes much simpler since it only contains the defected processes. Level 3 process map has made vital progress in locating the defects and it will be the major focus of Analysis phase. In addition to this level 3 process map, the baseline map, which consists of the outperforming processes, is also developed. As the graph in Figure 23 shows, there are also some MJVs that go through less than 5 activities. The procedural patterns of these MJVs can also be sorted out as the baseline process.

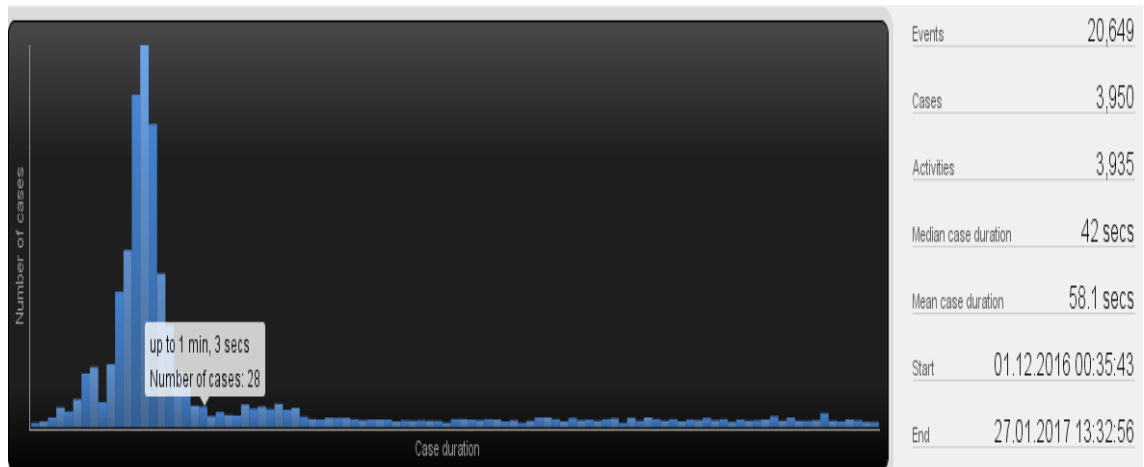


Figure 25 Baseline process Analysis

Figure 25 shows the statistical summary of such baseline process. As it is shown, the average duration is only 58.1 seconds and the median duration is 42 seconds, both of which are far superior in performance than the general MJV performance. The detail regarding to those MJV can be investigated. In the event log, the source of MJV is also recorded. Those outperforming MJV are mainly from North American region and Benelux region. The detailed content of those MJVs in these two regions can be investigated.

#### 4.4 Analysis Phase

There are several patterns of procedural orders that can be identified in Level 3 process map relating to inefficient processes. Figure 26 is a typical pattern of an inefficient MJV posting. This pattern can be found through filter function of the software. According to event logs, the MJV experienced this pattern normally took longer than one day for posting.

```
Request Submitted for Automated Validation
Request submitted for first approval
Request Automated Validation Succeeded
Request Approved By First Approver
Request Submitted for Automated Posting
Posting-request succeeded with SAP DocumentNr: 100238537
Request Automated Posting Succeeded
Request Closed
```

Figure 26 Inefficient process pattern

All the MJV postings that consist of this pattern are then analyzed to track the source of inefficiency. The event logs indicate that the major source of those inefficient processes come from India, China and Netherlands (shown in Figure 27). The suppliers of those processes can also be tracked through the event log. An investigation team is

formed to interview those process suppliers and make surveys regarding to the detail of those processes.

470461\India Commercial (IN01)	224	23.93 %	
263667\P(C) Co. Ltd - Lighting (CN09)	155	16.56 %	
119792\Lighting Global Headquarter, Ehv (EX01)	94	10.04 %	
251021\Chilena SA-Lighting (CL00)	45	4.81 %	

Figure 27 Source of inefficient processes

Figure 28 shows the cause and effect diagram that derives from the investigation. From Information Technology perspective, Company ABC is apparently still implementing SAP modules in MJV posting functions. SAP modules are standardized platforms. In regions such as India, China and Netherlands, MJV posting is under customized regulations within Company ABC. However, Company ABC uniformly adapt the same posting system in all regions, in which case the standard SAP system cannot tailor to the unique regulations in these three regions. Inefficiencies appear and cause trouble.

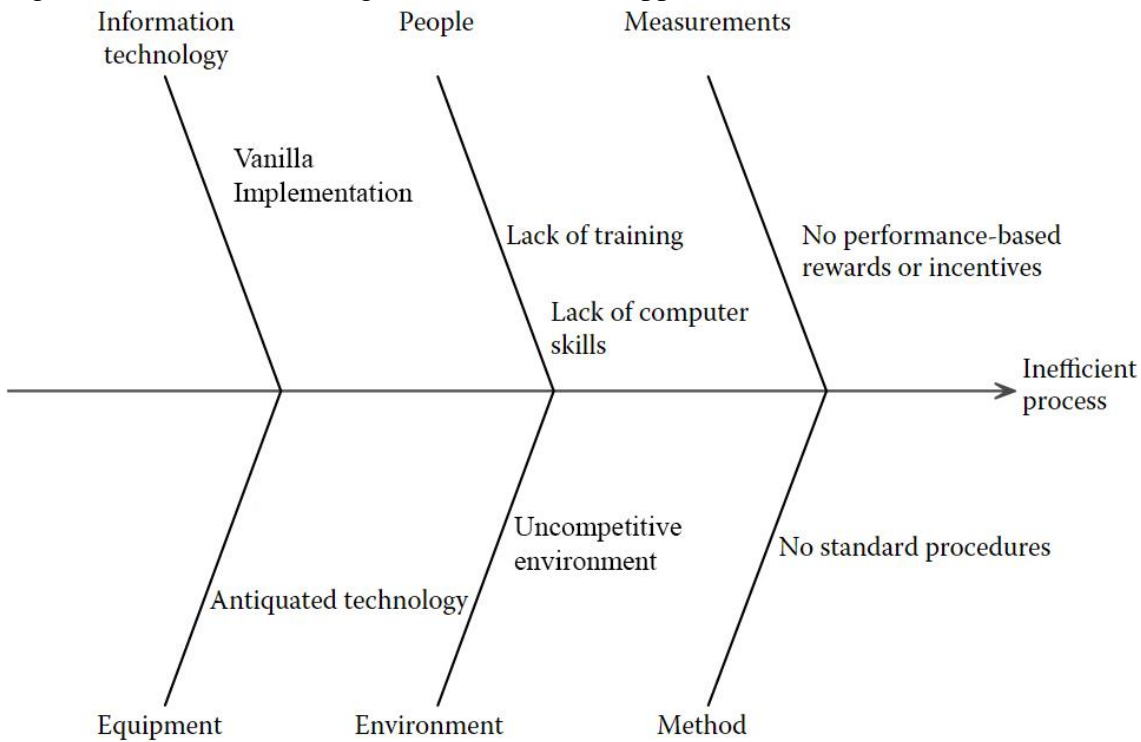


Figure 28 Cause and effect diagram

From the people’s perspective, especially in India personnel still needs trainings in the posting procedures. Company ABC outsourced part of posting processes to InfoSys in India and recently InfoSys recruited new employees to handle the transactions for Company ABC. The lack of training on these new employees significantly increased the errors in posting. In addition, the uncompetitive environment and no-performance-based incentives also gave rise to errors in posting. Personnel involved in posting processes know it is always possible to modify and re-post the MJV even if the posting failed. Company

ABC did not give any rewards or incentives to exceptional performance of MJV posting as well. Hence, employees do not attach too much importance to the MJV posting.

All the above causes collectively deteriorate the performance of MJV posting. The way those causes represent themselves in the level 3 process map (Figure 29) is via two forms. One is the number besides each path. Since those numbers represent the amount of MJV that experience the inefficient pattern of posting procedures, the decrease of the number is tantamount to increasing efficiency. Another form is the number within the node “Request Automated Failed” (Highlighted in red rectangle in Figure 29). This number represents the amount of MJV postings that are rejected or disapproved, which is also a perfect indicator of efficiency. In the capability development analysis, these two numbers can be used to substitute the defect parameter in the formula (see Appendix 8) for calculating the  $\sigma$ . This would set the standard to evaluate the improvement that is made in the project.

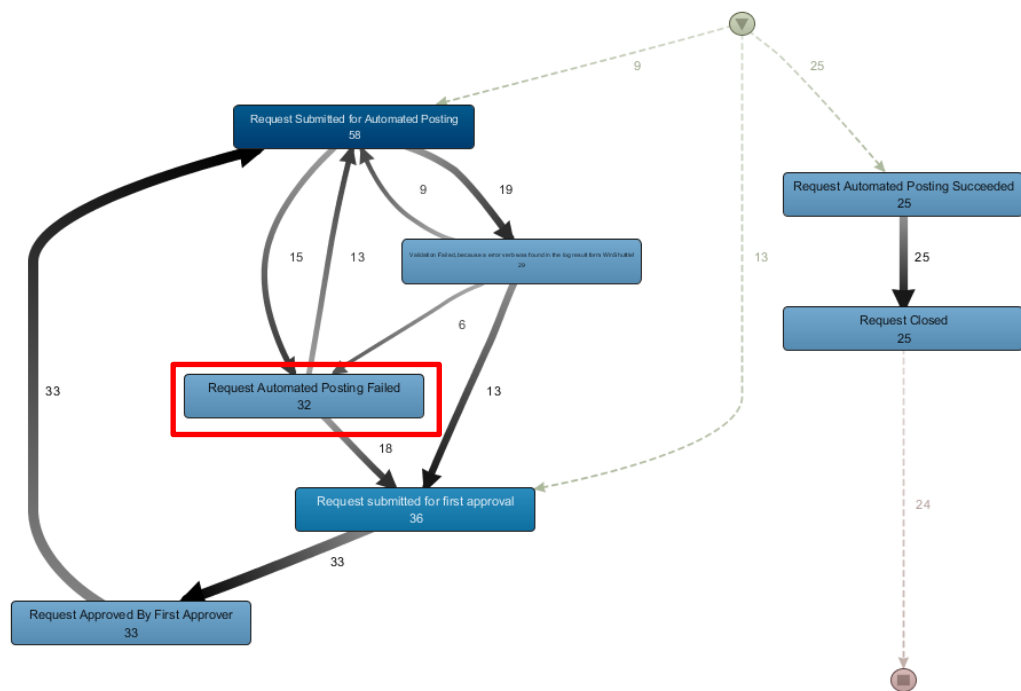


Figure 29 Inefficient process mapping

## 4.5 Improve Phase

As I was writing this thesis, the improvement plan barely started. A COPQ (Cost of Poor Quality) is analyzed and communicated with the stakeholders. Very specific improvement scheme is set up. It mainly incurred a thorough review over the system utilized in posting process in India, China and Netherlands region. Third party software is installed to tailor the specific demands in some unstandardized posting processes. Employees’

training project was scheduled. Furthermore, a MJV related KPI is also developed to monitor and encourage employees to make less errors in posting process. There were other measurements taking place as well. However, since it was irrelevant to process mining integration, they were not illustrated in detail here.

## **4.6 Control Phase**

Due to information confidentiality issues, the detail control plan cannot be listed in this thesis. In Company ABC, they utilized the level 3 process map as a reference model. After change measurement took effect, new event log is imported to the software and those originally defected processes were again filtered out to form the new level 3 process map. A thorough comparison between the new process map and the previous one contributed to constant tracking the effects of the changes. In addition, the suppliers were added to the level 3 process model as well. This newly added information can assist to monitor the accountability of the personnel for each change made on the process. So far, the project already started for one month and the effect is not statistically visible in the event log yet.

## **4.7 Interview**

Along the implementation of this business case, I have conducted several interviews concerning the validity of integration of process mining. The interviewees are all participants of this business case who implement this integrated methods throughout the whole project. In addition, they are all qualified Lean concept practitioners who have been systematically trained to do lean project. Therefore, their opinions and perspectives on this integrated method is valuable and legitimate in evaluating the effectiveness of the integration.

In total, there are ten questions being asked in the interview. Each of the questions investigates one perspective of the mechanism that connects the process mining to the Lean Six Sigma. Those mechanisms all derives from each of the 6 phases in DMAIC. A sample of the questions as well as the detailed result is attached in Appendix 9.

## 5 LIMITATIONS OF THE INTEGRATED METHODOLOGY

Existing process mining techniques have already been practiced in more than 100 organizations.<sup>40</sup> However, despite of the applicability of this technique there are still many limits on utilizing it as a tool.

### 5.1 Limitations on the project

The biggest limit is the concept drift, i.e., processes change while being observed.<sup>41</sup> Current process mining technique does not take those runtime change into account. In the introduced Lean Six Sigma methodology, once the hierarchical structure of the event log is set in the define phase, process mining will be based on that setting until the end. Hence, for the projects that target on the systems that are constantly encountering runtime changes, this integrated method may not be the best choice. But normally this phenomena does not happen very often. Even when a *scope creep*<sup>42</sup> happens, that hierarchical structure of the event log might still remain the same.

The business case illustrated in the thesis is dealing with MJV posting process, which is executed by the so-called PAIS (Process-Aware Information System). A *Process-Aware Information System* (PAIS) is a software *system* that manages and executes operational processes involving people, applications, and/or information sources on the basis of process models.<sup>43</sup> PAIS actually is a premise for process mining implementation. There must exist some sorts of patterns concerning the operational processes in order to implement the process mining techniques. As the business case shows, all analysis and discovery are completely based on the patterns that are summarized from the event log. Those patterns are then graphically represented in the process mapping.

---

<sup>40</sup> *Process support and knowledge representation in health care : BPM 2012 joint workshop, proHealth 2012/KR4HC 2012, tallinn, estonia, september 3, 2012, revised selected thesiss* (Lecture notes in artificial intelligence, 7738). (2013). Berlin: Springer. doi:10.1007/978-3-642-36438-9

<sup>41</sup> *Process mining : Discovery, conformance and enchancement of business processes*. (2011). Heidelberg etc.: Springer.

<sup>42</sup> Scope creep (also called requirement creep, function creep, feature creep, or kitchen sink syndrome) in project management refers to changes, continuous or uncontrolled growth in a project's scope, at any point after the project begins. This can occur when the scope of a project is not properly defined, documented, or controlled. It is generally considered harmful.

<sup>43</sup> Wil M. Aalst. 2009. Process-Aware Information Systems: Lessons to Be Learned from Process Mining. In *Transactions on Petri Nets and Other Models of Concurrency II*, Kurt Jensen and Wil M. Aalst (Eds.). Lecture Notes In Computer Science, Vol. 5460.

## 5.2 Limitations on the event log

Another limit is on the extracting method for the event logs. Process mining heavily relies on the extracted information from the event log. It is extremely critical that the scope and granularity of an event log match the questions that one would like to answer. Unfortunately, not all the event logs in systems are ready for use. Some systems do not contain the complete set of history information, e.g. when a record is updated, it simply overwrites the old values. Furthermore, the purpose of event logs in many systems is for debugging, hence, the information might be scattered over many separated tables. This situation might cause the event logs to be unusable for process mining or require substantial efforts in preparing the event log for process mining.

Even though the event log in the business case of this thesis is in excel format and very easily manageable, still the information recorded in this event log is limited to system behaviors. If readers pay attention to this MJV processes, readers might find out along the whole procedures from the MJV generation till MJV being posted in the database, only the posting processes are handled by the system. Therefore, the event logs can only track the posting processes. As for the information related to the content of MJV and manual editing of the MJV, they can only be investigated through analysis instead of process mining algorithm. In other words, process mining capability is strictly limited to the volume of information contained in event logs. At present, not every IT system can provide information-sufficient event logs since the original purpose of event log is for debugging. This is also the reason why process mining can only function as a tool to assist Lean Six Sigma but not as an individual methodology to support a lean project. However, this does not mean that process mining analysis is completely constrained by the quality of system event logs. Process mining itself is an algorithm that can quickly sort out dataset and produce the process stream based on this dataset. Event log is a very common type of business dataset. If practitioners cannot retrieve event logs from the system or do not have event logs at all, they just need to look for the suitable dataset that can be imported into process mining algorithms. Those datasets can be collected by investigation, interview etc. The structure of the dataset is in fact the most important prerequisite for applying process mining. Practitioners need to reconstruct the collected data into the previously introduced hierarchical structure of the event log (Figure 6).

To apply process mining does not necessarily require a formal event log, but it definitely requires the analyzed data to be in the correct data structure and contain all those necessary information. To summarize, if there are no useable event logs or only event logs that are not suitable for process mining, what practitioners can do is to create a event-log-like dataset themselves. But creating such kind of event-log-like dataset is extremely time consuming and troublesome, since it might incur the investigation upon thousand or even millions of transactions. That is why process mining is not limited by



the availability of event logs but its capability is somehow bonded to whether practitioners can find a suitable event log or at least an event-log-like dataset.

### **5.3 Limitations on the software tools**

Another limitation is related to the software tools. In the Lean Six Sigma methodology the constant tracking of processes in the process map is necessary. However, as new events happen and event logs are constantly recording new events, to keep tracking, event logs need to be continuously imported in the software tool to generate process maps. But now there are no software that can automatically draw event logs out of the system and import them to the process mining software. Hence, practitioners would have to manually draw data out of event logs and import them to the software. This manual processes can sometimes substantially limit the efficiencies in utilizing process mining tools.

## 6 CONCLUSION AND DISCUSSION

Each DMAIC phase has its own objectives and regardless of what kind of tools or techniques are implemented in each phase, as long as those objectives are achieved in order, the implementation of DMAIC can be considered successful. Process mining techniques are proven to be effective in accelerating and simplifying DMAIC processes by assisting to achieve every objective of DMAIC phases. This fact is validated by both theory induction and business case practice.

For validation from the theory, process mining concepts as well as Lean finance are introduced in Chapter 2, which provides the context for inferring the integrated method. Then Chapter 3 guide through the integration step-by-step and validate the integration based on the methodology framework. Both Chapter 2 and Chapter 3 are devoted to answer the first four sub-questions listed in Chapter 1.

Chapter 4 elaborates the business case, attempting to validate the integration from practical point of view. Even though the business case cannot show the result of the Company ABC's project in terms of the improvement upon the MJV processes, this is due to the fact that the methodology practiced in this thesis aims at the long-term improvement upon the performance, like any other lean projects. Long term is normally referring to longer than 2 years. In short term, however, lean project can hardly take any effects and in addition, it might even show signs of deteriorating performance. The business case in this thesis is conducted during the drafting of this thesis. Hence, the outcome of implementing this methodology cannot be presented through performance improvement. But this does not invalidate the proven compatibility between process mining techniques and Lean Six Sigma. I interviewed the practitioners of this business case right after the process mining techniques are implemented. The detail interview is attached in Appendix 9. They uniformly agreed that process mining as a tool substantially reduced the workload in terms of deficit-seeking time and proof-of-concept investigation. Compared to regular methodology, process mining technique can better consolidate the data for analysis in a much intuitive way. Data richness used to create troubles in filtering out the appropriate and useful information. Process mining offers simple accesses to extracting the essential features related to process data. Furthermore, the ease of utilizing process mining also create great convenience in understanding and communicating the project with the stakeholders. Instead of basing analysis on plain data table, process mining notations allow more explicit graphical expression of data to be presented for both analysis and monitoring. The visualization of data can effectively align the project team members' visions and perspectives upon the project. In addition, the biggest advantage of integrating process mining comes from process mapping. Process map is produced via a great volume of the event log data. The utilization of large volumes of data can, to some extent, cover some areas that cannot be reached through manual analysis. Simply speaking, analysts

must have a clearly defined objectives before collecting the data. So they know what kind of data should be collected to solve the objective question. This is so-called the collection of known unknown. However, an objective cannot always be clearly defined. Then analysts come to the predicament that they do not know what kind of data should be collected. In this case, the target information is called unknown unknowns.<sup>44</sup> Process mining provides algorithms to sort out large volume of data automatically and represent the patterns of data in a visually understandable way. In this way, practitioners would be able to decrease the blind point of dataset and make the best use of data.

In Chapter 4, process mapping is prioritized in every aspects of the project. Data collection and data analysis are all based on a complete process map that is directly generated via process mining algorithm. This process map is founded on the system event logs which capture the real runtime information, leading to precise alignment between the process map and reality. Simply speaking, process mining assists the implementation of Lean project in such a way that it precisely refine the raw data and use modeling language to represent those raw data. When changes start to take place, those changes can also be visualized in the model at the process level. As lean project heavily emphasize the continuous improvement and process streaming, process mining is such a natural tool to shape all the project-related data into the forms that a lean project exactly needs. To conclude, process mining contains all the necessary attributes to be integrated in Lean Six Sigma and it indeed substantially enhance the effectiveness and efficiency of practicing lean finance project. In theory, the synergy between process mining and Lean Six Sigma can be analogically applied to any other lean methodologies.

---

<sup>44</sup> Unkown unkown is referring to the information that people do not know that they do not know.

## 7 RECOMMENDATIONS

### 7.1 Recommendations for business practice

Process mining in nature is still a tool that supports the lean project implementation. Tools can only function when it is in the right hands. Therefore, the simplicity in implementing process mining cannot compensate for the lack of training. Practitioners of process mining must be systematically trained in lean thinking and organization structure should also be set up in a way that can accelerate the lean implementation. Process mining is a very flexible tool that does not only support activity-controlled process model (Like it is in the business case). It can be used to deduce the resource-centered process model as well. For instance, event logs also contains the information regarding to the owner of each process. They can also be used to generate process map, which reveals process information from another angle. Figure 30 is an example. Those names are the owners of different processes. Those arrows that connect those names represents the order of hand-in works. The thickness of the arrows indicate the amount of hand-in works. In Figure 30, Mike and Sara, for instance, are connected with two arrows, which means they hand works to each other. The arrows pointing to Sara is thicker than the other one, indicating that there are more works handing to Sara from Mike than vice versa.

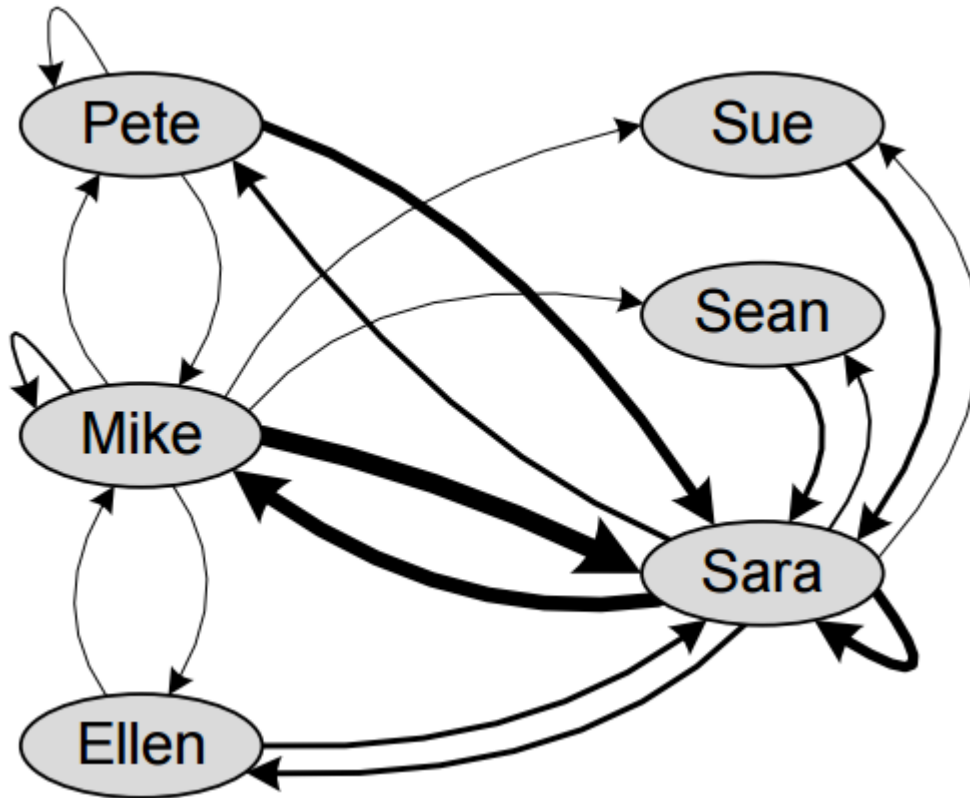


Figure 30 Resource-centered process mapping

Process mapping from different perspectives can render process mining revealing more information. Practitioners should utilize those different perspectives according to the purpose of the projects, for which it can optimize the values of process mining.

## **7.2 Recommendations for further research**

This thesis only focuses on lean finance project and the integration between process mining and Lean Six Sigma. Due to the time limits and scope of this thesis, only one business case is provided and it is insufficient to validate the integration via only one business case. More data analysis and statistical support are still required to prove the effectiveness of the integrated method. In addition, the logic of this integration can also be applied on much wider range of methodologies. For instance, the popular project management methodology, Scrum, can also integrate process mining almost in the same way. More research can be done upon whether this sort of integration is valid for other project management methods.

## REFERENCE

- Aalst, W. M. P. v. d. (2011). *Process mining : Discovery, conformance and enchancement of business processes*. Heidelberg etc.: Springer.
- Ahern, D. M., Clouse, A., & Turner, R. (2004). *CMMI distilled: a practical introduction to integrated process improvement*. Addison-Wesley Professional.
- Berger, A. (1997). Continuous improvement and kaizen: standardization and organizational designs. *Integrated manufacturing systems*, 8(2), 110-117.
- Buijs, J. C. A. M. (2010). *Mapping data sources to xes in a generic way*. Maters Thesis.
- Basu, R. (2009). *Implementing Six Sigma and Lean : A practical guide to tools and techniques (1st ed.)*. Amsterdam: Elsevier Butterworth-Heinemann.
- Carreira, B., & Trudell, B. (2006). *Lean Six Sigma that works : A powerful action plan for dramatically improving quality, increasing speed, and reducing waste*. New York, NY etc.: American Management Association.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.
- Carreira, B. (2006). *Lean six sigma that works : A powerful action plan for dramatically improving quality, increasing speed, and reducing waste*. New York, NY etc.: American Management Association.
- Cecil, B. (2014). FINANCE lean finance getting really fit means establishing common end-to-end process, technology, data and organization models. *Financial Executive*, 30(3), 53-53.
- Cima, R. R., Brown, M. J., Hebl, J. R., Moore, R., Rogers, J. C., Kollengode, A., ... & Team, S. P. I. (2011). Use of lean and six sigma methodology to improve operating room efficiency in a high-volume tertiary-care academic medical center. *Journal of the American College of Surgeons*, 213(1), 83-92.
- BPM (Conference) Beijing, China). (2013). *Business process management : 11th international conference, bPM 2013, beijing, china, august 26-30, 2013. proceedings* (Lecture notes in computer science, 8094). Berlin: Springer. doi:10.1007/978-3-642-40176-3
- De MEDEIROS, A. K., Weijters, A. J., & van der Aalst, W. M. (2007). *Genetic process mining: an experimental evaluation*. *Data Mining and Knowledge Discovery*, 14(2), 245-304.
- Dilton-Hill, K. (2015). Lean in finance. *Journal of Corporate Accounting and Finance*, 26(3), 15-25.
- Drew, J. (2004). *Journey to lean : Making operational change stick*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan. doi:10.1057/9781403948410

- Drohomeretski, E. (2013). Lean, six sigma and lean six sigma: An analysis based on operations strategy. *International Journal of Production Research*, 52(3), 804-824. doi:10.1080/00207543.2013.842015
- E. Badouel and P. Darondeau. Theory of Regions, In W. Reisig and G. Rozenberg, editors, *Lectures on Petri Nets I: Basic Models, volume 1491 of Lecture Notes in Computer Science*, pages 529-586. Springer, Berlin, 1998.
- El-Haik, B. (2006). *Simulation-based lean six-Sigma and design for six-Sigma*. Hoboken, NJ: Wiley-Interscience.
- Furterer, S. L. (2009). *Lean Six Sigma in service : Applications and case studies*. Boca Raton: CRC Press.
- Gómez-Pérez, J. M. (2010). *Acquisition and understanding of process knowledge using problem solving methods* (Studies on the Semantic Web, volume 007; Studies on the Semantic Web, v. 007). Amsterdam: IOS Press.
- Günther, C. W., & van der Aalst, W. M. (2006, September). *A generic import framework for process event logs*. In International Conference on Business Process Management (pp. 81-92). Springer Berlin Heidelberg.
- Hayler, R., & Nichols, M. D. (2007). *Six sigma for financial services : How leading companies are driving results using lean, six sigma, and process management*. New York: McGraw-Hill.
- Harrington, H. J. (1991). *Business process improvement: The breakthrough strategy for total quality, productivity, and competitiveness* (Vol. 1). New York, NY: McGraw-Hill.
- John, A., Dr.-Ing. (2008). *Six sigma Lean toolset : Executing improvement projects successfully* (S. Lunau, Ed.; A. Schmitz, Trans.). Berlin: Springer.
- Mader, D. (2009). Lean six sigma's evolution: Integrated method uses different deployment models. *Quality Control and Applied Statistics*, 54(1/2), 43-43.
- Macan, T. H. (1994). Time management: Test of a process model. *Journal of applied psychology*, 79(3), 381.
- Nisbet, R., Elder, J. F., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Amsterdam: Academic Press/Elsevier.
- Necessity and inefficiency in the generation of waste : A thermodynamic analysis.* (2003). *Journal of Industrial Ecology*, 7(2), 113-123. doi:10.1162/108819803322564389
- Pande, P., Neuman, R., & Cavanagh, R. (2001). *The Six Sigma way team fieldbook: An implementation guide for process improvement teams*. McGraw Hill Professional.

- Pepper, M. P. J., & S Oakland, J. (2008). *Statistical process control* (6th ed. ed.). Burlington, MA: Butterworth-Heinemann.
- pedding, T. A. (2010). The evolution of lean Six Sigma. *International Journal of Quality & Reliability Management*, 27(2), 138-155.
- Sharp, A., & McDermott, P. (2009). *Workflow modeling: tools for process improvement and applications development*. Artech House.
- Stelzer, D., & Mellis, W. (1998). Success factors of organizational change in software process improvement. *Software Process: Improvement and Practice*, 4(4), 227-250.
- Spendler, L. I. (2010). *Data mining and management (Computer science, technology and applications)*. New York: Nova Science Publishers.
- Taniar, D. (2007). *Research and trends in data mining technologies and applications* (IGI global research collection). Hershey, PA: Idea Group Pub.
- Taniar, D. (2008). *Data mining and knowledge discovery technologies* (Advances in data warehousing and mining, vol. 2). Hershey, PA: IGI Publishing.
- Tosini, P. (1977). *Leaning against the wind: A standard for managed floating* (Essays in international finance, 126). Princeton, N.J.: International Finance Section, Department of Economics, Princeton University.
- TomasYu-Lee, R. (2006). Determining the financial value of implementing lean. *Journal of Corporate Accounting & Finance*, 17(3), 79-88. doi:10.1002/jcaf.20200
- Van der Aalst, W. M., Reijers, H. A., Weijters, A. J., van Dongen, B. F., De Medeiros, A. A., Song, M., & Verbeek, H. M. W. (2007). *Business process mining: An industrial application*. *Information Systems*, 32(5), 713-732.
- Van der Aalst, W. M. (2011). *Getting the Data*. In *Process Mining* (pp. 95-123). Springer Berlin Heidelberg.
- Vashishth, A. (2017). Lean six sigma in financial services industry: A systematic review and agenda for future research. *Total Quality Management & Business Excellence*, 9(9), 1-19. doi:10.1080/14783363.2017.1308820
- Wirth, R., & Hipp, J. (2000, April). *CRISP-DM: Towards a standard process model for data mining*. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (pp. 29-39).
- Yang, K. (2009). *Design for six sigma : A roadmap for product development* (2nd ed. ed., McGraw hill professional). New York: McGraw-Hill.



## APPENDIX

### Appendix 1

List of organizations that successfully implemented Lean Finance Project <sup>45</sup>

Commercial and Savings Banks	Diversified Financials	Securities	Insurance
<ul style="list-style-type: none"> <li>• Abbey</li> <li>• ABN AMRO</li> <li>• Asian Development Bank</li> <li>• Bank of America</li> <li>• Bank Negara Indonesia</li> <li>• Bank Mandiri</li> <li>• Barclays</li> <li>• Capital One Financial</li> <li>• Citigroup</li> <li>• Commonwealth Bank of Australia</li> <li>• Credit Suisse</li> <li>• Development Bank Singapore</li> <li>• Deutsche Bank</li> <li>• Fifth Third Bancorp</li> <li>• Halifax Bank of Scotland</li> <li>• Household</li> <li>• HSBC</li> <li>• ICICI Bank</li> <li>• JPMorgan Chase</li> <li>• KeyCorp</li> <li>• Lloyds TSB</li> <li>• M &amp; S Money</li> <li>• Mellon Financial Corp</li> <li>• National City Corp</li> <li>• Nedcor</li> <li>• Overseas Chinese Banking Corporation</li> <li>• State Street Bank &amp; Trust Co.</li> <li>• UBS</li> <li>• Wachovia Corp</li> <li>• Washington Mutual</li> <li>• Wells Fargo</li> </ul>	<ul style="list-style-type: none"> <li>• American Express</li> <li>• Aon</li> <li>• Countrywide Financial</li> <li>• GE Capital</li> </ul>	<ul style="list-style-type: none"> <li>• Merrill Lynch</li> <li>• Goldman Sachs</li> <li>• Charles Schwab</li> <li>• Fidelity</li> </ul>	<ul style="list-style-type: none"> <li>• Aetna</li> <li>• AIG</li> <li>• Allstate</li> <li>• AXA Insurance</li> <li>• Berkshire Hathaway</li> <li>• CIGNA</li> <li>• Consecom</li> <li>• Jefferson Pilot</li> <li>• Lincoln National</li> <li>• Northwestern Mutual</li> <li>• Principal Financial</li> <li>• Prudential Financial</li> <li>• Unum Provident</li> <li>• USAA</li> <li>• Vertex</li> <li>• Zurich Financial Services</li> </ul>

Figure 31 List of organizations

<sup>45</sup> *Six sigma for financial services : How leading companies are driving results using lean, six sigma, and process management.* (2007). New York: McGraw-Hill.

## **Appendix 2**

### ***Discovery***

The first type of process mining is discovery. A discovery technique utilizes the event logs to produce process model via some algorithms without any a-priori information. For example,  $\alpha$ -algorithm can be applied on an event log to generate a Petri net representing the process streams that are recorded by event logs.

### ***Conformance***

The second type of process mining is called conformance. As indicated by the name, an existing pre-defined process model is compared with the event logs to check whether the event logs can indeed precisely perform the pre-defined process model. In other words, conformance checking is used to check if reality, as recorded in the log, conforms to the model and vice versa. For instance, in an accounting system, the pre-defined process model requires that all the transactions that are valued over 1 million euros have to be checked and approved twice. Then a conformance checking with the system's event log will show whether this rule is followed or not. In general, there are two input in conformance analysis, event logs which is the recorded reality and requirements which are pre-defined in the process. By scanning the event log using a model specifying these requirements, one can discover potential cases of fraud. Hence, conformance checking may be used to detect, locate and explain deviations, and to measure the severity of these deviations.

### ***Enhancement***

The third type of process mining is enhancement. Essentially it is a process searching for extending and improving the existing process model with the information regarding to the actual processes that are recorded in the event logs. Compared to conformance checking which measures the alignment between model and reality, enhancement targets at changing or extending the a-priori model. One type of enhancement is repair, i.e., modifying the model to better reflect reality. For example, in a pre-defined process model some activities are defined to follow certain order, but in reality event logs show that those activities can take place in other orders as well. Then via process mining analysis, this reality can be added to the pre-defined process model. Another type of enhancement is

extension, i.e., adding a new perspective to the process model by cross-correlating it with the log. An example is the extension of a process model with performance data. For instance, event logs contains not only the information regarding to activities but also the timestamps relating to when the activities happen and the duration of it. Those information can provide valuable insight to spot the bottleneck and service levels of the process, adding more perspectives to the pre-defined process model.

All those three types of process mining introduced above are the three most important objectives of implementing process mining. Those objectives are achieved via several statistical algorithms which derived from data mining and machine learning. Not all those algorithms are relevant to this thesis. Here, only the most basic as well as the most fundamental algorithms will be explained in this thesis. However, in real implementation, the selection of algorithm is very flexible, depending on the purpose of process mining project.

## Appendix 3

A-algorithm forms the basics of process mining. It takes 8 steps to discover the patterns of a random set of activities captured in event logs. Those 8 steps are as followed:

**Let  $L$  be an event log over  $T$ .  $\alpha(L)$  is defined as follows.**

1.  $T_L = \{ t \in T \mid \exists_{\sigma \in L} t \in \sigma \}$ ,
2.  $T_I = \{ t \in T \mid \exists_{\sigma \in L} t = \text{first}(\sigma) \}$ ,
3.  $T_O = \{ t \in T \mid \exists_{\sigma \in L} t = \text{last}(\sigma) \}$ ,
4.  $X_L = \{ (A, B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B} a \rightarrow_L b \wedge \forall_{a_1, a_2 \in A} a_1 \#_L a_2 \wedge \forall_{b_1, b_2 \in B} b_1 \#_L b_2 \}$ ,
5.  $Y_L = \{ (A, B) \in X_L \mid \forall_{(A', B') \in X_L} A \subseteq A' \wedge B \subseteq B' \Rightarrow (A, B) = (A', B') \}$ ,
6.  $P_L = \{ p_{(A, B)} \mid (A, B) \in Y_L \} \cup \{ i_L, o_L \}$ ,
7.  $F_L = \{ (a, p_{(A, B)}) \mid (A, B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A, B)}, b) \mid (A, B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \}$ , and
8.  $\alpha(L) = (P_L, T_L, F_L)$ .

46

If readers are interested to look into the detail explanation of those 8 steps, please refer to Professor Wil Aalst's book, *Process Mining*, where he provides very comprehensive explanation.

The Apriori Algorithm is one of the best known algorithms in computer science, that also forms the fundamental elements of process mining algorithm ( $\alpha$ -algorithm). Apriori Algorithm, initially developed by Agrawal and Srikant, has the capability to accelerate the specification of association rules by exploiting the following two observations:

- If an item-set is frequent (i.e., an item-set with a support above the threshold), then all of its non-empty item-sets  $X, Y$ : if  $Y \subseteq X$  and  $N_X/N \geq \text{minsup}$ <sup>47</sup>, then  $N_Y/N \geq \text{minsup}$ .
- If, for any  $k$ ,  $I_k$  is the set of all frequent item-sets with cardinality  $k$  and  $I_1 = \emptyset$  for some  $l$ , then  $I_k = \emptyset$  for all  $k \geq l$ .

These two properties can be used to dramatically reduce the search-space when constructing the set of frequent item-sets. For example, if an item-set  $\{a, b\}$  is considered infrequent, according to Apriori Algorithm all the item-sets that contain  $\{a, b\}$  can also be

<sup>46</sup> *Process mining : Discovery, conformance and enchancement of business processes.* (2011). Heidelberg etc.: Springer.

<sup>47</sup> *Minisup* is a data mining concept, standing for minimal support. This is a parameter which indicates the minimum frequency ratio between a specific item set and all possible item sets that can define this specific item set as an association rule.

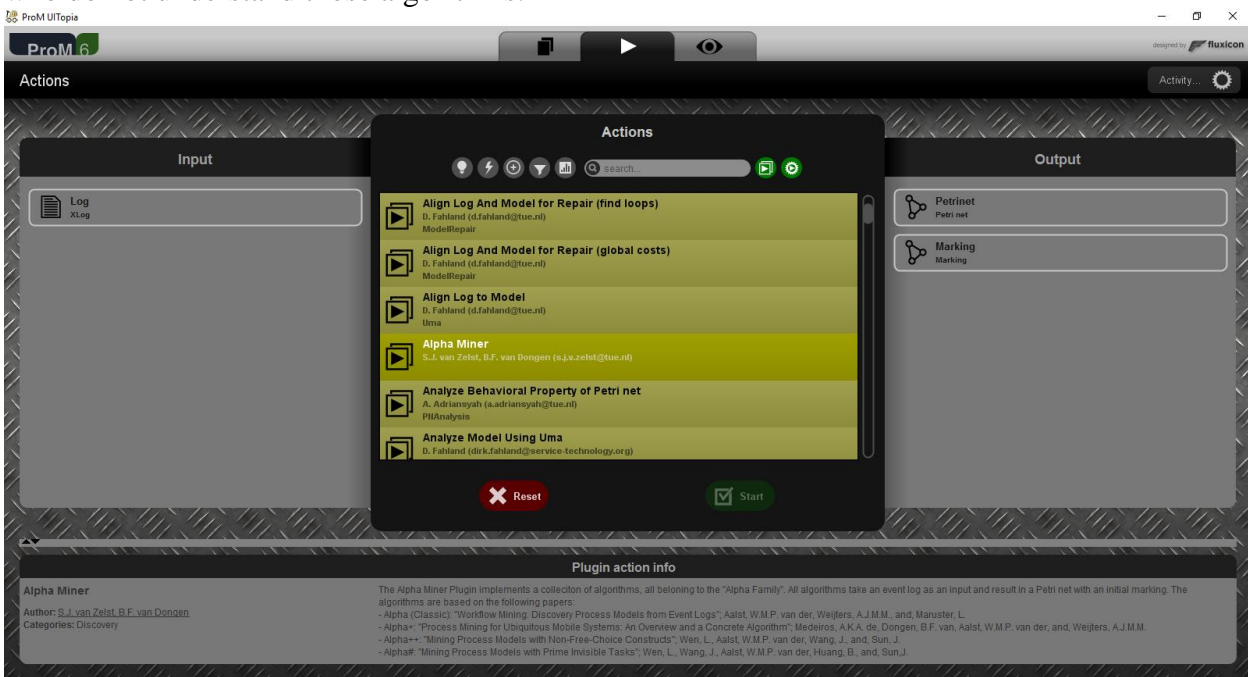
rendered infrequent item-sets. The Apriori algorithm uses the monotonicity property that all subsets of a frequent item-set are also frequent. Process mining utilizes this rule to increase the efficiency in defining the relevant activities in the process. For instance, in an event log there are 6 activities  $\{a,b,c,d,e,f\}$  that frequently appear in a specific process. Then all the subsets of those 6 activities, such as  $\{a,d\}$   $\{b,c\}$ , are also considered relevant to that specific process. Hence, analyzers do not need to check all the subsets of those 6 activities for their relevancy.

Apriori Algorithm can only discover the association rules of the activities but not the ordering. This algorithm can only assist to seek for all the events that are relevant to the process. However, a complete process model should also sort out the order of those events that leads to the final objective of the process. Therefore, sequence mining becomes another essential algorithm that specifically addresses the sequence of the item-set. To put it in simple terms, sequence mining is just another transformed version of Apriori Algorithm. It is the same trick to use Apriori Algorithm to explore item-set pattern in the dataset. But this time, when defining the subset of the frequent item-sets, it is the sequence that is focused on instead of the item object itself. For instance, the sequence  $\{\{x\}, \{x,y\}, \{y\}\}$  is a subsequence of  $\langle\{z\}, \{x\}, \{z\}, \{x,y,z\}, \{y,z\}, \{z\}\rangle$  because  $\{x\} \subseteq \{x\}$ ,  $\{x,y\} \subseteq \{x,y,z\}$ , and  $\{y\} \subseteq \{y,z\}$ . However,  $\langle\{x\}, \{y\}\rangle$  is not a subsequence of  $\langle\{x,y\}\rangle$  and vice versa, even though  $\langle\{x\}, \{y\}\rangle$  does contain the same item-set as  $\langle\{x,y\}\rangle$ . Since it is the sequence that matters in this analysis, as long as the sequence is not the same, item-sets are considered different regardless of the fact that they might contain the same items. By defining the item-set in this way, it is possible to exploit Apriori Algorithm again to discover the frequency pattern of sequence, in the light of which the sequence pattern can be spotted.

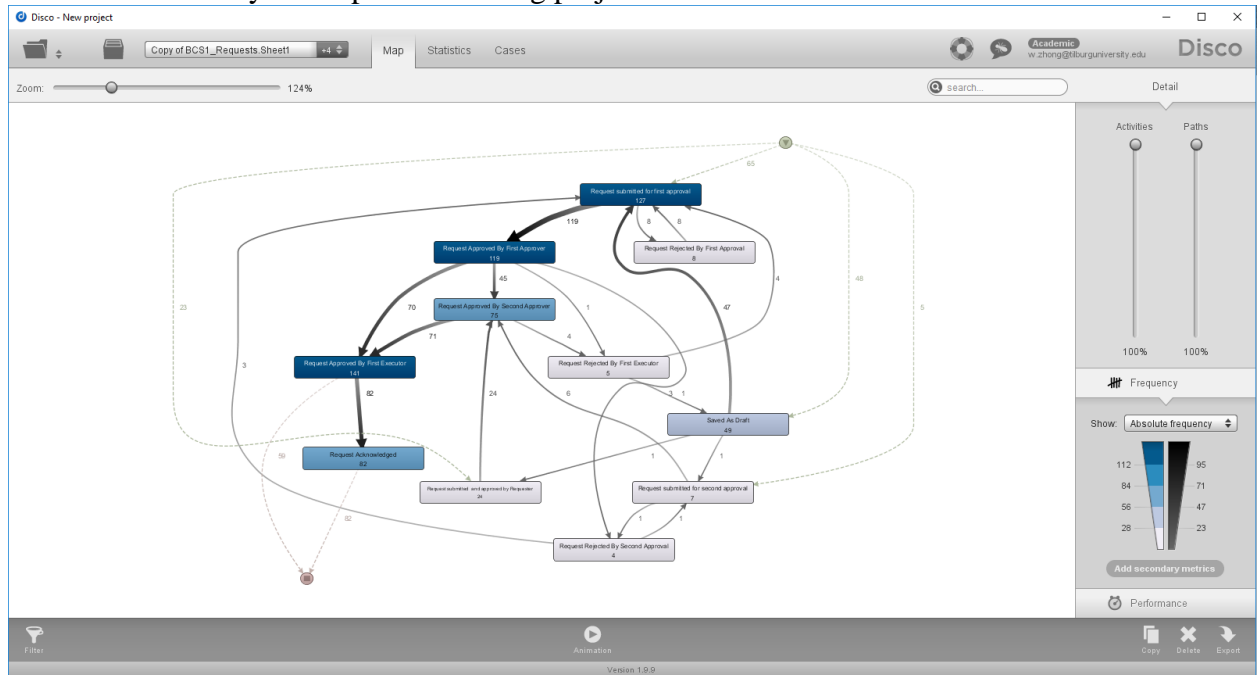
## Appendix 4

### Disco vs. ProM

ProM is an open source software that contains a lot of external packages. Those packages give ProM tremendous capability in handling different event logs and generating process models with a large portfolio of algorithms. Hence, ProM is more powerful in terms of functionality and it is also capable to customize the analysis according to the purpose. But ProM's interface is not very user-friendly. (As it is shown in the screen shot below) A tremendous amounts of options on algorithms also make it hard to practice for the people who do not understand those algorithms.



Disco, on the other hand, is already a mature commercialized software. It is user-friendly and does not have too many glitches (The screenshot of its interface is shown below). But All algorithms are built-in and users do not get the option to choose which algorithm to implement. Disco simply does the so-called heuristic mining for which the forms of analysis result is not customizable. In spite of the lack of customizability Disco can still generate valuable analysis for process mining project.



## Appendix 5

Petri Nets<sup>48</sup> are the oldest and best investigated process modeling language allowing for the modeling of concurrency. Although the graphical notation is intuitive and simple, Petri nets are executable and many analysis techniques can be used to analyze them. A Petri Net is a triplet  $N = (P, T, F)$  where  $P$  is a finite set of places,  $T$  is a finite set of transitions such that  $P \cap T = \emptyset$ , and  $F \subseteq (P \times T) \cup (T \times P)$  is a set of directed arcs, called the flow relation. A marked Petri Net is a pair  $(N, M)$ , where  $N = (P, T, F)$  is a Petri Net and where  $M \in B(P)$  is a multi-set over  $P$  denoting the marking of the net. The set of all marked Petri nets is denoted  $\mathcal{N}$ . Figure below is an example of a Petri Net. The graph can be formalized as follows:  $P = \{start, c1, c2, c3, c4, c5, end\}$ ,  $T = \{a, b, c, d, e, f, g, h\}$ , and  $F = \{(start, a), (a, c1), (a, c2), (c1, b), (c1, c), (c2, d), (b, c3), (c, c3), (d, c4), (c3, e), (c4, e), (e, c5), (c5, f), (f, c1), (f, c2), (c5, g), (c5, h), (g, end), (h, end)\}$ . The marking shown in the graph is  $[start]$ , i.e., a multi-set containing only one token. The dynamic behavior of such a marked Petri net is defined by the so-called firing rule. A transition is enabled if each of its input places contains a token. An enabled transition can fire thereby consuming one token from each input place and producing one token for each output place. Hence, transition  $a$  is enabled at marking  $[start]$ . Firing  $a$  results in the marking  $[c1, c2]$ . Note that one token is consumed and two tokens are produced. This process is called AND-split, stating that two transition can happen concurrently. At marking  $[c1, c2]$ , transition  $a$  is no longer enabled. However, transition  $b$ ,  $c$ , and  $d$  have become enabled. From marking  $[c1, c2]$ , firing  $b$  results in marking  $[c2, c3]$ . Here,  $d$  is still enabled, but  $b$  and  $c$  not anymore. This is defined as XOR-split transition, in which condition only one of the multiple transitions can happen (in this case, either  $b$  or  $c$  can happen). Because of the loop construct involving  $f$  there are infinitely many firing sequences starting in  $[start]$  and ending in  $[end]$ . This sort of loop construct is very common in finance system. Failure of meeting approval standard should normally result in reiteration of the process and loop construct in Petri net can clearly reflect this situation.

---

<sup>48</sup> E. Badouel and P. Darondeau. Theory of Regions, In W. Reisig and G. Rozenberg, editors, *Lectures on Petri Nets I: Basic Models*, volume 1491 of *Lecture Notes in Computer Science*, pages 529-586. Springer, Berlin, 1998.



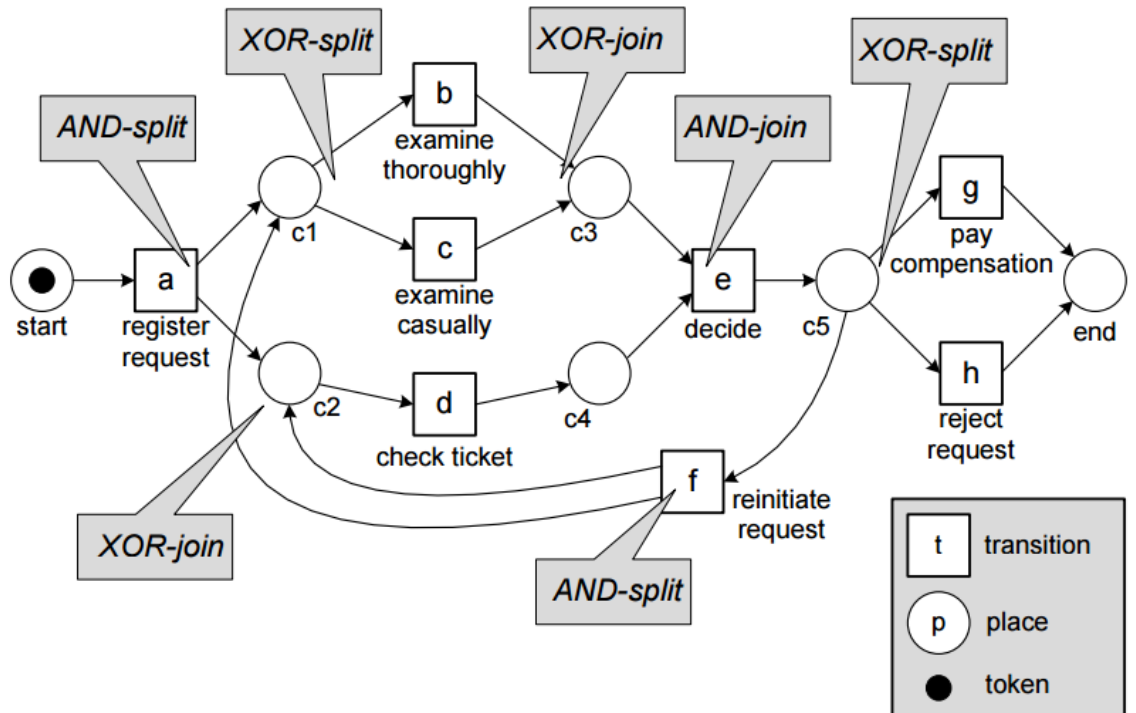
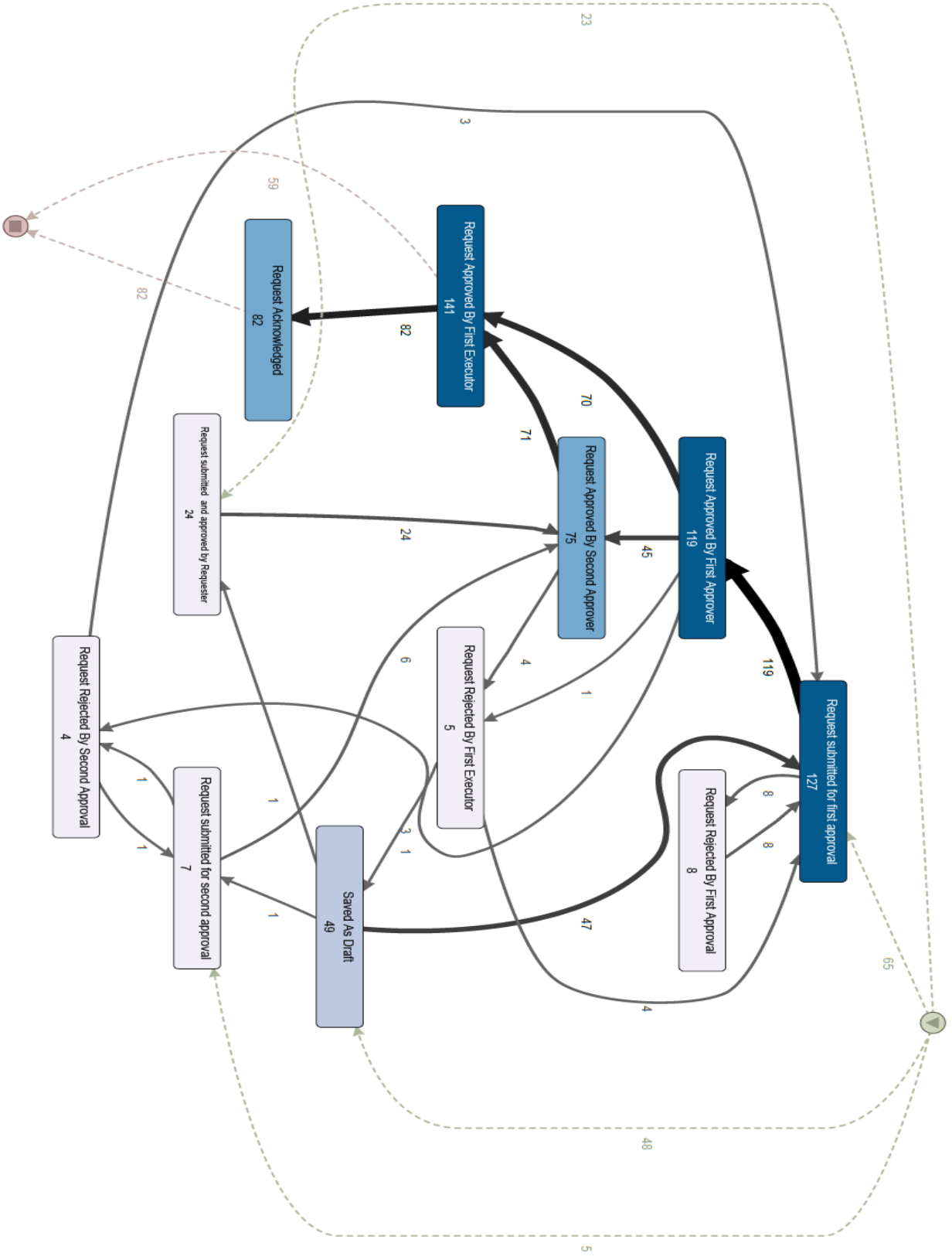


Figure 32 Petri net example

Petri Net is well known for its simplicity and representativeness. However, practitioners should choose whichever is more understandable to them. In the end, the purpose of notation is always communication and documentation.

# Appendix 6



## Appendix 7

VOC<sup>49</sup> is an expression for listening to external customers and understanding their requirements for the products or services. Detailed CTS and VOC need to be prepared in order to specify the corresponding measurement for the lean improvement. VOC can be investigated through interviews, surveys and warranty information. CTS is a little bit more tricky. A data collection plan need be developed to identify the data to be collected that relate to the CTS criteria.

The data collection plan ensures:

- Measurement of CTS metrics
- Identification of the right mechanisms to carry out data collection
- Collection and analysis of data
- Definition of how and who is responsible to collect the data

The steps for creating a data collection plan in the Measure phase are

- Define the CTS criteria: CTS is a characteristic of a product or service that fulfills a critical customer requirement or a customer process requirement. CTS measures are the basic elements in driving process measurement, improvement, and control. (George, 2005)
- Develop metrics: metrics are identified that help to measure and assess improvements related to the identified CTS measures. Some rule-of-thumb for selecting metrics are to (Lindsey, 2007):
  - Consider the vital few vs. the trivial many
  - Focus on the past, present and future
  - Link metrics to meet the needs of shareholders, customers, and employees
  - Identify data collection mechanisms
  - Identify analysis mechanisms
  - Develop sampling plans
  - Develop sampling instructions

---

<sup>49</sup> *Lean six sigma in service : Applications and case studies.* (2009). Boca Raton: CRC Press.

<b>Critical to Satisfaction (CTS)</b>	<b>Metric</b>	<b>Data collection mechanism (survey, interview, focus group, etc.)</b>	<b>Analysis mechanism (statistics, statistical tests, etc.)</b>	<b>Sampling plan (sample size, sample frequency)</b>	<b>Sampling instructions (who, where, when, how)</b>
Speed to market	Cycle time	Project management tool	Statistics (mean, variance); <i>t</i> -test	One year of projects	Collect data from project management system for last year

The table above is an example of typical CTS illustration table.

## Appendix 8

**DPMO<sub>50</sub>:** Six Sigma represents a stretch goal of six standard derivations from the process mean to the specification limits when the process is centered, but also allows for a 1.5 sigma shift toward either specification limit. This represents a quality level of 3.4 defects per million. This is represented in the Figure below. LSL means the lower specification limit and USL means the upper specification limit.

The greater the number of  $\sigma$  values, the smaller the variation (the tighter the distribution) around the average. DPMO provides a single measure to compare the performances of very different operations, giving an apple-to-apples, comparison, not apples-to-oranges. Table below shows a Sigma-to-DPMO conversion.

DPMO is calculated as (Brassard and Ritter 2001):

$$DPMO = \frac{Defects \times 1000000}{Units \times Opportunities}$$

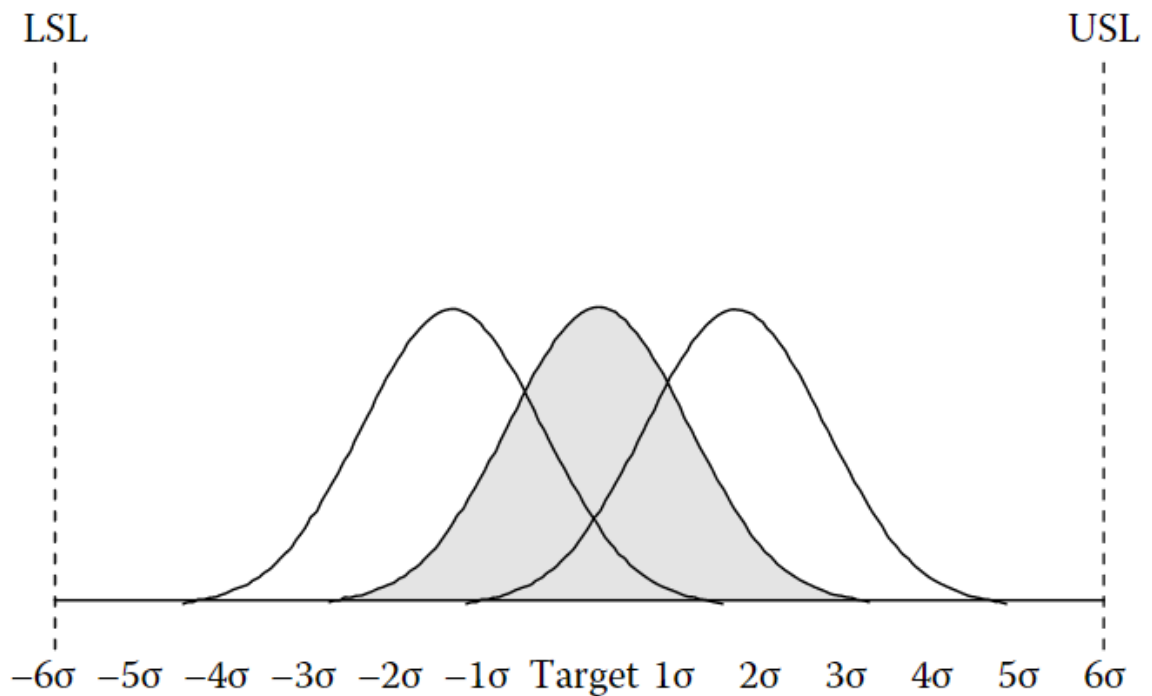


Figure 33 Six Sigma

---

<sup>50</sup> Furterer, S. L. (2009). *Lean Six Sigma in Service : Applications and Case Studies*. Boca Raton: CRC Press.

<b>Sigma level</b>	<b>DPMO</b>
6 $\sigma$	3.4 DPMO
5 $\sigma$	233 DPMO
4 $\sigma$	6,210 DPMO
3 $\sigma$	66,810 DPMO
2 $\sigma$	308,770 DPMO
1 $\sigma$	691,462 DPMO

Figure 34 Sigma to DPMO conversion

## Appendix 9

Survey with the practitioners:

There are in total 10 persons participating in this survey. They are all qualified lean project practitioners and they all have received sufficient systematical training in Lean concept. Hence, they are qualified to evaluate the integration of process mining and Lean Six Sigma.

In the below survey, the numbers in the answer represent the people who chooses the corresponding option.

1. To what extent do you think process mining helps to define the objectives of Lean project?

Very little help	Little help	Neutral	Some help	A lot of help
0	0	2	7	1

2. To what extent do you think process mining helps to collect and rationalize the data for Lean project?

Very little help	Little help	Neutral	Some help	A lot of help
0	0	0	3	7

3. To what extent do you think process mining helps to visualize the process model?

Very little help	Little help	Neutral	Some help	A lot of help
0	0	0	0	10

4. To what extent do you think process mining helps to track and monitor the de-fected processes?

Very little help	Little help	Neutral	Some help	A lot of help
0	0	0	5	5

5. To what extent do you think process mining helps to communicate the project objectives within the project team?

Very little help	Little help	Neutral	Some help	A lot of help
1	0	3	5	1

6. To what extent do you think process mining helps to trace the source of the de-fected processes?

Very little help	Little help	Neutral	Some help	A lot of help
0	0	0	4	6

7. To what extent do you think process mining helps to track and monitor changes made in the lean project?

Very little help	Little help	Neutral	Some help	A lot of help
0	0	0	3	7

8. To what extent do you think process mining helps to handle runtime changes that happen during the lean project?

Very little help	Little help	Neutral	Some help	A lot of help
3	2	2	3	0

9. To what extent do you think process mining is associated to the whole project management in a lean project?

Very closely	Closely	Neutral	Strongly	Very strongly
0	0	0	4	6

10. To what extent do you think process mining can be integrated in other similar lean finance projects?

Very badly	Badly	Neutral	Well	Very well
0	0	0	2	8